

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**Сумський державний університет**

**Навчально-науковий інститут бізнесу, економіки та менеджменту**

**Кафедра економічної кібернетики**

«До захисту допущено»

Завідувач кафедри

\_\_\_\_\_ Віталія КОБІЙЧУК

“ \_ ” \_\_\_\_\_ 2024 р.

**КВАЛІФІКАЦІЙНА РОБОТА**

на здобуття освітнього ступеня бакалавра  
(бакалавр / магістр)

зі спеціальності 051 «Економіка»,  
(код та назва)

освітньо-професійної програми «Економічна кібернетика та бізнес аналітика»


(освітньо-професійної / освітньо-наукової)

(назва програми)

на тему: «Прогнозування прибутку комерційного банку на прикладі «Райффайзен Банк»

Здобувача (ки) групи ЕК-01а  
(шифр групи)

Іващенко Єгор Сергійович  
(прізвище, ім'я, по батькові)

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело. 

\_\_\_\_\_ (підпис)

Єгор ІВАЩЕНКО

\_\_\_\_\_ (Ім'я та ПРІЗВИЩЕ здобувача)

Керівник асистентка кафедри економічної кібернетики, доктор філософії, Олена КОЛОТІЛІНА

(посада, науковий ступінь, вчене звання, Ім'я та ПРІЗВИЩЕ)

(підпис)

Суми – 2024

Міністерство освіти і науки України  
Сумський державний університет  
Навчально-науковий інститут бізнес-технологій «УАБС»  
Кафедра економічної кібернетики

ЗАТВЕРДЖУЮ

Завідувач кафедри

к.е.н., доцентка

\_\_\_\_\_ В.В. Койбічук

“ \_ ” \_\_\_\_\_ 2024 р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ БАКАЛАВРА  
спеціальність 051 «Економіка (Економічна кібернетика)

студенту IV курсу, групи ЕК-01а

Іващенко Єгору Сергійовичу

1. Тема роботи «Прогнозування прибутку комерційного банку на прикладі «Райффайзен Банк»», затверджена наказом по університету від «08» травня 2024 року № 0486-VI.
2. Термін подання студентом закінченої роботи «30» травня 2024 року.
3. Мета кваліфікаційної роботи – розробка моделі прогнозування прибутку комерційного банку, що дасть змогу отримати прогнозні значення на 4 квартали вперед.
4. Об'єкт дослідження – показники діяльності українського комерційного акціонерного банку «Райффайзен Банк».
5. Предмет дослідження – прогнозування прибутку комерційного банку на прикладі «Райффайзен Банк».
6. Кваліфікаційна робота виконується на матеріалах Міністерства фінансів України, сайту банку «Райффайзен», аналітичних оглядів та наукових публікацій вітчизняних та зарубіжних авторів, присвячених дослідженню

факторів впливу на прибуток комерційного банку та побудовам регресійних моделей.

7. Орієнтовний план кваліфікаційної роботи, терміни подання розділів керівникові та зміст завдань для виконання поставленої мети

Розділ 1. Аналіз об'єкта дослідження та побудова моделі – 14 травня 2024 року.

У розділі 1 необхідно розглянути фактори впливу на прибуток комерційного банку; дослідити стан банківської системи; розробити концептуальну постановку завдання та визначитися з вимогами до моделей.

Розділ 2. Реалізація моделювання та здійснення прогнозування – 30 травня 2024 року.

У розділі 2 провести побудову чотирьох моделей регресій на основі вхідних показників; оцінити їх адекватність; зробити прогнозування чистого прибутку за рахунок нових спостережень та визначити їх точність.

#### 8. Консультації з роботи:

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
1			21.03
2			14.05

9. Дата видачі завдання: «21» березня 2024 року.

Керівник кваліфікаційної роботи \_\_\_\_\_  
( підпис)

О.В. Колотіліна

Завдання до виконання одержав \_\_\_\_\_  
( підпис)

Є.С. Іващенко

## АНОТАЦІЯ

Іващенко Є. С. Прогнозування прибутку комерційного банку на прикладі «Райффайзен Банк»: робота на здобуття кваліфікаційного ступеня бакалавра : спец. 051 – економіка (економічна кібернетика) / наук. кер. О. В. Колотіліна. Суми: Сумський державний університет, 2024. 61 с.

Було розглянуто нинішній стан банківської системи, шлях її розвитку, а також елементи взаємодії в банківській діяльності. Визначено основні чинники впливу на зміну чистого прибутку, як зовнішні, так і внутрішні. Було описано переваги та недоліки кожної з регресійних моделей, які будуть застосовані. Наведено спектр вимог та припущень стосовно дослідження. Сформовано базу показників, які братимуть участь в аналізі.

Перед побудовою різних моделей, було зазначено основні алгоритми та постулати роботи регресій. За рахунок вхідних даних було сформовано чотири регресійні моделі, здійснено оцінку їх адекватності та точності прогнозів. Беручи за основу готові моделі виконувався прогноз чистого прибутку на 4 квартали вперед використовуючи нові спостереження.

Було висвітлено основні способи покращення продуктивності моделей для подальшого застосування. Наведено головні критерії, стосовно економічної значущості та цінності результатів для банку «Райффайзен».

Ключові слова: комерційні банки, прибуток банку, фактори впливу на прибуток, регресія, методики побудови регресійних моделей, прогнозування.

## ЗМІСТ

ВСТУП .....	1
РОЗДІЛ 1. АНАЛІЗ ПОКАЗНИКІВ КОМЕРЦІЙНИХ БАНКІВ ТА ВИЯВЛЕННЯ ВЗАЄМОЗВ'ЯЗКІВ .....	3
1.1 Дослідження стану банківської системи та визначення елементів для побудови математичної моделі.....	3
1.2 Опис зовнішнього середовища та характеру взаємодії елементів в банківській діяльності .....	7
1.3 Концептуальна постановка завдання, опис вимог до моделі та вхідних показників .....	12
РОЗДІЛ 2. РЕАЛІЗАЦІЯ МОДЕЛЮВАННЯ ТА ЗДІЙСНЕННЯ ПРОГНОЗУВАННЯ .....	14
2.1 Побудова моделей, оцінка їх адекватності та інтерпретація результатів.....	14
2.1.1 Лінійна регресія .....	14
2.1.2 Лассо регресія.....	18
2.1.3 Гребенева регресія .....	23
2.1.4 Поліноміальна регресія .....	27
2.2 Здійснення прогнозування значень чистого прибутку та вибір найбільш оптимальної моделі .....	30
ВИСНОВКИ.....	32
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ .....	34
ДОДАТОК А .....	39

## ВСТУП

На сьогодні на стан ринкової економіки значну роль відіграють прибутки з банківського сектору, основу якого складають комерційні банки. Отримання стабільно високого фінансового результату є головною задачею аналітичного відділу банку. Достатній прибуток в банку дозволяє формувати резерви, підвищити власний капітал, а також сприяє в утриманні конкурентної позиції та довготривалого функціонування. Пряма залежність між рівнем капіталізації та надійності банку веде до сильного зв'язку з усіма учасниками ринку, тобто від стабільності банку залежить і фінансова ситуація інших учасників.

Питання моделювання та прогнозування прибутків та інших фінансових показників в комерційних банках було розглянуто в декількох вітчизняних наукових роботах, серед яких можна виділити наступні статті[1; 2]. Міжнародна практика прогнозування фінансових результатів банку також знайшла своє висвітлення в працях[3; 4; 5]. Сама методика розрахунків прогнозу прибутку не відрізняється, формується на основі низки загальних принципів та методів прогнозування. Було звернено увагу на важливість виконання цього етапу в діяльності банку .зادля стабільного функціонування та отримання позитивних фінансових результатів.

Прогнозування прибутку є одним із найголовніших аспектів діяльності кожного бізнесу. Воно дає змогу компаніям обирати напрям розвитку, а також допомагає визначати потенційні проблеми з діяльності ще до моменту їх виникнення. Постійне використання практики на підприємстві з прогнозування доходів веде до покращення фінансових результатів та обрання найбільш вигідної інвестиційної стратегії. Важливою перевагою прогнозування управлінцем також є можливість коригування діяльністю підприємства в реальному часі. Воно дає змогу наявно побачити відхилення в роботі, що своєю чергою дає вам можливість внести корективи для покращення фінансової ситуації[6].

Актуальність проведення даного дослідження підтверджується необхідністю банків підтримання фінансової стабільності, оскільки прибуток банку відіграє ключову роль в фінансовій стійкості. Прогнозування допоможе керівникам контролювати здатність банку, щодо виконання своїх функцій. Окрім того прогнозування доходів банку дає змогу виявляти ризики та позбаватися від них заздалегідь, задля уникнення потенційних фінансових труднощів. Актуальність даної теми також проявляється у високій конкурентоспроможності в банківському сектору. Тому банкам необхідно також приділяти увагу прогнозним прибуткам конкурентів, щоб вчасно коригувати процентною ставкою або маркетинговою стратегією, щоб забезпечувати конкуренцію.

Метою роботи є розробка моделі прогнозування прибутку комерційного банку, що дасть змогу отримати прогнозні значення на 4 квартали вперед.

Завданнями дослідження є: обрання факторів впливу на прибуток банку; побудова моделі та оцінка її адекватності; використання моделі для отримання прогнозних значень прибутку банку.

Об'єктом дослідження будуть братися показники діяльності українського комерційного акціонерного банку «Райффайзен Банк», який є найбільшим українським банком з іноземним капіталом.

Предметом дослідження є прогнозування прибутку комерційного банку на прикладі «Райффайзен Банк». Увага буде зосереджена на виявленні факторів впливу на прибуток в банку, а також побудова моделі та її застосування для отримання прогнозного значення прибутку.

Для виконання розрахунків в нашому дослідженні було використано такі методики: лінійна регресія; поліноміальна регресія та LASSO та Ridge.

Розрахунки будуть проводитися на мові програмування Python в інтегрованому середовищі розробки Google Colaboratory.

## РОЗДІЛ 1. АНАЛІЗ ОБ'ЄКТА ДОСЛІДЖЕННЯ ТА ПОБУДОВА МОДЕЛІ

### 1.1 Дослідження стану банківської системи та визначення елементів для побудови математичної моделі

Попри невелику історію з початку формування, банківська система України вже зуміла пройти через критичні кризи, події яких відбувалися в 1998 р., 2004 р., 2008 р. та 2013-2015 рр. Серед основних причин формування цих криз можна виділити зовнішні фактори, приміром, світова фінансова криза. Проте варто також враховувати внутрішні неупереджені чинники, що сприяють до уразливості в банківській системі, ну і слід не забувати про специфіку функціонування вітчизняної системи в банківській сфері[7].

Всього варто виділити 4 етапи формування структури банківської системи України, починаючи з 1991 році. Останній на цей час четвертий етап почав формуватися в 2006 році і активний по теперішній час, хоча корективи стосовно змін почали вноситися ще в грудні 2000 року. Серед особливостей даного етапу варто виділити:

- банківська система України набуває дворівневої форми, де на першому місці Національний банк України. Другий же рівень розділяється на 4 блоки: державні банки, універсальні банки, спеціалізовані банки та філії іноземних банків. Також варто виділити підрівень банківські об'єднання;
- банківські установи можуть створювати або у формі відкритого АТ або кооперативного банку з організаційно-правової точки зору;
- за територіальною приналежністю банки розділяють: місцеві, регіональні, міжрегіональні та міжнародні;
- засновницький капітал на етапі реєстрації має бути не меншим за 10 мільйонів євро[8].

Незважаючи на те, що банки мають сплачувати податок в розмірі майже половини від загальних доходів, сумарний прибуток банків України в 2023 збільшився на 12%. Серед банків з іноземним капіталом «Райффайзен»



проявив найбільшу прибутковість, заробивши 4,78 млрд грн, що дорівнює 21,3% від сумарного прибутку банків з іноземним капіталом(рис.1.1)[9]. За розміром чистих активів банк займає 4 місце в банківській системі, збільшивши цей показник на 13,7% порівняно з минулим роком. Незважаючи на те, що він взяв паузу стосовно участі в державній програмі «Доступні кредити 5-7-9%», «Райффайзен» все одно найбільше з усіх зарубіжних кредиторів прокредитував економіку України. Також даний банк продовжує очолювати український ринок стосовно обсягів ввезеної готівкової іноземної валюти в країну.

Високий рівень довіри до банку підтверджується збільшенням коштів клієнтів, які зросли на 10%, а також позитивним трендом до збільшення активних клієнтів у порівняння з 2022 роком[10].

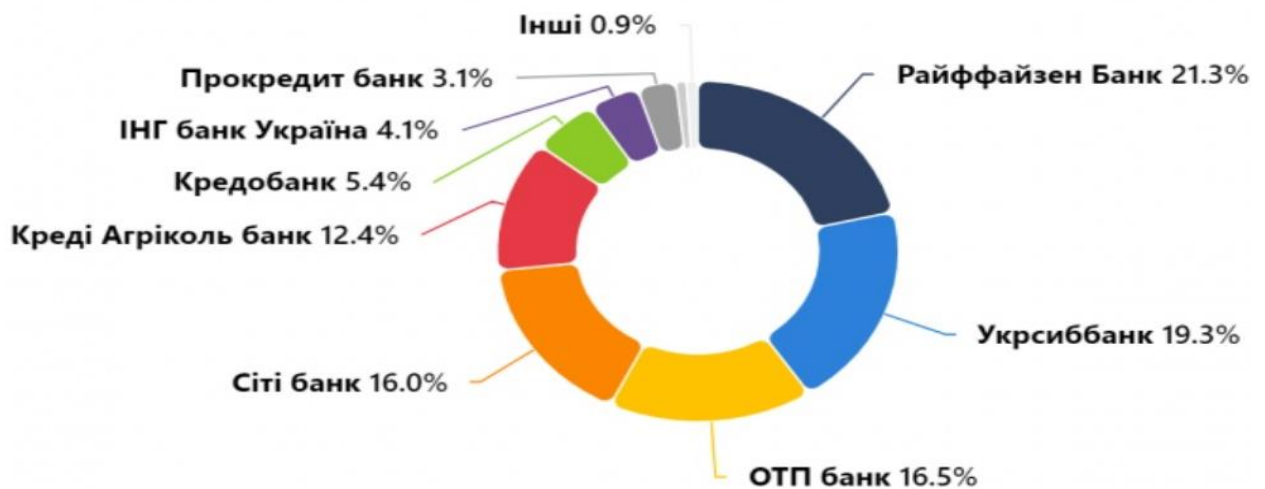


Рис.1.1 - Відсоткові частки прибутків банків з іноземним капіталом

Для регулювання стабільності банківської системи в Україні працює система нагляду від НБУ. За цим стоять відповідні органи банку, які мають чітко перевірений механізм контролю(рис.1.2). За визначення потенційних ризиків або проблем в діяльності банків відповідає рейтингова система CAMELSO, яка допомагає розраховувати показники реагування на початкових стадіях[7].

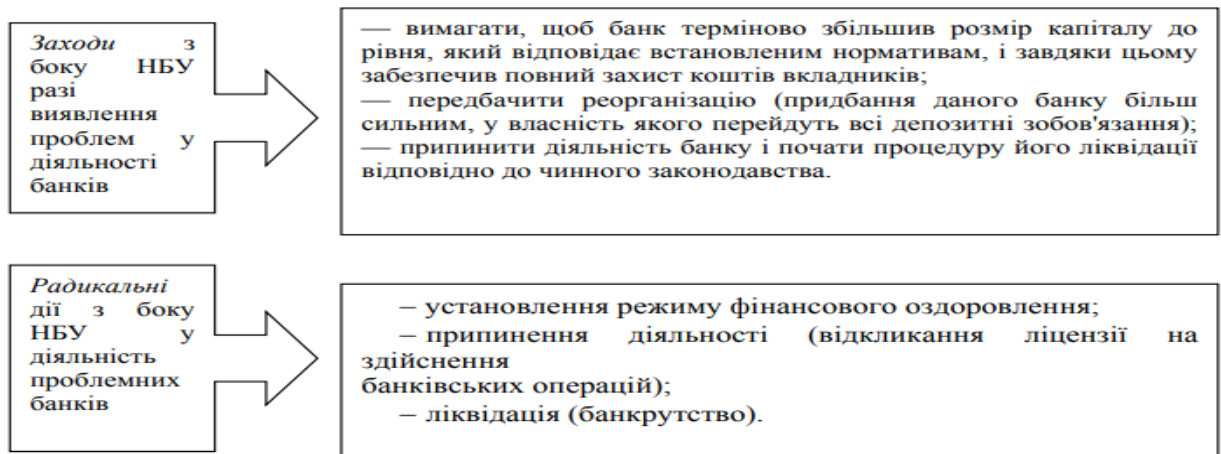


Рис.1.2 – Дії з боку Національного банку стосовно проблем в діяльності банківських установ

Серед тенденцій в розвитку вітчизняної банківської системи варто виділити наступне:

- перенесення уваги з кількісних та цінових характеристики на сервісні та якісні. Прикладом може слугувати індивідуальність в процесі роботи з клієнтом, час та рівень обслуговування;
- із-за проявлення бажання клієнтів стосовно цілодобової роботи системи відбулося зміщення фокуса роботи на засобах управління, при цьому не прив'язуючись до певного місця або часу[11].

Серед загальних показників стану фінансових результатів банківської установи можна зазначити прибуток, доходи, збитки та витрати. В результаті виконання базових задач, таких як грошові, кредитні та інші фінансові операції, відбувається генерація прибутку в комерційного банку.

Прибутковість комерційного банку виступає важливою складовою в описі життєвого циклу усіх ланок та структур в банківській діяльності в процесі переходу на новий етап розвитку та має залежність від групи факторів, які можна побачити на рис.1.3.

Не зважаючи на численні фактори впливу на фінансовий результат, в основному, залежність можна пояснити двома основними статтями: доходи та

витрати комерційного банку. Тобто проведення аналізу цих проміжних ознак дасть в теорії можливість точно змоделювати прибуток банку[12].

Для більш ефективного прогнозування, знання та розуміння факторів впливу на прибуток є невід’ємним етапом при підготовці до дослідження. Незважаючи на велику вибірку чинників та різне варіювання впливу, зазвичай їх можна поділити на групи: зовнішні та внутрішні. Ендогенні частіше всього асоціюються з набуванням доходів або витрат в процесі виконання банківських послуг і характеризуються залежністю від внутрішньої роботи банку. Стосовно екзогенних факторів варто зауважити про те, що у банків відсутня можливість впливу на їх формування або контроль над ними. Прикладом цього можуть слугувати стан міжнародних фондових і валютних ринків[13].

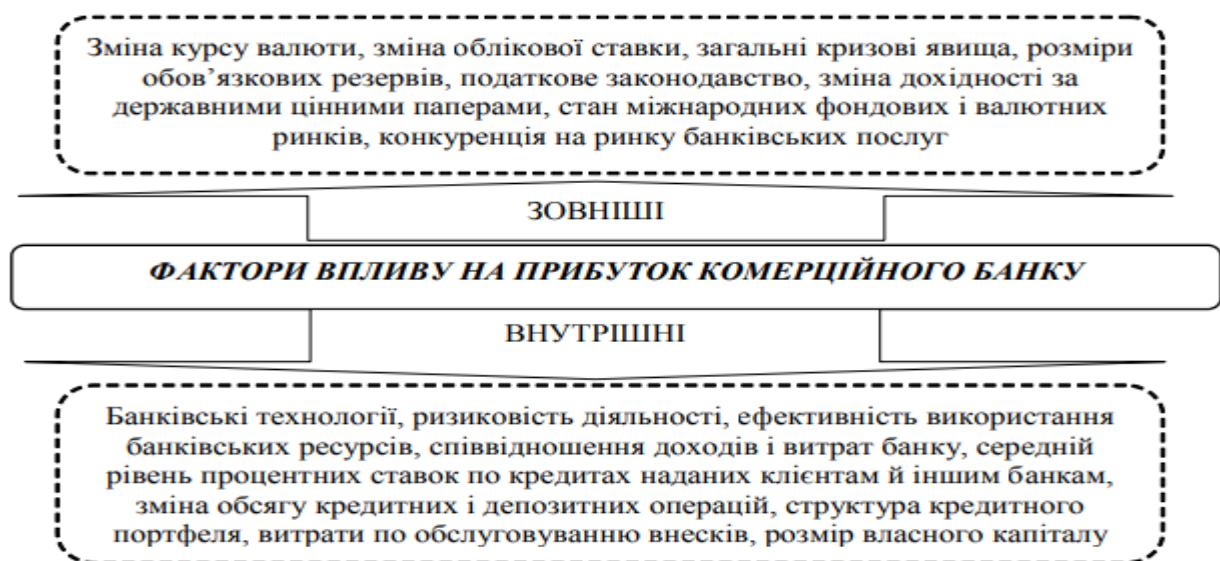


Рис.1.3 – Чинники впливу на прибуток комерційного банку[14]

Розглянемо внутрішні фактори. Серед множини ендогенних чинників особливу увагу варто приділити такому показнику, як рівень ризику. Даний показник висвітлює рівень нечіткості в процесі діяльності та можливі проблеми та наслідки у разі неуспіху певної стратегії розвитку. Задля отримання позитивного фінансового результату банкам рекомендується доглядати за перебігом ризиків, а також аналізувати їх. Приймання ризиків, які

відповідають фінансовій ситуації банку, а також з можливістю контролю – оптимальне рішення[15].

Значний вплив на прибутковість також мають активи. Оскільки доходи являються основною складовою прибутку, така характеристика активів, як дохідність, має безпосередній зв'язок з прибутковістю в банках. Оборотноість активів має прямий зв'язок з прибутком, зі збільшенням швидкості оборотів відповідно і покращується фінансовий результат[16, с.71].

Якщо не брати до уваги внутрішні та зовнішні чинники впливу, прибутковість комерційного банку може бути розрахована за допомогою доходів та витрат. Позитивний результат після здійснення різниці між доходами та витратами свідчить про отримання прибутку банком. Окрім різниці між доходами та витратами важливу роль відіграє розмір капіталу банку, який дає можливість покращувати темпи зростання прибутковості. Прибуток та капітал банку взаємодоповнюються та мають сильний прямий зв'язок. Важливо також зазначити про нерозподілений прибуток, який є важливою складовою статутного капіталу.

Банк може бути третейським суддею між економічними індивідами на ринку, при наявності в них вільних коштів або потребі в них. Це сприяє виникненню нових унікальних чинників впливу на прибуток[16, с.76]. Серед яких можна виділити відносні показники дохідності та витратності активних та пасивних операцій банку відповідно, які своєю чергою поділяються на процентні та непроцентні підгрупи[17, с.144; 18, с.139].

## 1.2 Опис зовнішнього середовища та характеру взаємодії елементів в банківській діяльності

Для повноти розуміння формування прибутку банку важливо також звертати увагу на зовнішнє середовище, знати загальний спектр факторів впливу. Тепер більш детально розберемося із зовнішніми чинниками. Економічну складову можуть пояснювати показники кризи в країні або зміни

курсу валют. Варто також зазначити про ступінь оподаткування, який має пряму залежність з прибутковістю банку[19].

Важливу роль також відіграє сектор соціально-політичних факторів. Головним показником із даної підгрупи варто виділити рівень довіри населення до банківського сектору, що є основою діяльності банків[20].

Не менш важливим сектором є природні умови. Прикладом може слугувати поганий врожай в сільськогосподарській діяльності, що в подальшому може впливати зміну платіжний баланс країни та відповідно на фінансові результати банків[16, с.70].

Інфляція також залишається важливим зовнішнім фактором впливу на грошовий стан банків. Серед основних цілей впливу інфляції можна виділити фінансову ситуацію кредиторів, фінансові активи та зобов'язання банку у валюті. Пріоритет в банках має бути сконцентрований на довгострокових активах.

Головною запорукою виконання планових задач та ефективного підтримання життєвого циклу роботи в банку є забезпечення взаємозв'язку між елементами. Для ефективного менеджменту цих взаємодій необхідно розбиратися в низці тем, а саме:

- важливим елементом в функціонуванні банку є взаємодія з клієнтами. Взаємозв'язок не обмежується лише під час виконання фінансових операцій, але й через постійне керування та формування відносин на довгий період з ними. Прикладом покращення регулювання взаємин з клієнтами може слугувати впровадження систем CRM, які допомагають глядіти за інформацією про клієнтів, що дає можливість прогнозувати їх потреби та поведінкові зміни[21];
- важливою взаємодією елементів в банківській сфері також залишається зв'язок між фінансовими продуктами та послугами. Прикладом залежності внутрішньої складової від зовнішньої може слугувати вплив зміни ставок на депозити, або в загальному зміни цін на ресурси, на

фінансовий результат банку. За рахунок оптимального регулювання загального розміру позик та заощаджень клієнтів в банку під відсоток можна уникнути або ж мінімізувати кредитний ризик[22];

- також вагомою складовою залишається участь банку в керуванні ризиками. В разі дотримання стабільного взаємозв'язку між відділами та виконання необхідних функцій, банк може створити унікальну результативну систему коригування ризиками[23];
- не варто забувати також про зв'язок між внутрішніми процесами в діяльності банку. Дотримання якісного контролю та ефективного управління таких внутрішніх процесів, як керування активами, пасивами, резервами або ж здійснення точної оцінки кредитоспроможності може впливати на результативність банку[24];
- ну і взаємозв'язок з ринком також відіграє ключову роль в функціонуванні банків. Важливо підтримувати конкурентоспроможність в банківському секторі, не зважаючи на регуляторні вимоги або ж процентні ставки. Це необхідно для формування оптимальних цілей розвитку та отримання необхідних фінансових результатів[25].

Аналізуючи основні підходи стосовно формування моделі прогнозування прибутку банку, можна виділити три основні напрямки, які мають свою плюси та мінуси. Методика з побудовою фінансових моделей бере за основу показники результатів з фінансових звітів. При використанні даного підходу може виникати проблема у зв'язку з певною обмеженістю методу контролювати зміни екзогенних факторів. Застосування нейронних мереж або інших методів штучного інтелекту потребує наявності багаточисельної бази даних, а також складних за тяжкістю розрахунків. Використання економетричних моделей в деяких випадках може призвести до неточності прогнозів, оскільки дана методика бере за основу лінійний взаємозв'язок між залежною та незалежними ознаками, що в реальності не завжди відповідає дійсності.

Схема на рис.1.4 висвітлює критичні економетричні методики аналізу прибутку.



Рис.1.4 – Основні підходи до прогнозування прибутку банку[26]

Також не варто забувати про методи математичної статистики, головною задачею яких є виявлення об'єктивного існуючого взаємозв'язку між факторами. Важливий розділ в даній групі методів є регресійний аналіз, який був обраний для проведення розрахунків дипломної роботи. Він висвітлює залежність однієї ознаки від іншої, і на основі сили та напряму зв'язку можна будувати модель або ж рівняння, для подальшого прогнозу певної характеристики.

Існує безліч різновидів регресій, проте для аналізу було обрано наступні: лінійна, поліноміальна, гребенева та Лассо. Основні недоліки та переваги методик висвітлені в табл.1.1.

Табл.1.1 – Характеристика регресійних підходів до моделювання прибутку банку

Назва підходу	Переваги	Недоліки
Лінійна	— порівняно легкий алгоритм, тому можна застосовувати навіть на системах з не високим	— досить часто виникає проблема недонавчання із-за неналежного опису та поганої адаптації

	<p>рівнем обчислювальної потужності;</p> <ul style="list-style-type: none"> <li>— математичне рівняння, як результат виконання лінійної регресії, легке для розуміння та трактування</li> </ul>	<p>рівняння функції до даних;</p> <ul style="list-style-type: none"> <li>— причиною низької точності моделі та поганої продуктивності являються викиди, із-за сильної чутливості до них[27]</li> </ul>
Поліноміальна	<ul style="list-style-type: none"> <li>— відсутня залежність від розміру набору даних;</li> <li>— наявність численних функцій, які можна пристосувати;</li> <li>— вирішення нелінійних проблем з хорошою точністю</li> </ul>	<ul style="list-style-type: none"> <li>— визначається високою чутливістю до викидів;</li> <li>— від вибору степеня поліному залежить продуктивність та точність моделі[28];</li> </ul>
Гребенева	<ul style="list-style-type: none"> <li>— надає унікальне та стабільне рішення у випадках, високо-розмірний набір даних;</li> <li>— добре працює в умовах мультиколінеарності;</li> </ul>	<ul style="list-style-type: none"> <li>— із-за наявності параметру штрафу більше 0, відбувається зміщення оцінки коефіцієнтів моделі;</li> <li>— оцінки коефіцієнтів моделі майже ніколи не дорівнюють нулю[29]</li> </ul>
Лассо	<ul style="list-style-type: none"> <li>— може ідентифікувати незначущі змінні;</li> <li>— при відповідному значенні регуляризатора не перенавчається при наближенні кількості ознак до кількості спостережень</li> </ul>	<ul style="list-style-type: none"> <li>— більш тяжкий процес обрахунку довірчих інтервалів або перевірку гіпотез;</li> <li>— у разі, коли значення регуляризатора не дорівнює 1, наявна похибка між фактичним значенням та прогнозом[30]</li> </ul>

Міра ефективності управління фінансами має пряму залежність від підходів, які будуть використовувати аналітики під час проведення дослідження на визначення джерел та факторів, які впливають на прибуток



банківської установи. Проблема виникає в точності відображення взаємозв'язків елементів структури, які виступають основою для утворення позитивного фінансового результату.

### 1.3 Концептуальна постановка завдання, опис вимог до моделі та вхідних показників

Постановка основних питань пов'язаних з темою дослідження, на які потрібно дати відповідь, є важливим етапом, який є основою для початку розрахунків та запорукою точних результатів. Стосовно питання факторів, які впливають на прибуток, та наявність зовнішніх чинників впливу на прибутковість банку було розглянуто в попередніх підрозділах. Ключовим питанням залишається саме політика банку «Райффайзен», яка безпосередньо має значну роль при формуванні прибутку. Відповідь на дане питання може бути такою: сам банк може прямо впливати на власну дохідність, за рахунок зміни вектору послуг, керуючись попитом клієнтів та відповідного сектору ринку. Окрім цього, банк може змінити прибутковість перекваліфікувавшись в інший тип комерційних банків, тобто обрати спеціалізацію в певних операціях[31].

Стосовно гіпотез особливостей поведінки показників фінансової діяльності банку варто виділити наступні тези:

1. Рівень відсоткових ставок виставлений НБУ має прямий вплив на прибутковість банку.
2. Між розміром зобов'язань та прибутком існує обернена залежність.

Зазначення низки припущень є також обов'язковим елементом підготовки до аналізу, серед яких опускання дії зовнішніх соціально-політичних чинників та природних факторів та неврахування їх в моделі. Обмеження в дослідження, в основному, стосуються збору фінансових результатів, пов'язані з політикою банку та конфіденційністю даних.

Перед початком розрахунків також критично важливо визначити вимоги до моделі. Сформована модель має чітко описувати природу змін прибутку та роботи прогноз за рахунок вхідних змінних. Цінність дослідження має бути значною для відділу фінансів банку для коригувань ціль розвитку, а також Національному банку для порівняння статистичних результатів банків та регулювання політикою в банківському секторі. Довірчі межі отриманого прогнозного значення мають дорівнювати  $\pm 15\%$  від фактичних значень. Для наочного розуміння, результати мають бути відображенні в графічному вигляді з коментарями трактування стосовно них.

Побудова моделей буде виконуватися за допомогою умовного алгоритму, наведеного на рис.1.5 Із-за ручного вводу та формування даних, етап обробки був пропущений.



Рис.1.5 – Етапи формування моделі[32]

Для дослідження було взято ряд ендогенних факторів: розмір кредитного портфеля; резерви під кредитні ризики; активи; капіталізація банку; кошти банку; загальний обсяг внесків; депозити фізичних осіб; зобов'язання; чистий прибуток. Також в аналізі буде брати участь низка зовнішніх чинників, серед яких: рівень інфляції, відсоткова ставка, зміна курсу валюти. Інтервал даних буде знаходитися в межах від 2012 до 2024 року. Значення показників було отримано за допомогою наступних ресурсів: НБУ, міністерство фінансів України та державна служба статистики України.

## РОЗДІЛ 2. РЕАЛІЗАЦІЯ МОДЕЛЮВАННЯ ТА ЗДІЙСНЕННЯ ПРОГНОЗУВАННЯ

### 2.1 Побудова моделей, оцінка їх адекватності та інтерпретація результатів

#### 2.1.1 Лінійна регресія

Суть лінійної регресії полягає у відображенні лінійного зв'язку між факторними ознаками та залежною за допомогою рівняння виду:

$$y = a + b * x + \varepsilon , \quad (2.1)$$

де  $a$ ,  $b$  – параметри лінійної регресії;

$\varepsilon$  – значення похибки.

Коефіцієнт  $a$  є константою, значення якого відображає вплив ознак, які не були включені в модель. Параметр  $b$  визначає силу впливу предиктора на залежну змінну. Орієнтуючись на величину коефіцієнта можна визначити на скільки зміниться  $Y$  в середньому, при зміні відповідного чинника на 1. Напрямок впливу встановлюється за допомогою знаку біля відповідного параметру регресії. Знак мінус біля оцінки визначає обернений зв'язок між  $Y$  та  $X_i$ , знак плюс висвітлює прямий зв'язок.

Після висунення гіпотези про опис залежності ознак за допомогою рівняння, наступним етапом розрахунків буде оцінка коефіцієнтів. Існує декілька способів виконання цього завдання, а саме: графічний метод з використанням первинних даних, або більш класична методика найменших квадратів. Сутність першого полягає у проведенні прямої через вказані точки (значення незалежних ознак). В такому випадку коефіцієнти оцінюються методом підбору, чим ближче розташовуються всі точки до прямої, тим оптимальніші будуть значення параметрів.

МНК також базується на підборі оцінок параметрів, тільки тепер пріоритетом буде мінімізація відхилень розкиду значень фактичних  $Y$  та розрахованих  $Y$  навколо їх середніх (SSE). Формула 2.2 аналітично висвітлює алгоритм даної методики[33]:

$$SSE = \sum_{i=1}^n (y_i - (a + bx_i))^2 \quad (2.2)$$

Узагальнюючи, другу частину рівняння можна охарактеризувати як нев'язку, тобто сильну відмінність варіативності даних.

Якість побудованої моделі, в основному, оцінюється по визначеному значенні коефіцієнта детермінації. Чим ближче ця величина до 0, тим слабкіший зв'язок, чим ближче до 1 – тим сильніший. В загальному, характеризує величину поясненої варіації результативної ознаки наявними предикторами моделі.

Серед найбільш частих проблем варто виділити недолік множинного коефіцієнта детермінації, який чітко висвітлюється в прикладі «Квартет Анскомбе», та вірогідність високого рівня залежності між факторними характеристиками, тобто мультиколінеарність[34].

Прогнозування результативної ознаки здійснюється за рахунок підстановки наявних значень незалежних ознак до відповідних параметрів в рівнянні регресії.

Перед початком моделювання відбувається завантаження в середовище розробки необхідних бібліотек мови програмування Python, методи яких направлені на маніпуляції з даними, статистичне моделювання та відповідну візуалізацію результатів. Використавши метод «read» з бібліотеки pandas, зчитались дані з Excel файлу в датафрейм, з яким надалі будемо працювати(див. ДОДАТОК А).

Для полегшення розуміння результатів моделі та попереднього виявлення аномалій або викидів був розрахований тест Шапіро-Уїлка та р-значення для визначення нормальності чи ненормальності розподілу цільової змінної. Знайдене значення статистики Шапіро-Уїлка дуже близьке до 1, що свідчить про нормальність розподілу чистого прибутку. Це підтверджується також і р-значенням, яке значно вище за 0,05. Крива розподілу також нормально описує вхідні дані(див. Додаток А).

Оскільки наявність мультиколінеарності має вагомий вплив на вірогідну оцінку коефіцієнтів рівняння та точність t-критерію Стьюдента формуємо візуалізацію кореляційної матриці для відображення попарних коефіцієнтів кореляції між незалежними ознаками(рис.2.1). Як бачимо, наявна сильна залежність між багатьма факторами, тому до переходу на наступний етап розрахунків варто виключити з аналізу декілька предикторів.

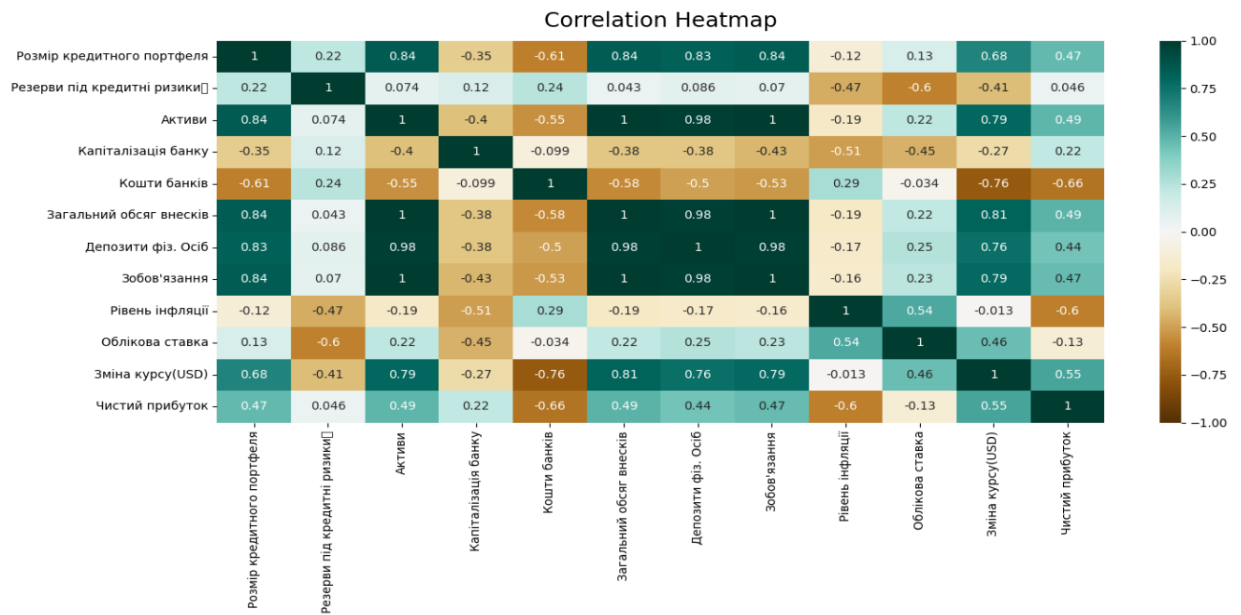


Рис.2.1 – Візуалізація кореляційної матриці

Після виключення необхідних ознак формуємо матрицю факторів та вектор залежної ознаки та присвоюємо їх до створених змінних. Для висвітлення впливу предикторів поза моделлю на результативну ознаку та задля врахування зміщення, що покращить точність прогнозування в майбутньому, також включаємо в рівняння константу(див. ДОДАТОК А).

Результати регресійного аналізу зображені на рис.2.2 цілком відповідають заданим вимогам. Адекватність та статистична значущість моделі перевіряється за допомогою F-статистики, в нашому випадку розраховане значення критерію Фішера більше за теоретичне, що говорить про пройдену перевірку. Нульова гіпотеза відкидається та приймається альтернативна.

Dep. Variable:	Чистий прибуток	R-squared:	0.705
Model:	OLS	Adj. R-squared:	0.641
Method:	Least Squares	F-statistic:	11.03
Date:	Mon, 20 May 2024	Prob (F-statistic):	8.57e-08
Time:	18:15:30	Log-Likelihood:	-389.16
No. Observations:	46	AIC:	796.3
Df Residuals:	37	BIC:	812.8
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
const	8752.1114	4071.994	2.149	0.038	501.467	1.7e+04
Розмір кредитного портфеля	0.0542	0.039	1.403	0.169	-0.024	0.133
Резерви під кредитні ризики	-0.0170	0.041	-0.415	0.681	-0.100	0.066
Капіталізація банку	3065.7608	1.06e+04	0.290	0.774	-1.84e+04	2.45e+04
Кошти банків	0.0551	0.177	0.311	0.757	-0.304	0.414
Депозити фіз. Осіб	-0.0550	0.032	-1.720	0.094	-0.120	0.010
Рівень інфляції	-1.065e+04	2693.258	-3.953	0.000	-1.61e+04	-5188.560
Облікова ставка	-2528.1139	4676.413	-0.541	0.592	-1.2e+04	6947.198
Зміна курсу(USD)	164.6195	72.304	2.277	0.029	18.117	311.122

Рис.2.2 – Результати лінійної регресії

Твердження про гарну відповідність моделі та наявних даних підтверджується коефіцієнтом детермінації, значення якого становить 0,705. Тобто вагома частка варіації, а саме 70,5%, чистого прибутку пояснюється наявними ознаками, все інше факторами невиключеними в модель.

Серед коефіцієнтів регресії статистично значущими є константа, рівень інфляції та зміна курсу(USD), що дає нам впевненість говорити про низьку ймовірність випадковості впливу на  $Y$ . Стосовно всіх інших факторів можемо констатувати відсутність доказів, щодо відхилення нульової гіпотези про відсутність впливу на  $Y$ . Тобто між ознаками і залежною змінною немає лінійного зв'язку.

Значення константи доволі велике, що може свідчити про високий вплив факторів поза моделлю. Серед найбільш впливових змінних на залежну змінну можна виділити капіталізацію банку та облікову ставку. При зміні капіталізації на 1, чистий прибуток зростає на 3065 мільйонів гривень. При збільшенні облікової ставки на 1, чистий прибуток зменшується на 2528 мільйонів гривень. Також варто відзначити вплив зміни курсу(USD), при зміні якого на 1 прибуток збільшується на 164 мільйонів гривень. Всі інші фактори мають не вагоми вплив на зміну  $Y$ .

Для з'ясування коректності, надійності та точності оцінок коефіцієнтів моделі був проведений тест Шапіро-Уїлка. В результаті було виявлено, що залишки моделі нормально розподілені(див. Додаток А).

Значення середньої абсолютної похибки становить 883, це значить, що в середньому прогнозоване значення чистого прибутку відрізняється від фактичного на 883 мільйона гривень. Тобто прогнозні значення знаходяться в допустимих межах  $\pm 10\%$ .

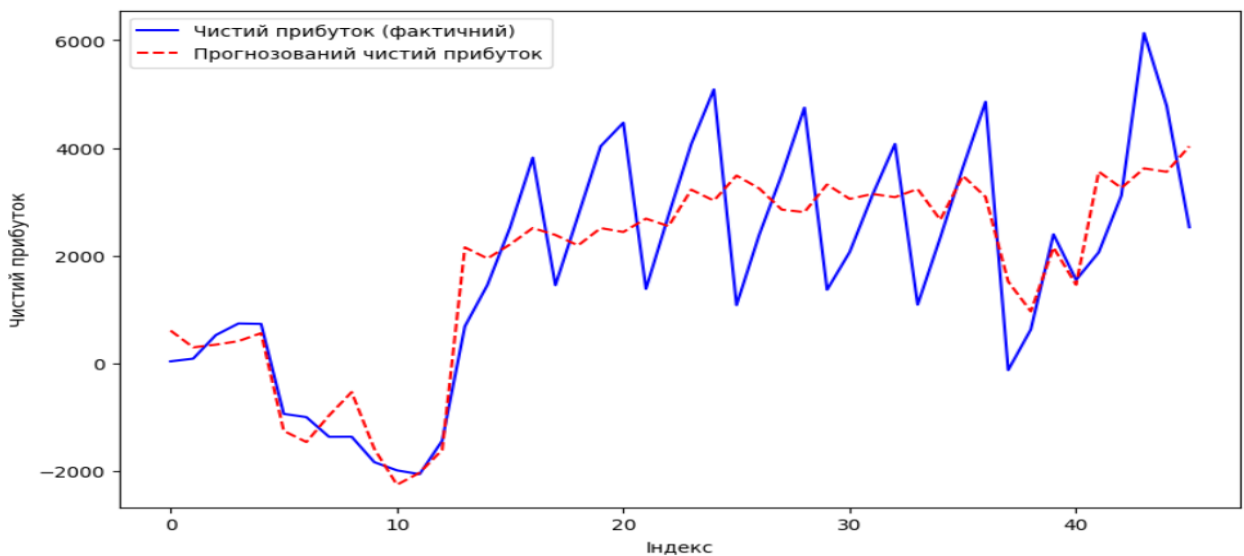


Рис.2.3 – Порівняння тенденції фактичних значень чистого прибутку з теоретичними

Схожість кривих розвитку значень фактичного чистого прибутку та теоретично розрахованого каже про хорошу якість побудованої моделі. Розрахована модель відтворила загальну тенденцію, проте була неточною на деяких ділянках графіку. Причиною цього виступає хаотичність та нестійкість чистого прибутку під дією факторів зовнішнього впливу, які не брали участь в аналізі, такими чинниками є політична, соціальна та природна складові(рис.2.3).

### 2.1.2 Лассо регресія

Дуже часто зустрічається проблема сильного пристосування моделі до поточних даних. Внаслідок чого відбувається точне прогнозування на основі

наявних даних, проте при використанні даної моделі для здійснення прогнозу на основі нових спостережень якість результатів значно падає. Тобто модель запам'ятовує значення існуючих ознак, проте не навчається на них.

При сильній розбіжності наборів критеріїв оцінки моделей на навчальних та тестових даних з'являється проблема пов'язана із сильним припасуванням моделі. Тобто виділяється надмірна увага налаштуванням параметрів для посилення узгодженості з даними та мінімізації відмінностей в значеннях фактичних та теоретичних.

Серед низки методів, сфокусованих на впорядкованості процесу навчання, використовуючи своєрідні характеристики у класичних наборах інструкцій, виділяється регуляризація.

Сутність даної стратегії полягає в нормалізації та зменшенні ноші певної ознаки задля позбавлення проблеми залежності алгоритмів від маленької кількості факторів та надмірного припасування при прогнозуванні .

Серед методів, які використовують стратегію регуляризації, варто виділити регресію Лассо. За рахунок обмеження на стан вектору параметрів рівняння моделі, приводячи значення деяких до 0. Застосування методу також ефективно в разі виникнення бажання зменшення рівня складності моделі у випадках, коли наявна велика кількість факторів та проблема сильної підгонки даних. В даному випадку в фінальну модель потрапляють лише ознаки з найбільшим значенням впливу на результат. Варто зазначити, що відбір предкторів в цьому методі є вбудованим, що забезпечує наявність розріджених коефіцієнтів[35].

Основна задача методу Лассо проявляється в додаванні у рівняння регресії умови зміщення абсолютного значення, або L1-регуляризацію. Це все робиться задля оптимізації та зменшення впливу мультиколінеарності на модель.

$$\min \|Xw - y\|^2 + z\|w\| \quad (2.3)$$



Якщо подивитися із середини на механізм боротьби методу із сильною залежністю між факторними ознаками, можна побачити, що Лассо керується силою зв'язку предиктора із залежною змінною, віддаючи перевагу тому фактору, який має більший вплив на зміну  $Y$ , зводячи інший фактор до 0 [36].

Особливу увагу при використанні методу Лассо варто виділяти значенню гіперпараметру, надто високе якого можна перетворити оцінки деяких параметрів на абсолютно нульові. Особливістю даної регресії є накладання штрафу на коефіцієнт, причиною якого може слугувати велике значення оцінки. Моделі побудовані за цією стратегією за рахунок мінімізації мінливості виділяються легкістю в їх читанні, точністю в прогнозах.

Варто висвітлити відсутність властивості розрядженості, що свідчить про наявність великої кількості параметрів зі значенням 0. Важливо також згадати про неможливість аналітично описати ефективність обчислених розрахунків.

Алгоритм формування моделі не відрізняється від попереднього. Так само підключаємось до попередніх бібліотек Python. Тільки в цей раз також завантажимо ще інші для обробки та розділення даних на набори, оцінки моделі та відповідну бібліотеку з необхідними методами для роботи з Лассо регресією (див. ДОДАТОК А).

Під час роботи з даною регресією варто не забувати про її чутливість до діапазону значень, які набувають різні змінні, тобто потрібно встановити однаковий масштаб змінних. З цим допоможе впоратись стандартизація, приводячи ознаки до нормального закону розподілу з середнім значенням 0 та стандартним відхиленням 1.

Для здійснення оцінки ефективності навченої моделі та виявлення тенденцій, взаємозалежностей в структурі даних ділимо загальний набір на тестовий та навчальний. Застосування даної стратегії також допомагає уникнути ризику перенавчання та покращення продуктивності роботи моделі на нових даних та узагальнення їх. Під тестову вибірку виділяємо чверть всього набору. Величина параметру «alpha» було виставлено на рівні 0.1, що є

досить низьким значенням регуляризації. Дане число обґрунтоване бажанням зберегти якомога більшу величину коефіцієнтів змінних в рівнянні моделі та покращити процес підлаштування під навчальний набір.

```
[ 0.          0.16722238 -0.          -0.          0.13725696 -0.10649646
 -0.          -0.04334922 -0.          -0.48892432  0.          0.40754264]
MSE train: 0.3032335, test: 0.3613371
R^2 train: 0.7015354, test: 0.6249078
```

Рис.2.4 – Результати Лассо регресії

Результати виконання Лассо регресії на рис.2.4 свідчать про високу якість побудованої моделі. За рахунок виконання однієї із задач L1-регуляризації відбулося надання абсолютно нульових значень коефіцієнтам наступних ознак: константа, резерви під кредитні ризики, активи, загальний обсяг внесків, зобов'язання та облікова ставка. Тобто сталося спрощення моделі із-за відкидання менш важливих змін, мають найменший вплив на чистий прибуток. Найбільший прямий зв'язок з результативною ознакою має зміна курсу(USD), зі збільшенням даного фактору на 1 – чистий прибуток зростає на 0,4(інтервал значень ознаки було визначено в межах від -1 до 1). Максимальну обернену залежність з Y серед усіх предикторів набув рівень інфляції. В разі збільшення рівня інфляції на 1 прибуток зменшується на 0,49. Також варто відзначити, що рівень депозитів фізичних осіб майже не має впливу на результат.

Значення середньої квадратичної похибки навчального набору доволі низьке та відносно близьке до 0, що говорить про хорошу здатність навчання моделі на вхідних даних. Про задовільну точність прогнозів на потенційних нових даних свідчить також низьке значення MSE тестового набору.

Значення коефіцієнтів детермінації обох наборів високі, що говорить нам про високу частку пояснення чистого прибутку наявними ознаками, в обох випадках це більше 60%. Можна стверджувати, що модель добре підходить під наявні дані та узагальнюється на нову інформацію.

Значення MAE становлять приблизно 0,5. Беручи до уваги масштабування залежної змінної, прогнозні значення чистого прибутку відрізняються від наявних в межах  $\pm 12,5\%$ , що відповідає заданим вимогам моделі.

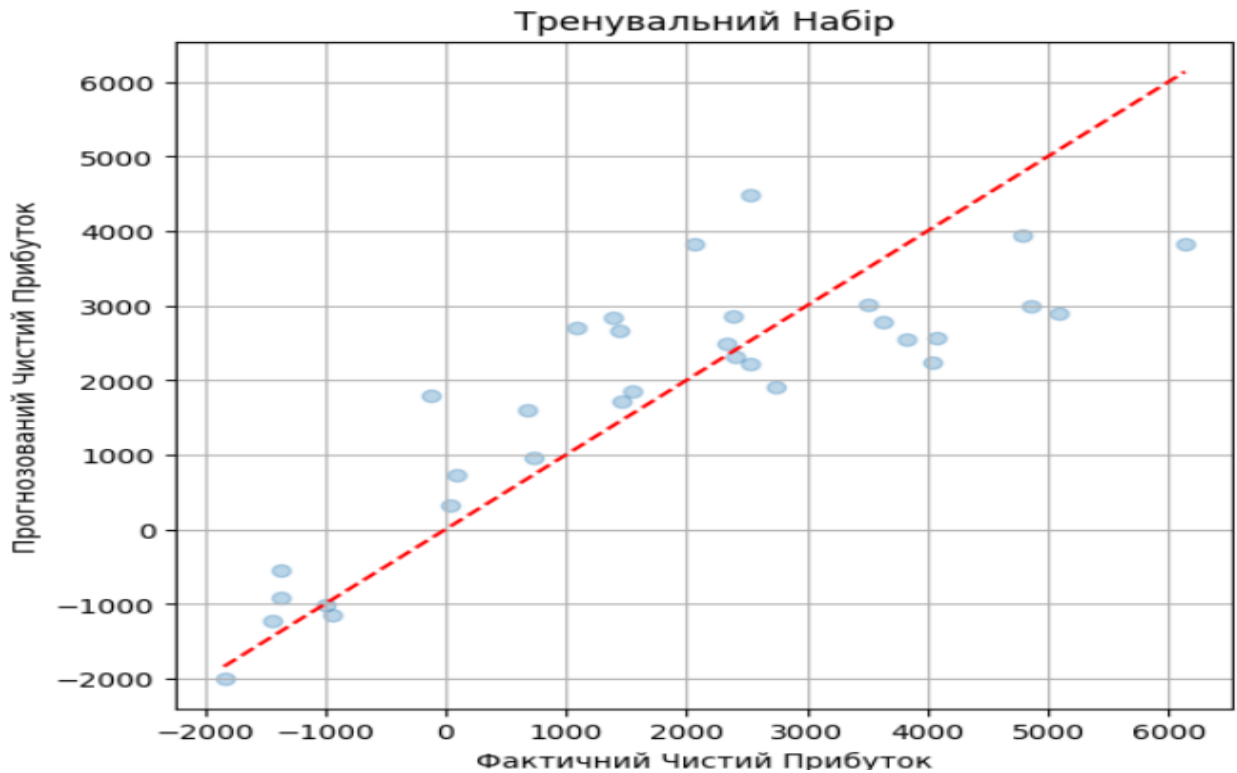


Рис.2.5 – Графічне порівняння фактичних та теоретичних значень на тренувальному наборі

На рис.2.5-2.6 червона пунктирна лінія вказує на збіги між прогнозними та фактичними значеннями чистого прибутку. Як бачимо, прогнозування на тренувальному наборі даних дало більш точні прогнози, можна побачити більше співпадінь. Прогнозування на тестовому наборі вийшло не досить вдалим із-за дуже низького значення гіперпараметру «alpha» регуляризації, як вже було зазначено вище, тому і теоретичні значення прибутку так відрізняються від наявних, оскільки не вистачає точності та узагальнення на нових даних.

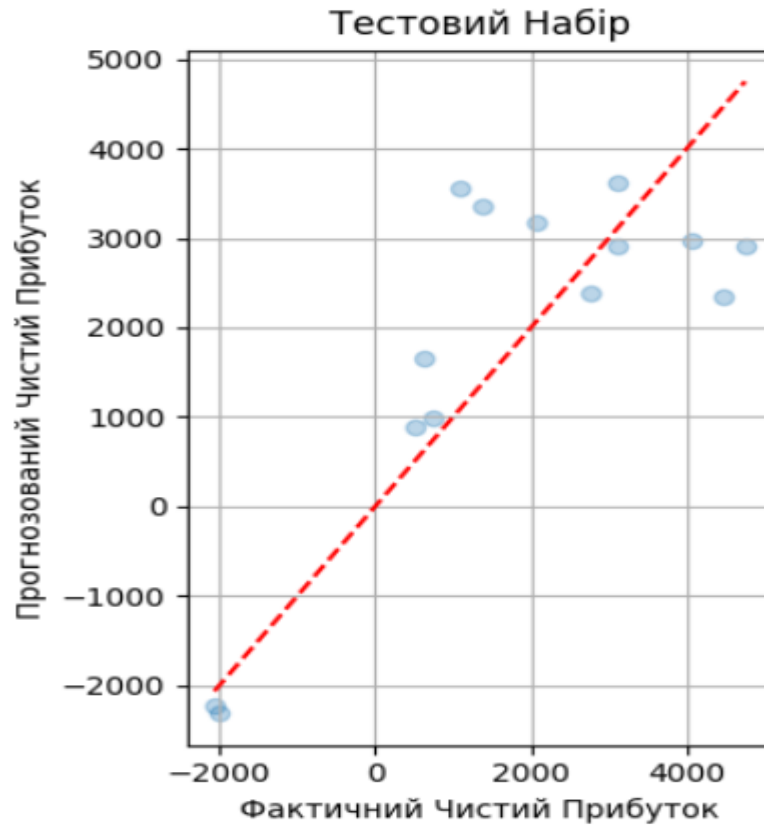


Рис.2.6 – Графічне порівняння фактичних та теоретичних значень на тестовому наборі

### 2.1.3 Гребенева регресія

Відомо, що зі зменшенням в моделі ступенів свободи ускладнюється підбір даних. Наявність в моделі факторів, які сильно між собою залежать впливає на точність та ефективність результатів. Гребенева регресія, або як її по іншому називають «Рідж», набуває важливості в ситуаціях з наявною мультиколінеарністю, або в при виникненні потреби в упорядкуванні моделі та зменшені розмірності. Існує два шляхи реалізації рідж регресії: через лінійне рівняння, використовуючи закриту форму, або здійснення градієнтної оптимізації[37].

Основною особливістю даного методу є включення в функцію оптимізації квадратичного зміщення зменшення розсіювання, з метою послаблення зв'язку між колінеарними факторами.

$$\min \|Xw - y\|^2 + z\|w\|^2 \quad (2.4)$$

Відбувається зменшення значення дисперсії за рахунок зняття жорсткості та обмежень, накладених на параметри факторів, та паралельно впроваджуючи зміщення. Це квадратичне зміщення призводить до виникнення незначної помилки, хоча напроці вагу цьому зменшується варіювання моделі. Цей зсув називають «штрафом» або ж L2-регуляризацією, аналітична форма якого –  $z\|w\|^2$  [38].

На відміну від Лассо регресії, гребенева не перетворює значення коефіцієнтів в нульові і внаслідок чого немає можливості відібрати предиктори, із-за чого відбувається зменшення величини середньої квадратичної похибки в теорії. Також варто зазначити механізм боротьби з колінеарністю двох факторів, який зосереджений на паралельному зменшенні значень оцінок параметрів.

```
[ 0.          0.50504496 -0.20171913  1.59217676  0.2487316  0.0354514
 -1.85365148 -0.37940195  0.31215176 -0.62012941 -0.0485748  0.53537478]
MSE train: 0.2543171, test: 0.3739430
R^2 train: 0.7496826, test: 0.6118220
```

Рис.2.7 – Результати виконання гребеневої регресії

Результати виконання гребеневої регресії на рис.2.7 свідчать про високу якість побудованої моделі. Як бачимо, нульових оцінок не спостерігається, так як алгоритм роботи L2-регуляризації зосереджений на паралельному зменшенні оцінок параметрів колінеарних ознак. Найкращу позитивну оцінку із залежною ознакою показує коефіцієнт показника «Активи», зі збільшенням даного фактору на 1 – чистий прибуток зростає на 1,59. Найбільшу реверсивність зв'язку з результативною змінною можна побачити з предиктором «Загальний обсяг внесків», в разі збільшення значення обсягів внесків на 1 прибуток зменшується на 1,85. Також зазначимо, що кошти банку та облікова ставка майже не мають впливу на зміну чистого прибутку.

Чим ближче значення середньоквадратичного відхилення до 0, тим краще модель навчається на наявних даних. Значення 0,25 показника на тренувальній вибірці підтверджує це. Також модель буде прогнозувати результат на основні даних доволі точно, це твердження базується на значення MSE тестового набору, що є близьким до 0.

Про високий рівень відповідності моделі вхідних даним та узагальненості на нові дані говорять значення коефіцієнтів детермінації навчальної та тестової вибірок відповідно. В обох випадках більше 60% варіації чистого прибутку пояснюється наявними ознаками.

Розраховані значення середніх абсолютних відхилень свідчать про хорошу якість прогнозів. Значення тренувальної вибірки в середньому відрізняються від фактичних на  $\pm 10\%$ , тестової вибірки –  $\pm 14\%$ .

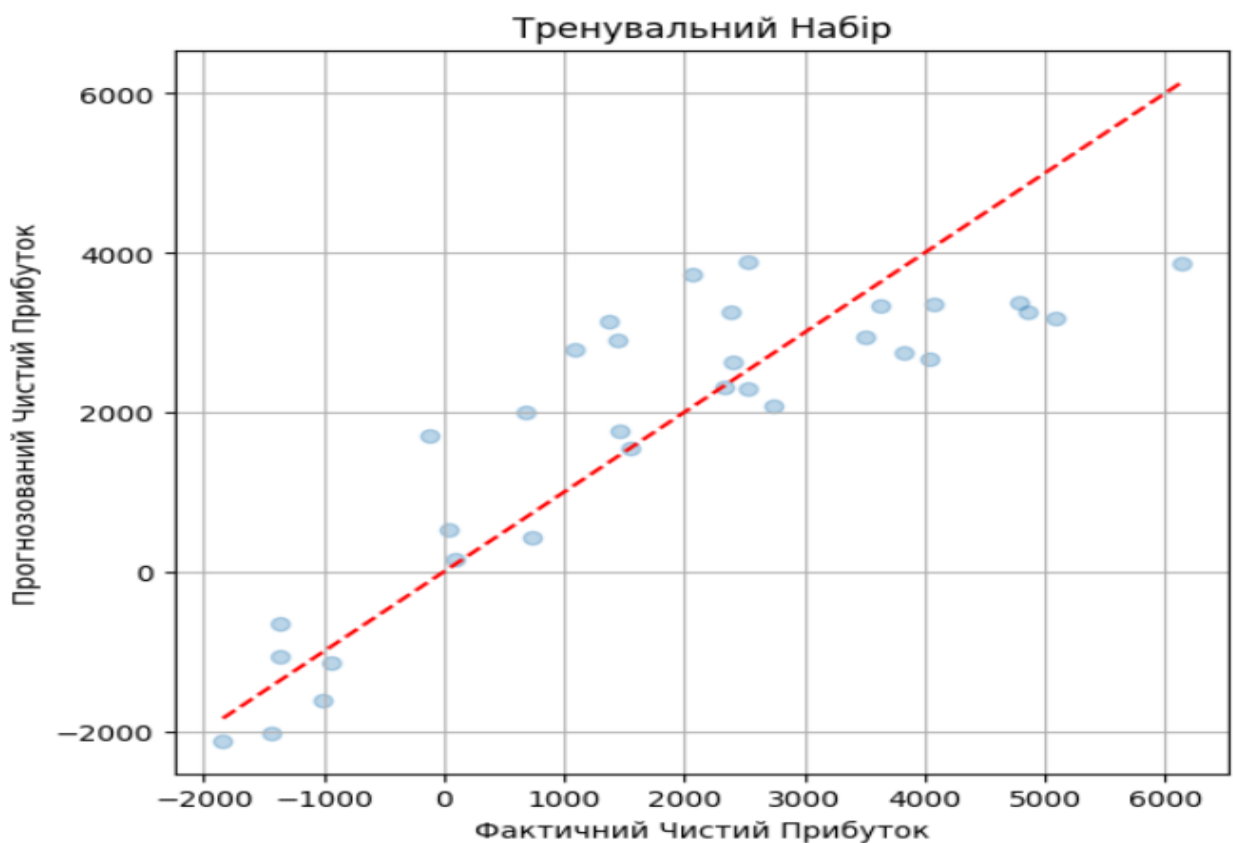


Рис.2.8 – Графічне порівняння фактичних та теоретичних значень на тренувальному наборі( рідж регресія)

На рис.2.8-2.9 можна побачити візуалізацію співпадінь прогнозних та фактичних значень чистого прибутку керуючись результатами розділених вибірок. Результати навчальної вибірки показують нормальну точність прогнозів. Результати тестової вибірки можна вважати незадовільними, оскільки наявна мала кількість співпадінь на графіку. Причиною цього явища може бути низьке значення гіперпараметру регуляризації при формуванні моделі, що дає більші оцінки параметрів та кращу здатність навчання на вхідних даних, проте погіршує можливість узагальнення нових спостережень змінних.

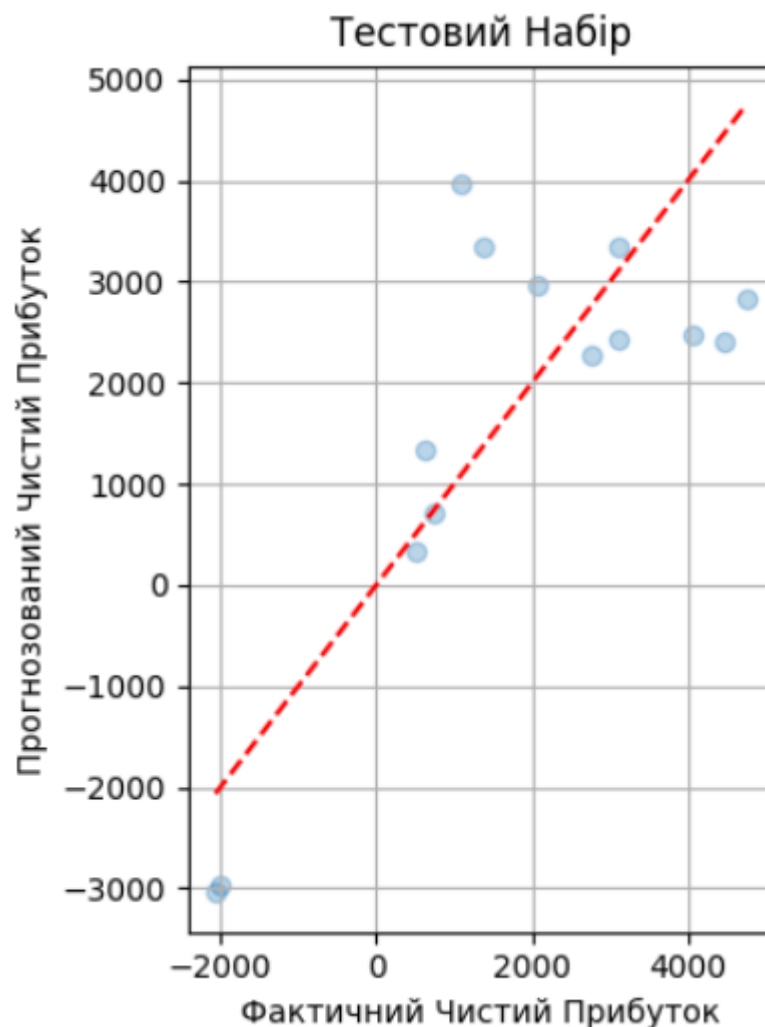


Рис.2.9 – Графічне порівняння фактичних та теоретичних значень на тестовому наборі(рідж регресія)

#### 2.1.4 Поліноміальна регресія

Важливо також розглянуто варіант, при якому дані розділені нелінійно. В такому випадку варто застосовувати поліноміальну регресію, результатом виконання якої буде крива, яка наближено описує розподіл точок, тобто значень чистого прибутку на графіку[39]. Дане наближення здійснюється якраз таки через поліном, аналітичний опис якого наступний:

$$y = \sum_{j=0}^k b_j x^j \quad (2.5)$$

Опис рівняння схожий з лінійною регресією, де  $b_j$  це оцінки параметрів, а  $x^j$  – самі значення факторів. Алгоритм знаходження оцінок коефіцієнтів також відповідний, метод найменших квадратів. Принцип залишається незмінним, пріоритет на мінімізації квадрату різниці між розрахованим та фактичним значеннями[40].

Оскільки поліноміальна регресія також чутлива до наявності колінеарних ознак, важливо видалити предиктори, які мають сильну залежність один з одним. Використовуючи кореляційну матрицю, побудовану раніше, з аналізу вилучаються наступні фактори: активи, зобов'язання, депозити фізичних осіб, загальний обсяг внесків, зміна курсу(USD). Перед розрахунками було приведено всі дані до одного масштабу для покращення продуктивності моделі.

Вибір ступеня полінома є важливим підготовчим етапом. Надто високий ступінь може ідеально відображати зміну наявних значень змінної, проте при прогнозуванні будуть вже неточності. Проблема перенавчання залишається актуальною перед виконанням поліноміальної регресії. Вирішення ситуації полягає в знаходженні оптимуму між надто простою та складною моделями.

Для порівнюваності результатів, було вибрано декілька моделі для побудови, серед яких лінійна, поліноміальні регресії 2 та 3 ступенів. В результаті, поліноміальна модель 2 ступеня вийшла більш значимою, на



відміну від полінома 3 ступеня, де спостерігається перенавчання, що в майбутньому призведе до погіршення точності прогнозування на нових даних.

Результати на 2.10 свідчать про високу якість та відповідність вимогами моделі. Оскільки розраховане значення критерію Фішера більше за теоретичне, можна казати про статистичну значущість та адекватності моделі. У підсумку приймається альтернативна гіпотеза. Значення  $p$  F-статистики свідчить також про статистично значимий вплив наявних предикторів на залежну змінну.

Висока відповідність моделі даним підтверджується значенням коефіцієнта детермінації. В нашому випадку, 86% варіації чистого прибутку пояснюється наявними факторами, і тільки 14% пояснюється факторами поза моделлю.

OLS Regression Results						
=====						
Dep. Variable:	y	R-squared:	0.864			
Model:	OLS	Adj. R-squared:	0.660			
Method:	Least Squares	F-statistic:	4.230			
Date:	Mon, 27 May 2024	Prob (F-statistic):	0.00122			
Time:	14:34:52	Log-Likelihood:	-19.409			
No. Observations:	46	AIC:	94.82			
Df Residuals:	18	BIC:	146.0			
Df Model:	27					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]
-----						
const	0.4641	0.388	1.197	0.247	-0.351	1.279
x1	1.0633	0.639	1.663	0.114	-0.280	2.406
x2	-0.6040	0.485	-1.244	0.229	-1.624	0.416
x3	0.2756	0.314	0.877	0.392	-0.384	0.936
x4	0.0887	0.423	0.210	0.836	-0.800	0.977
x5	-0.4046	0.352	-1.151	0.265	-1.143	0.334
x6	0.2865	0.476	0.602	0.555	-0.714	1.287
x7	1.0476	0.649	1.613	0.124	-0.317	2.412
x8	-2.4061	1.328	-1.811	0.087	-5.197	0.385
x9	1.1609	0.645	1.799	0.089	-0.195	2.517
x10	2.4556	1.567	1.567	0.134	-0.836	5.747
x11	-0.7290	0.517	-1.411	0.175	-1.815	0.357
x12	-0.9050	0.857	-1.056	0.305	-2.705	0.895
x13	1.1188	1.086	1.030	0.316	-1.162	3.400

Рис.2.10 – Результати поліноміальної регресії 2 ступеня

Серед оцінених коефіцієнтів регресії немає статистично значущих параметрів. Тобто приймається нульова гіпотеза про високу ймовірність випадковості впливу ознаки на результат та можливий нелінійний зв'язок. Значення оцінки константи вказує на вплив факторів, яких немає в аналізі. Перші шість оцінок параметрів  $x(x_1, \dots, x_6)$  показують зміну в середньому залежної ознаки при збільшенні на 1 вхідного фактору. Тобто при зміні розміру кредитного портфелю на 1, чистий прибуток зменшиться на 1,06 (враховуючи змінений масштаб змінної). Наступні шість оцінок параметрів  $x(x_1^2, \dots, x_2^2)$  враховують зміну результату при збільшенні квадрату вхідного фактору на 1. При збільшенні квадрату значення резервів під кредитні ризики на 1, чистий прибуток зменшується 2,4. Інші п'ятнадцять оцінок коефіцієнтів вказують на зміну  $Y$  при зміні добутку певної пари вхідних факторних ознак на 1. При збільшенні добутку резервів під кредитні ризики та коштів банку на 1, чистий прибуток збільшується на 1,45 (див. Додаток А).

Точність та надійність прогнозних значень підтверджена значення статистики Шапіро-Уїлка, яке дуже близьке до 1, та значенням р-статистики, що більше за 0,05, що говорить про нормальність розподілу залишків.

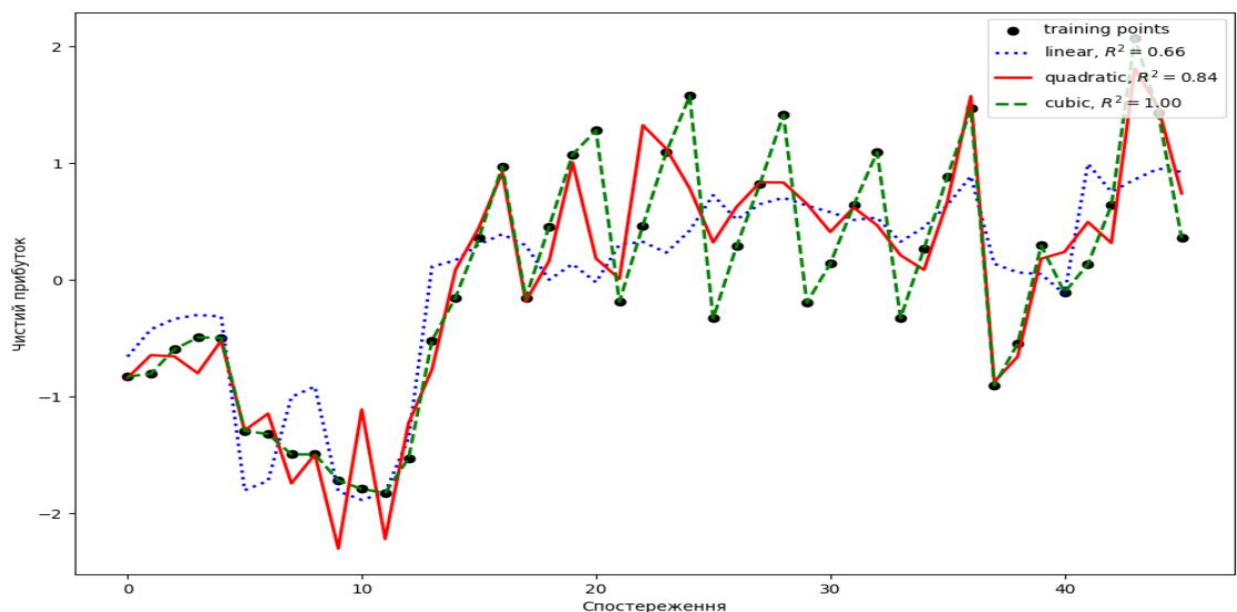


Рис.2.11 – Візуалізація результатів поліноміальних моделей

В середньому прогнозні значення чистого прибутку відрізняються від фактичних на  $\pm 7,5\%$ , що відповідає поставленим вимогам до формування моделі(див. Додаток А).

Візуалізація результатів зображена на рис.2.11 підтверджує висновки. Незважаючи на ідеальний опис кривою наявного розкиду тренувального набору поліномом 3 ступеня, модель буде погано узагальнюватись та адаптуватись під нові спостереження із перенавчання. Крива поліному 2 ступеня також характеризується високою здатність до опису вхідних значень, проте її можна буде використовувати для прогнозування.

## 2.2 Здійснення прогнозування значень чистого прибутку та вибір найбільш оптимальної моделі

Оскільки нові дані наших факторних змінних нам невідомі, для прогнозування буде застосовуватись методика сценарного моделювання, сутність в якій в аналізі прогнозних результатів за рахунок виставлення припущень стосовно шляху розвитку певного предиктора, їх змін.

Робимо припущення, що наявні незалежні ознаки набуватимуть таких трендів в наступних чотирьох кварталах: розмір кредитного портфеля буде збільшуватись на 1,25% з кожним кварталом; з таким же відсотком буде зменшуватись резерви під кредитні ризики; рівень капіталізації буде збільшуватись на 5%; активи на 2,5%; кошти банків збільшуватись на 10%; загальний обсяг внесків на 2,5; депозити фізичних осіб на 1,5%; зобов'язання на 1,25%; рівень інфляції та облікова ставка не будуть змінюватись; зміна курсу(USD) буде зменшуватись на 2%.

Після розрахунку нових значень варто створити підставити їх в побудовані моделі. Важливим етапом буде також зведення всіх результатів до одного масштабу, оскільки тільки розраховані значення чистого прибутку за лінійної регресією не відповідали початковому вигляду. Надалі лише залишилось відобразити результати графічно(рис.2.12).

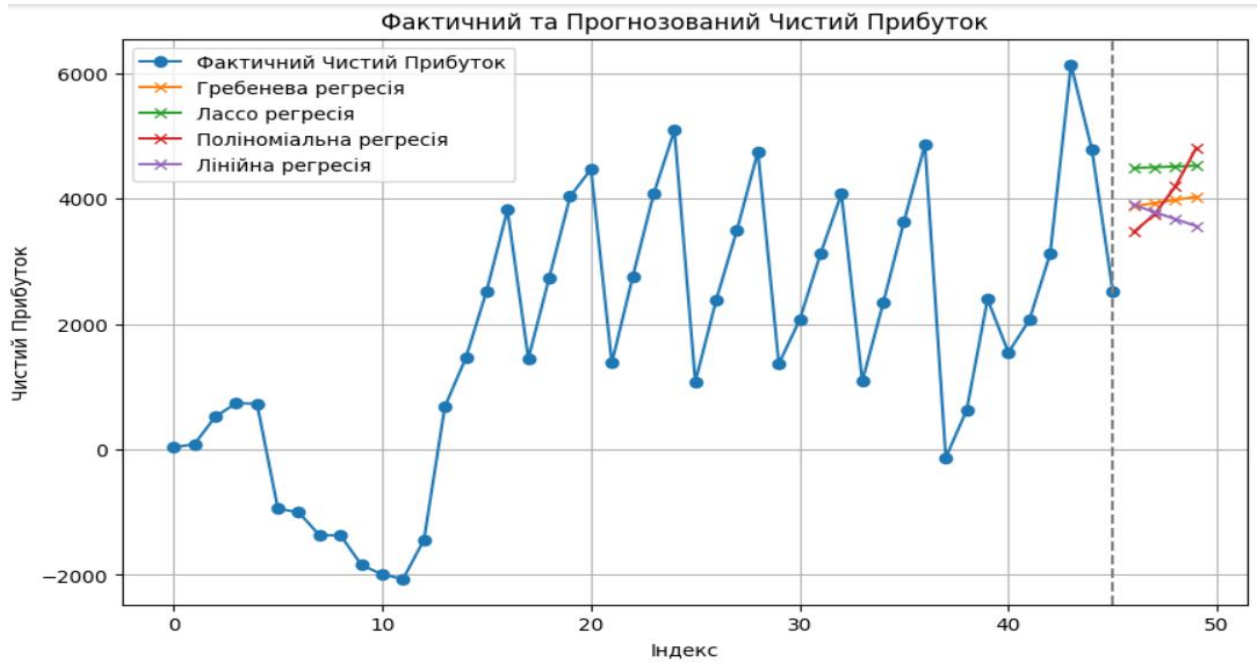


Рис.2.12 – Графічне відображення прогнозів

Перед обранням найбільш кращої моделі варто зауважити, що всі з них дотрималися поставлених вимог стосовно точності прогнозів, а саме  $\pm 15\%$ . Також було дотримано пункт з високою відповідністю даним моделями, в усіх випадках варіація чистого прибутку пояснювалася більше ніж 70%. Звернемо увагу звичайно ж на статистичну значущість кожної моделі, а також нормальність розподілу залишків лінійної та поліноміальної регресії.

Як бачимо з результатів прогнозування на нових даних, гребенева та Лассо регресії показали себе не дуже. Розраховані значення взагалі не відповідають наявному тренду розвитку залежної ознаки. Все ж низьке значення гіперпараметру регуляризації в обох випадках призвело до поганих результатів пристосованості моделей до нових спостережень в тестових вибірках, що і підтверджується на графіку.

Обираючи оптимальну модель серед інших двох, вибір все ж падає на поліноміальну. Дана модель добре відображає істинний тренд, а також має високу точність прогнозів за рахунок кращої здатності до опису вхідних змінних. Це підтверджується також значенням абсолютної середньої похибки, прогнозні значення найменше відрізняються від фактичних з усіх моделей.

## ВИСНОВКИ

В ході виконання роботи було висвітлено основні фактори впливу на зміну фінансового результату комерційних банків. Використовуючи вхідний набір показників було побудовано чотири моделі регресії, для відображення зв'язку між ними та чистим прибутком. Також робилася перевірка моделей на адекватність та статистичну значущість. На основі результатів прогнозувалися значення чистого прибутку банку «Райффайзен» на 4 квартали вперед. Для отримання нових значень показників використовувався метод сценарного планування.

Всі побудовані моделі відповідають поставленим вимогам, а саме: точність прогнозних значень варіюється від фактичних не більше 15%; висока відповідність моделей вхідних даних. В результаті здійснення прогнозування на нових даних найкраще себе показали результати лінійної та поліноміальної регресій. Найбільш оптимальною моделлю для подальшого застосування було обрано поліноміальну. Вона відзначається найбільшою точністю прогнозів та кращою здатністю моделі описувати дані.

Для використання цієї моделі для прогнозування в майбутньому рекомендується приділити увагу покращенню властивості узагальнення моделі на нових даних, для підсилення точності результатів, за рахунок розширення бази даних, проведення крос-валідаційних тестів, включення регуляризації або ж приділення особливої уваги тестовій вибірці. Також можливим покращенням моделі може бути включення факторів зовнішнього впливу, такі як природні та соціально-політичні.

Економічну значущість результатів для банку «Райффайзен» можна висвітлити тезами:

- Оптимізація управління щодо чинників, які мають сильний вплив на зміну чистого прибутку, задля покращення фінансових результатів та діяльності загалом;

- Розраховані прогностні значення визначають стратегічний напрям діяльності банку, краще розуміння майбутніх значень прибутку полегшує виконання стратегічних цілей розвитку;
- Прогнозування прибутку дозволяє завчасно виявляти та боротися з можливим ризиками, формувати комплект заходів для їх усунення;
- Модель дає можливість визначати оптимальні значення показників, які беруть участь в формуванні ціни, за рахунок коригування вхідних даних;
- Презентування прогностних фінансових результатів дозволяє підвищити рівень довіри інвесторів, що в майбутньому відгукується в збільшенні капіталізації.

Загалом, результати моделювання можна вважати хорошими. За рахунок балансування простоти та фокусування на максимізації здатності моделі описувати вхідні дані, було досягнуто високої точності при прогнозуванні значень прибутку. Актуальність результатів підтверджується необхідністю кожного банку контролювати фінансову стійкість та покращувати ефективність виконання операцій, задля досягнення необхідних фінансових цілей.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. The Importance of Financial Forecasting. *Constructus INTERNATIONAL TM*. Building a Better Business. URL: <https://constructus.com/wp-content/uploads/Financial-forecasting.pdf> (date of access: 04.05.2024).
2. Норік Л. О. Прогнозування прибутку банку за показниками активів / Л. О. Норік, М. С. Бріль // Сучасні проблеми управління підприємствами: теорія та практика : матеріали міжнарод. науково-практ. конф., 29-30 бер. 2018 р. : тези доповід. – Х. : Вид. ФОП Мезіна В. В., 2018. – С. 336–339.
3. Бондаренко В. МОДЕЛІ ПРОГНОЗУВАННЯ ЧИСТОГО ПРИБУТКУ БАНКУ. *Розвиток європейського простору очима молоді: економічні, соціальні та правові аспекти* : матеріали Всеукр. науково-практ. конф. докторантів, молодих уч. та студентів, м. Харків, 17 трав. 2019 р. Харків, 2019. URL: <https://cdn.hneu.edu.ua/rozvitok19/thesis05-03.html> (дата звернення: 04.05.2024).
4. Polyanskaya N. M. Profit of the Commercial Bank: A Retrospective Analysis and Short-Term Forecast. *Finance and Credit*. 2018. Vol. 24, no. 2. P. 388–413. URL: <https://doi.org/10.24891/fc.24.2.388> (date of access: 04.05.2024).
5. Kao C., Liu S.-T. Predicting bank performance with financial forecasts: A case of Taiwan commercial banks. *Journal of Banking & Finance*. 2004. Vol. 28, no. 10. P. 2353–2368. URL: <https://doi.org/10.1016/j.jbankfin.2003.09.008> (date of access: 04.05.2024).
6. Giroux G. A. FINANCIAL FORECASTING IN COMMERCIAL BANKS AN INDUSTRY SURVEY : DISSERTATION. Texas. 143 p. URL: <https://ttu-ir.tdl.org/server/api/core/bitstreams/24b3d1b7-c3b7-48a9-b001-6a012c7198fd/content> (date of access: 04.05.2024).
7. Руцишин Н., Костак З. Банківська система України: сучасний стан та перспективи розвитку. *ЕКОНОМІКА І СУСПІЛЬСТВО*. 2018. № 16. С. 783–789.

- URL: [https://economyandsociety.in.ua/journals/16\\_ukr/119.pdf](https://economyandsociety.in.ua/journals/16_ukr/119.pdf) (дата звернення: 04.05.2024).
8. Онищенко Ю. І. Основні етапи формування структури банківської системи України в залежності від структурних характеристик / Ю. І. Онищенко // *Інноваційна економіка*. – 2009. - №1 – С. 264-271.
  9. Тоцька К. Торік прибутки банків виросли вдвічі, ніж до великої війни. *Investory News*. URL: <https://investory.news/torik-pributki-bankiv-virosli-vdvichi-nizh-do-velikoi-vijni/> (дата звернення: 04.05.2024).
  10. 4,8 млрд грн чистого прибутку та 6,6 млрд грн податків і зборів задекларував Райффайзен Банк за 2023 рік | Райффайзен Банк. *Raiffeisen*. URL: <https://raiffeisen.ua/news/rezultati-roku-2005> (дата звернення: 04.05.2024).
  11. Vincenti E., Kushnir N. CURRENT TRENDS OF THE MARKET OF BANKING SERVICES USING INFORMATION TECHNOLOGIES. *Herald UNU. International Economic Relations And World Economy*. 2022. No. 43. URL: <https://doi.org/10.32782/2413-9971/2022-43-5> (date of access: 04.05.2024).
  12. Банківські операції : підручник / [А. М. Мороз, М. І. Савлук, М. Ф. Пуховкіна та ін.] ; за заг. ред. А. М. Мороза ; М-во освіти і науки України, ДВНЗ «Київ. нац. екон. ун-т ім. Вадима Гетьмана». – 3-тє вид., переробл. і доповн. – К. : КНЕУ, 2008. – 603 с.
  13. Катан Л. І. Прибутковість комерційного банку та її підвищення / Л. І. Катан, М. С. Плахотник // *Агросвіт*. – 2018. - № 21. - С. 3 – 6. - Режим доступу : URI (Уніфікований ідентифікатор ресурсу): <http://dspace.dsau.dp.ua/jspui/handle/123456789/2947>
  14. Криклій О.А. Управління прибутком банку : монографія / О.А. Криклій, Н.Г. Маслак. – Суми : ДВНЗ “УАБС НБУ”, 2008. – 136 с.
  15. Фінансова безпека банку як основа його стійкості / Л. О. Добрик, Г. В. Запорожець. // *Ефективна економіка*. - 2014. - № 6. - Режим доступу: [http://nbuv.gov.ua/UJRN/efek\\_2014\\_6\\_27](http://nbuv.gov.ua/UJRN/efek_2014_6_27)



16. Аналіз банківської діяльності: навч.посібник / Національний банк України; За заг. Ред.. канд..екон.наук У. Я. Груздевич. — К.:УБС НБУ, 2007. — 222 с.
17. Ключко Л. А. Фінансова стабільність банків та її визначальні характеристики / Л. А. Ключко // Збірник наукових праць Національного університету державної податкової служби України. 2009. — № 1. — С. 143-150
18. Базилінська О. Я. Фінансовий аналіз: теорія та практика : [підручник] / О. Я. Базилінська. — К. : ЦУЛ, 2011. — 328 с
19. Іванова В. О., Кантур С. Ф. Економічна сутність прибутку комерційного банку та джерела його формування[Електронний ресурс] / В. О. Іванова. —Режим доступу: <http://dspace.nbu.gov.ua/bitstream/handle/123456789/91628/15-Ivanova.pdf?sequence=1>
20. Шумейко П.Є. Оцінка прибутковості комерційних банків та розробка шляхів її забезпечення. Кваліфікаційна робота першого (бакалаврського) рівня вищої освіти. Національний університет харчових технологій, 2021
21. Tiwari, Abhishek, et al. "Customer Relationship Management in Banking Sector." *International Journal of Marketing & Business Communication*, 2016.
22. Lavrentieva, Svetlana V., et al. "Bank Products and Services." *Journal of Advanced Research in Dynamical and Control Systems*, 2019.
23. Jain, Utkarsh, et al. "Risk Management in Banking Sector." *International Journal of Research in Finance and Marketing*, 2018.
24. Zimovkina, Marina S., et al. "Efficiency of Internal Processes in Commercial Banks." *European Research Studies Journal*, 2017.
25. Săraru, Cătălin-Silviu, et al. "Market Environment Analysis in Banking Sector." *Procedia Economics and Finance*, 2015.

26. Сало І.В. Необхідність аналізу показників прибутку банку [Текст] / І.В. Сало, О.О. Лисянська. – 2010. Збірник наукових праць УАБС НБУ
27. I. Forbrich, L. Kutzbach, A. Hormann, and M. Wilmking, “A comparison of linear and exponential regression for estimating diffusive CH<sub>4</sub> fluxes by closed-chambers in peatlands,” *Soil Biology and Biochemistry*, vol. 42, no. 3, pp. 507–515, 2010.
28. Application of Regression Techniques with their Advantages and Disadvantages. MA Iqbal. *Elektron Magazine* 4 (1), 11-17, 2021. 17, 2021.
29. A. E. Hoerl and R. W. Kennard. Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1):55.
30. R. Tibshirani. *The Lasso: A brief Review and a new significance test*, 2014 (accessed June 1, 2018). <http://statweb.stanford.edu/~tibs/ftp/ubctalk.pdf>
31. Комерційні банки: функції та типи. Реферат. *Освіта.UA*. URL: [https://web.posibnyky.vntu.edu.ua/firen/3zlepko\\_osnovy\\_biomedychnogo\\_radioelektronnogo\\_aparatobuduvannya/6.html](https://web.posibnyky.vntu.edu.ua/firen/3zlepko_osnovy_biomedychnogo_radioelektronnogo_aparatobuduvannya/6.html) (дата звернення: 14.05.2024).
32. 6 МАТЕМАТИЧНІ МОДЕЛІ: ОЗНАЧЕННЯ, ХАРАКТЕРИСТИКИ, ЕТАПИ ПОБУДОВИ. *web.posibnyky.vntu.edu.ua*. URL: [https://web.posibnyky.vntu.edu.ua/firen/3zlepko\\_osnovy\\_biomedychnogo\\_radioelektronnogo\\_aparatobuduvannya/6.html](https://web.posibnyky.vntu.edu.ua/firen/3zlepko_osnovy_biomedychnogo_radioelektronnogo_aparatobuduvannya/6.html) (дата звернення: 14.05.2024).
33. МНОЖИННА ЛІНІЙНА РЕГРЕСІЯ. Статистичне моделювання та прогнозування. Українські підручники та статті – Бібліотека Posibniki.com.ua. *Українські підручники, посібники та статті - Бібліотека Posibniki. Електронна бібліотека підручників онлайн*. URL: [https://posibniki.com.ua/post-mnozhinna-liniina-regresiya\\_1](https://posibniki.com.ua/post-mnozhinna-liniina-regresiya_1) (дата звернення: 29.05.2024).

34. Ципліцька О., Антонова Л. ТЕМА 2 Методи побудови загальної лінійної моделі. *Економетрика*. Миколаїв, 2011. С. 232. URL: <https://lib.chmnu.edu.ua/pdf/posibnuku/299/5.pdf> (дата звернення: 29.05.2024).
35. Marinamath. Kurshach\_Math Modeling. *Kaggle: Your Machine Learning and Data Science Community*. URL: <https://www.kaggle.com/code/marinamath/kurshach-math-modeling> (date of access: 29.05.2024).
36. What is lasso regression? | IBM. *IBM in Deutschland, Österreich und der Schweiz*. URL: <https://www.ibm.com/topics/lasso-regression> (date of access: 29.05.2024).
37. Довідник по Machine Learning – Regularization Algorithms | База знань IT. *База знань IT технологій*. URL: <https://itwiki.dev/data-science/ml-reference/ml-glossary/regularization-algorithms> (дата звернення: 29.05.2024).
38. What Is Ridge Regression? | IBM. *IBM in Deutschland, Österreich und der Schweiz*. URL: <https://www.ibm.com/topics/ridge-regression> (date of access: 29.05.2024).
39. Abhigyan. Understanding Polynomial Regression!!!. *Medium*. URL: <https://medium.com/analytics-vidhya/understanding-polynomial-regression-5ac25b970e18> (date of access: 29.05.2024).
40. Прохоренко, О. С. Огляд регресійних підходів для прогнозування місць гравців на кінець матчу у відеогрі : дипломна робота ... бакалавра : 113 Прикладна математика / Прохоренко Олександр Сергійович. – Київ, 2022. – 47 с.

## ДОДАТОК А

```

▶ import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import statsmodels.api as sm
import sklearn
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split

```

Рис.1 – Імпортування необхідних бібліотек

```
[ ] df = pd.read_excel('Дані.xlsx')
df
```

	Розмір кредитного портфеля	Резерви під кредитні ризиків	Активи	Капіталізація банку	Кошти банків	Загальний обсяг внесків	Депозити фіз. Осіб	Зобов'язання	Рівень інфляції	Облікова ставка	Зміна курсу(USD)	Чистий прибуток
0	26071	9916	47694	0.133057	9944	27804	15436	43133	0.998	0.075	7.993	34
1	25866	10193	45977	0.146747	6120	28973	16399	39229	1.005	0.075	7.993	84
2	26618	9436	44929	0.156647	5652	28761	16768	37891	1.005	0.070	7.993	522
3	26348	8914	44804	0.161057	5358	28735	15653	37588	1.005	0.065	7.993	738
4	27047	8849	43460	0.166521	6012	26419	14633	36223	1.005	0.065	7.993	730
5	29667	11508	44987	0.150910	9307	23851	13490	38197	1.249	0.065	10.955	-943
6	28837	12209	44413	0.142256	7282	25768	14935	38094	1.249	0.095	11.823	-1003
7	28448	-13657	44676	0.131592	6139	26452	13848	38796	1.249	0.125	12.949	-1369
8	29705	-13866	46859	0.131202	5378	28800	14207	40711	1.249	0.140	15.769	-1367

Рис.2 – Створення датафрейму на основі зчитування Excel файлу з даними

```

▶ from scipy import stats
  from scipy.stats import shapiro, anderson
  target = df['Чистий прибуток']

  # Виконуємо тест Шапіро-Уїлка
  stat, p_value = shapiro(target)
  print('Статистика Шапіро-Уїлка для цільової змінної:', stat)
  print('p-значення:', p_value)

  if p_value > 0.05:
      print('Цільова змінна має нормальний розподіл (не відхиляємо нульову гіпотезу)')
  else:
      print('Цільова змінна не має нормального розподілу (відхиляємо нульову гіпотезу)')

  # Q-Q графік для цільової змінної
  plt.figure(figsize=(6, 6))
  stats.probplot(target, dist="norm", plot=plt)
  plt.show()

  # Гістограма і KDE для цільової змінної
  sns.histplot(target, kde=True)
  plt.show()

```

Рис.3 – Перевірка цільової змінної на нормальність розподілу

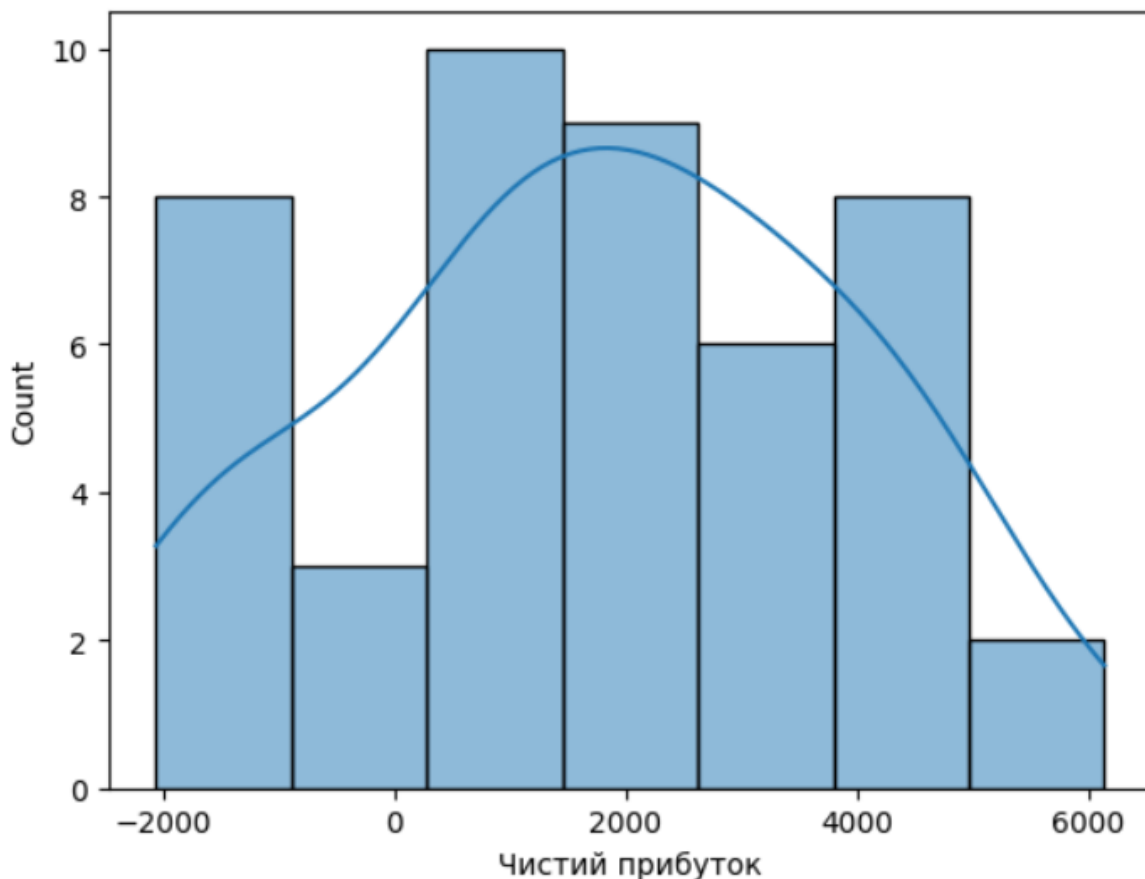


Рис.4 – Діаграма розподілу

```

▶ plt.figure(figsize=(16,6))

heatmap = sns.heatmap(df.corr(), vmin=-1, vmax=1, annot=True, cmap='BrBG')
heatmap.set_title('Correlation Heatmap', fontdict={'fontsize':18}, pad=12);

```

Рис.5 – Створення кореляційної матриці

```

▶ X= df.drop(['Чистий прибуток', 'Загальний обсяг внесків', 'Активи', "Зобов'язання"], axis=1)
y=df['Чистий прибуток']
X = sm.add_constant(X)
model = sm.OLS(y, X).fit()
predictions1 = model.predict(X)
print_model = model.summary()
print(print_model)

```

Рис.6 – Побудова лінійної регресії

```

▶ from sklearn.metrics import mean_absolute_error
mae = mean_absolute_error(y, predictions1)
print("MAE: train", mae)

```

Рис.7 – Розрахунок середньої абсолютної похибки

```

▶ residuals = model.resid

# Перевіряємо нормальність залишків
stat, p_value = stats.shapiro(residuals)
print('Статистика Шапіро-Уїлка для залишків:', stat)
print('p-значення:', p_value)

if p_value > 0.05:
    print('Залишки мають нормальний розподіл (не відхиляємо нульову гіпотезу)')
else:
    print('Залишки не мають нормального розподілу (відхиляємо нульову гіпотезу)')

# Q-Q графік для залишків
plt.figure(figsize=(6, 6))
stats.probplot(residuals, dist="norm", plot=plt)
plt.title('Q-Q графік для залишків')
plt.show()

# Гістограма і KDE для залишків
sns.histplot(residuals, kde=True)
plt.title('Гістограма і KDE для залишків')
plt.show()

```

Рис.8 – Перевірка на нормальність розподілу залишків

Статистика Шапіро-Уїлка для залишків: 0.9804201126098633

p-значення: 0.623162567615509

Залишки мають нормальний розподіл (не відхиляємо нульову гіпотезу)

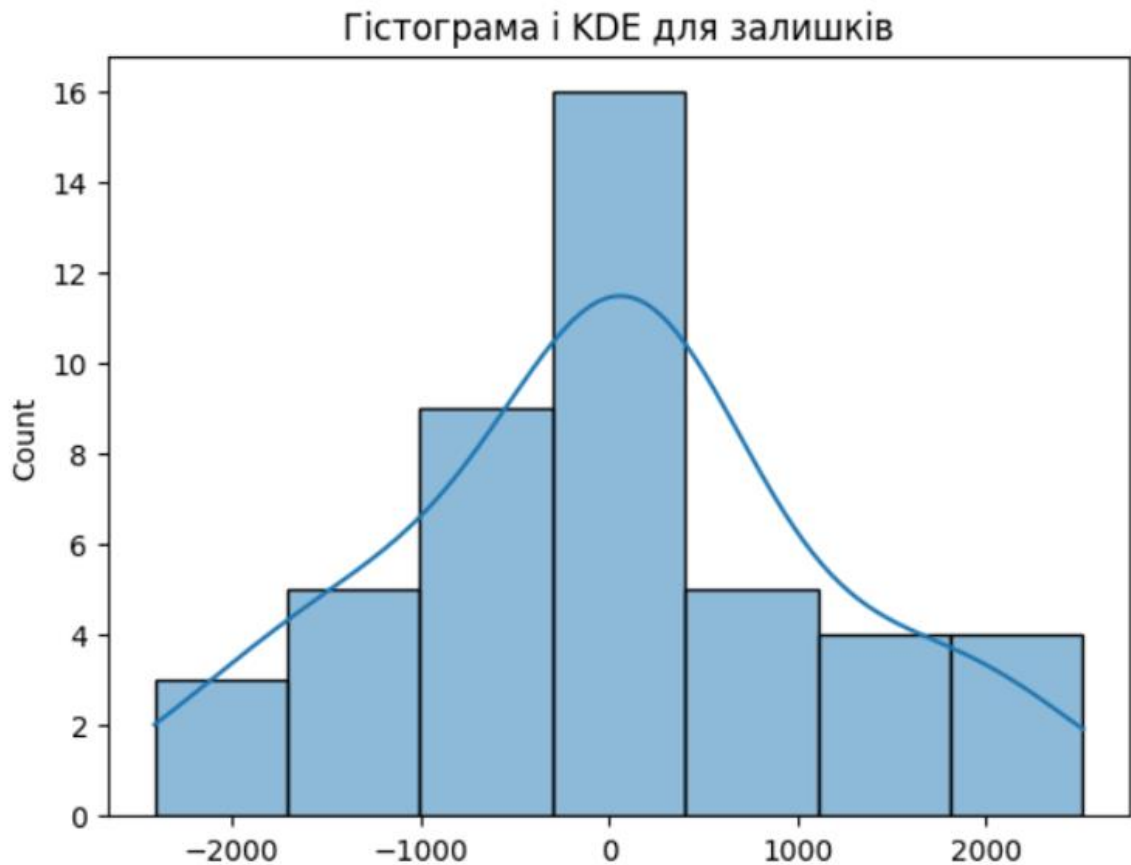


Рис.9 – Результати перевірки на нормальність розподілу залишків

```

▶ plt.figure(figsize=(10, 6))
plt.plot(df.index, y, label='чистий прибуток (фактичний)', color='blue')
plt.plot(df.index, predictions1, label='прогнозований чистий прибуток', color='red', linestyle='--')
plt.xlabel('Індекс')
plt.ylabel('Чистий прибуток')
plt.title('Порівняння фактичного та прогнозованого чистого прибутку')
plt.legend()
plt.show()

```

Рис.10 – Візуалізація результатів побудованої моделі

```

▶ new_data = pd.DataFrame({
    'const': 1,
    'Розмір кредитного портфеля': [53497, 54166, 54843, 55529],
    'Резерви під кредитні ризики': [-2612, -2645, -2678, -2711],
    'Капіталізація банку': [0.132042, 0.139542, 0.147042, 0.154542],
    'Кошти банків': [386, 425, 467, 514],
    'Депозити фіз. осіб': [61408, 62329, 63264, 64212],
    'Рівень інфляції': [1.014, 1.014, 1.014, 1.014],
    'Облікова ставка': [0.150, 0.150, 0.150, 0.150],
    'Зміна курсу(USD)': [38.436, 37.667, 36.914, 36.175],
})

# Додавання константи до нових даних
new_data = sm.add_constant(new_data)

# Прогнозування для нових даних
predictions_new = model.predict(new_data)

# Вивід прогнозних значень
print(predictions_new)

```

```

⇌ 0    3899.999487
   1    3784.735413
   2    3671.934283
   3    3561.486283
dtype: float64

```

Рис.11 – Прогнозування чистого прибутку на основі нових даних

```

▶ import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import statsmodels.api as sm
import sklearn
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import Lasso
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score

```

Рис.12 – Імпортування необхідних бібліотек для Лассо регресії



```

X= df.drop(['Чистий прибуток'], axis=1)
y=df['Чистий прибуток']
X = sm.add_constant(X)
sc_x = StandardScaler()
sc_y = StandardScaler()
X_std = sc_x.fit_transform(X)
y_std = sc_y.fit_transform(y.values.reshape(-1, 1)).flatten()
X_train_scaled, X_test_scaled, y_train_scaled, y_test_scaled = train_test_split(X_std, y_std, test_size=0.25, random_state=0)

from sklearn.linear_model import Lasso
from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
from sklearn.metrics import mean_absolute_error

lasso = Lasso(alpha=0.1)
lasso.fit(X_train_scaled, y_train_scaled)
y_train_pred = lasso.predict(X_train_scaled)
y_test_pred = lasso.predict(X_test_scaled)
print(lasso.coef_)

print('MSE train: {:.7f}, test: {:.7f}'.format(
    mean_squared_error(y_train_scaled, y_train_pred),
    mean_squared_error(y_test_scaled, y_test_pred)))
print('R^2 train: {:.7f}, test: {:.7f}'.format(
    r2_score(y_train_scaled, y_train_pred),
    r2_score(y_test_scaled, y_test_pred)))

mae = mean_absolute_error(y_train_scaled, y_train_pred)
print("MAE: train", mae)
mae1 = mean_absolute_error(y_test_scaled, y_test_pred)
print("MAE: test", mae1)

```

Рис.13 – Побудова Лассо регресії

```

▶ y_train_pred_original = sc_y.inverse_transform(y_train_pred.reshape(-1, 1)).flatten()
y_test_pred_original = sc_y.inverse_transform(y_test_pred.reshape(-1, 1)).flatten()
y_train_original = sc_y.inverse_transform(y_train_scaled.reshape(-1, 1)).flatten()
y_test_original = sc_y.inverse_transform(y_test_scaled.reshape(-1, 1)).flatten()

# Побудова графіків порівняння прогнозованих та фактичних значень
plt.figure(figsize=(14, 6))

# Тренувальний набір
plt.subplot(1, 2, 1)
plt.scatter(y_train_original, y_train_pred_original, alpha=0.3)
plt.plot([y_train_original.min(), y_train_original.max()], [y_train_original.min(), y_train_original.max()], 'r--')
plt.xlabel('Фактичний Чистий Прибуток')
plt.ylabel('Прогнозований Чистий Прибуток')
plt.title('Тренувальний Набір')
plt.grid(True)

```

Рис.14 – Візуалізація результатів тренувальної вибірки

```

▶ plt.subplot(1, 2, 2)
plt.scatter(y_test_original, y_test_pred_original, alpha=0.3)
plt.plot([y_test_original.min(), y_test_original.max()], [y_test_original.min(), y_test_original.max()], 'r--')
plt.xlabel('Фактичний Чистий Прибуток')
plt.ylabel('Прогнозований Чистий Прибуток')
plt.title('Тестовий Набір')
plt.grid(True)

plt.tight_layout()
plt.show()

```

Рис.15 – Візуалізація результатів тестової вибірки

```

▶ X_new = pd.DataFrame({
    'const': 1,
    'Розмір кредитного портфеля': [53497, 54166, 54843, 55529],
    'Резерви під кредитні ризики\t': [-2612, -2645, -2678, -2711],
    'Активи': [200524, 205652, 210911, 216304],
    'Капіталізація банку': [0.132042, 0.139542, 0.147042, 0.154542],
    'Кошти банків': [386, 425, 467, 514],
    'Загальний обсяг внесків': [170081, 174430, 178891, 183466],
    'Депозити фіз. осіб': [61408, 62329, 63264, 64212],
    'Зобов\`язання': [173312, 175478, 177672, 179892],
    'Рівень інфляції': [1.014, 1.014, 1.014, 1.014],
    'Облікова ставка': [0.150, 0.150, 0.150, 0.150],
    'Зміна курсу(USD)': [38.436, 37.667, 36.914, 36.175]})

# Додавання константи до нових даних
X_new = sm.add_constant(X_new)
# Масштабування нових даних
X_new_scaled = sc_x.transform(X_new)
# Прогноз на основі нових даних
y_new_pred_scaled = lasso.predict(X_new_scaled)
# Зворотнє масштабування передбачених значень
y_new_pred = sc_y.inverse_transform(y_new_pred_scaled.reshape(-1, 1))
print('Прогнозовані значення для нових даних:', y_new_pred)

⇒ Прогнозовані значення для нових даних: [[4119.96557526]
 [4061.40742227]
 [4003.52377502]
 [3945.83494326]]

```

Рис.16 – Прогнозування чистого прибутку на основі нових значень

```

▶ X= df.drop(['Чистий прибуток'], axis=1)
y=df['Чистий прибуток']
X = sm.add_constant(X)
sc_x = StandardScaler()
sc_y = StandardScaler()
X_std = sc_x.fit_transform(X)
y_std = sc_y.fit_transform(y.values.reshape(-1, 1)).flatten()
X_train_scaled, X_test_scaled, y_train_scaled, y_test_scaled = train_test_split(X_std, y_std, test_size=0.25, random_state=0)

▶ from sklearn.linear_model import Ridge
from sklearn.metrics import mean_absolute_error
ridge = Ridge(alpha=0.1)
ridge.fit(X_train_scaled, y_train_scaled)
y_train_pred1 = ridge.predict(X_train_scaled)
y_test_pred1 = ridge.predict(X_test_scaled)
print(ridge.coef_)

print('MSE train: {:.7f}, test: {:.7f}'.format(
    mean_squared_error(y_train_scaled, y_train_pred1),
    mean_squared_error(y_test_scaled, y_test_pred1)))
print('R^2 train: {:.7f}, test: {:.7f}'.format(
    r2_score(y_train_scaled, y_train_pred1),
    r2_score(y_test_scaled, y_test_pred1)))
mae = mean_absolute_error(y_train_scaled, y_train_pred1)
print("MAE: train", mae)
mae1 = mean_absolute_error(y_test_scaled, y_test_pred1)
print("MAE: test", mae1)

```

Рис.17 – Побудова гребеневої регресії

```

▶ y_train_pred_original = sc_y.inverse_transform(y_train_pred1.reshape(-1, 1)).flatten()
y_test_pred_original = sc_y.inverse_transform(y_test_pred1.reshape(-1, 1)).flatten()
y_train_original = sc_y.inverse_transform(y_train_scaled.reshape(-1, 1)).flatten()
y_test_original = sc_y.inverse_transform(y_test_scaled.reshape(-1, 1)).flatten()

```

```

[ ] plt.figure(figsize=(14, 6))

# Тренувальний набір
plt.subplot(1, 2, 1)
plt.scatter(y_train_original, y_train_pred_original, alpha=0.3)
plt.plot([y_train_original.min(), y_train_original.max()], [y_train_original.min(), y_train_original.max()], 'r--')
plt.xlabel('Фактичний Чистий Прибуток')
plt.ylabel('Прогнозований Чистий Прибуток')
plt.title('Тренувальний Набір')
plt.grid(True)

```

Рис.18 – Візуалізація результатів тренувальної вибірки

```

▶ plt.subplot(1, 2, 2)
plt.scatter(y_test_original, y_test_pred_original, alpha=0.3)
plt.plot([y_test_original.min(), y_test_original.max()], [y_test_original.min(), y_test_original.max()], 'r--')
plt.xlabel('Фактичний Чистий Прибуток')
plt.ylabel('Прогнозований Чистий Прибуток')
plt.title('Тестовий Набір')
plt.grid(True)

plt.tight_layout()
plt.show()

```

Рис.19 – Візуалізація результатів тестової вибірки

```

▶ X_new = pd.DataFrame({
    'const': 1,
    'Розмір кредитного портфеля': [53497, 54166, 54843, 55529],
    'Резерви під кредитні ризики\t': [-2612, -2645, -2678, -2711],
    'Активи': [200524, 205652, 210911, 216304],
    'Капіталізація банку': [0.132042, 0.139542, 0.147042, 0.154542],
    'Кошти банків': [386, 425, 467, 514],
    'Загальний обсяг внесків': [170081, 174430, 178891, 183466],
    'Депозити фіз. осіб': [61408, 62329, 63264, 64212],
    'Зобов\язання': [173312, 175478, 177672, 179892],
    'Рівень інфляції': [1.014, 1.014, 1.014, 1.014],
    'Облікова ставка': [0.150, 0.150, 0.150, 0.150],
    'Зміна курсу(USD)': [38.436, 37.667, 36.914, 36.175]
})
# Додавання константи до нових даних
X_new = sm.add_constant(X_new)
# Масштабування нових даних
X_new_scaled = sc_x.transform(X_new)
# Прогноз на основі нових даних
y_new_pred_scaled = ridge.predict(X_new_scaled)
# Зворотнє масштабування передбачених значень
y_new_pred1 = sc_y.inverse_transform(y_new_pred_scaled.reshape(-1, 1))

print('Прогнозовані значення для нових даних:', y_new_pred1)

```


 Прогнозовані значення для нових даних: [[ 3892.38680564]
  
[ 3935.39592839]
  
[ 3980.14048197]
  
[ 4026.58924392]]

Рис.20 – Прогнозування чистого прибутку на основі нових значень

```

from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import PolynomialFeatures
X = df.drop(['Чистий прибуток', 'Загальний обсяг внесків', 'Активи', 'Зобов'язання', 'Зміна курсу(USD)', 'Депозити фіз. осіб'], axis=1).values
y = df['Чистий прибуток'].values
# Нормалізація даних
sc_x = StandardScaler()
sc_y = StandardScaler()
X_std = sc_x.fit_transform(X)
y_std = sc_y.fit_transform(y.reshape(-1, 1)).flatten()
# Лінійна регресія
regr = LinearRegression()
# Створення квадратичних та кубічних ознак
quadratic = PolynomialFeatures(degree=2)
cubic = PolynomialFeatures(degree=3)
X_quad = quadratic.fit_transform(X_std)
X_cubic = cubic.fit_transform(X_std)
# Навчання квадратичної моделі
regr_quad = regr.fit(X_quad, y_std)
y_quad_fit = regr_quad.predict(X_quad)
quadratic_r2 = r2_score(y_std, y_quad_fit)
regr_linear = regr.fit(X_std, y_std)
y_lin_fit = regr_linear.predict(X_std)
linear_r2 = r2_score(y_std, y_lin_fit)
# Навчання кубічної моделі
regr_cubic = regr.fit(X_cubic, y_std)
y_cubic_fit = regr_cubic.predict(X_cubic)
cubic_r2 = r2_score(y_std, y_cubic_fit)
print("R^2 for Quadratic Regression:", quadratic_r2)
print("R^2 for Cubic Regression:", cubic_r2)
print("R^2 for Linear Regression:", linear_r2)

```

Активация W  
Перейдіть до роз

Рис.21 – Побудова поліноміальної регресії

```

▶ import statsmodels.api as sm

# Створення моделі поліноміальної регресії
model = sm.OLS(y_std, X_quad)

# Виконання підгонки моделі
results = model.fit()

# Виведення статистичної інформації
print(results.summary())

```

Рис.22 – Виведення результатів побудованої поліноміальної регресії

```

▶ from scipy.stats import shapiro

# Визначення залишків
residuals = y_std - y_quad_fit

# Перевірка нормальності залишків
stat, p = shapiro(residuals)
print('Statistics=%.3f, p=%.3f' % (stat, p))
# Інтерпретація результатів
alpha = 0.05
if p > alpha:
    print('Зразок видається нормально розподілим (не відкидається H0)')
else:
    print('Зразок не видається нормально розподілим (відкидається H0)')

```

```

⇒ Statistics=0.978, p=0.523
Зразок видається нормально розподілим (не відкидається H0)

```

```

[ ] from sklearn.metrics import mean_absolute_error
mae = mean_absolute_error(y_std, y_quad_fit)
print("MAE: train", mae)

```

```

⇒ MAE: train 0.28947239608571734

```

Рис.23 – Перевірка залишків на нормальність розподілу та розрахунок середньої абсолютної похибки

```
▶ plt.figure(figsize=(12, 8)) # Задайте більші розміри фігури
indexes = range(len(y_std))
plt.scatter(indexes, y_std, label='training points', color='black')
plt.plot(indexes, y_lin_fit,
         label='linear,  $R^2={:.2f}$ '.format(linear_r2),
         color='blue',
         lw=2,
         linestyle=':')

plt.plot(indexes, y_quad_fit,
         label='quadratic,  $R^2={:.2f}$ '.format(quadratic_r2),
         color='red',
         lw=2,
         linestyle='-')

plt.plot(indexes, y_cubic_fit,
         label='cubic,  $R^2={:.2f}$ '.format(cubic_r2),
         color='green',
         lw=2,
         linestyle='--')

plt.xlabel('Спостереження')
plt.ylabel('Чистий прибуток')
plt.legend(loc='upper right')
plt.show()
```

Рис.24 – Візуалізація результатів

```

# Creating new data for prediction
new_data = pd.DataFrame({
    'Розмір кредитного портфеля': [53497, 54166, 54843, 55529],
    'Резерви під кредитні ризики': [-2612, -2645, -2678, -2711],
    'Капіталізація банку': [0.132042, 0.139542, 0.147042, 0.154542],
    'Кошти банків': [386, 425, 467, 514],
    'Рівень інфляції': [1.014, 1.014, 1.014, 1.014],
    'Облікова ставка': [0.150, 0.150, 0.150, 0.150],
})

# Normalizing the new data
new_data_std = sc_x.transform(new_data.values)

# Creating quadratic features for the new data
new_data_quad = quadratic.transform(new_data_std)

# Predicting the profit for the new data
predicted_profit_std = regr_quad.predict(new_data_quad)

# Inverting the normalization of the predicted profits
predicted_profit = sc_y.inverse_transform(predicted_profit_std.reshape(-1, 1)).flatten()

# Displaying the predicted profits
print(predicted_profit)

```

[3465.3134122 3748.60482619 4194.0703105 4807.86837414]

Рис.25 – Прогнозування чистого прибутку на основі нових значень

```

# Припустимо, що у вас є дані фактичних значень і спрогнозованих значень для різних моделей
actual_profit = df['Чистий прибуток'].values
# Приклад прогнозів різних моделей
ridge_pred = np.array([3878.58625437, 3927.5364874, 3978.07110631, 4030.10792779])
lasso_pred = np.array([4488.2287449, 4499.04445332, 4511.34909178, 4524.80140297])
polynomial_pred = np.array([3465.3134122, 3748.60482619, 4194.0703105, 4807.86837414])
linear_pred = np.array([3899.999487, 3784.735413, 3671.934283, 3561.486283])
# Визначаємо індекси для нових прогнозованих значень
start_index = len(actual_profit)
future_indices = np.arange(start_index, start_index + 4)
# Створимо графік
plt.figure(figsize=(10, 6))
# Фактичні значення
plt.plot(np.arange(len(actual_profit)), actual_profit, label='Фактичний Чистий Прибуток', marker='o')
# Прогнозовані значення для різних моделей
plt.plot(future_indices, ridge_pred, label='Гребенева регресія', marker='x')
plt.plot(future_indices, lasso_pred, label='Лассо регресія', marker='x')
plt.plot(future_indices, polynomial_pred, label='Поліноміальна регресія', marker='x')
plt.plot(future_indices, linear_pred, label='Лінійна регресія', marker='x')

# Додаткові налаштування графіку
plt.axvline(x=start_index - 1, color='grey', linestyle='--') # Вертикальна лінія для відокремлення фактичних і прогнозованих значень
plt.xlabel('Індекс')
plt.ylabel('Чистий Прибуток')
plt.title('Фактичний та Прогнозований Чистий Прибуток')
plt.legend()
plt.grid(True)
plt.show()

```

Рис.26 – Візуалізація прогнозів на нових спостереженнях усіх моделей