

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

**ФАКУЛЬТЕТ МЕНЕДЖМЕНТУ ТА МАРКЕТИНГУ
КАФЕДРА ЕКОНОМІЧНОЇ КІБЕРНЕТИКИ**

«На правах рукопису»
УДК 330.46

До захисту допущено:
Завідувач кафедри
_____ Катерина БОЯРИНОВА
« 13 » _____ грудня _____ 2022 р.

МАГІСТЕРСЬКА ДИСЕРТАЦІЯ

**на здобуття ступеня магістра
за освітньо-професійною програмою
«Економічна кібернетика»
спеціальності 051 Економіка**

**на тему: «Економіко-математичне моделювання динаміки
ф'ючерсних контрактів на фінансових ринках»**

Виконала:

Студентка 2-го курсу, групи УК-11мп
ДЕЙНЕКО Мирослава Богданівна _____

Науковий керівник:

Доцент кафедри економічної кібернетики,
к.т.н., доц. ЦЕСЛІВ Ольга Володимирівна _____

Рецензент:

Завідувачка кафедри економіки і підприємництва,
д.е.н., проф. ТУЛЬЧИНСЬКА Світлана Олександрівна _____

*Засвідчую, що у цій магістерській дисертації
немає запозичень з праць інших авторів без
відповідних посилань*

Студентка _____

Київ – 2022 року

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

**ФАКУЛЬТЕТ МЕНЕДЖМЕНТУ ТА МАРКЕТИНГУ
КАФЕДРА ЕКОНОМІЧНОЇ КІБЕРНЕТИКИ**

Рівень вищої освіти – другий (магістерський)

Спеціальність – 051 Економіка

Освітньо-професійна програма «Економічна кібернетика»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Катерина БОЯРИНОВА

« 01 » вересня 2022 р.

ЗАВДАННЯ

на магістерську дисертацію студентці

ДЕЙНЕКО МИРОСЛАВІ БОГДАНІВНІ

1. Тема дисертації «Економіко-математичне моделювання динаміки ф'ючерсних контрактів на фінансових ринках»,

науковий керівник дисертації Цеслів Ольга Володимирівна, к.т.н., доц.,
затверджені наказом по університету від 10.11.2022 року № 4143-с.

2. Термін подання студентом дисертації 13 грудня 2022 року.

3. Об'єкт дослідження: дослідження методів і моделей для прогнозування цін на ф'ючерсні контракти.

4. Вихідні дані: Динаміка цін на нафтові ф'ючерси з 2021 по 2022 рр.
Прогнозовані дані на 1 місяць.

5. Перелік завдань, які потрібно розробити:

а) теоретико–методологічна частина:

- Дослідити теоретичний аспект ф'ючерсного контракту;
- Проаналізувати поведінку ф'ючерсних контрактів на фінансових ринках.

б) дослідницько–аналітична частина:

- Дослідити статистичні моделі тимчасових рядів;
- Проаналізувати методи та моделі для прогнозування;

в) проектно–рекомендаційна частина:

- Практично реалізувати обрані методи та моделі;
- Проаналізувати отримані результати;
- Порівняти результати з реальними даними, зробити відповідні висновки.

6. Орієнтовний перелік ілюстративного матеріалу:

- 1) теоретичні аспекти роботи ф'ючерсних контрактів на фінансових ринках;
- 2) опис роботи прогнозової моделі ARIMA;
- 3) динаміка цін на ф'ючерсні контракти на сиру нафту у 2021-2022 рр.;
- 4) динаміка отриманих значень в результаті роботи методу GBM;
- 5) програмне впровадження моделей;
- 6) порівняння отриманих значень в результаті роботи методу GBM з історичними даними;
- 7) динаміка отриманих значень в результаті роботи моделі ARIMA;
- 8) порівняння отриманих значень в результаті роботи методу ARIMA з історичними даними;
- 9) динаміка отриманих значень в результаті роботи моделі Random Forest; порівняння отриманих значень в результаті роботи методу Random Forest з історичними даними.

7. Орієнтовний перелік публікацій за напрямом роботи:

- 1) Стаття «Економіко-математична модель ф'ючерсних контрактів» у науковому журналі «Причорноморські економічні студії» (фахове видання категорії Б);
- 2) Тези «Економіко-математичне моделювання динаміки ф'ючерсних контрактів на фінансових ринках», участь у XVI Всеукраїнській науково-практичній конференції з міжнародною участю «Моделювання та прогнозування економічних процесів», м. Київ.

8. Дата видачі завдання: 01 вересня 2022 року.

9. Календарний план

<i>№ з/п</i>	<i>Назва етапів виконання магістерської дисертації</i>	<i>Термін виконання етапів магістерської дисертації</i>	<i>Примітка</i>
1.	Збір необхідної інформації, вивчення та аналіз наукових джерел щодо досліджуваної тематики	01.09.2022-22.09.2022	
2.	Дослідження теоретичних засад обігу ф'ючерсів на фінансових ринках, аналіз відповідних процесів, надання на перевірку першого розділу	23.09.2022-09.10.2022	
3.	Аналіз математичних методів і моделей для дослідження динаміки ф'ючерсних контрактів на фінансових ринках	10.10.2022-24.10.2022	
4.	Визначення основних методів і моделей для реалізації поставлених завдань	25.10.2022-05.11.2022	
5.	Програмна реалізація моделей і методів	06.11.2022-24.11.2022	
6.	Аналіз отриманих результатів	25.11.2022-03.12.2022	
7.	Оформлення магістерської дисертації. Підготовка презентації.	04.12.2022-12.12.2022	

Студент

Мирослава ДЕЙНЕКО

Науковий керівник

Ольга ЦЕСЛІВ

РЕФЕРАТ

Магістерська дисертація на тему: «Економіко-математичне моделювання динаміки ф'ючерсних контрактів на фінансових ринках» містить 91 сторінок, 4 таблиці, 21 рисунок. Перелік посилань нараховує 60 найменувань.

Актуальність теми. Актуальність даної роботи полягає у необхідності побудови максимально точних прогнозів динаміки цін на ф'ючерсні контракти за умови, що невідома динаміка зовнішніх факторів, що впливають на ціноутворення цих ф'ючерсів, а також необхідність адаптації нових статистичних підходів до прогнозування.

Зв'язок роботи з науковими програмами, планами, темами. Магістерська дисертація на здобуття ступеня магістра виконувалась в Національному технічному університеті України «Київський політехнічний інститут імені Ігоря Сікорського» відповідно до планів наукових досліджень кафедри економічної кібернетики за темою «Глобалізація напрямів формування промислового потенціалу в умовах постіндустріальних трансформацій» (№ ДР 0112U007817). Роль автора полягає в дослідженні прогнозних методів і моделей.

Мета та завдання роботи. Метою даної роботи є побудова моделі прогнозування прибутковості ф'ючерсів і вибір оптимального методу чи моделі.

Об'єкт та предмет дослідження. Об'єктом дослідження являється фондовий ринок. Предметом дослідження є ф'ючерсний контракт і ф'ючерс на сиру нафту.

Методи дослідження. Протягом виконання роботи були застосовані такі методи дослідження як регресійний та кореляційний аналіз, статистичні методи, методи прогнозування часових рядів, використання інформаційних технологій та методів машинного навчання.

Наукова новизна. Була побудована моделі для прогнозування прибутковості ф'ючерсів, які базуються на методі геометричного броунівського руху, ARIMA та Random Forest. Була визначена оптимальна модель для прогнозування.

Практична значущість. Завдяки отриманим результатам можна зробити висновки і порівняти результати від прогнозування за допомогою моделей. В майбутньому можна застосовувати модель з кращим результатом наприклад при моделюванні інвестиційного портфелю чи при дослідженні коливань цін.

Апробація результатів роботи:

Стаття: Цеслів О.В., Цеслів О.С., Дейнеко М.Б. Економіко-математична модель ф'ючерсних контрактів. *Причорноморські економічні студії*. 2022. Вип. 77. С. 185-189. DOI: <https://doi.org/10.32782/bses.77-30>;

Тези: Цеслів О.В., Дейнеко М.Б. Економіко-математичне моделювання динаміки ф'ючерсних контрактів на фінансових ринках. *Моделювання та прогнозування економічних процесів* : матеріали XIV Всеукраїнської наук.-практ. конф. з міжнародною участю, 17 листопада 2022 року. Київ: КПІ ім. Ігоря Сікорського, Вид-во «Політехніка», 2022. С. 45-47.

Ключові слова: ціна, індекси цін, ф'ючерси, фінансовий ринок, прогнозування, модель ARIMA, геометричний броунівський рух, машинне навчання.

ABSTRACT

Master's dissertation on the topic «Economic-mathematical modeling of the dynamics of futures contracts on financial markets» includes 85 pages, 4 tables, 21 drawings. The bibliography list consists of 60 items.

The relevance of the topic. The relevance of this work lies in the need to build the most accurate forecasts of price dynamics for futures contracts, provided that the dynamics of external factors affecting the pricing of these futures are unknown, as well as the need to adapt new statistical approaches to forecasting.

Connection of work with scientific programs, plans, topics. The master's thesis for obtaining a master's degree was carried out at the National Technical University of Ukraine "Ihor Sikorskyi Kyiv Polytechnic Institute" in accordance with the scientific research plans of the Department of Economic Cybernetics on the topic "Globalization of directions for the formation of industrial potential in the conditions of post-industrial transformations" (No. DR 0112U007817). The role of the author is to research predictive methods and models.

Purpose and objectives of the work. The purpose of this work is to study the theoretical and practical aspects of the dynamics of futures contracts on financial markets.

The object and subject of research. The object of research is the financial market. The subject of the study is the futures contract and crude oil futures.

Research methods. The research used such research methods as regression and correlation analysis, statistical methods, time series forecasting methods, the use of information technology and machine learning methods.

Scientific novelty. Models were built for predicting futures returns, based on the method of geometric Brownian motion, ARIMA and Random Forest. The optimal model for forecasting was determined.

Practical significance. Analyzing the obtained results, you can draw conclusions and compare the results of forecasting using models.

Approbation of work results:

Article: Tsesliv O.V., Tsesliv O.S., Deineko M.B. (2022) Ekonomiko-matematychna model fiuchersnykh kontraktiv. [Economic-mathematical model of futures contracts.]. *Prychornomorski ekonomichni studii*. Vol. 77. P. 185-189. (in Ukrainian);

Theses: Tsesliv O.V., Deineko M.B. (2022) Ekonomiko-matematychne modeliuvannia dynamiky fiuchersnykh kontraktiv na finansovykh rynkakh. [Economic and mathematical modeling of the dynamics of futures contracts on financial markets]. *Modeliuvannia ta prohnozuvannia ekonomichnykh protsesiv*. XIV Vseukrainska nauk.-prakt. konf. z mizhnarodnoiu uchastiu, November 17th, 2022. Kyiv: KPI im. Ihoria Sikorskoho, Vyd-vo «Politekhnik». P. 45-47. (in Ukrainian).

Keywords: price, price indices, futures, financial market, forecasting, ARIMA model, geometric Brownian motion, machine learning.

ЗМІСТ

ВСТУП	9
1 ТЕОРІЯ Ф'ЮЧЕРСНОГО РИНКУ	12
1.1 Історія та еволюція ф'ючерсних бірж	12
1.2 Поняття ф'ючерсного контракту	20
1.3 Процес визначення ціни	23
1.4 Ф'ючерсні трейдери	27
1.5 Ф'ючерсні контракти на нафту	29
Висновки до розділу 1	32
2 ОГЛЯД МЕТОДІВ І МОДЕЛЕЙ ПРОГНОЗУВАННЯ	33
2.1. Моделі класу ARIMA	33
2.2 Побудова моделі ARIMA	35
2.3 Оцінка параметрів моделей ARIMA	38
2.2 Геометричний броунівський рух.....	42
2.4 Метод випадкових лісів.....	43
Висновки до розділу 2	49
3 ПРАКТИЧНЕ ЗАСТОСУВАННЯ ПРОГНОЗНИХ МОДЕЛЕЙ.....	50
3.1 Геометричний броунівський рух.....	50
3.2 Побудова моделі GBM з дискретним часом.....	56
3.3 Реалізація моделі	58
3.4 Реалізація моделі ARIMA.....	62
3.5 Модель Random Forest.....	67
Висновки до розділу 3	76
ВИСНОВКИ.....	78
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	80
ДОДАТОК А.....	86
ДОДАТОК Б.....	89

ВСТУП

Фондові біржі за свою історію стали не тільки інструментом, який може генерувати дохід для інвесторів у вигляді спекулятивних прибутків і дивідендів, а й механізмом, який має значний вплив на економіку.

Це спонукає не лише аналітиків, брокерів, банкірів, інвесторів, а й науковців приділяти все більше уваги цій галузі, намагаючись виявити закономірності та правила оцінки фінансових активів.

Часові ряди, що стосуються фінансових ринків, наприклад щоденних ціни на ф'ючерсні контракти, є дуже нестаціонарними та шумними. Значний інтерес полягає в тому, чи можна робити прогнози щодо майбутніх значень фінансових часових рядів, не в останню чергу через винагороду, яку принесе така можливість.

Іншою особливістю фінансових часових рядів є те, що в багатьох випадках вони є багатовимірними, так що, наприклад, замість скалярного часового ряду цін на валюту можна мати десятивимірний векторний часовий ряд, що представляє ціни угод, які ще не виконано. кожен екземпляр часового ряду.

Нестационарний, шумний, але потенційно багатовимірний характер фінансових часових рядів робить їх придатною мішенню для різних прогнозних методів або моделей. Однак усі ці методи мають свої недоліки, і жоден із них не може дати абсолютно точний прогноз. Більше того, багато методів, особливо щодо нових підходів, розглядаються лише у зарубіжних дослідженнях. Таким чином, прогнозування динаміки цін на ф'ючерсні контракти ще мало вивчено.

Проблема прогнозування цін є дуже складною через високу волатильність ринку та безліч факторів, що впливають на зміну вартості активів. Через множинність і непередбачуваність цих факторів поведінка ф'ючерсних цін набуває характеру випадкових коливань. У цьому контексті

котирування ф'ючерсів можна розглядати як часові ряди, а моделі можна використовувати для аналізу часових рядів для прогнозування їх майбутньої динаміки.

Актуальність роботи. Актуальність даної роботи полягає в необхідності максимально точного прогнозування динаміки цін на ф'ючерси, через невідомість динаміки зовнішніх факторів, що впливають на оцінку цих ф'ючерсів, а також у необхідності адаптації нових статистичних підходів до прогнозів.

Зв'язок роботи з науковими програмами, планами, темами. Магістерська дисертація на здобуття ступеня магістра виконувалась в Національному технічному університеті України «Київський політехнічний інститут імені Ігоря Сікорського» відповідно до планів наукових досліджень кафедри економічної кібернетики за темою «Глобалізація напрямів формування промислового потенціалу в умовах постіндустріальних трансформацій» (№ ДР 0112U007817). Роль автора полягає в дослідженні прогнозних методів і моделей.

Мета та завдання роботи. Метою даної роботи є дослідження теоретичних і практичних аспектів динаміки ф'ючерсних контрактів на фінансових ринках.

Об'єкт дослідження. Об'єктом дослідження є фінансовий ринок.

Предмет дослідження. Предметом дослідження є ф'ючерсний контракт і ф'ючерс на сиру нафту.

Теоретична основа дослідження. Розробкою та дослідженням аспектів створення прогнозів на фінансових ринках займалися такі економісти та математики, як: Г. Бокс, Г. Дженкінс, М. Броді, П. Глассерман, Ф.Блек, М.Шоулз, П. Самуельсон, М. Трейнор, Ш.Кассуф, Дж.Бонес, Е.Торп, Фама, Френч.

Методи дослідження. У дослідженні використовувалися такі методи дослідження, як регресійний і кореляційний аналіз, статистичні методи,

методи прогнозування часових рядів, використання інформаційних технологій і методів машинного навчання.

Наукова новизна. Була побудовані моделі для прогнозування прибутковості ф'ючерсів, які базуються на методі геометричного броунівського руху, ARIMA та Random Forest. Була визначена оптимальна модель для прогнозування.

Практична значущість. Завдяки отриманим результатам можна зробити висновки і порівняти результати від прогнозування за допомогою моделей. В майбутньому можна застосовувати модель з кращим результатом наприклад при моделюванні інвестиційного портфелю чи при дослідженні коливань цін.

1 ТЕОРІЯ Ф'ЮЧЕРСНОГО РИНКУ

1.1 Історія та еволюція ф'ючерсних бірж

Витоки використання угод щодо постачання певного товару в якийсь майбутній час за узгодженою ціною сягають стародавніх грецьких і римських ринків. Крім центрів політичного і культурного життя, Агора в Афінах і Форум в Римі були торговими центрами, де товари, привезені з далеких країн, обмінювалися на валюту або дорогоцінні метали, такі як золото і срібло. Торговці, торговці, виробники та покупці стародавнього світу стикалися з багатьма тими ж вимогами, що й сучасні торговці, комерсанти, виробники та споживачі, зокрема потреба в управлінні цінами та ризиками доставки, а також потреба в своєчасній та прозорій ринковій інформації та цінах. Хоча сьогоднішня інформація, доступна в електронному вигляді, не відповідає своєчасності, ці торговці минулих років встановлювали та отримували доступ до інформації та цін через неформальні мережі, які вони створили. Однак першою формалізованою ф'ючерсною біржею зазвичай вважають рисову біржу Додзіма в Осаці, Японія, на початку 1700-х років [1].

Протягом багатьох століть товарні ринки процвітали скрізь, де виробники та споживачі торгували товарами в обмін на готівку. На якомусь етапі між контрагентами почали укладатися угоди про доставку товару в обумовлений час у майбутньому за ціною, узгодженою в даний момент. Оплата буде здійснена після отримання обумовленої кількості товару зазначеної якості. Ці угоди, які стали відомі як форвардні контракти, означали, що виробник, як правило, фермер, міг посадити свій урожай, знаючи, що йому гарантується віддача від витраченого часу та інвестицій, а покупець був впевнений в тому, що отримає від виробника договірний товар.

Спочатку ці форвардні контракти часто не виконувалися тією чи іншою стороною, якщо ринкові умови, а саме ціна, суттєво змінилися з моменту укладення контракту. Якщо ціна різко зросла через, наприклад, несприятливі

кліматичні умови, продавець може відмовитися від угоди: нездатність доставити договірну кількість або якість через наслідки посухи чи повені, навіть якщо вони, можливо, не зазнали жодних негативних наслідків для свого власного виробництва, але тепер змогли продати за вищою ціною іншому покупцю. Подібним чином, у роки перевиробництва, коли ціни падали, покупець міг відступити, оскільки ціни на врожай зараз були набагато нижчими на відкритому ринку [2]. Інші проблеми з цими ранніми форвардними контрактами включали брак ліквідності, проблеми з якістю тощо, практичні труднощі покупців і продавців у пошуку один одного, а потім укладання контракту, індивідуальний характер кожного форвардного контракту та відсутність будь-якого формального способу забезпечення виконання цих контрактів.

Форвардні контракти можуть наражати обидві сторони на декілька ризиків, включаючи кредитний ризик, якщо будь-яка зі сторін не виконує угоду або не має достатньо коштів для виконання угоди; ризик переговорів, якщо будь-яка сторона діє недобросовісно; і ризик того, що майбутні події, такі як несприятлива погода, політичні чи економічні події, можуть перешкодити одній або обом сторонам виконати свої зобов'язання. Очевидно, потрібна була більш формалізована та безпечна комерційна домовленість.

У 1840-х роках Чикаго став ринковим центром для зернових фермерів у Сполучених Штатах. Щороку під час жнив виробники приходили продавати зерно. У рясні роки було б надто мало покупців, ціни були б низькими, і за відсутності можливості зберігати надлишок продукції значна її частина була б викинута на демпінг. Взимку ціни зростали, оскільки пропозиція падала. Між покупцями та продавцями укладалися форвардні контракти, укладені в приватному порядку. У 1848 році група торговців зерном об'єдналася, щоб створити організовану біржу зерна, де покупці та продавці могли зустрічатися та вести справи в одному місці. Була створена Чиказька торгова біржа (CBOT), і 13 березня 1851 року було укладено перший офіційний форвардний контракт.

У 1865 році СВОТ представив стандартизовані ф'ючерсні контракти на кукурудзу, пшеницю та овес. Ці контракти вказували кількість і якість товару, що продається, і дату поставки, але не ціну. Вони також були взаємозамінними або «взаємозамінними», що означало, що їх можна було обмінювати між покупцями та продавцями багато разів до визначеної дати доставки, оскільки ціни коливалися. З наявністю офіційної біржі можливість доступу до продукції протягом усього року стала пріоритетом, а силоси для зберігання були побудовані для обробки надлишків, які виникали щороку під час збору врожаю. Це сховище також допомогло згладити різкі коливання ціни, які мали місце раніше [3].

У 1898 році, натхненна тим, що сталося на зернових ринках, група торговців сформувала Чиказьку раду з виробництва масла та яєць, щоб дозволити купівлю та продаж ф'ючерсних контрактів на інші сільськогосподарські продукти, включаючи яйця, масло, шкіру, цибулю та картоплю. У 1919 році її назву було змінено на Чиказьку товарну біржу (СМЕ). З роками було додано більше контрактів, починаючи з заморожених свинячих шлунків у 1961 році, потім фінансові та валютні ф'ючерси в 1970-х роках, процентні ставки в 1981 році та ф'ючерси та опціони на індекс акцій у 1980-х роках. Серед них перший ф'ючерсний контракт на євродолар з розрахунком готівкою в 1992 році та перший ф'ючерсний контракт на індекс акцій для індексу S&P500.

Зростання фінансових ф'ючерсів збіглося з руйнуванням наприкінці 1970-х років Бреттон-Вудського механізму обмінного курсу, згідно з яким світові валютні курси були фактично прив'язані до вартості долара США, яка сама була прив'язана до вартості золота через золотий стандарт. Після того, як ці прив'язки були усунені, розвинувся вільний ринок для виявлення значень обмінних курсів і дозволу хеджерам і спекулянтам брати участь у цьому процесі відкриття цін. Вільно плаваючі обмінні курси наражали виробників і споживачів на коливання курсів валют, а ф'ючерсні ринки забезпечували цінний механізм хеджування. Спекулянти були залучені до цих контрактів,

оскільки обсяги та ліквідність зростали. Фінансові ф'ючерси, надані для впливу на процентні продукти як для хеджування, так і для спекулятивних цілей [4]:

У 1850-х роках форвардні контракти на бавовну також продавалися в Нью-Йорку. У 1870 році на Нью-йоркській бавовняній біржі розпочалася організована торгівля стандартизованими ф'ючерсними контрактами з правилами та процедурами, які формалізували процес, оголошений у 1872 році. Того ж року в Нью-Йорку було створено Масло-сирну біржу (пізніше New York Mercantile). Біржа, або NYMEX), яку придбала CME Group у 2008 році. У 1882 році формалізована ф'ючерсна торгівля почалася на бавовняній біржі New Orleans.

Протягом 19-го сторіччя металеві товари та кава торгувалися в Англії на неофіційній основі на грандіозній «Королівській біржі». Найчастіше ці «торги» проводилися в одній із багатьох кав'ярень, де трейдери сиділи в фіксованих точках навколо кола, накресленого крейдою на підлозі, що породило термін «кільцева торгівля», еквівалент англійських трейдерів торговій підлозі або пит трейдинг в США. У 1877 році група англійських купців заснувала Лондонську металургійну компанію, яка торгувала оловом, міддю та чавуном.

Пізніше назва була змінена на Лондонську біржу металів, яка сьогодні залишається провідною ф'ючерсною біржею для низки контрактів на метали.

Після скасування валютного контролю та успішного запуску CME валютних ф'ючерсів на своїй братській біржі Міжнародному валютному ринку (IMM). у 1982 році була створена Лондонська міжнародна біржа фінансових ф'ючерсів (LIFFE). У 1996 році LIFFE об'єдналася з Лондонською товарною біржею, а в 2002 році була придбана біржовою групою Euronext (яка була створена шляхом злиття Амстердамської фондової біржі, Брюссельської фондової біржі та Паризької біржі). У 2007 році Euronext об'єдналася з Нью-Йоркською фондовою біржею, утворивши NYSE Euronext, яка включає LIFFE, хоча остання продовжує працювати під власним управлінням.

Історія ф'ючерсної торгівлі не обійшлася без суперечок, і донині є ті, хто стверджує, що це не більше ніж прославлена азартна гра. Здебільшого ці напади походять від необізнаних і неосвічених людей, тих, хто просто не розуміє, або, найгучніше, тих, хто або пробував і зазнав невдачі, або ніколи не намагався будь-якої форми торгівлі, але хто засуджує це з певних вищих моральних підстав. З цього приводу мало що змінилося за понад 150 років, коли торгівля ф'ючерсами була легальною та легітимною формою як хеджування для захисту прибутку, так і спекуляції для отримання прибутку. Як і сьогодні, багато хто в 1800-х і на початку 1900-х років були водночас зачаровані та жахані поняттям торгівлі ф'ючерсами. Навколо його легітимності точилися нескінченні судові процеси та публічні дебати. Різниця між відвертими азартними іграми та спекуляціями ф'ючерсними контрактами була (і досі залишається) втраченою для значної частини широкої громадськості. Хоча азартні ігри – це, по суті, азартна гра, у якій «удача» відіграє значну роль у результаті, а шанси зіставляються проти «гравця», торгівля ф'ючерсами, як і всі інші торговельні заходи, вимагає освіти, детального плану та суворого правила торгівлі та управління капіталом. Спираючись на розуміння цих концепцій і тверде розуміння ймовірності, торгівля будь-якого типу є законним бізнесом.

Протягом 19 століття було багато спроб обмежити або заборонити ф'ючерсну торгівлю в Сполучених Штатах. Наведу кілька прикладів: у 1812 році короткі продажі стали незаконними в Нью-Йорку (це законодавство було скасовано в 1858 році); у 1841 році неспроможність покрити короткі продажі протягом п'яти днів вважалася проступком згідно із законом Пенсільванії (скасованим у 1862 році); у 1879 р. Конституція Каліфорнії визнала ф'ючерсні контракти недійсними (скасовано в 1908 р.); а в 1882 році Огайо та Іллінойс безуспішно намагалися обмежити готівкові розрахунки за ф'ючерсними контрактами.

Окрім постійних юридичних нападів на ф'ючерсну торгівлю, дві інші проблеми наприкінці 1800-х років серйозно загрожували новонародженим

ф'ючерсним ринкам. Першим був рух проти опціонів, який прагнув заборонити торгівлю як ф'ючерсами, так і опціонами. Кульмінацією цієї кампанії стало прийняття обома палатами Конгресу США двох законопроектів проти опціонів. Якби будь-який із законопроектів був прийнятий законом, це означало б кінець торгівлі ф'ючерсами в США, але обидва зрештою провалилися через юридичні нюанси.

Другою проблемою була суперечка навколо «відерних магазинів». По суті, це були групи гравців, які збиралися в приміщеннях за межами визнаних бірж і робили ставки на напрямок руху цін. Хоча вони покладалися на дані про ціни з біржі, вони не торгували на офіційній біржі. Епоха «bucket shop» була яскравою главою в історії ф'ючерсів, яка завершилася в 1915 році після рішення Верховного суду США про те, що біржі, і зокрема СВОТ, можуть обмежувати розповсюдження інформації про ціни третім особам.

З розширенням використання Інтернету на початку 1990-х років біржі почали запроваджувати електронну торгівлю деякими ф'ючерсними контрактами у позабіржовий час, і перехід від торгівлі на базі до електронної торгівлі через комп'ютерні термінали почав набирати обертів. У 1992 році CME запустила електронну торгову платформу CME Globex. Відтоді електронна торгівля розширилася й охопила паралельну електронну торгівлю, коли ринки торгуються одночасно в ямі та на електронних платформах, а нещодавно - контракти, які торгуються виключно на електронних ринках, таких як E-mini S&P. Електронна торгівля дозволяє ринкам торгувати протягом більш тривалого часу, ніж це можливо у відкритій торговій ямі. E-mini S&P, наприклад, торгує протягом 23 годин і 15 хвилин кожного робочого дня. Електронна торгівля не тільки дозволяє збільшити час торгівлі, але й забезпечує одночасну роботу ринків у всьому світі, надаючи переваги всім учасникам ринку.

У 1998 році німецька біржа деривативів Deutsche Terminbourse (DTB) і Швейцарська біржа опціонів і фінансових ф'ючерсів (SOFFEX) об'єдналися, щоб створити біржу Eurex. Eurex була однією з перших бірж, яка пропонувала

торгівлю лише через свою електронну торгову платформу (на Eurex взагалі немає біржової торгівлі).

Сьогодні CME Group є найбільшою ф'ючерсною біржею в США з річним обсягом торгів, що перевищує 3 мільярди контрактів на суму понад 10 трильйонів доларів. Це одна з трьох найбільших ф'ючерсних бірж у світі разом із європейською біржею Eurex із річним обсягом торгів понад 2 мільярди контрактів та біржею NYSE Euronext (включаючи LIFFE) із річним обсягом торгів понад 1 мільярдів контрактів.

Зараз у світі існує понад 100 бірж, на яких торгують ф'ючерсними контрактами, починаючи від гігантів, таких як CME Group, і закінчуючи набагато скромнішими підприємствами, такими як Mercantile Exchange of Madagascar та біржа Sibex у Румунії.

Сучасні ф'ючерсні ринки далеко переросли своє сільськогосподарське походження, і зараз на фінансові контракти припадає понад 80 відсотків щоденного обсягу ф'ючерсних торгів. Більшість ф'ючерсних контрактів зараз торгуються в електронному вигляді через онлайн-платформи, які пропонують справді глобальний ринок. Наприклад, більше 83 відсотків усіх угод на CME Group і всіх угод на Eurex здійснюються в електронному вигляді.

Як обговорювалося, ранні форвардні контракти були вразливі до дефолту. Прямі угоди між брокерами були відносно простими, коли покупки та продажі були збалансовані, що дозволяло брокерам розраховуватися готівкою за різницю між обсягами контрактів. Проблеми почалися, коли, як це найчастіше траплялося, кількість не збалансувалася. Якщо, для

Наприклад, брокер А придбав контракт на пшеницю на 5000 бушелів у брокера В, а пізніше брокер В купив контракт на пшеницю на 6000 бушелів у брокера А, обидва контракти можуть бути зведені на дату розрахунку, якщо брокер В просто продасть контракт на 1000 бушелів пшениці для брокера А. Проблеми виникли, якби брокер В також продав контракт на 1000 бушелів брокеру С, який продав контракт на 1000 бушелів брокеру А. У цьому випадку чиста позиція кожного брокера була компенсована, але всі троє повинні були

зустрітися, щоб погасити свої зобов'язання один перед одним. Це було названо «кільцевим поселенням». «Трансферні розрахунки», за яких брокери передавали свої взаємні зобов'язання іншим брокерам, щоб збалансувати свої зобов'язання за форвардними контрактами, також були поширеними.

Сучасні ф'ючерсні контракти розвинулися зі стандартизованих форвардних контрактів. Хоча вони зберігають основні риси попередніх контрактів, вони є набагато більш конкретними та охоплюють надзвичайний діапазон ринків і пов'язаних продуктів. Ф'ючерсні контракти торгуються на зареєстрованих біржах, де між контрагентами кожного ф'ючерсного контракту діє розрахункова палата. Будь-який контракт є взаємозамінним і може здійснюватися або з поставкою, або з оплатою готівкою. Єдиною змінною в сучасному ф'ючерсному контракті є ціна. Саме постійний процес визначення ціни призводить до змін ринкових цін на величезну кількість ф'ючерсних контрактів на біржах у всьому світі, і саме це робить ф'ючерсну торгівлю динамічною глобальною індустрією, якою вона є сьогодні. У наступному розділі ми більш детально розглянемо специфіку ф'ючерсних контрактів і роль розрахункової палати [5, 6, 7].

Витоки використання угод щодо доставки певного товару в якийсь майбутній час за узгодженою ціною сягають стародавніх грецьких і римських ринків.

Форвардні контракти часто не виконувалися тією чи іншою стороною, якщо ринкові умови, а саме ціна, суттєво змінилися з моменту укладення контракту.

У 1865 році Чиказька торгова біржа запровадила стандартизовані ф'ючерсні контракти на кукурудзу, пшеницю та овес. Ці контракти вказували кількість і якість товару, що продається, і дату поставки, але не ціну. За ними йшов ряд сільськогосподарських продуктів, потім інші природні товари.

Зростання фінансових ф'ючерсних ринків можна пов'язати з розривом Бреттон-Вудської угоди, плаванням основних світових валют і зростанням вільного ринку, що дозволило вільно торгувати фінансовими інструментами.

Історія ф'ючерсної торгівлі не обійшлася без суперечок, і донині є ті, хто стверджує, що це не більше ніж прославлена азартна гра. Здебільшого ці атаки походять від тих, хто просто не розуміється на ф'ючерській торгівлі.

Спочатку всі ф'ючерсні контракти торгувалися відкритим протестом у торговій «ямі», де брокери розробили унікальну систему жестів руками для полегшення виконання торгів.

У міру розширення використання Інтернету на початку 1990-х років біржі почали запроваджувати електронну торгівлю ф'ючерсними контрактами у позабіржовий час, і почався перехід від торгівлі на базі до електронної торгівлі через комп'ютерні термінали.

Сучасні ф'ючерсні ринки далеко переросли своє сільськогосподарське походження, і зараз на фінансові контракти припадає понад 80 відсотків щоденного обсягу ф'ючерсних торгів. Більшість ф'ючерсних контрактів зараз торгуються в електронному вигляді через онлайн-платформи, які пропонують справді глобальний ринок.

Ф'ючерсні контракти торгуються на зареєстрованих біржах, де між контрагентами кожного ф'ючерсного контракту діє розрахункова палата. Будь-який контракт є взаємозамінним і може здійснюватися або з поставкою, або з оплатою готівкою. Єдиною змінною в сучасному ф'ючерсному контракті є ціна.

1.2 Поняття ф'ючерсного контракту

Ф'ючерсний контракт – це юридично зобов'язуюча стандартизована угода про купівлю чи продаж товару чи фінансового інструменту певної якості та кількості на визначену майбутню дату поставки у певному місці. Єдиною змінною є ціна, узгоджена під час угоди. Деякі ф'ючерсні контракти, такі як пшениця та жива худоба, вимагають фізичної доставки, а інші, такі як євродолари та S&P 500, розраховуються готівкою [8].

Усі ф'ючерсні контракти базуються на базовому фізичному інструменті – фізичному товарі, такому як пшениця чи золото, або фінансовому інструменті, такому як індекс капіталу, валюта чи процентна ставка. Ф'ючерсний контракт – це угода між двома сторонами про купівлю або продаж стандартизованої кількості та якості базового активу на визначену майбутню дату, але за ціною, узгодженою сьогодні. Контракти укладаються на зареєстрованій біржі, яка діє як посередник між двома сторонами. Покупця ф'ючерсного контракту називають «довгим», що відображає його думку про зростання цін. Продавця ф'ючерсного контракту називають «коротким», що відображає його думку про зниження цін. Ф'ючерсні контракти можуть розраховуватися готівкою або передбачати фізичну поставку основного товару чи інструменту.

Важливим фактором для будь-якого товару чи інструменту, який розглядається у ф'ючерсному контракті, є взаємозамінність. Будь-який товар або інструмент, який можна замінити, вважатиметься взаємозамінним. Наприклад, покупець контракту на кукурудзу на 5000 бушелів не зацікавлений у покупці 5000 бушелів кукурудзи з певного поля на певній фермі в Огайо; Йому просто потрібно бути впевненим, що кукурудза, яку він купує, відповідає контракту, незалежно від її точного походження. Так само покупцеві вартістю 100 000 канадських доларів не потрібно надавати точні купюри та номінали, необхідні під час доставки; доки він є фактичним законним платіжним засобом у розумінні контракту, договірні зобов'язання продавця виконуються. Обидва базові інструменти в цих двох прикладах є обмінними або взаємозамінними, оскільки кожні 5 000 бушелів кукурудзи або кожні 100 000 канадських доларів, які відповідають специфікації контракту, можуть бути використані для розрахунків згідно з умовами контракту [9].

Оскільки ф'ючерсні контракти взаємозамінні, багато трейдерів, які ніколи не збираються виробляти або доставляти основний товар, можуть купувати та продавати ф'ючерсні контракти. Трейдери та спекулянти купують і продають ф'ючерсні контракти, щоб скористатися коливаннями цін у

мінливих ринкових умовах. Вони отримують вигоду від іншої ключової функції ф'ючерсної торгівлі – можливості ліквідувати відкриту позицію шляхом купівлі або продажу протилежної сторони існуючої відкритої позиції до закінчення терміну дії, процес, відомий як «взаємозалік». Кожен ф'ючерсний контракт має дату останньої операції, до якої всі відкриті позиції мають бути закриті.

Ф'ючерсні контракти (як от Wheat і Lean Hogs), ф'ючерсні контракти з розрахунком готівкою (як от євродолари або E-mini S&P), відкриті позиції закриваються шляхом офсетної торгівлі. Наприклад, якщо ви продаєте ф'ючерсний контракт, ви просто продаєте ф'ючерсний контракт і користуєтеся різницею між сплаченою ціною та ціною запиту або зазнаєте збитків, якщо ціна змінилася проти вашої відкритої позиції. Таким чином ви «збалансували» свою відкриту позицію.

Завдяки взаємозаліку трейдери та спекулянти, які абсолютно не зацікавлені в отриманні базового активу, можуть активно брати участь у ф'ючерсній торгівлі, знаючи, що до дати поставки вони можуть просто закрити будь-які існуючі позиції та або зберегти прибуток, або збитки, які вони можуть прийняти від обмін .

Ф'ючерсні контракти можна розділити на ті, які вимагають фізичної поставки конкретного товару або фінансового інструменту, і ті, які розраховуються готівкою. Ви можете прийняти або виконати поставки за контрактами на фізичну доставку або закрити свою позицію шляхом остаточного коригування готівки за ринковою вартістю для розрахунку готівкою

Контракти, які кредитують ваш рахунок, коли ви отримуєте прибуток, або дебетують ваш рахунок, коли ви отримуєте збитки. Менше 5 відсотків усіх ф'ючерсних контрактів фактично закінчуються фізичною доставкою. Переважна більшість учасників ринку вважають за краще закрити або ліквідувати свої відкриті позиції до закінчення терміну дії, оскільки дуже мало спекулянтів готові взяти або відправити, скажімо, 112 000 фунтів цукру або 40

000 фунтів свинячих туш. Звичайно, деякі виробники та кінцеві користувачі вибирають доставку.

Незважаючи на те, що небагато ф'ючерсних контрактів призводять до фактичної фізичної доставки, забезпечення фізичної доставки відіграє дві важливі ролі на ф'ючерсних ринках [9].

По-перше, це дозволяє покупцям і продавцям брати або доставляти товар або базовий актив, якщо вони того вимагають. Таким чином, наприклад, масовий споживач кукурудзи може використовувати ф'ючерсний ринок для накопичення кукурудзи для фізичної доставки, якщо він вірить, що ціна на кукурудзу зросте до того, як йому справді знадобиться зерно для його виробничих потреб. Подібним чином фермер, який продав ф'ючерси на кукурудзу, яку він справді виробляє, може вибрати вищу ціну, ніж на переважаючому ринку. По-друге, той факт, що його можна доставити або отримати, гарантує, що ф'ючерсні ціни точно відображають грошову ціну, сплачену за товар або базовий товар на момент закінчення терміну дії контракту. З наближенням терміну дії контракту обидві ціни будуть зближуватися. Ф'ючерсні контракти з розрахунком готівкою не мають контракту на поставку, але розраховуються готівкою, коли термін дії договору закінчується. Наприклад, ф'ючерси на фондові індекси розраховуються готівкою, оскільки було б недоцільно організувати доставку акцій, які складають індекс. Коли термін дії контракту закінчується, з рахунку трейдера просто списується або кредитується різниця між вартістю базового активу та ціною ф'ючерсного контракту. У випадку контрактів, розрахунок яких здійснюється готівкою, зближення цін відбувається автоматично [10].

1.3 Процес визначення ціни

Ціни ф'ючерсів визначаються шляхом конкурентних торгів або аукціону. Спочатку це визначалося за допомогою методу «відкритого протесту» на біржових торгових майданчиках або в ямах, де брокери та їхні

агенти зустрічалися для проведення аукціону. З кінця 1970-х років процес також здійснювався через електронні торгові платформи, коли ціни були доступні в електронному вигляді негайно та постійно по всьому світу. Низка контрактів (таких як асортимент продуктів e-mini) і бірж (таких як Eurex) тепер здійснюють лише електронну торгівлю. Трейдери можуть доручити своєму брокеру виконувати замовлення від їх імені або, як це все частіше трапляється, вони можуть виконувати власні замовлення на покупку чи продаж через надані брокером електронні платформи на своїх власних комп'ютерах з будь-якого місця, де є доступ до Інтернету. На рис. 1.1 показано типовий екран угод, доступний роздрібним клієнтам на їхньому персональному комп'ютері.

На цьому екрані ми бачимо електронний аукціон ф'ючерного контракту E-mini S&P 500 у дії. Остання котирована ціна була 1408,75. За цією ціною на аукціоні виставлено 12 контрактів і виставлено на продаж 65 контрактів по 1409,00. Щоб відбулася нова транзакція, продавець повинен знизити ціну, щоб відповідати ціні пропозиції 1408,75, або покупець повинен скорегувати свою ціну, щоб відповідати ціні пропозиції або пропозиції 1409,00.

Ціна товару «відкривається» на основі змін попиту та пропозиції, і ф'ючерсні ціни не є винятком. З боку пропозиції, продавці, швидше за все, продадуть, коли ціни високі, тоді як вони менш імовірно продадуть, коли ціни нижчі. З боку попиту все навпаки: покупці готові купувати більше за нижчими цінами і менше за вищими цінами. Коли криві попиту та пропозиції для будь-якого продукту чи фінансового інструменту розміщені на одному графіку, точкою, де вони перетинаються, є ринкова або рівноважна ціна та ціна, за якою здійснюється транзакція [11]. Це показано на рис. 1.2. У цьому прикладі ціна спочатку знаходиться в рівновазі з ціною P_1 і кількістю Q_1 , де крива попиту D_1 перетинає криву пропозиції S_1 . Зменшення пропозиції до Q_2 спричиняє підвищення ціни до нової, вищої рівноваги на рівні P_2 .



Рисунок 1.1 – Типовий екран угоди
Джерело. [25]

Криві попиту та пропозиції, а отже, і рівноважні ціни, постійно змінюються, коли на ринки надходить нова інформація. Учасники ф'ючерсного ринку реагують на ці зміни, досягається нова рівноважна ціна і відбуваються нові угоди. Ця рівноважна ціна може змінюватися щохвилини або навіть частіше на швидкозмінних, нестабільних ринках.

Кожен ф'ючерсний контракт має мінімальну суму, яку він може підвищити або знизити. Ці мінімальні зміни ціни встановлюються біржею та відрізняються залежно від інструменту. Рухи низьких цін називаються «тіками». Наприклад, у ф'ючерсному контракті на золото один тік становить 10 центів за унцію, тому один тік у контракті на 100 унцій коштує 10 доларів (10 центів за унцію $\times$ 100 унцій). Для порівняння, один тик ф'ючерсного контракту на зерно коштує 0,0025 доларів (1/4,1 цента) за бушель. Для стандартного ф'ючерсного контракту на пшеницю на 5000 бушелів це 12,50

доларів США за тик ($5000 \text{ бушелів} \times 0,0025 \text{ доларів США за бушель}$). Щоб знати, як зміна ціни вплине на вартість контракту, важливо знати зміни тиків для кожного ф'ючерсного контракту, яким ви торгуєте.

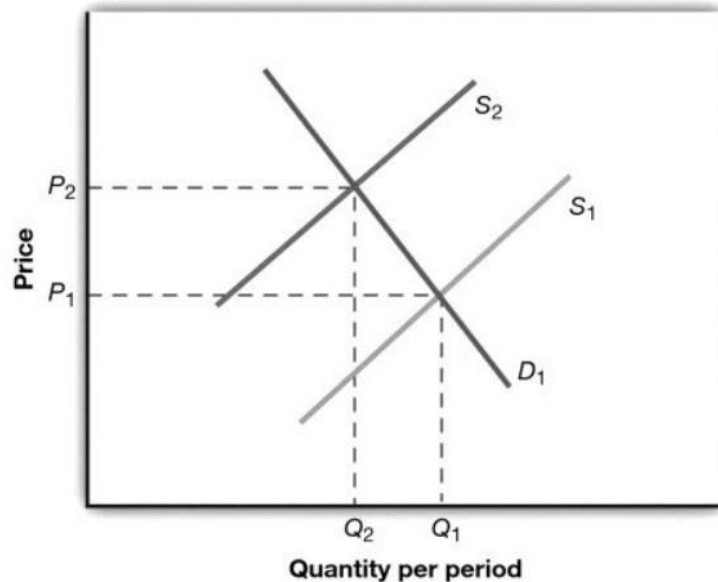


Рисунок 1.2 – Попит і пропозиція та цінова рівновага

Джерело. [25]

Товарні біржі також встановлюють максимальні денні ліміти цін на ф'ючерсні контракти. Ці обмеження ціни котируються відносно ціни закриття попереднього дня плюс або мінус визначена сума в центах або доларах за контрактну одиницю. Коли ціна ф'ючерсного контракту зростає або знижується на свій денний ліміт, подальша торгівля не може відбуватися до наступної торгової сесії; це відоме як «ліміт блокування». Наприклад, якщо денний ліміт для ф'ючерсів на худобу становить 3 центи за фунт для ф'ючерсного контракту на 40 000 фунтів стерлінгів, а розрахункова ціна попереднього дня становила 92 центи за фунт, тоді протягом поточної торгової сесії не можна торгувати ціною вище 95 центів за фунт. або менше 89 центів за фунт. Ціна може зростати або знижуватися щодня в межах обмеження. Ці щоденні цінові обмеження можуть призвести до днів, коли вихід із існуючої ф'ючерсної позиції буде неможливим, оскільки

Ринок заблокований. Це може бути неприємним і дорогим, коли ринок рухається проти вас, або надзвичайно корисним, коли він наближається до вашої відкритої позиції. У наведеному вище прикладі з живою худобою щоденний ліміт у 3 центи за фунт був еквівалентним 1200 доларів США за ф'ючерсний контракт (3 центи за фунт $\times$ 40 000 фунтів стерлінгів).

Ці денні ліміти цін скасовуються для деяких контрактів у місяць закінчення контракту (також відомий як місяць доставки або спотовий місяць), що дозволяє торгувати цінами відповідно до попиту та пропозиції замість того, щоб їх штучно обмежувати денними лімітами цін. Хоча біржі встановлюють щоденні ліміти, щоб привести ринок у порядок, вони неефективні та можуть розчаровувати спекулянтів. Було б набагато краще не застосовувати обмеження до цих ринків і дозволити ціні природним чином знайти свій власний рівень. Професійне використання трейдерами стоп-лосс ордерів забезпечило б набагато ефективнішу ліквідацію позицій [12].

Біржі та регулятори також встановлюють обмеження на максимальну спекулятивну позицію, яку будь-яка особа може мати в будь-який момент часу в будь-якому окремому ф'ючерсному контракті. Обмеження максимальної позиції встановлюються, щоб запобігти можливості будь-якого покупця чи продавця впливати на ринкову ціну шляхом купівлі та продажу. Ці обмеження позиції вказуються у загальній кількості контрактів або загальній кількості одиниць інструменту чи товару [13].

1.4 Ф'ючерсні трейдери

Учасники ф'ючерсних ринків варіюються від дрібних роздрібних продавців, які торгують окремими контрактами, до гігантських хедж-фондів, які торгують великою кількістю контрактів. Ф'ючерсні трейдери охоплюють дуже широкий діапазон, але кожен має власні причини для участі в цьому глобальному ринку. У цьому розділі ми визначаємо цих учасників ринку та досліджуємо деякі причини, чому вони торгують ф'ючерсами. Підходи на

ф'ючерсному ринку можна грубо розділити на два типи - хеджування або спекуляція. У той час як деякі займаються як хеджуванням, так і спекуляціями, їхні причини для активності в кожній сфері часто зовсім різні.

Однією з основних економічних функцій ф'ючерсних ринків є управління ціновим ризиком. Хеджування – це практика компенсації цінового ризику на спотовому ринку шляхом зайняття рівних і протилежних позицій на ф'ючерсному ринку. Хеджування базується на принципі, згідно з яким спотова ринкова ціна товару чи фінансового інструменту та ринкова ціна ф'ючерсного контракту на той самий продукт мають тенденцію зростати та знижуватися більш-менш одночасно. Займаючи протилежні позиції, хеджування дає змогу компенсувати втрати на одному ринку прибутками на іншому. Таким чином, хеджування може зафіксувати відомий рівень ціни для спотових угод на тиждень або місяць наперед.

Діапазон можливостей безпеки майже безмежний. Фермер, який вирощує пшеницю, може застрахуватися від падіння цін на пшеницю, тоді як виробник печива може застрахуватися від підвищення цін на пшеницю. Позикодавець може застрахуватися від можливого зниження процентних ставок, тоді як позичальник може застрахуватися від підвищення процентних ставок. Наведені нижче приклади безпеки показують, як стратегія використовується в різних сценаріях [14].

Ф'ючерсні ринки полегшують передачу ризику від хеджерів до спекулянтів. Спекулянти беруть на себе ризики, яких хеджери хочуть уникнути або принаймні мінімізувати, прагнучи отримати прибуток від очікуваного підвищення чи зниження ф'ючерсних цін. Бажаючи зайняти й утримати протилежну сторону позиції хеджування, спекулянти додають капітал на ринок, що допомагає обмежити вплив хеджера на коливання ціні ціновий ризик. Без спекулянтів та їхньої схильності до ризику хеджерам було б важко реалізувати свої стратегії, оскільки короткі хеджери могли торгувати лише з довгими хеджерами. Спекулянти додають ф'ючерсним ринкам ліквідність. Ліквідність означає кількість незавершених контрактів або

одиниць в будь-який момент часу. Чим більш ліквідний ф'ючерсний контракт, тим легше здійснювати великі операції без значної зміни ціни. Якби ринки були обмежені тими, хто займається хеджуванням, ліквідність була б набагато меншою на всіх ринках. Діяльність спекулянтів сприяє безпрограшному характеру ф'ючерсних ринків; Хеджери отримують вигоду від збільшення ліквідності, а спекулянти отримують вигоду від можливостей брати участь у ф'ючерсних ринках, щоб отримати прибуток від своєї торгової та інвестиційної діяльності.

Спекулянти використовують різні методи аналізу, методи та стратегії для участі на ринках. Вони рідко беруться за отримання або доставку основного товару чи інструменту, що також сприяє ліквідності ринків. Балансуючи довгі позиції шляхом продажу та короткі позиції шляхом зворотного викупу, спекулянти можуть входити на ринки та виходити з них швидко та легко, надаючи хеджерам більшу ліквідність.

Спекулянти приносять додаткову та різну інформацію на ф'ючерсні ринки, тим самим посилюючи роль ціноутворення цих ринків; і оскільки вони мають багато різних причин купувати та продавати за різними цінами, вони сприяють процесу визначення ціни.

1.5 Ф'ючерсні контракти на нафту

Сучасне суспільство значною мірою залежить від невідновлюваних джерел енергії як основного джерела енергії для приведення в рух автомобілів, опалення наших будинків, забезпечення світла та електрики та електроенергетики. Чотири основні ф'ючерсні контракти в енергетичному секторі – це сира нафта, бензин RBOB, мазут і природний газ. Хоча є й інші, знову ж таки простір не дозволяє нам охопити їх усіх.

Сучасне суспільство значною мірою залежить від невідновлюваних джерел енергії як основного джерела енергії для приведення в рух автомобілів, опалення наших будинків, забезпечення світла та електрики та

електроенергетики. Чотири основні ф'ючерсні контракти в енергетичному секторі – це сира нафта, бензин RBOB, мазут і природний газ. Хоча є й інші, знову ж таки простір не дозволяє нам охопити їх усіх.

Нафта – це природна складна органічна сполука, що утворюється під землею. Земна поверхня зазнавала надзвичайного тиску та тепла протягом мільйонів років у результаті розкладання морських і рослинних організмів. Перша нафтова свердловина, пробурена спеціально для видобутку сирої нафти для комерційних цілей, була зроблена в Польщі в 1853 році після того, як раніше було винайдено процес дистиляції гасу із сирої нафти. Видобуток і виробництво нафти швидко поширилися по всьому світу, в тому числі Сполучені Штати стали провідним виробником в 1960-х рр. Відкриття нафти на Близькому Сході зробило країни «арабської п'ятірки» - Саудівську Аравію, Об'єднані Арабські Емірати, Ірак, Катар і Кувейт швидко перетворилися на найбільші нафтовидобувні країни світу. Сполучені Штати досі є третім за величиною виробником нафти після Саудівської Аравії, а також найбільшим імпортером [15].

Організація країн-експортерів нафти (ОПЕК), картель з 12 основних країн-виробників нафти: Алжиру, Анголи, Еквадору, Ірану, Іраку, Кувейту, Лівії, Нігерії, Катару, Саудівської Аравії, Об'єднаних Арабських Еміратів і Венесуели як основної ціни нафта в 1970-х рр. Члени ОПЕК регулярно зустрічаються для обговорення квот на видобуток для країн-членів, щоб утримувати ціни на сприятливому рівні.

Нафта, яку часто називають «чорним золотом», є, мабуть, найбільш політично та економічно чутливим товаром у світі. Через постійно зростаючий попит на сиру нафту для палива, а також пластмаси, добрива, інсектициди, фармацевтичні препарати та багато інших продуктів, а також обмежену та скорочену пропозицію, ціни на нафту мають вирішальне значення для всіх економік у всьому світі.

Світове споживання нафти становить понад 88 мільйонів барелів на добу. Сполучені Штати, найбільший споживач у світі, споживають понад 19

мільйонів барелів нафти на день, або близько 23 відсотків. Китай, друга за величиною країна-споживач, споживає близько 9 мільйонів барелів на день, або 10 відсотків. Світовий видобуток нафти оцінюється приблизно в 90 мільйонів барелів на день. Очікується, що до 2030 року споживання нафти зросте до понад 107 мільйонів барелів на день [16].

Фактори, що впливають на ціну сирої нафти, включають наступне:

- Міцність економіки США та світової економіки є ключовим фактором зростання цін на сиру нафту. Коли світова економіка розширюється, попит на сиру нафту зростає, і у відповідь зростає ціна. Коли світова економіка скорочується, попит на сиру нафту сповільнюється, а ціна падає або, у кращому випадку, стабілізується.

- Геополітичні події або загроза таких подій, як війни, повстання та політична нестабільність у нафтовидобувних країнах, можуть спричинити раптові стрибки цін на нафту у відповідь на побоювання щодо скорочення пропозиції.

- Заяви та оголошення ОПЕК щодо рівнів видобутку, квот і цінової політики можуть спричинити швидкі зміни ціни [19].

- Ціна на сиру нафту включає складну взаємодію кількох факторів попиту та пропозиції. Якщо попит високий, ціни високі, а складські запаси низькі, країни ОПЕК можуть розглянути можливість збільшення виробництва. Збільшення пропозиції може знизити ціни або може бути просто поглинено ринком, якщо попит залишається високим [17, 18, 20]. Якщо попит низький, запаси відносно великі, а ціни низькі, вони можуть вирішити скоротити виробництво. Це може призвести до підвищення ціни, особливо якщо це сприймається як спроба обмежити пропозицію або відбувається в період політичної невизначеності.

- Загрози довгостроковому постачанню сирої нафти, уявні чи реальні, також можуть спричинити раптові коливання цін. Це часто спровоковано оголошеннями про припинення глобального видобутку сирої нафти, оскільки це невідновлюване джерело енергії врешті-решт вичерпується. Дебати про

«під нафти» зосереджуються на тому, чи глобальні запаси сирої нафти перевищили межу, коли попит перевищує пропозицію, і не відкриваються нові родовища нафти, що призводить до постійного зменшення пропозиції «чорного золота» .

Висновки до розділу 1

1. В першому розділі дані теоретичні засади виникнення, історії та еволюції ф'ючерсного ринку.

2. Дано визначення ф'ючерсний контрактів, як юридично зобов'язуючих стандартизованих угод про купівлю чи продаж товару чи фінансового інструменту певної якості та кількості на визначену майбутню дату поставки у певному місці.

3. Проаналізовано шляхи визначення цін на контракти ф'ючерсів визначаються шляхом конкурентних торгів або аукціону.

4. Визначено учасників ф'ючерсних ринків, які варіюються від дрібних роздрібних продавців, які торгують окремими контрактами, до гігантських хедж-фондів, які торгують великою кількістю контрактів.

5. Розглянуті ф'ючерсні контракти на нафту, як одні з чотирьох основних ф'ючерсні контракти в енергетичному секторі.

2 ОГЛЯД МЕТОДІВ І МОДЕЛЕЙ ПРОГНОЗУВАННЯ

В даний час комп'ютерна обробка даних займає в рази менше часу, ніж на початку розвитку теорій прогнозування цін акцій, що призвело до відносної доступності методів математичного моделювання на вирішення поточних завдань. При прогнозуванні цін акцій почали набирати популярності статистичні методи, у тому числі й тимчасові лави. Основні статистичні методи прогнозування описані далі.

Перед тим, як перейти до методів, слід зазначити визначення ключових термінів.

Метод прогнозування – послідовність дій, необхідні отримання моделі прогнозування. Поняття «метод прогнозування» ширше, ніж «модель прогнозування».

Модель прогнозування – функціональне уявлення, адекватно описує тимчасові ряди, і є основою отримання майбутніх значень процесу [21].

2.1. Моделі класу ARIMA

Моделі класу ARIMA, тобто моделі авторегресійного інтегрованого ковзного середнього або авторегресійного інтегрованого ковзного середнього, є одними з найвідоміших і популярних методів аналізу та прогнозування часових рядів різних структур, різниці d -порядку яких є стаціонарними процесами через середнє значення (інтегрований порядку d). Завдяки новаторському внеску Бокса та Г.М. Дженкінса (1970) у літературі також називають моделями Бокса-Дженкінса. Моделі характеризуються високою гнучкістю та універсальністю, але певним обмеженням їх використання є необхідність мати відносно довгий, щонайменше п'ятдесятиелементний, а на практиці часто значно більший ряд спостережень [22].

У моделях ARIMA поточне значення змінної представлено як лінійна функція її минулих значень і випадкових коливань:

$$Y_t = f[Y_{t-k}, e_{t-k}] + e_t, \quad k > 0, \quad (2.1)$$

де Y_t — значення змінної в момент часу або період t ,

e_t — випадкові коливання в час або період t .

Побудова моделей ARIMA базується на явищі автокореляції в часових рядах, тобто лінійному зв'язку між значеннями даної змінної та значеннями тієї ж змінної з відставанням у часі.

Існує два основних типи залежностей, на основі яких будуються інші більш складні моделі:

– модель авторегресії порядку 1: AR(1), у якій кожне спостереження Y_t корелюється з попереднім спостереженням Y_{t-1} , а отже, меншою мірою, з попередніми спостереженнями:

$$Y_t = \phi_1 Y_{t-1} + e_t. \quad (2.2)$$

– модель ковзного середнього порядку 1: MA(1), у якій кожне значення спостереження Y_t корелюється з випадковим відхиленням (залишком) e_{t-1} попереднього спостереження Y_{t-1} і не корелюється з попередніми спостереженнями (Y_t залежить від випадкового коливання, що відбуваються в період t і $t-1$):

$$Y_t = e_t - \theta_1 e_{t-1}. \quad (2.3)$$

Цих двох базових моделей, AR(1) і MA(1), як правило, недостатньо для опису автокореляційної структури рядів реального часу. Вони є лише основою, на якій будуються складніші моделі ARIMA. Загальні моделі Бокса-Дженкінса також включають можливість введення довготривалих залежностей (тобто тих, у яких відставання більше 1) [23].

Модель авторегресії порядку p - AR(p) можна записати так:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t, \quad (2.4)$$

де ϕ_j — параметр моделі авторегресії біля члена Y_{t-j} ,

e_t — залишок за момент або період t .

Загальна модель ARIMA є комбінацією моделей AR(p) і MA(q), а також дозволяє диференціювати із затримкою 1 або вище, щоб усунути нестационарність даних через середнє значення. Іншими словами, модель ARIMA – це модель із різницевим перетворенням, яка зводить нестационарний ряд відносно середнього до стаціонарної форми. Це означає, що нестационарний процес можна диференціювати в d -кратному порядку, щоб стати стаціонарним, і застосувати модель ARMA до цього нового стаціонарного часового ряду або безпосередньо побудувати модель ARIMA вихідного нестационарного ряду [24].

Усі несезонні моделі класу ARIMA записуються з використанням універсальної нотації ARIMA(p,d,q):

де p – порядок авторегресії AR,

d – ступінь диференціації (інтеграції),

q – кількість параметрів ковзного середнього MA.

Якщо будь-який із параметрів p , d або q дорівнює 0, назву моделі можна записати, опустивши відповідну частину в назві, наприклад: $ARIMA(2,0,0) = AR(2)$, $ARIMA(1,0,1) = ARMA(1,1)$. Вказівка 0 може знадобитися при роботі з програмним забезпеченням, тоді як в описі або інтерпретації прийнято опускати нульові параметри в назві моделі, тобто процес $AR(p)$, а не $ARIMA(p,0,0)$ [25, 26].

2.2 Побудова моделі ARIMA

Процедура побудови моделей ARIMA за методологією Бокса-Дженкінса складається з п'яти основних етапів:

- 1.Визначення складових ряду та перевірка стаціонарності.
- 2.Перетворення ряду до стаціонарного вигляду.
- 3.Ідентифікація, тобто визначення типу моделі та вибір кількості параметрів процесу AR та MA на основі трансформованого ряду.
- 4.Оцінка параметрів, тобто оцінка значень параметрів.

5.Діагностична перевірка моделі, тобто перевірка адекватності та відповідності розрахункової моделі.

Таблиця 2.1 – Порівняння властивостей автокореляції та часткової автокореляції авторегресії, ковзного середнього та змішаних процесів

	AR(p) процеси авторегресії	MA(q) процеси ковзного середнього	Змішані процеси ARMA(p,q), $p>0, q>0$
Автокореляційна функція ACF	плавню зменшується (згасаюча експоненціальна функція та/або затухаюча синусоїда)	розривається після затримки q	плавню зменшується (після перших інтервалів q-p, що спадають експоненціальні функції та/або затухаючі синусоїди)
Часткова автокореляційна функція PACF	обривається після затримки p	плавню зменшується (згасаюча експоненціальна функція та/або затухаюча синусоїда)	плавню зменшується (після перших інтервалів p-q, спадаючих експоненціальних функцій та/або затухаючих синусоїд)

Джерело: [36]

Процес побудови починається з аналізу стаціонарності ряду та прийняття рішення про можливе перетворення та диференціювання даних. Потім автокореляція та часткова автокореляція використовуються для визначення порядків p і q операторів авторегресії та ковзного середнього.

Автокореляційна функція процесу авторегресії p-порядку зменшується плавно (експоненціально), тоді як часткова автокореляційна функція закінчується після p-го інтервалу. У випадку процесу ковзного середнього порядку q функція автокореляції закінчується після інтервалу q, тоді як функція часткової автокореляції плавно зменшується. Якщо і ACF, і PACF розпадаються експоненціально, це свідчить про змішаний процес. Автокореляційна функція змішаного процесу, що містить компонент авторегресії порядку p і компонент ковзного середнього порядку q, після перших інтервалів q-p, є сумою експоненціальних функцій і затухаючих

синусоїд. Подібним чином, функція часткової автокореляції змішаного процесу після перших інтервалів p - q домінує комбінацією експоненціальних функцій і затухаючих синусоїд [27].

Вищевказана інформація підсумована в табл. 2.1.

У практичних прикладах часових рядів найбільш поширеними є авторегресійні процеси та ковзне середнє першого та другого порядку.

Виведення висновків про перебіг функцій ACF і PACF можна здійснити на основі корелограм – графічно представленої кореляційної функції. Порівняння графіків автокореляції реального часового ряду та теоретичних форм функцій ACF і PACF різних процесів ARIMA є простим способом відхилити задану модель або залишити її для подальшої діагностики.

На рис. 2.1 показано хід виконання функцій ACF і PACF для процесу AR(1).

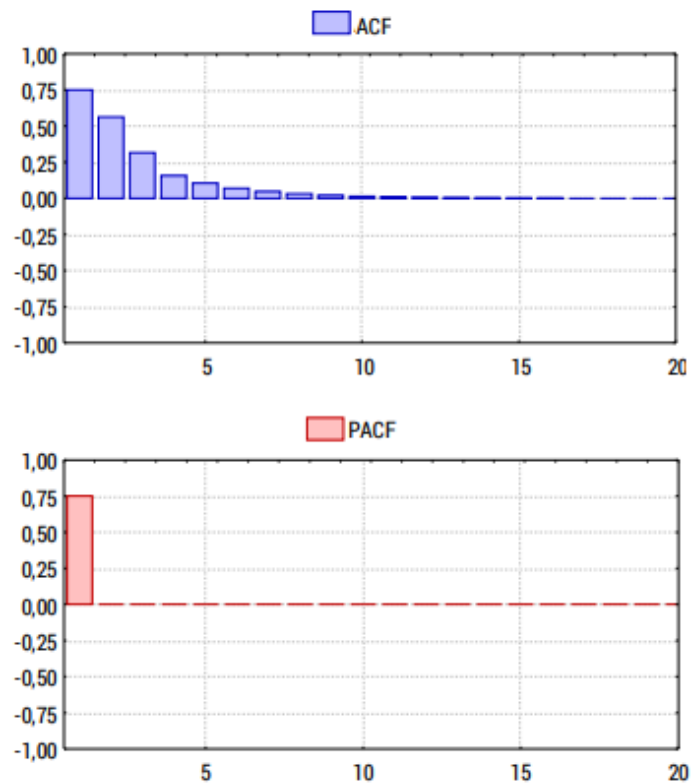


Рисунок 2.1 – Хід автокореляції та часткової автокореляції процесу AR(1)

Джерело. [36]

Функції автокореляції спадають експоненціально, а функція часткової автокореляції є найбільшою на лазі 1 і не показує кореляції для інших лагів. Проте існує принципова різниця в ході автокореляційної функції для моделей з $\phi_1 < 0$ і $\phi_1 > 0$. У випадку, коли $\phi_1 > 0$, значення автокореляційної функції для всіх лагів додатні, якщо $\phi_1 < 0$ вони змінюються по чергові [28].

Якщо в процесі присутні як сезонні, так і несезонні складові, їх досить складно розділити та чітко визначити модель. Загалом, значення автокореляції рядів реального часу дуже рідко збігаються з описом теоретичних ходів автокореляційних функцій процесів АР або МА. На практиці дуже часто випробувано декілька різних моделей і лише після етапу оцінки та діагностичної перевірки приймається остаточне рішення. Це стало можливим завдяки програмним пакетам, таким як: R, Matlab або SAS [29].

2.3 Оцінка параметрів моделей ARIMA

Отримання ефективних оцінок параметрів є наступним кроком у побудові моделі ARIMA. Оцінка параметрів моделі ARIMA є досить складним завданням. Параметри AR-моделей можна оцінити за допомогою методу найменших квадратів. В інших моделях (тобто, коли $q > 0$) необхідно використовувати ітераційні процедури мінімізації суми квадратів залишків. Швидше за все також використовуються методи. Слід пам'ятати про обмеження значень параметрів процесу, які впливають з необхідності забезпечення стаціонарності та оборотності моделей [30].

Для визначення початкових значень параметрів для ітераційних методів можна використовувати автокореляційні та часткові автокореляційні функції.

Для процесу ковзного середнього МА(q) функція автокореляції може бути виражена як:

$$\rho_k = \frac{-\theta_k + \theta_1\theta_{k+1} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2}, k = 1, 2, \dots, q. \quad (2.5)$$

Підставляючи ρ_k замість їхніх оцінок r_k і розв'язуючи q нелінійних рівнянь, отримуємо попередні оцінки параметрів θ_k . У загальному випадку рівняння (2.5) мають багато розв'язків, але лише один задовольняє умову оборотності. Припускаючи, що досліджуваний ряд є процесом AR, початкові оцінки параметрів можна отримати за допомогою автокореляційних рівнянь Юла-Уокера:

$$\phi = P_p^{-1} \rho_p, \quad (2.6)$$

де

$$\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \dots \\ \phi_p \end{bmatrix}, P_p = \begin{bmatrix} 1 & \rho_1 & \dots & \rho_{p-1} \\ \rho_1 & 1 & \dots & \rho_{p-2} \\ \dots & \dots & \dots & \dots \\ \rho_{p-1} & \rho_{p-2} & \dots & 1 \end{bmatrix}, \rho_p = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \dots \\ \rho_p \end{bmatrix}. \quad (2.7)$$

Розрахункові початкові значення параметрів є основою для визначення оптимальних значень в ітераційних процедурах. Проблема оцінки параметрів є предметом багатьох наукових публікацій (її обговорюють, серед інших: Caines і Rissan, Gardner Harvey і Phillips, Lindquist). Мелард, Рох і Саїді цитують різні підходи як вступ до представлення власного рішення. Основна складність полягає у визначенні початкових значень (особливо залишків у моделях МА, тому що у випадку AR ряд можна скоротити [31].

Діагностика побудованих моделей ARIMA може базуватися на аналізі залишків. Якщо модель була правильно підібрана, то значення модельних залишків повинні наближатися до білого шуму та ін. Таким чином, дослідження залишків моделі може виявити існування та характер неадекватності моделі. Зокрема, сприймана форма оціненої залишкової автокореляційної функції $e^\wedge$ може вказувати на необхідність проведення відповідної модифікації моделі.

Якщо відомі справжні значення параметрів ϕ і θ , оцінені автокореляції ряду e_t будуть некорельованими і матимуть приблизно нормальний розподіл із середнім рівним 0 і стандартною помилкою (n – кількість спостережень у ряді).

На практиці справжні значення параметрів невідомі, можна розрахувати лише оцінки, на основі яких .

Тому використання в якості стандартної помилки для може призвести до недооцінки значущості відхилень автокореляції від нуля на малих інтервалах, але може використовуватися для середніх і великих інтервалів.

На практиці для оцінки автокореляції визначається рівень ймовірності, з яким тестовані залишкові автокореляції повинні потрапляти між граничними значеннями (z – критичне значення стандартизованого нормального розподілу). Ці інтервали, нанесені на корелограми, дозволяють швидко розпізнати ситуацію, коли серія залишків відрізняється від білого шуму [32].

Графічне представлення значення AIC (інформаційний критерій Акаїке) також може допомогти у виборі найкращої моделі з кількох альтернатив з різною кількістю параметрів. Ця кількість визначається як:

$$AIC = -2\ln\hat{L} + 2m, \quad (2.8)$$

де m – кількість оцінюваних параметрів ($m = p + q + P + Q$),

$\hat{L}$ – функція максимальної ймовірності.

Індикатор AIC надає інформацію про якість підгонки різних моделей для одних і тих же даних з урахуванням ряду параметрів моделі, т.зв. "штраф". Коли значення функції L невідоме, зручним способом обчислення AIC є наближення її значення за формулою:

$$AIC \approx n(1 + \ln(2\pi)) + n\ln\hat{\sigma}^2 + 2m, \quad (2.9)$$

де $\hat{\sigma}^2$ – дисперсія модельних залишків,

n – кількість спостережень у серії.

Оскільки для тих самих даних: $n(1 + \ln(2\pi))$ має однакове значення для всіх моделей, цей член рівняння часто опускається. Іноді дисперсію залишків у формулі замінюють сумою квадратів залишків. На рис. 3.15 показаний гіпотетичний хід такої функції, побудованої за значеннями AIC для

моделі, в якій присутні як процес ковзного середнього $MA(q)$, так і авторегресія $AR(p)$ а $0 \leq p, q \leq 2$.

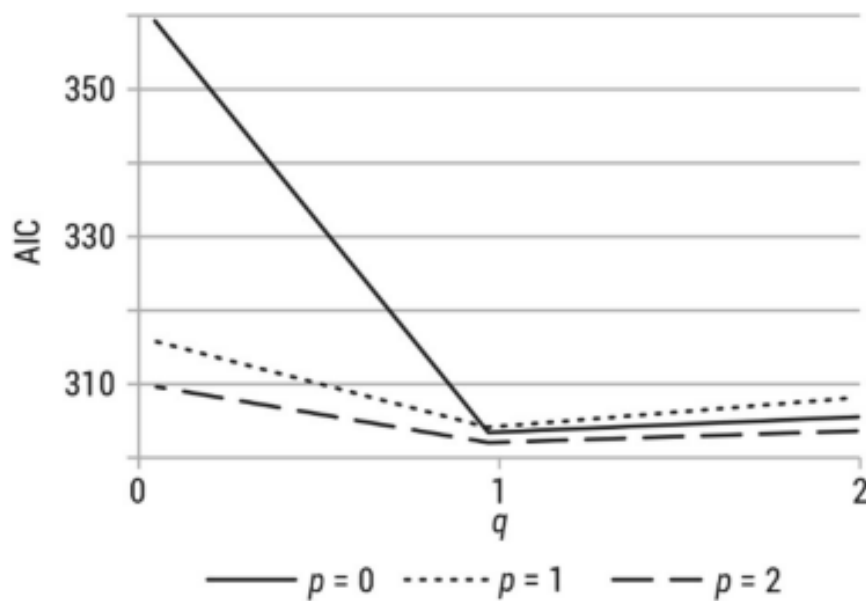


Рисунок 2.2 – Графік значень AIC для гіпотетичних моделей

Джерело: [36]

Значення AIC інтерпретується тільки шляхом порівняння зі значеннями цього коефіцієнта для інших моделей для тих же даних. Найкраща модель - це модель з найнижчим AIC. Однак різниця в розмірі 2 або менше не має значення - ви можете вибрати модель, яка простіше або краще підходить за даними. Значення AIC моделі з $q = 1$, представленої на рис. 2.2, менші, ніж у випадку, коли $q = 2$. Крім того, при $q = 1$ AIC практично ідентичний, незалежно від складності авторегресії процес. Тому можна зробити висновок, що в цій гіпотетичній моделі необхідний лише один параметр ковзного середнього [33].

Критерієм для порівняння різних моделей ARIMA також можуть бути похибки: середнє значення модуля похибки (середня абсолютна похибка), середньоквадратична похибка або середня абсолютна похибка у відсотках, які обговорювалися в попередніх розділах посібника «Прогнозування в управлінні бізнесом». При оцінці моделі ARIMA особливо часто використовується стандартна помилка оцінки моделі:

$$\hat{\sigma}_k = \sqrt{\frac{1}{n-p-q} \sum_{t=1}^n \hat{e}_t^2}, \quad (2.10)$$

де $\hat{e}_t^2$ – рештки моделі.

Ця помилка також використовується для визначення дисперсії оцінки параметрів і оцінки їх значущості. Загалом, чим менше значення помилки, тим краще відповідає модель даним. Щоб перевірити випадковість залишків моделі, можна використати статистичний тест, запропонований Ljung-Box, який є розширенням статистики Box Pierce. Гіпотеза H_0 перевірена: залишки незалежні, по відношенню до H_1 : залишки незалежні [34, 35].

2.2 Геометричний броунівський рух

Дробовий броунівський рух є узагальненням відомого броунівського процесу. Це гаусівський процес зі стаціонарними приростами, але його прирости не є незалежними (за винятком звичайного броунівського руху $H = 0,5$). Структура залежності приростів залежить від значення показника Херста, тобто коваріації дробового броунівського руху ($H \in (0,1)$; випадок $H = 1$ нецікавий, оскільки реалізації цього процесу є прямими лініями з випадковим нахилом).

Дифузійні процеси з неперервним часом є незамінним інструментом для моделювання багатьох явищ у сучасних фінансах. У деяких роботах, як [37, 38], випадкові процеси встановлюють певний стандарт для моделювання економічних процесів.

Розглянемо модель ціни акцій, засновану на геометричному Броунівському русі:

$$dS = \mu S dt + \sigma S dB^H, \quad (2.11)$$

де B^H – геометричний броунівський рух.

На жаль, довга пам'ять означає, що аналітичні вирази для розподілу курсів акцій, створені рівнянням (2.11) (не кажучи вже про розподіл максимального збитку), невідомі. Це пов'язано з тим, що вихідна задача складається з нескінченної кількості «початкових» величин, тобто вихідна величина - це вся історична реалізація вартості інструменту.

У подальших міркуваннях автор опускає питання, пов'язані з оцінкою параметрів рівняння, тобто передбачається, що значення параметрів відомі.

Незнання розподілу не виключає можливості зробити, наприклад, неупереджений прогноз.

2.4 Метод випадкових лісів

Дерева рішень – це логічні схеми, які дають змогу прийняти остаточне рішення щодо класифікації об'єкта шляхом відповіді на ієрархічно організовану систему запитань, яка найкоротшим шляхом веде до істинної відповіді. Щоб побудувати дерево, алгоритм проходить усі можливі тести та знаходить той, який є найпотужнішим з точки зору прогнозування значення цільової змінної [39]. Рекурсивне поділ даних повторюється, доки всі точки даних у кожному діапазоні поділу не належатимуть до одного класу цільової змінної.

Алгоритми побудови дерева рішень засновані на принципі жадібної максимізації отримання інформації, тобто на кожному кроці вибирається ознака, яка при поділі на неї дає більший приріст інформації. Щоб розбити вершини дерева, сформулюйте умови

$$x_j \leq t, \quad (2.12)$$

де x_j – j -й атрибут об'єкта,

t – поріг.

Якщо нерівність вірна, об'єкт направляється в один бік, якщо ні – в інший. Розглянемо будь-який кут дерева g . Шляхом до нього є набір об'єктів

X_g . Позначимо критерій помилки як $Q(X_g, j, t)$, потім параметри j і t вибираються перерахуванням, щоб мінімізувати цю помилку:

$$Q(X_g, j, t) \rightarrow \min_{j, t}. \quad (2.13)$$

Позначимо вибірки, надіслані до лівого та правого піддерев для X_u та X_v відповідно на цьому кроці m , тоді ми визначимо критерій відмови наступним чином:

$$Q(X_g, j, t) = \frac{|X_u|}{|X_g|} H(X_u) + \frac{|X_v|}{|X_g|} H(X_v), \quad (2.14)$$

де $|X_j|$ – розмір вибірки X_j , а H вимірює, наскільки розсіяними є відповіді, коли вони потрапляють у відповідне піддерево. Чим менше розкид відповідей, тим менше значення H [40].

Введемо допоміжне значення p_k , яке показує, який відсоток об'єктів класу k є в будь-якій вибірці X :

$$p_k = 1/|X| \sum_{i \in X} [y_i = k], \quad (2.15)$$

де y_i – значення цільової ознаки на i -му об'єкті, k – значення з набору класів (у випадку бінарної класифікації є тільки два класи – 0 і 1), а функція $[y_i = k]$ приймає значення 1, якщо умова в квадратних дужках істинна, 0 в іншому випадку.

Існує два різні критерії важливості для проблем класифікації, які можна вказати за допомогою параметра критерію під час використання методу дерева рішень.

Нехай K буде максимальним значенням цільової ознаки (1 у нашому випадку). Перший критерій – це критерій Джіні, де функція H обчислюється за формулою

$$H(X) = \sum_{k=0}^K p_k(1 - p_k), \quad (2.16)$$

де p_k — визначається наведеними вище формулами,

X — будь-який вибір об'єктів.

Зокрема, формула показує, що якщо $p_0 = 1, p_1 = \dots = p_K = 0$, то $H = 0$, тобто якщо вибірка містить елементи лише одного з класів, то функція H досягне мінімального значення [41].

Другим критерієм є ентропія, де H визначається формулою

$$H(X) = - \sum_{k=0}^K p_k \ln p_k. \quad (2.17)$$

Як правило, така деревна структура призводить до моделей, які є дуже складними та переобладнаними. Одним із способів вирішення цієї проблеми, є зупинка процесу побудови дерева на певній глибині. Це призводить до нижчої точності на навчальному наборі, але підвищує точність на тестовому наборі. Коли використовується зупинка, клас нової точки даних, яка закінчується у вершині g після розбиття, визначається за формулою:

$$r = \operatorname{argmax}_{y \in Y} \sum_{i \in X_g} [y_i = y], \quad (2.18)$$

де y_i — значення цільової ознаки на i -му об'єкті із вибірки X_g ,

Y — набір усіх можливих значень цільової ознаки,

y — значення з множини Y ,

argmax — функція, результатом якої є аргумент при якому досягається найбільше значення функції, тобто y , при якому сума з формули приймає найбільше значення, тобто, простіше кажучи, клас, який є найпопулярнішим у вибірці.

Основна ідея випадкового лісу [47] полягає у використанні великого набору дерев рішень, які самі по собі не забезпечують дуже якісну класифікацію, але забезпечують хороший результат завдяки своїй великій кількості. Random Forest вирішує проблему схильності дерев рішень до переобладнання.

Щоб створити модель випадкового лісу, вкажіть кількість дерев, які будуть створені незалежно. Під час побудови кожного дерева спочатку створюється випадкова вибірка з копіями того самого розміру, що й вихідний набір даних. Тому деякі зразки влучають у нього два і більше разів. Цей процес називається провисанням. Крім того, алгоритм випадковим чином вибирає функції для створення кожного дерева для ще більшої кількості різних дерев. Даний параметр відповідає за кількість характеристик на вибір. Якщо в деревах рішень ознака номер j була обрана з усієї множини ознак з мінімізацією критерію помилки, то у випадкових лісах вона береться з випадкової підмножини заданого розміру. У задачах класифікації розмір даної підмножини зазвичай береться як значення

$$q = \sqrt{n}, \quad (2.19)$$

де n — вихідна кількість ознак.

Алгоритм призначає об'єкт, який потрібно класифікувати, до класу, до якого він був призначений більшістю дерев у композиції.

Бустінг — це побудова ансамблю, в якому дерева, що входять до нього, будуються не незалежно один від одного, як у випадковому лісі, а послідовно [48]. Кожне дерево намагається виправити помилки попереднього. Тут використовуються переважно дерева невеликої глибини. Це робить модель більш ефективною з точки зору пам'яті та швидкості обчислень прогнозу.

Ми позначаємо через $L(y,z)$ функцію втрат, яку потрібно мінімізувати, де y — істинна відповідь, а z — результат алгоритму, тобто прогноз.

Позначимо композицію з N базових алгоритмів (в даному випадку дерев)

$$a_N(x) = \sum_{n=1}^N b_n(x), \quad (2.20)$$

де N — кількість алгоритмів у композиції,

$b_1 \dots b_n$ — алгоритми, що входять до композиції,

$b_j(x)$ – результат роботи (прогноз) j -го алгоритму.

Спочатку побудуємо перший базовий алгоритм $b_1(x)$. Як такий алгоритм можна, наприклад, взяти найпопулярніший клас на навчальному наборі даних:

$$b_1(x) = \underset{y \in Y}{\operatorname{argmax}} \sum_{i=1}^m [y_i = y], \quad (2.21)$$

де y_i – значення цільової ознаки на i -му об'єкті з навчальної множини,

Y – множина всіх можливих значень цільової ознаки, у приймає значення з множини Y ,

m – кількість об'єктів у вибірці.

Базові алгоритми $N-1$ вже створені. Тепер потрібно вибрати наступний алгоритм $b(x)$ так, щоб мінімізувати помилку композиції L на навчальному наборі:

$$\sum_{i=1}^m L(y_i, a_{N-1}(x^{(i)}) + b(x^{(i)})) \rightarrow \min_b, \quad (2.22)$$

де $x^{(i)}$ – вектор ознак i -го об'єкта з навчальної множини,

y_i – значення цільової ознаки на цьому об'єкті,

m – кількість об'єктів у навчальній множині.

Спочатку розв'яжемо простішу задачу: давайте спробуємо зрозуміти, які конкретні значення s_i повинен приймати новий алгоритм на об'єктах навчальної вибірки $x^{(i)}$, щоб функція втрат приймала найменше значення [49]. Тому попередня формула переписується у форму

$$F(s) = \sum_{i=1}^m L(y_i, a_{N-1}(x^{(i)}) + s_i) \rightarrow \min_s, \quad (2.23)$$

де $s = (s_1, \dots, s_m)$ – вектор зрушень, який потрібно знайти.

Як вектор s потрібно брати антиградієнт, так як він спрямований у бік якнайшвидшого зменшення функції:

$$s = -\nabla F = (-L'_z(y_1, a_{N-1}(x^{(1)})), \dots, -L'_z(y_m, a_{N-1}(x^{(m)}))) \quad (2.24)$$

За цією формулою можна знайти всі елементи вектора s , тому тепер відомо, які значення має набувати новий алгоритм на об'єктах набору, щоб мінімізувати помилку. Тоді навчання нового алгоритму $b_N(x)$ є завданням навчання на розмічених даних, в якій $\{(x_i, s_i)\}_{i=1}^m$ – навчальна вибірка та використовується, наприклад, квадратична функція помилки:

$$b_N(x) = \operatorname{argmin}_b \frac{1}{m} \sum_{i=1}^m (b(x^{(i)}) - s_i)^2, \quad (2.25)$$

де $s = (s_1, s_2, \dots, s_m)$ – вектор зміщення, знайдений вище,

argmin – це функція, результатом якої є аргумент, при якому досягається найменше значення функції, тобто b , при якій сума з формули найменшого значення.

Щоб забезпечити точний рух до мінімуму функції помилки, введено таке значення, як розмір кроку, тобто при додаванні знайденого алгоритму $b_N(x)$ відбувається лише незначний зсув у напрямку цього вектора:

$$a_N(x) = a_{N-1}(x) + \lambda b_N(x), \quad (2.26)$$

де $\lambda \in (0, 1]$ – довжина кроку. Зрозуміло, що менше розмір кроку, тим більше часу займає процес, але тим кращої якості можна досягти.

Висновки до розділу 2

1. В другому розділі проведений аналіз методів та моделей прогнозування.
2. Описана процедура побудови моделей ARIMA за методологією Бокса-Дженкінса, яка складається з п'яти основних етапів.
3. Щоб перевірити випадковість залишків моделі, був використаний статистичний тест Ljunga-Vox, який є розширенням статистики Vox Pierce.
4. Розглянута модель ціни акцій, заснована на геометричному Броунівському русі.
5. Застосований алгоритм побудови дерева рішень заснований на принципі жадібної максимізації отримання інформації.
6. Зазначено, що прогнозування може бути якісним чи кількісним, залежно від зібраної інформації та її характеру, який зазвичай буває суб'єктивним чи об'єктивним і залежить від математичних розрахунків чи взагалі залежить від математичних розрахунків.
7. При обробці інформації про поведінку фінансових часових рядів необхідно застосовувати відповідні математичні методи.
8. Було розглянуто 3 методи для прогнозу показників. Було визначено основні аспекти прогнозування за допомогою цих інструментів, а також те, що ці методи підходять для прогнозування цін на ф'ючерсні контракти.

3 ПРАКТИЧНЕ ЗАСТОСУВАННЯ ПРОГНОЗНИХ МОДЕЛЕЙ

3.1 Геометричний броунівський рух

У дослідженні використовуються ціни на нафтові ф'ючерси як приклад дослідження динаміки цін активів. Ця задача буде реалізована за допомогою мови програмування Python у середовищі Jupyter Notebook. Досліджуватися будуть ціна на ф'ючерси сирої нафти (Crude Oil), а дані отримуються із Yahoo Finance через пакет Yfinance.

На наведеному нижче лінійному графіку вісь X вказує дні з 1 січня 2022 р. по 31 серпня 2022 р., а вісь Y вказує ціну ф'ючерсів у доларах. Як можна побачити на рисунку нижче, то графік має хвилеподібну траєкторію. Він має тенденцію збільшення з січня до середини березня, потім знижується до середини травня, після чого слідує ще один ряд збільшення до середини червня і, нарешті, поступове зменшення до кінця серпня [42, 43].

Розглянемо коливання за липень. Тренд графіку низхідний. Ціна падає плавно, але під час свого коливання до кінця серпня графік безперервно піднімається і опускається, що вказує на те, що в цьому русі немає інтерпретації закономірності, що піддається.

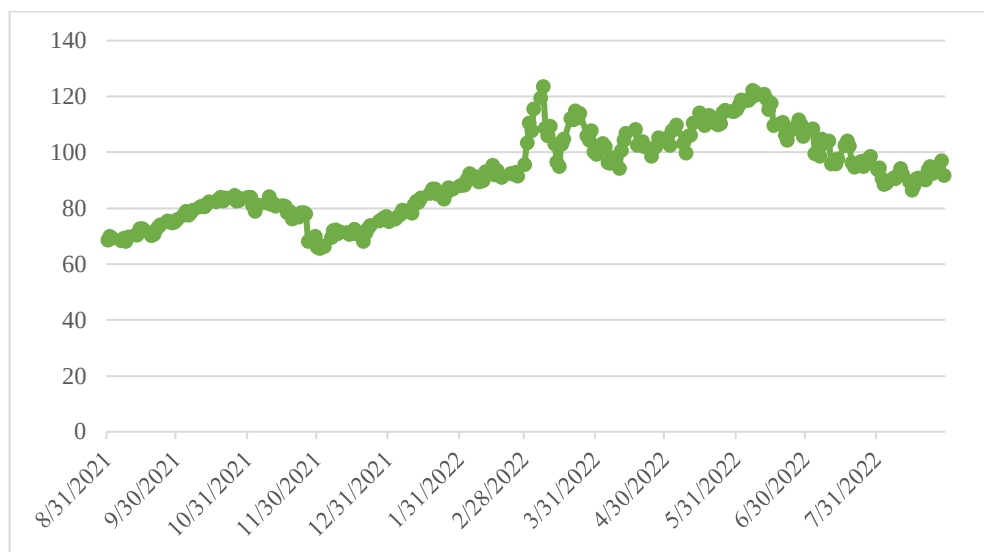


Рисунок 3.1 – Візуалізація динаміки цін
(розраховано автором на основі [27])

Можна зробити висновок що курс ф'ючерсів у липні має тенденцію до падіння, а також щодня відбуваються випадкові потрясіння падіння курсу ф'ючерсів, що призводить до нерегулярної лінії [44, 45, 46].

У результаті використаємо відповідну модель, яка враховує обидва типи руху курсу ф'ючерсів. Для побудови моделі візьмемо за основу геометричний броунівський рух. GBM має дві компоненти, які виконують цю роботу. Перший компонент включає довгострокову тенденцію, тоді як інший компонент застосовує випадкові шоки.

В цьому дослідженні використовувались ціни на ф'ючерси нафти за липень, щоб зробити прогноз на вересень. Крім того припускається, що тиждень складається лише з будніх днів, тобто те, що йде після п'ятниці, є понеділком. Нижче наведено вхідні параметри, які прийматиме модель GBM:

- 1) S_0 — початкова ціна;
- 2) dt — приріст часу, день;
- 3) T — довжина часового горизонту передбачення, день;
- 4) N — кількість часових точок у часовому горизонті прогнозування
- 5) t — масив часових точок у час.горизонті прогнозування [1,2,..,N];
- 6) μ — середнє історичне;
- 7) σ — стандартне відхилення щоденних історичних даних;
- 8) b — масив для броунівських приростів;
- 9) W — масив для броунівського шляху.

Отримуємо історичні ціни ф'ючерсів між 1 серпня 2022 р. та 31 серпня 2022 р. Потім, використовуючи модель GBM, отримаємо наші симуляції до прогнозної дати 30 вересні 2022 р.

	Date	Close
0	2022-08-01	93.889999
1	2022-08-02	94.419998
2	2022-08-03	90.660004
3	2022-08-04	88.540001
4	2022-08-05	89.010002
	Date	Close
17	2022-08-24	94.889999
18	2022-08-25	92.519997
19	2022-08-26	93.059998
20	2022-08-29	97.010002
21	2022-08-30	91.639999

Рисунок 3.2 – Візуалізація динаміки цін
(розраховано автором на основі [27])

Виведемо значення початкової ціни ф'ючерсів. Симуляції форвардної ціни беруть свій початок із цього початкового значення ціни. Ми помножимо це початкове значення на деякі вирази, щоб отримати прогнози для кожного торгового дня в серпні. У нашому випадку ціна ф'ючерсів на момент закриття 30 серпня 2022 року виглядає наступним чином:

$$S_0 = 91,63999938964844$$

Приростом часу в моделі є dt . Крок часу в нашій моделі становить 1 день, оскільки в даних є ціни на закриття для кожного торгового дня, тобто це значення впливає з часових кроків у історичних даних.

Довжину прогнозованого часового горизонту позначає величина T . В цьому випадку T має становити 22 дні, оскільки потрібні прогнози для 22 торгових днів серпня.

Параметр N обчислюється після знаходження dt і T . Це кількість часових точок у часовому горизонті прогнозування. У нашому випадку приріст часу становить 1 день, тоді отримаємо прогнози на 22 торгові дні. Це

означає, що у нас є 22 різні моменти часу (дні), і матимемо 22 передбачення в кінці [50].

Масив, де показується прогресія часу в моделі є значення t . Ця величина слугує своєрідним таймером, де вимірюється час, підраховується кількість моментів часу, що минув. Побудова масиву t впливає з обчислення N .

Ціни на ф'ючерси витягуються лише за торгові дні, як можна зрозуміти з наведених вище даних. Розрахуємо середню прибутковість цін ф'ючерсів у вибраному діапазоні дат. У нашому випадку ми вибираємо торгові дні серпня як історичні дані, і ми повинні обчислити μ , використовуючи ціни ф'ючерсів серпня. Перш ніж розрахувати μ , ми повинні розрахувати прибутковість для кожного торгового дня за такою формулою:

$$r_k = \frac{S_k - S_{k-1}}{S_{k-1}}, \forall k, \quad (3.1)$$

де k – період часу.

$$\mu = \frac{1}{|k|} \times \sum_{all\ k} r_k, \quad (3.2)$$

$$\mu = -0,00432.$$

Після обчислення доходів за всі дні серпня потрібно обчислити середнє арифметичне, використовуючи ці прибутки, щоб отримати μ . Потім використаємо μ у обчисленні компонента дрейфу. Це матиме вплив на довгостроковий рух курсу ф'ючерсів [51]. Крім того, можемо помітити, μ становить приблизно -0,08%, що вказує на те, що в середньому в серпні спостерігається від'ємна прибутковість, і це необхідно врахувати під час розрахунку прогнозів на серпень:

$$\mu = -0,0008126039885336756.$$

Розрахуємо волатильність історичних даних за наступною формулою:

$$\sigma = \sqrt{\frac{1}{|k|} \sum_{all\ k} (r_k - \mu)^2}. \quad (3.3)$$

Це стандартне відхилення прибутковості цін на акції в серпні. Волатильність важлива, тому що буде корисно включити випадкові поштовхи в прогнози. На лінійному графіку цін на нафтові ф'ючерси, наведеному вище, ціна безперервно зростає та падає день за днем. Крім того, величина та напрямок цього невеликого руху здаються випадковими. Тут волатильність допоможе визначити величину руху. Це значення, саме по собі, не додає до моделі потрібної випадковості [58]. Буде використовуватись стандартна нормальна випадкова змінна під час вибору випадкових значень. Вона сприятиме масштабуванню величини випадкового шоку таким чином, щоб невеликі коливання відбувалися відповідно до історичної волатильності цін на ф'ючерси. Отримаємо наступне значенні волатильності:

$$\sigma = 0,02606285344615231.$$

Масив для броунівських приростів (b) є масивом, у який ми додаємо випадковість до нашої моделі. Тут важливою є змінна, яка показує кількість прогнозних сценаріїв. Оскільки це імітаційна модель, ми хочемо мати багато гіпотетичних шляхів для ціни ф'ючерсів у межах прогнозованого часового горизонту. Припустимо, що прогнозується 2 сценарії коливання значень. Логіка кожного сценарію така ж, як і іншого. Масив t є масивом для прогресії в часі, який підраховує моменти часу, що минув. Тут масив b для кожної відповідної точки часу передбачення зберігає випадкове число, отримане зі стандартного нормального розподілу. Ці випадкові числа додадуть до моделі випадкові поштовхи. Для цього будемо використовувати *numpy.random.normal()* для генерації випадкових значень зі стандартного нормального розподілу. Знову ж таки, ми не множимо це випадкове значення на будь-яке число для коригування, дотримуючись тих самих міркувань з мю та сигмою.

Припускаємо, що наші історичні дані про ціни акцій є щоденними, тому ми призначаємо 1 для dt . Оскільки потрібні прогнози на 5 днів, що означає, що значення T дорівнює 5. Після цього N обчислюється як 5 часових точок (5 одноденних періодів). Таким чином, для кожного з цих часових моментів генерувало випадкове число, отримане зі стандартного нормального розподілу (середнє=0, стандартне значення=1) за допомогою методу `numpy.random.normal()`. Ці випадкові значення склали масив b .

У нашому випадку T має бути 22 дні. Нижче наведено код, у якому ми створюємо 2 масиви випадкових значень довжиною 22 для 2 різних сценаріїв.

```
{'1': array([ 1.41190114,  0.41344187,  0.33044612, -1.04199243, -0.85222131,
 2.56946067, -0.81051665, -0.47150252,  2.07712116,  0.38003216,
 1.91165461,  0.87429765, -0.20990656,  0.25497569,  0.31296387,
-0.24595873, -2.30824808,  0.054108 , -1.07081308,  0.19322023,
-0.07919112,  0.56605886,  0.72963734]), '2': array([-0.35504397,  0.63927438,  0.853574 , -0.48815241, -1.56651374,
-0.44190734, -0.12992412,  0.37761515,  0.18450937,  1.43291316,
 0.71128427,  0.79893706, -0.37519268, -0.7333454 , -1.44217034,
-0.04178473, -0.08328814, -2.05981085, -0.15583914,  0.29992471,
 0.45541165, -1.35019696,  0.64244117])}}
```

Рисунок 3.3 – Масиви випадкових значень

((розраховано автором на основі [27]))

W – це броунівський шлях, і він визначає, як ціни на акції коливаються від початкової точки часу (S_0) до іншої точки часу t . Припустимо, що в момент часу 3 ціна акцій становить S_3 . Під час прогнозування 4 точки часу $b(4)$ застосовується до S_3 як випадковий удар. W , з іншого боку включає наслідки всіх випадкових потрясінь з початку часового горизонту прогнозування. Це загальний ефект випадковості, включений у S_0 (початкова ціна акцій) до певного моменту часу.

$$W_k = \sum_{i=1}^k b_i, k = 1, \dots, N. \quad (3.4)$$

3.2 Побудова моделі GBM з дискретним часом

В розділі 3.1.1 було визначено дві компоненти геометричного броунівського руху. Перша – це довгострокова тенденція цін на акції, друга – короткострокові випадкові коливання. Призначимо цим компонентам імена. Нехай довгострокові тенденції називються *Drift*, а короткострокові коливання – *Diffusion*. Припустимо, що в певний момент часу $(k-1)$ і передбачаємо момент часу (k) . Однак ті самі припущення, які описувалися в попередньому розділі діють в цьому випадку.

Припущення 1: тривалість періоду часу між $(k-1)$ і (k) , яка є dt , відповідає частоті історичних даних.

Припущення 2: час у нашій симуляції прогресує через підрахунок періодів часу.

Ми знаходимося в точці часу $(k-1)$. Ціна акцій у момент часу $(k-1)$ повинна підкорятися довгостроковій тенденції цін на акції, піддаючись випадковому шоку, коли його використовують для створення прогнозу для моменту часу (k) . Отже, потрібно застосувати два компоненти GBM до цієї ціни акцій.

Розглянемо компоненти з математичної точки зору:

1) *Drift*

$$drift_k = \mu - \frac{1}{2}\sigma^2 \rightarrow S_k = S_{k-1} \times e^{drift_k} = S_{k-1} \times e^{\mu - \frac{1}{2}\sigma^2}. \quad (3.5)$$

Drift відображає довгострокову тенденцію цін на акції. *Drift* є постійною величиною, і якщо припустити, що немає випадкового шоку, то можна застосувати його до ціни акцій у момент часу $(k-1)$. Якщо застосовувати дрейф без будь-яких випадкових потрясінь, ціни на акції плавно піднімаються вгору, якщо дрейф позитивний, або знижуються, якщо дрейф негативний. Зауважимо, що ціни на акції ніколи не можуть стати рівними 0, оскільки *Drift* застосовується через експоненціальну функцію. Тепер перейдемо до дифузії.

2) *Diffusion*

$$diffusion_k = \sigma b_k. \quad (3.6)$$

Diffusion відображає короточасні коливання. Масив b зберігає потрібну інформацію про випадкові поштовхи та отримує інформацію про випадкові поштовхи зі стандартної нормальної випадкової змінної b . При обчисленні дифузійної складової множимо випадкове значення $b(k)$ на сигму. Компонент *diffusion* дозволяє створювати різні сценарії прогнозування курсу акцій.

Отже, роблячи прогноз для моменту часу (b), враховуючи ціну акцій у момент часу ($b-1$), потрібно додати ефекти обох компонентів. Загальний сукупний ефект дає прогноз для моменту часу (b). Якби був лише компонент *Drift*, ми б завжди знали, якою стане ціна акцій у наступний момент часу, але компонент *Diffusion* порушує цю плавність, вводячи випадковий шок. Таким чином, не спостерігаються монотонні тенденції зростання або зниження цін на акції. Нижче показано, як загальний комбінований ефект застосовується до ціни акцій у момент часу ($b-1$).

$$S_k = S_{k-1} \times e^{(drift_k + diffusion_k)} = S_{k-1} \times e^{\left(\mu - \frac{1}{2}\sigma^2 + \sigma b_k\right)}. \quad (3.7)$$

Побудуємо загальне рівняння замкнутої форми геометричного броунівського руху з поправкою на контекст дискретного часу. Зпрогнозуємо значення $S(k)$ за допомогою S_0 (відома початкова ціна ф'ючерсів).

$$S_k = S_{k-2} \times e^{\left(\mu - \frac{1}{2}\sigma^2 + \sigma b_{k-1}\right)} \times e^{\left(\mu - \frac{1}{2}\sigma^2 + \sigma b_k\right)}; \quad (3.8)$$

$$S_k = S_{k-3} \times e^{\left(\mu - \frac{1}{2}\sigma^2 + \sigma b_{k-2}\right)} \times e^{\left(\mu - \frac{1}{2}\sigma^2 + \sigma b_{k-1}\right)} \times e^{\left(\mu - \frac{1}{2}\sigma^2 + \sigma b_k\right)}. \quad (3.9)$$

Якщо продовжимо це робити, то зрештою буде отримане значення S_0 , помножене на багато експоненціальних членів. Розширимо рівняння, виражаючи члени *Diffusion* та *Drift* цього разу окремо.

$$S_k = S_0 \times e^{\left(\mu - \frac{1}{2}\sigma^2\right)} \times \dots \times e^{\left(\mu - \frac{1}{2}\sigma^2\right)} \times e^{\sigma b_1} \times \dots \times e^{\sigma b_k}; \quad (3.10)$$

$$S_k = S_0 \times e^{\left(\left(\mu - \frac{1}{2}\sigma^2\right) \times k\right)} \times e^{\sigma \sum_{i=1}^k b_i}; \quad (3.11)$$

$$S_k = S_0 \times e^{\left(\left(\mu - \frac{1}{2}\sigma^2\right) \times k + e^{\sigma \sum_{i=1}^k b_i}\right)}. \quad (3.12)$$

Вище наведено передостаннє рівняння для $S(k)$. Це означає, що можна передбачити $S(k)$ з S_0 , додавши сукупний ефект k багатьох дрейфів і кумулятивну дифузію до k . Від S_0 до $S(k)$ є k багато часових точок, які називаються шляхом. Щоб передбачити k моментів часу, треба розглянути загальний вплив усіх випадкових поштовхів, які зустрічаються на шляху. На останньому кроці побудови рівняння змінюємо вищезгадане рівняння, роблячи заміни нижче.

$$t_k = k; \quad (3.13)$$

$$W_k = \sum_{i=1}^k b_i. \quad (3.14)$$

Це дає нам остаточну форму рівняння з усіма знайомими параметрами.

Отже, знаючи μ , σ , масив t і масив W можна легко створювати прогнози для всіх часових точок у часовому горизонті прогнозування за один крок.

$$S_k = S_0 \times e^{\left(\left(\mu - \frac{1}{2}\sigma^2\right) \times k + \sigma W_k\right)}. \quad (3.15)$$

3.3 Реалізація моделі

Розрахувати дрейф для всіх часових точок у часовому горизонті прогнозування. Для цього помножимо масив t на дрейф і отримаємо масив дрейфів. Цей масив дрейфу містить загальний дрейф для всіх часових точок у часовому горизонті прогнозування.

Маємо масив дифузії для кожного зі сценаріїв. За допомогою змінної *scen_size* вказано 2 сценарії для дослідження цієї проблеми. Тому отримаємо 2 різні дифузійні масиви, позначені «1» і «2» відповідно. Але для кожного з них однакова логіка. Помножимо броунівський шлях (масив W) на сигму.

```

drift:
[-0.00496266 -0.00992531 -0.01488797 -0.01985062 -0.02481328 -0.02977593
-0.03473859 -0.03970125 -0.0446639 -0.04962656 -0.05458921 -0.05955187
-0.06451452 -0.06947718 -0.07443984 -0.07940249 -0.08436515 -0.0893278
-0.09429046 -0.09925311 -0.10421577 -0.10917843 -0.11414108]
diffusion:
{'1': array([0.05047556, 0.06525614, 0.07706961, 0.03981831, 0.00935134,
0.10120973, 0.07223371, 0.05537746, 0.12963468, 0.14322086,
0.21156264, 0.24281885, 0.23531467, 0.24443007, 0.25561855,
0.24682551, 0.16430548, 0.16623985, 0.1279582 , 0.13486584,
0.13203475, 0.1522714 , 0.17835598]), '2': array([-0.01269285, 0.01016126, 0.04067659, 0.0232251 , -0.03277787,
-0.04857609, -0.05322088, -0.03972112, -0.03312489, 0.01810185,
0.04353031, 0.07209237, 0.0586792 , 0.03246205, -0.01909564,
-0.02058944, -0.023567 , -0.09720538, -0.10277664, -0.09205431,
-0.07577331, -0.12404295, -0.10107563])}

```

Рисунок 3.4 – Масиви випадкових значень
(розраховано автором на основі [27])

Таким чином отримаємо прогнозні значення. Оскільки у нас є 2 сценарії, маємо дві серії прогнозів цін на акції в результаті операції, наведеної нижче.

Створимо сюжет передбачень. Як можна побачити нижче, є 2 сценарії прогнозування ціни акцій. Один піднімається, а інший опускається. Дрейф у нашій задачі негативний, як можна спостерігати з масиву дрейфу. Це означає, що довгострокова тенденція до зниження цін на акції, але цей висновок зроблений на основі історичних даних за серпень [52, 54, 55]. Однак, незважаючи на негативний дрейф, маємо прогноз, який йде вгору. Це відбувається через випадкові поштовхи. У сценарії, коли ціни зростають, зрозуміло, що ці випадкові шоки перевищили ефект дрейфу, і це призвело до такої серії.

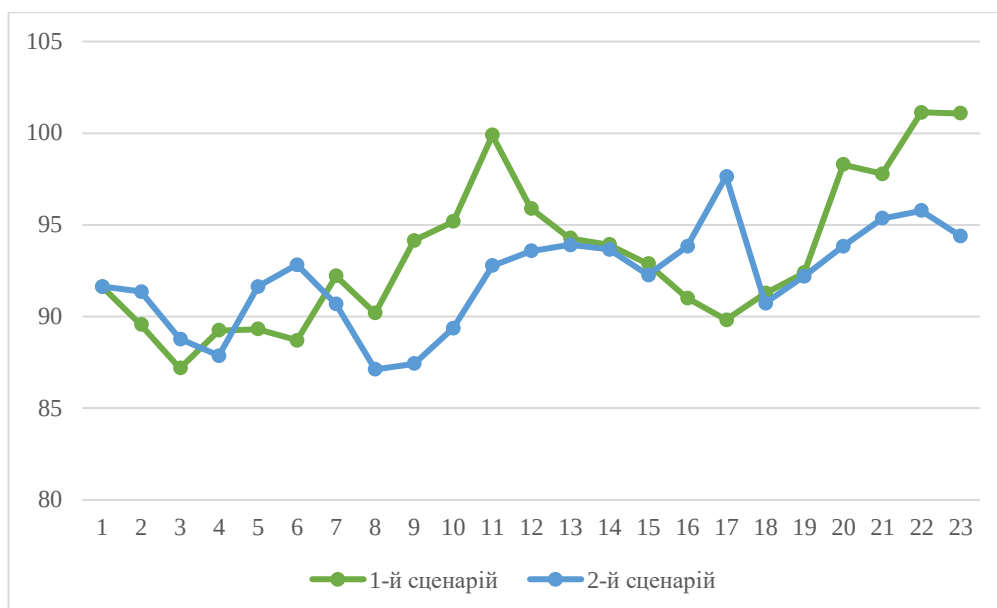


Рисунок 3.5 - Отримані результати
(розраховано автором на основі [27])

Таблиця 3.1 – Отримані результати

Прогнозні значення за 1 сценарієм, дол. США	Прогнозні значення за 2 сценарієм, дол. США
91,64	91,64
89,55965	91,35755
87,17542	88,78041
89,23601	87,86628
89,31965	91,63452
88,69896	92,8213
92,20541	90,66681
90,18032	87,11855
94,15319	87,43481
95,17315	89,37244
99,89942	92,76516
95,90043	93,5668
94,26504	93,91054
93,92633	93,65337
92,88328	92,26044
91,00309	93,82273
89,82188	97,64137
91,26797	90,72256
92,39434	92,19107
98,28695	93,84356
97,77693	95,35324
101,1417	95,77284
101,0872	94,39382

Змодельовано автором на основі даних

Графічний аналіз отриманого набору прогнозів показує, що модель загалом фіксує волатильність, але погано прогнозує її амплітуду. Крім того, чим нижча волатильність початкового часового ряду, тим слабша відповідь моделі [56].

Отримані прогнози досить точно відображають тренд і загальну довгострокову динаміку, але не чутливі до великих коливань.

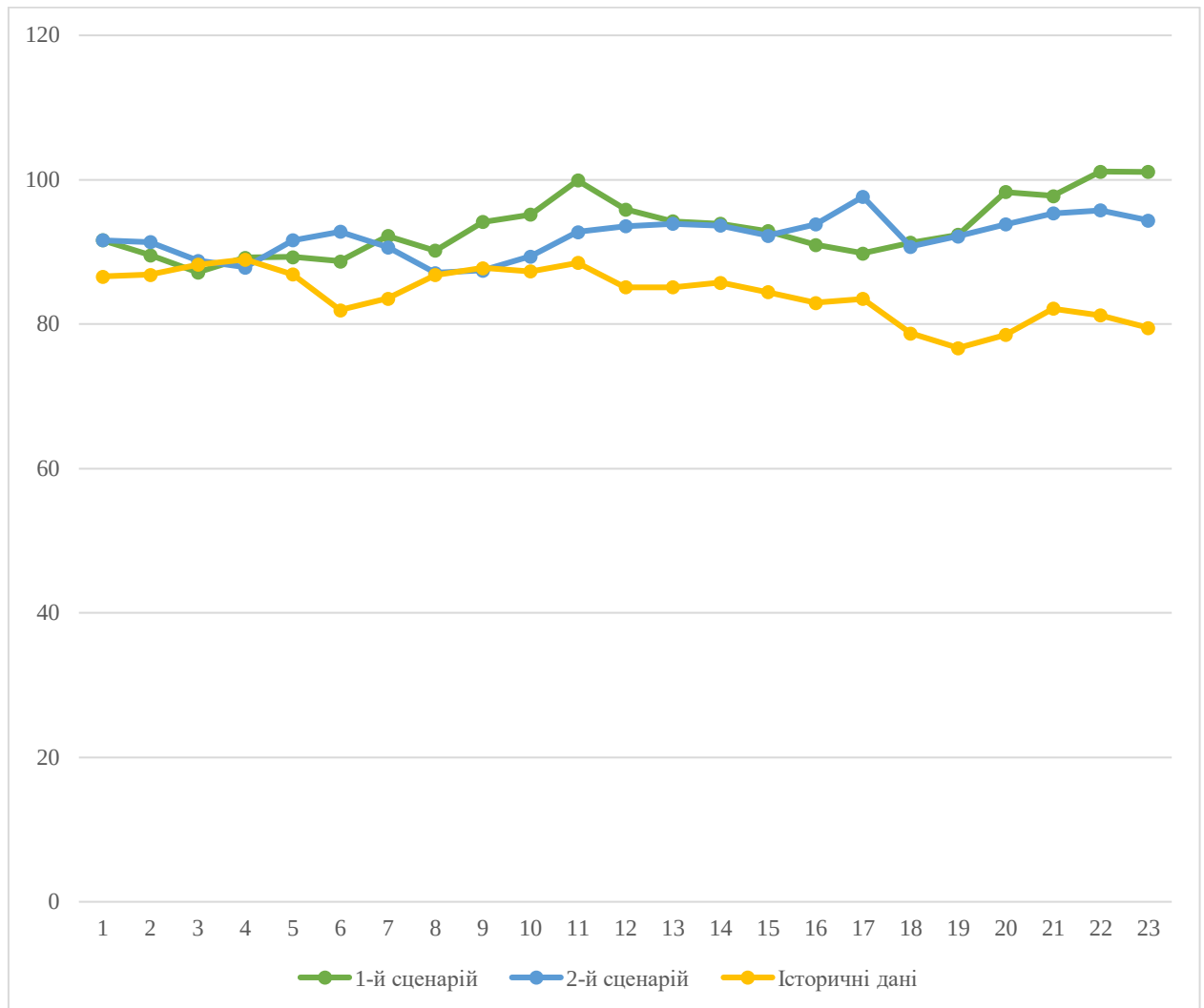


Рисунок 3.6 – Порівняння отриманих результатів з історичними даними
(розраховано автором на основі [27])

Геометричний броунівський рух допомагає побачити, якими шляхами можуть йти ціни на акції, і дозволяє бути готовими до майбутнього. Звичайно, ніколи неможливо передбачити точне майбутнє, але ці статистичні методи дають шанс створити надійні стратегії торгівлі та хеджування, на які можна покластися.

3.4 Реалізація моделі ARIMA

Модель ARIMA застосовується до даних часових рядів із властивістю нестационарності шляхом застосування різниці в необроблених спостереженнях, щоб зробити дані часових рядів стаціонарними. Після розрізнення модель ARIMA для послідовно корельованих даних може бути виражена у формі

$$y_t = \sum_{i=1}^p \alpha_i y_{t-i} + \sum_{j=1}^q \beta_j \varepsilon_{t-j} + \varepsilon_t, \quad (3.16)$$

де y_t – залежна змінна в період t (в нашому дослідженні - Crude Oil Futures);

p – кількість лагів залежної змінної;

α_i – коефіцієнт перед залежною змінною з лагом i ;

q – кількість лагів випадкової помилки;

β_j – коефіцієнт перед випадковою помилкою з лагом j ;

ε_t – випадкова помилка моделі, $\varepsilon_t \sim WN(0, \sigma^2)$.

Для дослідження збираються ціни на ф'ючерси з вересня 2021 року по серпень 2022 року. Ми можемо спостерігати, що ціна на ф'ючерси зростала в період з січня 2022 року по червень 2022 року, а потім почала різко спадати, як показано на рис. 3.1. Через різкі варіації традиційним моделям ARIMA або моделям SVM важко вловити ці зміни [57].



Рисунок 3.7 – Динаміка цін
(розраховано автором на основі [27])

Ф'ючерсні ціни, як і ціна будь-якого іншого фінансового активу, як правило, являють собою нестационарний ряд, як показано за допомогою графічного аналізу (рис. 3.7), і це підтверджується результатами тесту Дікі-Фуллера, який має p -значення $= 0,304523$. Вибір першої різниці випадків не вирішує проблему стаціонарності. Тому при аналізі фінансових даних зазвичай використовуються не самі ціни, а скоріше щоденна прибутковість, побудована на основі цих цін.

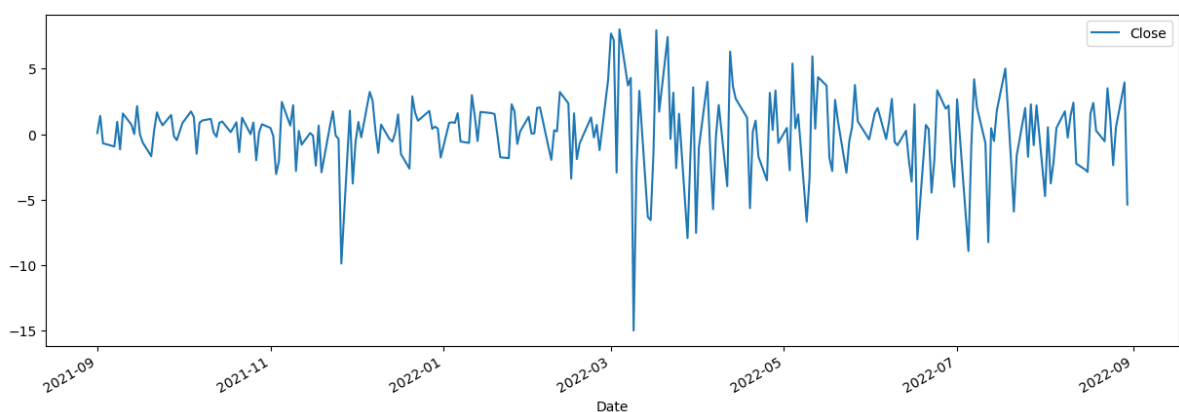


Рисунок 3.8 – Стаціонарність даних
(розраховано автором на основі [27])

Для побудови ряду щоденних прибутків можна взяти натуральний логарифм цін або наступну формулу:

$$r_t = \frac{(p_t - p_{t-1})}{p_{t-1}}, \quad (3.17)$$

де r_t – доходність в день t ,

p_t – ціна в день t ,

p_{t-1} – ціна в день $t-1$ (тобто в попередній день).

У цій роботі часовий ряд було отримано шляхом логарифмування, але слід зазначити, що обидва методи дають практично ідентичні результати.

Графічно уявлення тимчасового ряду денних доходностей ф'ючерсу нагадує білий шум, однак у деякі моменти спостерігаються сплески, що характеризують кластеризацію волатильності [58].

Тест Дікі-Фуллера для низки доходностей показав стаціонарність низки всім розглянутих рівнів значимості.

```
ADF Statistic: -1.959689
p-value: 0.304523
Critical Values:
    1%: -3.456
    5%: -2.873
   10%: -2.573
```

Рисунок 3.9 – Результати тесту Дікі - Фуллера
(розраховано автором на основі даних)

Тому немає необхідності брати різницю, параметр інтеграції моделі d ARIMA(p,d,q) дорівнює 0. Далі розглянемо модель ARIMA($p,0,q$), яка є розширеною нотацією ARMA (p ,q) модель.

Для часового ряду щоденних прибутків LM тест показав, що значення p прагне до нуля, значно нижче 0,05. Таким чином, нульова гіпотеза відхиляється на рівні значущості 5%, що означає, що розглянутий часовий ряд прибутків містить умовну гетероскедастичність.

Подальша оцінка моделі ARIMA та створення прогнозів здійснюється за допомогою мови програмування python. На першому кроці алгоритм вибирає найкращу специфікацію ARIMA(p,0,q) та створює різні моделі на цій основі вибір комбінацій параметрів p і q, кожен зі значеннями від 0 до 5, а потім порівнює інформаційні критерії AIC отриманих моделей. Найкращою вважається модель з найменшим критерієм.

Щоб побудувати динамічну прогностичну модель на основі гібридної моделі ARIMA, описаний вище алгоритм повторюється для кожного наступного спостереження за принципом ковзного вікна, де формується нова вибірка величини для прогнозування кожного окремого спостереження.

Ця модель вирішує проблему стандартного регресійного прогнозування, яка полягає в тому, що прогнозування часових рядів базується на одному рівнянні, параметри якого оцінюються з усіх історичних даних, введених у модель. Однак при аналізі фінансових інструментів ця стратегія неефективна, оскільки, згідно з теорією ефективного ринку, ціни на фінансові інструменти є випадковими коливаннями і не можуть бути передбачені. Все, що можна побачити, це деякий тренд, який може тривати кілька періодів, але не є постійним, і закономірності спаду волатильності після зовнішніх шоків. Модель моментуму також не може передбачити точні значення ціни, але вона здатна розпізнати і передбачити наступні розвороти трендів, тобто короткострокові періоди, в яких зростання змінюється спадом або навпаки.

У цій роботі розмір ковзного вікна становить 255. Оскільки вихідна вибірка складається з 225 спостережень, було виконано 225 ітерацій, тобто вибрано 225 оптимальних гібридних моделей ARIMA, на основі яких робляться прогнози на майбутній день, тобто на період після останнього в підвибірці.

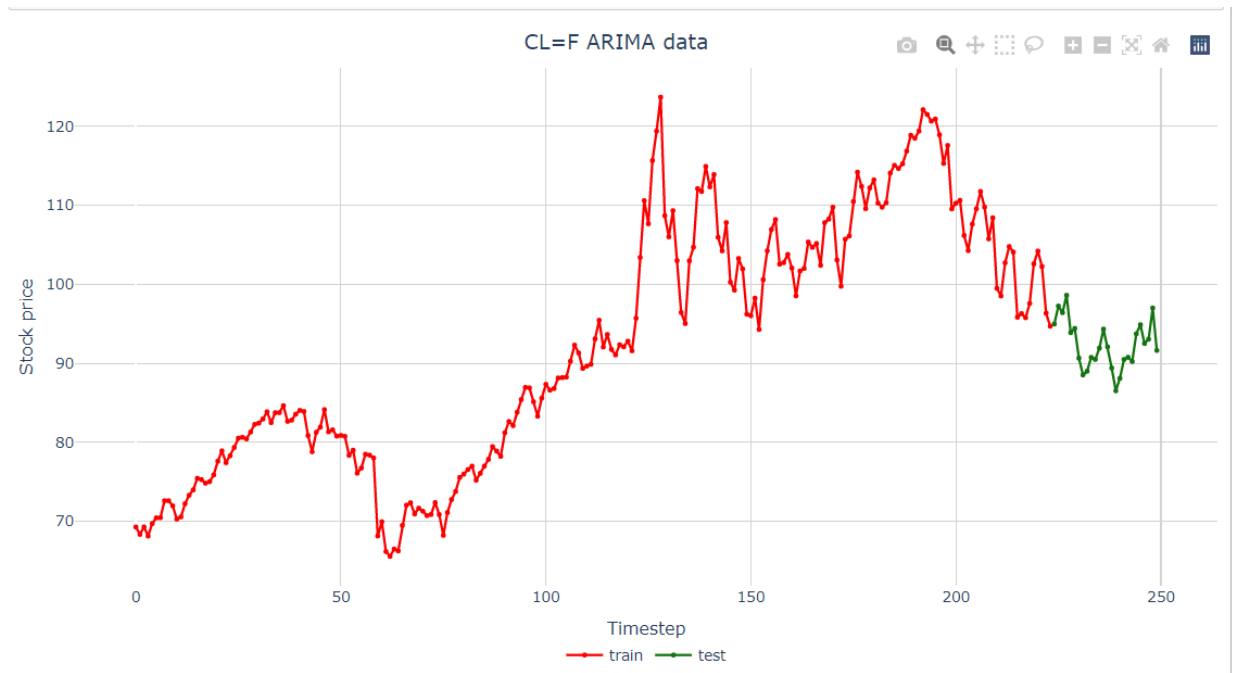


Рисунок 3.10 – Візуалізація тренувального та тестового сету
(розраховано автором на основі [27])

Результатом роботи алгоритму став набір прогнозів на період з 01.09.2022 по 30.09.2022 (рис. 3.11).

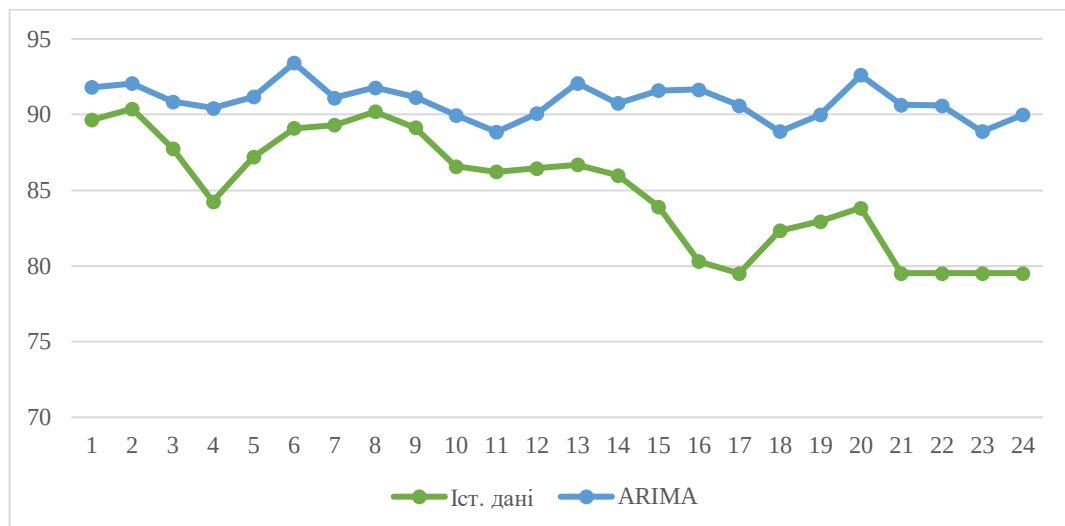


Рисунок 3.11 – Отримані прогнози
(розраховано автором на основі [27])

Графічний аналіз отриманого ряду прогнозів показує, що загалом модель вловлює волатильність, але слабо прогнозує її амплітуду. Причому що менше волатильність початкового часового ряду, то слабкіша реакція моделі.

Ціна, побудована на основі спрогнозованих гібридною ARIMA моделлю доходностей, досить точно відображає тренд та загальну довгострокову динаміку, проте малочутлива до сильних коливань.

3.5 Модель Random Forest

Для прогнозування використовується регресійна модель випадкового лісу. Це передбачить мінімуми та максимуми на наступні кілька торгових днів, включаючи майбутні ціни на один місяць на ф'ючерси сировини нафти. Результат купівлі, продажу або утримання базується на прогнозованих значеннях.

Історичні біржові дані часових рядів спочатку експоненціально згладжуються. Застосовується експоненціальне згладжування збільшення ваги для останніх спостережень і експоненціальне зменшення ваг для минулих спостережень. Експоненціально згладжену статистику ряду Y можна рекурсивно обчислити як:

$$S_0 = Y_0,$$

$$\text{для } t > 0, S_t = \alpha * Y_t + (1 - \alpha) * S_{t-1}, \quad (3.18)$$

де α – коефіцієнт згладжування, $0 < \alpha < 1$.

Більші значення α зменшують ступінь згладжування. Коли $\alpha = 1$, згладжена статистика стає рівною фактичному спостереженню. S_t можна обчислити, якщо доступні два спостереження. Це згладжування усуває випадкові коливання або шум із історичних даних, дозволяючи моделі легко ідентифікувати довгострокову цінову тенденцію в поведінці ціни на ф'ючерси. Потім технічні показники обчислюються на основі експоненціально згладжених даних часового ряду, які потім організовуються в матрицю ознак. Прогнозована ціль на день і розраховується таким чином:

$$target_i = Sign(close_{i+d} - close_i), \quad (3.19)$$

де d – кількість днів, за якими повинен бути зроблений прогноз.

Якщо цільове значення дорівнює +1, це вказує на позитивний зсув ціни через d днів, а -1 вказує на негативний зсув через d днів. Цільові значення призначаються як мітки до i -го рядка в матриці ознак.

Таблиця 3.2 – Дані, необхідні для прогнозування

Назва значення	Опис
Date	Дата ціни.
Open	Open price of the stock, at the beginning of trading day.
High	Найвища точка курсу акції в торговий день.
Low	Найнижча ціна акцій у торговий день.
Close	Ціна акції наприкінці торгового дня.
Adj close	Коригована ціна закриття.
Volume	Кількість проданих акцій на ринку за період.

Джерело: [46]

Перед запуском необхідно інсталиювати наступні бібліотеки:

- 1) Pandas – бібліотека Python, яка завантажує файл даних як фрейм даних Pandas для аналізу даних.
- 2) Matplotlib – пакет Python для побудови.
- 3) Scikit-learn – бібліотека Python з відкритим вихідним кодом для науки про дані, яка підтримує моделі машинного навчання, попередню обробку, оцінку моделей і навчальні програми. Він також діє як підбібліотека для `train_test_split`, `RandomForestRegressor`, `StandardScaler`, `RandomizedSearchCV` і метрик.
- 4) Numpy – бібліотека Python, яка працює з масивами.
- 5) Yfinance – бібліотека Python з відкритим кодом для доступу до фінансових даних.

Інформація, необхідна для цього проекту, є історичними даними. Він включає дату, ціну відкриття, ціну закриття, високу ціну, низьку ціну та обсяг торгівлі для кожного торгового дня. Трейдери використовують ці дані, щоб оцінити волатильність акцій (табл.3.2).

Дані витягуються за допомогою скрипту python. Тут скрипт python використовується для отримання даних за допомогою yfinance. Дані цін на ф'ючерси завантажуються з 1 вересня 2021 р. до 1 вересня 2022 р. Завантажені

біржові дані завантажуються у фрейм даних і перетворюються на файл CSV (значення, розділені комами).

```
clfdata = yf.download("CL=F", start="2021-09-01", end="2022-08-31")
clf_df = pd.DataFrame(clfdata)
clf_df.to_csv("clf_data.csv")
```

Рисунок 3.11 – Код для створення дата фрейму
(впроваджено автором на основі методів програмування)

Графік, який відображає коливання скоригованої ціни закриття виглядає наступним чином:

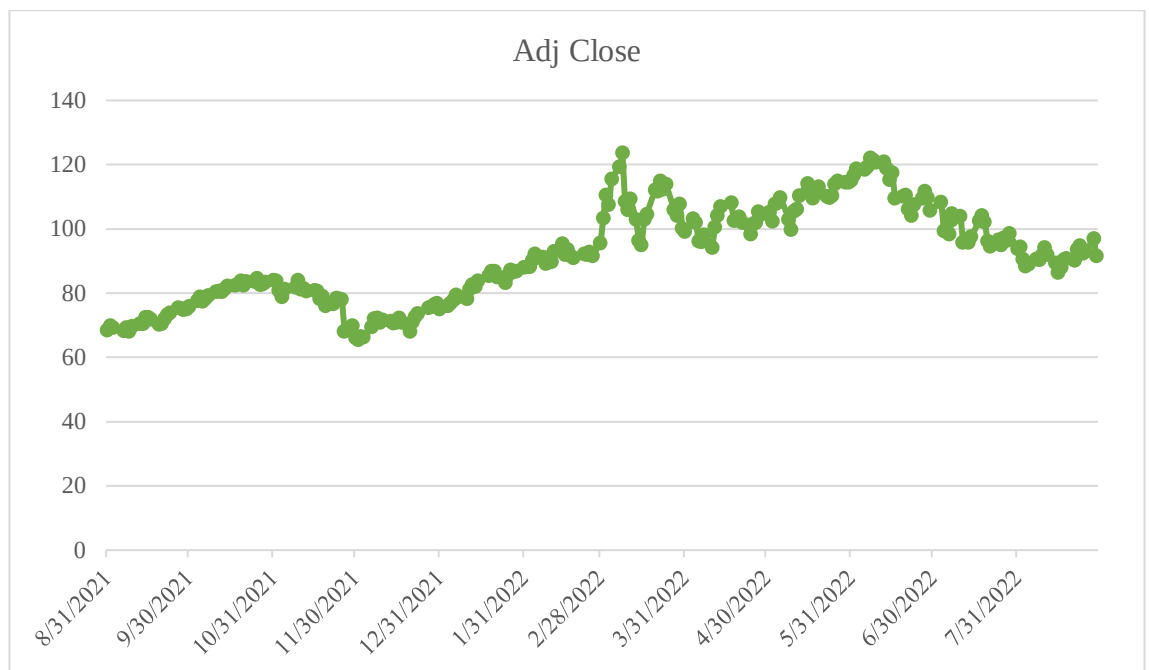


Рисунок 3.12 – Візуалізація даних
(розраховано автором на основі [27])

Попередня обробка даних – це етапи підготовки даних для моделі машинного навчання. Попередня обробка перетворює необроблені дані у формат, який модель може прийняти та працювати з ним. Цей дизайн має на меті мати набір даних, який приймає модель і розуміє алгоритм. У наборі даних можуть бути відсутні значення, а інформація може бути надлишковою, нерелевантною або шумною. Очищення даних – це форма попередньої

обробки, яка видаляє відсутні або неузгоджені значення та змінює індекс. На додаток до вибору функцій, налаштування гіперпараметрів і стандартизації даних.

```
df = pd.read_csv("clf_data.csv")
df.set_index("Date", inplace=True)
df.dropna(inplace=True)
```

Рисунок 3.13 – Очищення даних

(впроваджено автором на основі методів програмування)

Тут вибираються функції x і y для отримання набору даних для моделі. Функції x і y оголошуються для набору даних для навчання та тестування.

Функції – це стовпці в наборі даних. Вибір функцій є однією з важливих концепцій програм машинного навчання, які мають величезний вплив на продуктивність моделі. У майбутньому виборі не кожен стовпець потрібен. Ці вибрані функції мають вплив і сприяють отриманню прогнозу. Непотрібні функції знижують загальну продуктивність набору тестів. Спосіб вибору ф'ючерсів полягає в з'ясуванні найважливіших ознак, важливості ознаки. Sklearn має важливість функцій і модуль вибору функцій, який можна реалізувати. За допомогою модуля важливості ознак оцінка надається кожній ознаці в даних. Функції з найвищими балами є найбільш релевантними, а змінні вихідного звуку гарантовані. Переваги використання вибору функцій включають підвищення точності, зменшення переобладнання, скорочення часу навчання та покращення візуалізації даних. Чим більше функцій, тим швидше за все модель постраждає від переобладнання.

X містить значення для стовпців *open*, *high*, *low*, *close* і *adj close*. y містить значення для стовпця *adj-close*. Інші стовпці, включно з об'ємом, не вибрано для процесу, оскільки вони не знадобляться. Було використано п'ять функцій.

Перед моделюванням набір даних потрібно розділити на набір даних для навчання та тестування.

Навчальний набір: це підмножина набору даних, що використовується для створення та підгонки прогнозних моделей. Навчальний набір створюється шляхом побудови сценарію навчального набору даних, який генерує функції навчального набору з параметрів введення та вихідних даних про ціну акцій. Дані вводяться в модель для навчання. Модель навчається на основі цих даних.

Тестовий набір: це підмножина набору даних для оцінки ймовірної майбутньої продуктивності моделі. Це хороший стандарт для оцінки моделі. Тестовий набір використовується проти прогнозованого набору даних і тестування навченої моделі. Модель не бачила цю частину набору. Він використовується для оцінки.

Гіперпараметри - це налаштування моделі. Важливо мати параметри, налаштовані для оптимізації продуктивності. Гіперпараметри вирішують головну проблему в машинному навчанні, яка полягає в переобладнанні. У цьому проекті я використовував перехресну перевірку випадкового пошуку.

Гіперпараметри в моделі випадкового лісу використовуються або для збільшення передбачуваної потужності моделі, або для пришвидшення моделі.

Для регресійної моделі випадкового лісу найкраще враховувати такі параметри:

- 1) `n_estimators` – кількість дерев у лісі
- 2) `max_depth` – максимальна глибина в дереві
- 3) `min_samples_split` – мінімальна кількість точок даних перед поділом вибірки
- 4) `min_samples_leaf` – мінімальна кількість листових вузлів, які повинні бути вибілковими
- 5) `bootstrap` – вибірка для точок даних, істинних чи хибних
- 6) `random_state` – згенеровані випадкові числа для випадкового лісу.

Тепер набір даних готовий для моделі. Першим кроком є вибір значення для випадкового стану та побудова дерева на основі кількості

випадкових станів. Випадковий ліс запобігає переобладнанню, створюючи випадкові підмножини функцій і створюючи менші дерева з цими підмножинами. Щоб побудувати випадковий ліс, дані повинні бути навчені. Крім того, тут застосовуються параметри налаштування гіперпараметрів. Отримаємо прогнозовані значення для наступних 66 торгових днів.

```
[2802.86056944 2269.85750483 2477.74466762 2766.66301296
2374.42506466 4186.38753688 2634.95184493 3224.6825587 2642.15134695
2855.06226344 3131.29479236 2468.950334 4471.22437158 4372.35380957
2449.15157999 4226.55254199 2639.50649724 2433.13013304 2574.02973146
4442.45695459 3804.28322217 3462.75922217 3693.31345703 2953.07781528
2977.10962773 4365.6066488 3207.05099634 2557.40480942 4457.36409033
3246.82609651 3909.51829541 2895.40942318 2839.78738238 2721.42121206
2278.46803987 2782.98224145 3380.38209671 2514.85785607 2365.21841972
2941.80436097 3470.12492676 4515.03218359 4251.40260934 3916.72142383
2659.2645638 2976.54650081 3184.4597124 4439.67785669 2751.49107469
2279.15900854 4203.51592741 3727.24902539 2951.28400113 2811.7957286
2872.68410491 2426.05153603 2344.69569602 2737.90027594 2884.65725974
3286.42364014 3556.61905762 2395.49148071 4146.90240527 2874.4843484
3965.50093457 2931.24300486 4194.48325 3075.12630176 4523.85240527
3225.66883659 2882.05245894 2792.77252876 3809.20848584 3695.51101709
2475.63809759 2489.38631159 2789.6221085 2276.46002652 2659.14468915
2641.1087841]
```

Рисунок 3.14 – Отримані прогнози
(розраховано автором на основі [27])

Технічні індикатори – це важливі параметри, які розраховуються на основі часових рядів даних цін на ф'ючерси, щоб передбачити напрямок розвитку фінансових ринків. Це інструменти, які зазвичай використовуються трейдерами для пошуку сигналів вгору або вниз. Технічні індикатори, які ми використовували, перераховані нижче.

RSI – це індикатор моменту, який визначає, перекуплений чи перепроданий ф'ючерс. Ф'ючерс вважається викупленими, якщо попит викликає необґрунтоване підвищення ціни. Ця умова зазвичай не тлумачиться як ознака того, що акція переоцінена та ймовірно впаде. Ф'ючерси вважаються розпроданими, коли ціна падає значно нижче їх справжньої вартості. Це результат панічної розпродажі. RSI коливається від 0 до 100, і загалом RSI

вище 70 може означати, що акція перекуплена, а коли RSI нижче 30, це може означати, що акція перепродана.

Формула для розрахунку RSI така:

$$RSI = 100 - \frac{100}{1 + RS}; \quad (3.20)$$

$$RS = \frac{\text{Середній приріст за останні 14 днів}}{\text{Середня втрата за останні 14 днів}}. \quad (3.21)$$

Осцилятор – різновид технічного індикатора, що характеризується зміною певного діапазону, він вимірюється у відносних, а не абсолютних величинах.

Стохастичний осцилятор (%K) показує положення ціни закриття поточного дня відносно серії мінімумів і максимумів за певний період у минулому. Для дослідження використовувався період 14 днів.

Формула для розрахунку стохастичного осцилятора така:

$$\%K = 100 * \frac{(C - L_{14})}{(H_{14} - L_{14})}, \quad (3.22)$$

де C – поточна ціна закриття,

L_{14} – найнижчий мінімум за останні 14 днів,

H_{14} – найвищий максимум за останні 14 днів.

Стохастичний осцилятор слідує за швидкістю або імпульсом ціни. Зазвичай моментум змінюється перед зміною ціни. Він вимірює рівень ціни закриття відносно низько-високої області протягом певного періоду часу.

Відсотковий діапазон Вільямса показує положення ціни закриття відносно високої ціни за певний період у минулому. У цій роботі, як і у випадку з попередніми показниками, період становить 14 днів.

Значення індикатора змінюється від -100 до 0. Значення вище -20 вказує на перекуплений ринок і є сигналом до продажу ф'ючерсів, значення нижче -80 вказує на перепроданість і є сигналом до покупки.

Формула для розрахунку цього показника виглядає наступним чином:

$$\%R = 100 * \frac{(H_{14} - C)}{(H_{14} - L_{14})} * -100, \quad (3.23)$$

де C – поточна ціна закриття,

L_{14} – найнижчий мінімум за останні 14 днів,

H_{14} – найвищий максимум за останні 14 днів.

Williams %R коливається від -100 до 0. Коли його значення вище -20, це вказує на сигнал про продаж, а коли його значення нижче -80, це вказує на сигнал покупки.

Індикатор зміни ціни – це осцилятор, який показує зміну ціни у відсотках за певний період часу. У цій роботі період становить 20 днів і тому розраховується для часового ряду з 21-го спостереження.

Показник розраховується за такою формулою:

$$PROC(t) = \frac{C(t) - C(t - n)}{C(t - n)}, \quad (3.24)$$

де $PROC(t)$ – швидкість зміни ціни в момент t ,

$C(t)$ – ціна закриття в момент t .

Він вимірює останню зміну ціни відносно ціни за n днів тому.

Індикатор конвергенції та дивергенції ковзного середнього використовується для аналізу сили та напрямку тренду, а також для пошуку точок розвороту. MACD базується на різниці між двома експоненціальними ковзними середніми, розрахованими для різних періодів - довшого та коротшого [59-60]. Коротке середнє - це 12-денне експоненціальне ковзне середнє, довге - 26-денне експоненціальне ковзне середнє.

$$MACD = EMA_{12}(C) - EMA_{26}(C), \quad (3.25)$$

$$SignalLine = EMA_9(MACD), \quad (3.26)$$

де $MACD$ – розбіжність конвергенції ковзного середнього

C – завершальна серія

EMA_n – n -денне експоненціальне ковзне середнє.

Розрахуємо відповідні технічні індикатори. Остаточний набір даних складається з 5 змінних.

- 1) MAE – вимірює середню величину помилок у наборі прогнозів без урахування їхнього напрямку,
- 2) MSE – є сумою абсолютного значення похибки. Середня квадратична помилка також визначає ефективність моделі,
- 3) RMSE – є стандартним відхиленням залишків (помилки передбачення),
- 4) R^2 – вказує, наскільки добре модель відповідає даному набору даних

Було створено масив даних із прогнозованими значеннями на наступний місяць. В місяці 21 торговий день. Було взято потрібні майбутні дні з передбачених 66 торгових днів. Дати та ціни для цих майбутніх днів конвертуються у файли CSV.

Щоб знайти ціну пропозиції, пропозиції та утримати ціни, трейдери намагаються отримати прибуток, продаючи за найвищою ціною, купуючи за найнижчою ціною та утримуючи ціну, коли нічого не відбувається. Наприклад; Інвестор купує акцію на найнижчій точці за 5 доларів, а ціна, за прогнозами, досягне 20 доларів за 21 день, цей інвестор буде тримати свої акції до 21 дня, щоб продати з прибутком. У цьому контексті максимальна ціна — це ціна продажу, а мінімальна — ціна покупки.

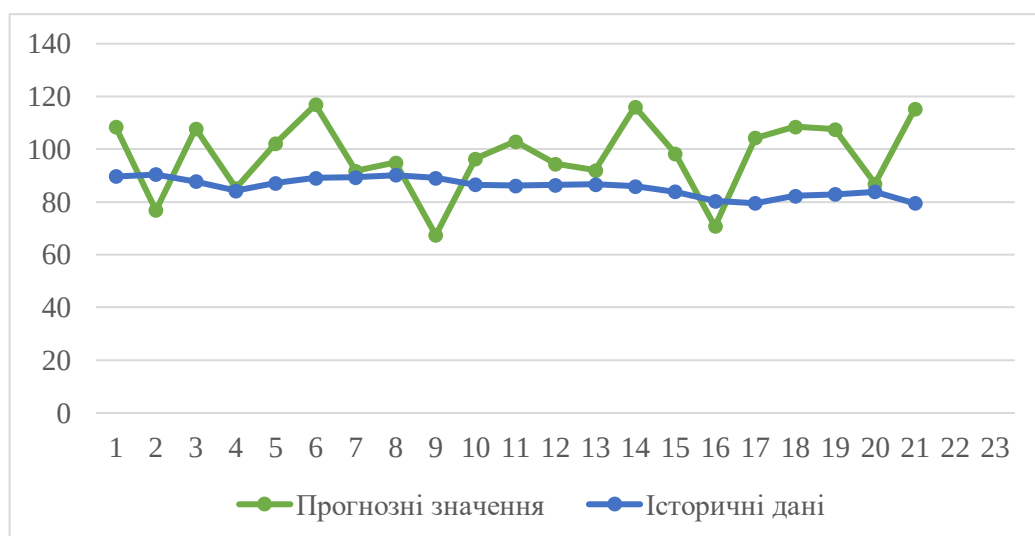


Рисунок 3.15 – Отримані прогнози
(розраховано автором на основі [27])

Таблиця 3.3 – Помилки моделей

Модель	MSE, %
Random Forest	15,25
ARIMA	6,85
GBM	6,99

Розраховано автором на основі даних

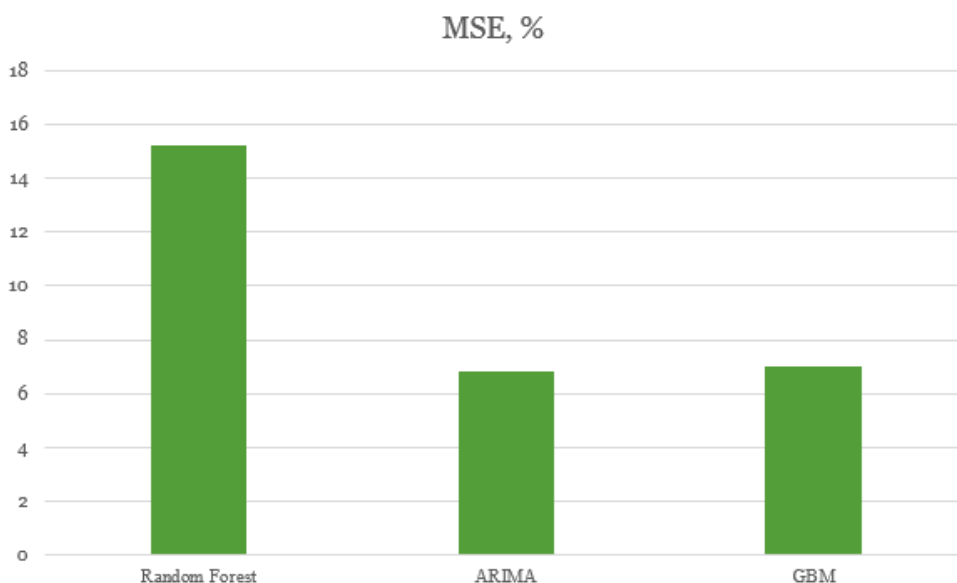


Рисунок 3.16 – Порівняння отриманих прогнозів
(розраховано автором на основі [27])

Таким чином, було отримано наступний графік. Значення були зпрогнозовані не зовсім точно, але через те, наскільки мінливим є фондовий ринок, не існує ідеальних чи точних систем прогнозування. Загалом, цей метод здебільшого швидкий, простий і гнучкий, але не позбавлений деяких обмежень.

Висновки до розділу 3

1. Були порівняні прогнози моделей за аналогічні періоди, що відповідають тестовій вибірці моделі, тобто з 01 серпня 2022 року по 31 серпня 2022 року.

2. Було розраховано статистичні показники точності моделей: середньоквадратична помилка (MSE). Результати представлено у таблиці 3.1.
3. Відзначено, що внаслідок різної чутливості моделей вони можуть по-різному показати себе в різні часові періоди.
4. Період, розглянутий для порівняння моделей, характеризується досить високою волатильністю.
5. Можливо, порівняння моделей на спокійнішій ділянці, тобто ділянці з меншою волатильністю, покаже перевагу моделі машинного навчання над статистичною моделлю часових рядів.

ВИСНОВКИ

Фондова біржа є складним інструментом, який може приносити своїм учасникам величезні прибутки. Але в той же час це і дуже ризикований інструмент, торгівля яким цілком може виявитися не вигідною. Незважаючи на високий рівень ризику, можливість отримання високих доходів завжди привертала увагу інвесторів до фінансових активів, що торгуються на біржах.

Уникнути збитків можна, якщо знати майбутню динаміку цін інструментів фондового ринку. Тому з самого його створення, незважаючи на гіпотези випадкового блукання та ефективності ринку, згідно з якими вартість акції - непрогнозована випадкова величина, розробляються різні методи та моделі прогнозування даної динаміки. Деякі намагаються передбачити значення ціни, деякі - значення прибутковості, деякі - напрямок руху цін і трендів у цих цінах. Якись моделі будують прогнози з урахуванням зовнішніх чинників, інші - з урахуванням історичної динаміки вартості.

У цій роботі для прогнозування було досліджено три моделі, засновані на аналізі історичної динаміки цін на ф'ючерсні контракти. Всі три моделі показують прогнозну динаміку цього значення, за рахунок аналізу трендів та коливань.

Перша модель заснована на процесі в якому логарифм величини, що випадково змінюється, слідує броунівському руху (також відомому як вінеровський процес) з дрифтом. В результаті реалізації цієї моделі було отримано 2 сценарії і один з них був порівняний з історичними даними. Різниця між прогнозними та історичними даними була не великою, а це означає, що модель є адекватною та може застосовуватись у майбутніх дослідженнях чи для виконання певних завдань.

Друга модель базується на статистичних моделях часових рядів, таких як ARIMA для прогнозування загального тренду. Результатом моделі є низка спрогнозованих цін на 22 торгових дні. Ця модель показала хороші результати. Незважаючи на деякі помилки у прогнозуванні, вона чутлива до змін трендів і

дуже швидко вловлює розвороти ринку, даючи сигнали до купівлі чи продажу. Також було встановлено, що прогнозна точність цієї моделі по-різному поводить ся на різних ділянках ринку і більш ефективна за високої волатильності, тобто за великих стрибків вартості фінансових активів.

Третя модель, заснована на методі машинного навчання Random Forest, ґрунтується на індикаторах технічного аналізу, які використовуються для виявлення наявності та динаміки трендів у цінах акцій, а також для аналізу відповідності ціни її справедливому рівню. Результатом роботи моделі є бінарний ряд із прогнозами напрямів зміни ціни. Іншими словами, прогноз моделі - це індикатор того, чи зросте ціна в наступному періоді, чи впаде.

Хоча прогнози моделей не є повністю точними і не ліквідують ризик інвестування у фондовий ринок, вони можуть призвести до отримання прибутковості, яка в деякі періоди навіть випереджає прибутковість ринку.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Арбузов С. Ф'ючерний контракт. *Банківська енциклопедія*. Київ, 2011. С. 504.
2. Базилевич В., Шелудько В., Вірченко В. Фондовий ринок: Практикум. Знання, 2017. 718 с.
3. Еш С. Фінансовий ринок. Центр навч. літ., 2019. 528 с.
4. Кравченко Ю. Ринок цінних паперів. Дакор КНТ, 2009. 672 с.
5. Примостка Л. Фінансовий менеджмент у банку. 2-ге вид. Київ : КНЕУ, 2004. 468 с.
6. Сохацька О. Біржова справа: підручник. 2-ге вид. Тернопіль : Кондор, 2008. 632 с.
7. Товарні біржі в Україні: аналіз діяльності, законодавче поле, перспективи розвитку. Київ : За ред. Саблука П. Т., Шпичака О. М., 1997. 424 с.
8. Цінні папери та фондовий ринок в Україні. Нормативно-правове регулювання. КНТ, 2008. 392 с.
9. Alexakis P. On the Effect of Index Futures Trading on Stock Market Volatility. *International Research Journal of Finance*. 1995. No. 11. P. 7-20.
10. Bae S., Kwon T., Park J. "Futures Trading, Spot Market Volatility and Market Efficiency: The Case of the Korean Index Futures Markets. *The Journal of Futures Markets*. 2004. Vol. 24, no. 12.
11. Butterworth D. The Impact of Futures Trading on Underlying Stock Index Volatility: The Case of the FTSE Mid 250 Contract. *Applied Economics Letters*. 2004. Vol. 7. P. 439-442.
12. Chernenko S., Schwarz K., Wright J. The Information Content of Forward and Futures Prices: Market Expectations and the Price of Risk. Board of Governors of the Federal Reserve System International Finance Discussion Paper. 2004. No. 808.

13. Fama E. F., French K. R. Commodity Futures Prices: Some Evidence on Forecast Power, Premiums, and the Theory of Storage. *The Journal of Business*. 1987. Vol. 60 (1), no. 55-73. P. 55-73.
14. Gentile T., Fontanills G. *The Stock Market Course*. The Stock Market Course. 2001.
15. Gorton G., Rouwenhorst. Fact and Fantasies about Commodity Futures. *Financial Analysts Journal*. 2006. Vol. 62, no. 2.
16. McKenzie M., Brailsford T., Faff R. New Insights into the Impact of the Introduction of Futures Trading on Stock Price Volatility. *Journal of Futures Markets*. 2001. Vol. 21, no. 3. P. 237-255.
17. Schaede U. Forwards and futures in Tokugawa-period Japan: A new perspective on the Dōjima rice market". *Journal of Banking & Finance*. *Journal of Banking & Finance*. 1989. Vol. 13, no. 4-5. P. 487-513.
18. Valbuzzi B. "Swaps, Forwards, and Futures Strategies". CFA Institute. 2020. P. 23.
19. Working H. The theory of the price of storage. *The theory of the price of storage*. Vol. 39, no. 1254-1262.
20. Yu S. Index futures trading and spot price volatility. *Applied economics letters*. 2001. No. 8. P. 183-186.
21. Дзендзелюк О., Костів Л., Рабик В. Побудова arima моделей часових рядів для прогнозування метеоданих на мові програмування r. *Електроніка та інформаційні технології*. 2013. № 3. С. 211-219.
22. Любіч О., Домрачев В. Застосування моделі ARIMA для прогнозування доходів зведеного бюджету України. *Екон.-мат. моделювання соц.-екон. систем: Зб. наук. пр.* 2011. № 16. С. 33-37.
23. Цеслів О., Цеслів О., Дейнеко М. Економіко-математична модель ф'ючерсних контрактів. *Причорноморські економічні студії*. 2022. № 77. С. 185-189.

24. Awan T., Aslan F. Prediction of daily COVID-19 cases in European countries using automatic ARIMA model. Journal of Public Health Research. 2020. Vol. 9, no. 3. P. 227-233.
25. Butcher K. Futures Made Simple. Wrightbooks, 2013. 320 p.
26. Chodakowska E. Analiza wartości sprzedaży w przedsiębiorstwie handlowym z zastosowaniem modeli ARIMA. Zeszyty Naukowe Politechniki Białostockiej. Ekonomia i Zarządzanie. 2004. No. 9. P. 37-52.
27. Crude Oil Jan 23 (CL=F) Stock Price, News, Quote & History - Yahoo Finance. Yahoo Finance - Stock Market Live, Quotes, Business & Finance News. URL: https://finance.yahoo.com/quote/CL=F/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAIvB7TSbITdWApAgTotrUaJf4lYApTLimtJzNs7Gu10Cf2gpoa0oKoEXNiGPSRVjoalJ3BXkOUTZJtLipqCxwgi-BNxS5amkYzfvy20zsRa8dqzoNVvfzq324vzyTzYS-X_UaCX4TiN76fw1UJdvYI2lMBxj17FVimkWjBbnYqB4 (дата звернення: 23.11.2022).
28. Dharavath R., Khosla E. Seasonal ARIMA to Forecast Fruits and Vegetable Agricultural Prices. IEEE International Symposium on Smart Electronic Systems (iSES). 2019. P. 47-52.
29. Dickey D., Fuller W. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. Journal of American Statistical Association. 1979. Vol. 74, no. 366. P. 427-431.
30. Forecasting of demand using ARIMA model / J. Fattah et al. International Journal of Engineering Business Management. 2018. No. 10. P. 1-9.
31. Geiger J. U.S. Oil And Gas Rig Count Falls Slightly | OilPrice.com. OilPrice.com. URL: <https://oilprice.com/Energy/Crude-Oil/US-Oil-And-Gas-Rig-Count-Falls-Slightly.html> (дата звернення: 11.11.2022).
32. Hamilton J. Time Series Analysis. Princeton University Press, 1994. 820 p.

33. Heiberger R., Teles P. Displays for Direct Comparison of ARIMA Models. *The American Statistician*. 2002. Vol. 56, no. 2. P. 134.
34. *Time Series Analysis: Forecasting and Control* / G. Box et al. 5th ed. Wiley, 2015. 712 p.
35. Iordache R. China, Saudi Arabia stress importance of oil market stability as Xi visits Riyadh. *CNBC*. URL: <https://www.cnb.com/2022/12/09/china-saudi-arabia-stress-importance-of-oil-markets-stability.html> (дата звернення: 11.10.2022).
36. Nazarko J., Chodakowska E. Prognozowanie na podstawie modeli Arima. Białystok : Oficyna wydawnicza Politechniki Białostockiej, 2022. 2022 p.
37. Nazarko J., Rybaczuk M., Jurczuk A. Wpływ poziomu zakłóceń losowych na możliwość identyfikacji modeli ARIMA. *Taksonomia*. 2005. Vol. 12. P. 238-247.
38. Yildiz U. Simulating stock prices in Python using Geometric Brownian Motion. *Medium*. URL: <https://towardsdatascience.com/simulating-stock-prices-in-python-using-geometric-brownian-motion-8dfd6e8c6b18> (дата звернення: 17.10.2022).
39. Shadab A., Shadab S., Said S. Spatial forecasting of solar radiation using ARIMA model. *Remote Sensing Applications: Society and Environment*. 2020. Vol. 20, no. 100427.
40. Shumway R. H., Stoffer D. S. *ARIMA Models. Time Series*, 2019. 30 p.
41. Wallace J. Oil Market Takes Russian Price Cap in Stride. *WSJ*. URL: https://www.wsj.com/articles/oil-market-takes-russian-price-cap-in-stride-11670599278?mod=markets_lead_pos7 (дата звернення: 15.10.2022).
42. Мозговий О. Фондовий ринок. Київ : КНЕУ, 1999. 316 с.
43. Amit Y., Geman D. Shape Quantization and Recognition with Randomized Trees. *Neural Computation*. 1997. Vol. 9, no. 7. P. 1545-1588. URL: <https://doi.org/10.1162/neco.1997.9.7.1545> (дата звернення: 25.10.2022).

44. Chaudhary M. Random Forest Algorithm - How It Works and Why It Is So Effective. Hire the World's Most Deeply Vetted Remote Developers | Turing. URL: <https://www.turing.com/kb/random-forest-algorithm> (дата звернення: 17.11.2022).

45. Contributors to Wikimedia projects. Geometric Brownian motion - Wikipedia. Wikipedia, the free encyclopedia. URL: https://en.wikipedia.org/wiki/Geometric_Brownian_motion (дата звернення: 21.10.2022).

46. Contributors to Wikimedia projects. Predictive modelling - Wikipedia. Wikipedia, the free encyclopedia. URL: https://en.wikipedia.org/wiki/Predictive_modelling (дата звернення: 26.10.2022).

47. Contributors to Wikimedia projects. Random forest - Wikipedia. Wikipedia, the free encyclopedia. URL: https://en.wikipedia.org/wiki/Random_forest (дата звернення: 01.12.2022).

48. Generalised Geometric Brownian Motion: Theory and Applications to Option Pricing / V. Stojkoski et al. Entropy. 2020. Vol. 22, no. (12).

49. Hayes A. Futures Contract Definition: Types, Mechanics, and Uses in Trading. Investopedia. URL: <https://www.investopedia.com/terms/f/futurescontract.asp> (дата звернення: 21.11.2022).

50. Karatzas I. Brownian motion and stochastic calculus. 2nd ed. New York : Springer, 1996. 470 p.

51. Lawton G., Carew J. M., Burns E. What is Predictive Modeling?. Enterprise AI. URL: <https://www.techtarget.com/searchenterpriseai/definition/predictive-modeling#:~:text=Predictive%20modeling%20is%20a%20mathematical,forecast%20activity,%20behavior%20and%20trends> (дата звернення: 20.11.2022).

52. Löffler A. Brownian Motion. Springer Nature Switzerland AG, 2020. 125 p.

53. Piryonesi S. M., El-Diraby T. E. Role of Data Analytics in Infrastructure Asset Management: Overcoming Data Size and Quality Problems. *Journal of Transportation Engineering, Part B: Pavements*. 2020. Vol. 146, no. 2. P. 04020022. URL: <https://doi.org/10.1061/jpeodx.0000175> (дата звернення: 15.10.2022).

54. Predicting Your Business Future. Oracle NetSuite. URL: <https://www.netsuite.com/portal/resource/articles/financial-management/predictive-modeling.shtml> (дата звернення: 15.10.2022).

55. Smith C., Koning M. Decision Trees and Random Forests: A Visual Introduction For Beginners. Independently published, 2017. 168 p.

56. Suganthi K., Jayalalitha G. Geometric Brownian Motion in Stock Prices. *Journal of Physics: Conference Series*. 2019. Vol. 1377. P. 012016. URL: <https://doi.org/10.1088/1742-6596/1377/1/012016> (дата звернення: 15.11.2022).

57. Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1998. Vol. 20, no. 8. P. 832-844. URL: <https://doi.org/10.1109/34.709601> (дата звернення: 05.12.2022).

58. Topcu Guloksuz C. Geometric brownian motion approach to modelling stock prices. *Focus on Research in Contemporary Economics (FORCE)*. 2021. Vol. 2, no. 1. P. 53-63.

59. Understanding Futures. Schwab Brokerage. URL: <https://www.schwab.com/futures/what-are-futures> (дата звернення: 09.12.2022).

60. Woodward W. A., Sadler B. P., Robertson S. Time Series for Data Science: Analysis and Forecasting. Chapman and Hall/CRC, 2022. 506 p.

ДОДАТОК А

Код прогнозування ціна на ф'ючерси за допомогою GBM

```

import pandas as pd
import numpy as np
import yfinance as yf
import matplotlib.pyplot as plt
from pandas.plotting import register_matplotlib_converters
register_matplotlib_converters()
prices = yf.download("CL=F", start='2022-01-01', end='2022-08-31').reset_index(drop = False)[['Date', 'Close']]
plt.figure(figsize = (15, 5))
plt.plot(prices['Date'], prices['Close'])
plt.xlabel('Days')
plt.ylabel('Stock Prices, USD')
plt.show()
start_date = '2022-08-01'
end_date = '2022-08-31'
pred_end_date = '2022-09-30'
S_eon = yf.download("CL=F", start='2022-08-01', end='2022-08-31').reset_index(drop = False)[['Date', 'Close']]
print(S_eon.head())
print(S_eon.tail())
So = S_eon.loc[S_eon.shape[0] - 1, "Close"]
print(So)
dt = 1
print(dt)
n_of_weekdays = pd.date_range(start = pd.to_datetime(end_date,
format = "%Y-%m-%d") + pd.Timedelta('1 days'),
end = pd.to_datetime(pred_end_date,

```

```

format = "%Y-%m-%d")).to_series(
).map(lambda x:
1 if x.isoweekday() in range(1,6) else 0).sum()
T = n_of_wkdays
print(T)
N = T / dt
print(N)
t = np.arange(1, int(N) + 1)
print(t)
returns = (S_eon.loc[1:, 'Close'] - \
S_eon.shift(1).loc[1:, 'Close']) / \
S_eon.shift(1).loc[1:, 'Close']
print(returns.tolist())
mu = np.mean(returns)
print(mu)
sigma = np.std(returns)
print(sigma)
scen_size = 2
b = {str(scen): np.random.normal(0, 1, int(N)) for scen in range(1, scen_size + 1)}
print(b)
W = {str(scen): b[str(scen)].cumsum() for scen in range(1, scen_size + 1)}
print(W)
drift = (mu - 0.5 * sigma**2) * t
print("drift:\n", drift)
diffusion = {str(scen): sigma * W[str(scen)] for scen in range(1, scen_size + 1)}
print("diffusion:\n", diffusion)
S = np.array([So * np.exp(drift + diffusion[str(scen)]) for scen in range(1,
scen_size + 1)])
S = np.hstack((np.array([[So] for scen in range(scen_size)]), S))
print(S)

```

```
plt.figure(figsize = (20,10))

for i in range(scen_size):
    plt.title("Daily Volatility: " + str(sigma))
    plt.plot(pd.date_range(start = S_eon["Date"].max(),
        end = pred_end_date, freq = 'D').map(lambda x:
        x if x.isoweekday() in range(1, 6) else np.nan).dropna(), S[i, :])
    plt.ylabel('Prices, usd')
    plt.xlabel('Prediction Days')

plt.show()
```

ДОДАТОК Б

Код прогнозування ціна на ф'ючерси за допомогою Random Forest

```
import yfinance as yf
import datetime as dt
import pandas as pd
import numpy as np
from numpy import arange
import matplotlib.pyplot as plt
from pandas import read_csv
from sklearn import metrics
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import RandomizedSearchCV

sp500data = yf.download("CL=F", start="2021-09-01", end="2022-08-31")
sp500_df = pd.DataFrame(sp500data)
sp500_df.to_csv("sp500_data.csv")
read_df = pd.read_csv("sp500_data.csv")
read_df.set_index("Date", inplace=True)
read_df['Adj Close'].plot()
plt.ylabel("Adjusted Close Prices")
plt.show()
df = pd.read_csv("sp500_data.csv")
df.set_index("Date", inplace=True)
df.dropna(inplace=True)
x = df.iloc[:, 0:5].values
y = df.iloc[:, 4].values
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size=0.26,
random_state=0)
```

```

scale = StandardScaler()
x_train = scale.fit_transform(x_train)
x_test = scale.transform(x_test)
model = RandomForestRegressor(n_estimators=500, random_state=42,
min_samples_split=2, min_samples_leaf=1, max_depth=10, bootstrap=True)
model.fit(x_train, y_train)
predict = model.predict(x_test)
print(predict)
print(predict.shape)
print("Mean Absolute Error:", round(metrics.mean_absolute_error(y_test,
predict), 4))
print("Mean Squared Error:", round(metrics.mean_squared_error(y_test, predict),
4))
print("Root Mean Squared Error:",
round(np.sqrt(metrics.mean_squared_error(y_test, predict)), 4))
print("(R^2) Score:", round(metrics.r2_score(y_test, predict), 4))
print(f'Train Score : {model.score(x_train, y_train) * 100:.2f}% and Test Score :
{model.score(x_test, y_test) * 100:.2f}% using Random Tree Regressor.')
errors = abs(predict - y_test)
mape = 100 * (errors / y_test)
accuracy = 100 - np.mean(mape)
print('Accuracy:', round(accuracy, 2), '%.')
predictions = pd.DataFrame({"Predictions": predict},
index=pd.date_range(start=df.index[-1], periods=len(predict), freq="D"))
predictions.to_csv("Predicted-price-data.csv")
onemonth_df = pd.DataFrame(predictions[:21])
onemonth_df.to_csv("one-month-predictions.csv")
onemonth_df_pred = pd.read_csv("one-month-predictions.csv")
buy_price = min(onemonth_df_pred["Predictions"])
sell_price = max(onemonth_df_pred["Predictions"])

```

```
onemonth_buy = onemonth_df_pred.loc[onemonth_df_pred["Predictions"] ==  
buy_price]  
onemonth_sell = onemonth_df_pred.loc[onemonth_df_pred["Predictions"] ==  
sell_price]  
print("Buy price and date")  
print(onemonth_buy)  
print("Sell price and date")  
print(onemonth_sell)  
onemonth_df_pred["Predictions"].plot(figsize=(10, 5), title="Forecast for the next  
1 month", color="blue")  
plt.xlabel("Date")  
plt.ylabel("Price")  
plt.legend()  
plt.show()
```