

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА**

**Навчально-науковий інститут
«Інститут інформаційних технологій в економіці»**

Кафедра математичного моделювання та статистики

Освітньо-професійна програма

Галузь знань

Спеціальність

Економічна кібернетика

05 Соціальні та поведінкові науки

051 Економіка

Форма навчання: очна (денна)

КВАЛІФІКАЦІЙНА БАКАЛАВРСЬКА РОБОТА

на тему «Моделювання банківського кредитного ризику»

(назва теми)

здобувача Василенко Дениса Павловича

(ПІБ, підпис)

Науковий керівник: кандидат економічних наук,
доцент Іщук Я. В.

(науковий ступінь, учене звання, ПІБ)

(підпис)

**Робота допущена до захисту перед екзаменаційною
комісією з атестації здобувачів вищої освіти (ЕК)**

Завідувач кафедри кандидат фізико-математичних наук,
професор Великоіваненко Г.І.

(підпис)

Київ 2023

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАДИМА ГЕТЬМАНА**

**Навчально-науковий інститут
«Інститут інформаційних технологій в економіці»
Кафедра математичного моделювання та статистики**

Освітньо-професійна програма
Галузь знань
Спеціальність

Економічна кібернетика
05 Соціальні та поведінкові науки
051 Економіка

ПОГОДЖЕНО

Керівник проектної групи (гарант)
освітньо-професійної програми

(підпис)
_____ 20__ р.

ЗАТВЕРДЖУЮ

Завідувач кафедри математичного
моделювання та статистики
Г.І. Великоіваненко

(підпис)
_____ 20__ р.

ІНДИВІДУАЛЬНЕ ЗАВДАННЯ

**здобувачу вищої освіти Василенко Денису Павловичу
очної (денної) форми навчання**

на підготовку кваліфікаційної бакалаврської роботи

на тему «Моделювання банківського кредитного ризику»

Тему затверджено наказом ректора Університету від "20" березня 2023 р. №

Кваліфікаційна бакалаврська робота виконується на матеріалах

ТОВ «ЕПІЦЕНТР К»

План кваліфікаційної бакалаврської роботи

Розділ 1	ЕКОНОМІЧНА СУТНІСТЬ БАНКІВСЬКОГО РИЗИКУ
(назва розділу)	
Розділ 2	ОГЛЯД МЕТОДІВ АНАЛІЗУ І ПРОГНОЗУВАННЯ, ЩО ЗАСТОСОВУЮТЬСЯ ДЛЯ АНАЛІЗУ БАНКІВСЬКОГО РИЗИКУ
(назва розділу)	
Розділ 3	ПРОГНОЗУВАННЯ ДОХОДІВ КЛІЄНТІВ БАНКУ ДЛЯ ФОРМУВАННЯ ОЦІНКИ РИЗИКОВОСТІ З ВИКОРИСТАННЯМ СКОРИНГОВОЇ МОДЕЛІ
(назва розділу)	

Об'єкт дослідження:	Процеси банківського кредитування і ризику, які виникають при цьому.
Предмет дослідження:	Математичні методи і моделі для аналізу кредитного ризику.
Мета кваліфікаційної бакалаврської роботи:	Дослідження кредитного ризику на основі теоретичних засад та практичних аспектів діяльності клієнтів банку або фінансової установи, використати на практиці методи прогнозування часових для прогнозу доходів клієнтів і структурувати результати прогнозу в скоринговій системі.

Конкретні завдання, які здобувач повинен виконати для досягнення поставленої мети:

У розділі 1	Проаналізувати поняття банківського ризику, визначити його характер і класифікацію і методи його регуляції. Дослідити досвід застосування методів прогнозування для аналізу банківських ризиків. Визначити методи управління банківським ризиком, описати скорингову систему як один із методів.
У розділі 2	Описати методологію і математичне пояснення застосування моделей прогнозування часових рядів.
У розділі 3	Сформулювати алгоритм прогнозування банківського кредитного ризику на основі описаних методів прогнозування. Розробити скорингову модель для розрахунку оцінки ризикованості клієнта на основі моделей прогнозування.

**Завдання підготував
науковий керівник**

(підпис)

Я. В. Ішук

(ініціали, прізвище)

«____» _____ 2023 р.

**Завдання одержав
здобувач**

Д. П. Василенко

(підпис)

(ініціали, прізвище)

«____» _____ 2023 р.

РЕФЕРАТ

Об'єктом дослідження є процеси банківського кредитування і ризику, які виникають при цьому.

Предметом дослідження є математичні методи і моделі для аналізу кредитного ризику.

Метою роботи є дослідження кредитного ризику на основі теоретичних засад та практичних аспектів діяльності клієнтів банку або фінансової установи, використати на практиці методи прогнозування часових для прогнозу доходів клієнтів і структурувати результати прогнозу в скоринговій системі.

Для досягнення поставленої мети необхідно виконати наступні завдання:

1. дослідити характер банківського ризику і методи його аналізу;
2. визначити роль прогнозування як методу управління ризиками;
3. розглянути методи оцінки варіації даних в часових рядах
4. розглянути методологію розрахунку ковзної середньої й прогнозування на її основі для управління кредитним ризиком;
5. дослідити роботу методів простого експоненційного згладжування і їх використання для аналізу часового ряду;
6. розглянути методологію прогнозування за допомогою моделі Хольта при управлінні банківським ризиком;
7. розглянути методи прогнозування часових рядів на основі моделі Хольта – Вінтерса;
8. створення скоринг системи для підрахунку оцінки ризикованості клієнта;

Методами дослідження в кваліфікаційній роботі є методи аналізу варіації даних, методи згладжування часових рядів, модель Хольта і моделі на основі методу Хольта – Вінтерса.

Впродовж цього дослідження роботи було ретельно досліджено сучасні методи оцінки ризикованості клієнтів банків, використання прогнозування для розрахунку ризику і методології прогнозування часових рядів за різними моделями. Також були проаналізовані реальні данні кредитного портфелю фінансової міжнародної компанії з надання кредитних послуг, сформовані моделі прогнозування доходів клієнтів за методами згладжування мовою R і створена

скорингова система в середовищі Excel. Побудовані прогнози розділились на 2 групи. Моделі які не враховували сезонність прогнозували подальше зростання доходів протягом трьох місяців. Були використані моделі на основі ковзної середньої, ETS і модель Хольта. Всі вони спрогнозували приріст доходів. Проте моделі які враховують сезонність, такі як моделі Хольта Вінтерса, Хольта Вінтерса адативна та мультиплікативна моделі і модель Хольта Вінтерса з демпфованим трендом, показали прогноз на спад доходів.

Обсяг кваліфікаційної роботи складає 66 сторінок. У складі дипломної роботи 3 розділи.

Кваліфікаційна бакалаврська робота містить 23 ілюстрації, 9 формул та 39 літературних джерел.

Рік виконання кваліфікаційної бакалаврської роботи – 2023.

Рік захисту роботи – 2023.

Ключові слова: банківський ризик, скоринг, прогнозування, аналіз, модель, часовий ряд, дохідність, Хольт.

В і д г у к
про кваліфікаційну бакалаврську роботу
здобувача навчально-наукового інституту /факультету
освітньо-професійної програми «Економічна кібернетика»
Василенка Дениса Павловича
на тему «**Модельовання банківського кредитного ризику**»

1. Актуальність теми. Тема модельовання банківського кредитного ризику залишається дуже актуальною в сучасному світі. Кредитний ризик є одним з основних ризиків, з якими зіштовхуються банки та інші фінансові установи. Модельовання кредитного ризику дозволяє оцінити його ймовірність та можливі наслідки. Це допомагає фінансовим установам приймати інформовані рішення щодо надання кредитів, установлення відсоткових ставок, а також при розробці стратегій управління ризиками.

2. Позитивні риси кваліфікаційної бакалаврської роботи. Кваліфікаційна бакалаврська робота демонструє, що здобувач здатний дібрати та опрацювати необхідну наукову літературу, зібрати необхідну для дослідження інформацію, аналізувати її на основі обраних методів дослідження, а також виявляє його навички самостійної роботи з використанням наукових методів та підходів. Робота також вказує на його вміння чітко та логічно викладати результати своїх досліджень.

3. Наявність самостійних розробок автора. Автор роботи продемонстрував навички самостійних розробок, сформувавши алгоритм прогнозування банківського кредитного ризику на основі моделі Хольта, адитивної та мультиплікативної моделі Хольта-Вінтерса, демпфованого тренду, також була розроблена скорингова модель для розрахунку оцінки ризикованості клієнта на основі моделей прогнозування.

4. Цінність теоретичних висновків та практичних рекомендацій. На основі отриманих результатів, можна зробити висновок, що використання скорингової моделі, де враховані результати прогнозування потоків грошових коштів на основі різних моделей є достатньо новим і ефективним підходом до модельовання кредитного ризику. Ці рекомендації можуть бути використані банками для поліпшення своїх процедур оцінки кредитоспроможності клієнтів, розроблення стратегій управління кредитним ризиком та виявлення потенційних ризикових ситуацій.

5. Наявність недоліків. Попри наявність самостійних розробок, у даній роботі можна відзначити деякі недоліки. Наприклад, відсутність розбиття вибірки на тренувальну та тестову, що дозволило б використовувати тренувальну вибірку для навчання моделі і тестову вибірку для оцінки її прогностичної здатності для нових невідомих даних. Для оцінки надійності та ефективності скорингової моделі доцільно було провести оцінку та валідацію запропонованих моделей.

6. Загальна оцінка кваліфікаційної бакалаврської роботи та її допущення до захисту перед ЕК. Проведений аналіз відповідає в цілому вирішенню поставлених завдань, є окремі недоліки та некоректності в опрацюванні отриманих результатів, текст оформлено з незначними порушеннями вимог оформлення таблиць, рисунків. Робота допущена до захисту.

Науковий керівник
доцент кафедри математичного моделювання та статистики, к.е.н.
“16” червня 2023 р.

Ішук Я.В.

ЗМІСТ

ВСТУП.....	2
РОЗДІЛ 1. ЕКОНОМІЧНА СУТНІСТЬ БАНКІВСЬКОГО РИЗИКУ	5
1.1 Сутність та зміст управління банківського ризику	5
1.2 Застосування методів прогнозування для аналізу банківських ризиків	14
1.3 Методи управління банківськими ризиками, застосування скорингової моделі для управління банківським ризиком.....	19
РОЗДІЛ 2. ОГЛЯД МЕТОДІВ АНАЛІЗУ І ПРОГНОЗУВАННЯ, ЩО ЗАСТОСОВУЮТЬСЯ ДЛЯ АНАЛІЗУ БАНКІВСЬКОГО РИЗИКУ	22
2.1 Оцінка якості даних за допомогою коефіцієнту Херста, стандартного відхилення і коефіцієнту варіації	22
2.2 Методологія прогнозування даних на основі ковзної середньої та простого експоненціального згладжування	28
2.3 Методологія прогнозування методом Хольта, Хольта – Вінтерса і демпфованого тренду	31
РОЗДІЛ 3. ПРОГНОЗУВАННЯ ДОХОДІВ КЛІЄНТІВ БАНКУ ДЛЯ ФОРМУВАННЯ ОЦІНКИ РИЗИКОВОСТІ З ВИКОРИСТАННЯМ СКОРИНГОВОЇ МОДЕЛІ	41
3.1 Опис початкових даних і технічної реалізації	41
3.2 Підготовка та оцінка якості часових рядів	43
3.3 Прогнозування доходів клієнтів	47
3.4 Формування скоринг моделі	56
ВИСНОВКИ.....	59
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	61

ВСТУП

Аналіз і моделювання банківського кредитного ризику є важливою частиною роботи будь-якої фінансової організації. Від коректної оцінки кредитного ризику залежать витрати компаній і можливість клієнтів сплачувати відсотки згідно з договором без прострочення платежів.

В сучасних умовах повномасштабної війни України з'явилося багато невизначеності стосовно фінансової стабільності в країні. Багато людей втратило постійну роботу і при цьому потребують фінансової допомоги. В таких кризових умовах управління банківським кредитним ризиком є складним завданням, і класичні методи управління ризиком потребують покращення щодо точності та застосування нових методів. І тому дослідження питання поліпшення управління банківським ризиком є актуальним на сьогодні.

Моделювання та управління кредитним банківським ризиком привернуло увагу численних вчених та професорів, оскільки ця тема є критично важливою для фінансової стабільності. Один з відомих дослідників у цій галузі — турецький професор Герчек Озпарлак, який проводив дослідження, спрямовані на розробку нових методик оцінки кредитного ризику та моделювання банківських портфелів. Інші вчені, такі як Джоель Сусс та Генрі Трайтель, фокусувалися на аналізі впливу макроекономічних факторів на кредитний ризик та розробці стратегій управління ризиком для банківських установ. Дослідницькі роботи таких вчених сприяли розумінню складності кредитного ризику та сприяли розвитку більш точних та надійних моделей прогнозування та управління ризиком у банківському секторі.

Загальне поняття кредитного ризику, як їм керувати та розраховувати було сформовано користуючись Basel III. Цей документ є міжнародними стандартами регулювання банківської системи, розробленими Базельським комітетом з банківського нагляду (Basel Committee on Banking Supervision). Він був впроваджений з метою зміцнення стійкості банків та зменшення ризиків,

пов'язаних з їх діяльністю. Basel III було впроваджено поетапно з 2010 року до 2019 року. Ці стандарти були розроблені відповідно до рекомендацій Групи з фінансової діяльності (Financial Stability Board) та були підтримані багатьма країнами та банківськими органами нагляду.

Кредитний ризик є одним із важливих аспектів, який розглядає Basel III. Він визнає, що кредитний ризик є одним з найважливіших ризиків, з якими стикаються банки, а також встановлює ряд вимог та рекомендацій, що стосуються оцінки та управління кредитним ризиком.

Метою є дослідження кредитного ризику на основі теоретичних засад та практичних аспектів діяльності клієнтів банку або фінансової установи, використати на практиці методи прогнозування часових для прогнозу доходів клієнтів і структурувати результати прогнозу в скоринговій системі.

Для досягнення поставленої мети необхідно виконати такі завдання:

- дослідити характер банківського ризику і методи його аналізу;
- визначити роль прогнозування як методу управління ризиками;
- розглянути методи оцінки варіації даних в часових рядах
- розглянути методологію розрахунку ковзної середньої й прогнозування на її основі для управління кредитним ризиком;
- дослідити роботу методів простого експоненційного згладжування і їх використання для аналізу часового ряду;
- розглянути методологію прогнозування за допомогою моделі Хольта при управлінні банківським ризиком;
- розглянути методи прогнозування часових рядів на основі моделі Хольта – Вінтерса;
- створення скоринг системи для підрахунку оцінки ризикованості клієнта;

Об'єктом дослідження є процеси банківського кредитування і ризику, які виникають при цьому.

Предметом дослідження є математичні методи і моделі для аналізу кредитного ризику.

Методами дослідження в кваліфікаційній роботі є методи аналізу варіації даних, методи згладжування часових рядів, модель Хольта і моделі на основі методу Хольта – Вінтерса.

Інформаційна база дослідження містить реальні дані фінансової міжнародної компанії з надання кредитних послуг для авторів контенту на онлайн платформі YouTube, наукові праці українських і закордонних авторів і технічну документацію з використання мови R.

Структура роботи включає вступ, три розділи, висновки та список використаних джерел.

РОЗДІЛ 1

ЕКОНОМІЧНА СУТНІСТЬ БАНКІВСЬКОГО РИЗИКУ

1.1 Сутність та зміст управління банківського ризику

Проблемам дослідження кредитного ризику присвячена велика кількість наукових робіт як українських, так і закордонних вчених та професорів.

У статті "Проблеми управління кредитним ризиком в комерційних банках України" автором Л. Б. Штефан розглядаються основні проблеми, пов'язані з управлінням кредитним ризиком в банківському секторі України. Автор аналізує особливості кредитного ризику та його вплив на фінансову стабільність банків, а також розглядає різні методики та стратегії управління кредитним ризиком. Дослідження автора спрямоване на виявлення основних проблем та пошук ефективних рішень у сфері управління кредитним ризиком в банківському секторі України [1].

У роботі "Counterparty credit risk in Basel III" автор аналізує проблему контрагентського кредитного ризику в контексті впровадження стандартів Базель III. Він досліджує вплив нових регуляторних вимог на оцінку та управління контрагентським кредитним ризиком у фінансовому секторі. Дослідження автора зосереджене на розумінні основних положень Базель III, а також на аналізі методів виміру та контролю контрагентського кредитного ризику, що дозволяє зрозуміти вплив цих нових регуляторних вимог на банківську систему та її стійкість [2]. Своєю чергою в дослідженні "5 C's кредиту в кредитній індустрії" автором Джоном Е. Байденом розглядаються основні аспекти, що визначають кредитну спроможність в позиковій галузі. Дослідження автора спрямоване на

надання практичних вказівок та рекомендацій щодо оцінки кредитного ризику та прийняття рішень з позикової діяльності на основі аналізу цих п'яти складових [3].

Дослідження з прогнозування банківського провели Генрі Трейтел і Джоел Сасс. В їх дослідженні "Прогнозування фінансових проблем банків у Великобританії за допомогою машинного навчання" проведено аналіз використання методів машинного навчання для прогнозування фінансової нестабільності банків у Великобританії. Автори досліджують різні моделі та алгоритми машинного навчання з метою розробки ефективних інструментів для передбачення можливих проблем та фінансових ризиків у банківській системі Великобританії [4].

Схоже дослідження провів Герчек Озпарлак під назвою "Прогнозування банкрутства корпоративних підприємств з використанням методів машинного навчання: випадок США". У своїй роботі автором було проведено аналіз застосування методів машинного навчання для прогнозування банкрутства корпоративних підприємств у США. Автор досліджує різні моделі та алгоритми машинного навчання з метою розробки ефективних інструментів для передбачення можливих фінансових проблем та банкрутства у корпоративному секторі США [5].

На сьогодні, банківський ризик є невід'ємною частиною діяльності будь-якого комерційного банку чи схожих структур з надання фінансового кредитування. Своєю чергою, причини ризиків банкрутства банків можна розкласти на 2 категорії: внутрішні й зовнішні [1].

До внутрішніх відносяться такі причини, вплив яких банк здатен мінімізувати використовуючи лише внутрішні можливості та мінімальні фінансові затрати. У цю категорію заведено відносити:

- правові ризики, які можуть виникнути в наслідок неправильних правових порад, що може призвести до помилкової оцінки вартості активів;
- ризик ліквідності, пов'язаний зі здатністю банку до забезпечення необхідної кількості готівки за допомогою конвертації активів;

- операційний ризик, що пов'язаний з помилками в корпоративному управлінні та внутрішньому адмініструванні, як наслідок цей тип ризику здатен причинити фінансові втрати пов'язанні з помилками працівників, шахрайством, або несправністю інформаційних систем;
- стратегічний ризик, що виникає в наслідок помилкових управлінських рішень, стратегій розвитку цілей банку і неправильний розподіл ресурсів для досягнення цих цілей.

Вплив внутрішніх ризиків банк здатен врегулювати за допомогою коректного адміністрування і шляхом прийняття коректних рішень, при цьому не витрачаючи велику кількість фінансів.

До зовнішніх причин банківського ризику відносяться ті, які банк не здатен врегулювати чи зменшити їх вплив, як правило ці причини залежать від загального економічного клімату в країні в певний момент, сюди можна віднести:

- ринковий ризик, який виникає внаслідок коливань ринкових цін і охоплює валютний ризик, який виникає в результаті зміни валютного курсу і впливає на фінансову позицію банку;
- відсотковий ризик, що виникає через мінливість відсоткових ставок і приводить до уразливості фінансової позиції банку
- фондовий ризик що характеризується коливанням цін на цінні папери в які інвестує банк, що може призвести до фінансових втрат [6].

Проте кредитний ризик, що виникає в результаті нестачі фінансової вигоди банку через невчасне, неповне чи неналежне дотримання своїх зобов'язань кредиторами, позичальниками чи контрагентами можна одночасно віднести та до внутрішніх і до зовнішніх причин банківського ризику. До зовнішніх причин його можна віднести оскільки зниження платоспроможності клієнта є результатом зовнішніх економічних чинників і банк не здатен на них вплинути. Проте його можна віднести й до внутрішніх, оскільки коливання фінансового становища банку виникає в результаті впливу цих зовнішніх чинників на клієнтську базу яку створив банк.

Саме через причини свого виникнення і впливу на фінансове положення компанії, кредитний ризик є великою проблемою для комерційних організацій і банків.

Кредитний ризик, згідно з Basel III, можна охарактеризувати як ймовірність виникнення втрат, викликаних невиконанням позичальником банку або контрагентом своїх зобов'язань відповідно до умов угоди [2]. Цей ризик відображає потенційні наслідки, які можуть виникнути, коли позичальник неспроможний або небажано не виконує свої зобов'язання з погашення кредиту чи іншої фінансової угоди. Визначення кредитного ризику засвідчує, що він включає не лише повне невиконання зобов'язань (дефолт), але й інші форми невідповідності, такі як прострочені платежі або часткове невиконання. Крім того, враховується можливість знецінення або втрати вартості активів, пов'язаних з кредитними операціями, а також можливість змін у кредитному рейтингу позичальника.

Оцінка кредитного ризику полягає у визначенні кредитного ризику контрагента відповідно до критеріїв кредитного прийняття фінансової установи з метою встановлення здатності та готовності контрагента виконувати свої кредитні зобов'язання, як при оформленні кредиту, так і на будь-якому етапі його дії. Метою управління кредитним ризиком є максимізація ризик-зваженої дохідності банку шляхом підтримки кредитного ризику в межах прийнятних параметрів. Банки повинні управляти кредитним ризиком, що є неодмінною складовою всього портфеля, а також ризиком в окремих кредитах або операціях. Стійке управління кредитним ризиком є неодмінною складовою довгострокової життєздатності фінансових установ. Це також є критичним фактором для сталого розвитку реальної економіки.

Оцінка і моделювання кредитного ризику є важливими аспектами фінансового управління підприємством згідно з вимогами МСФЗ (IFRS) 9 [7]. Підприємство повинно враховувати різні фактори та використовувати обґрунтовані методи для визначення очікуваних кредитних збитків. При цьому, враховуються об'єктивна і зважена на ймовірність сума, вартість грошей у часі

та інформація про минулі події, поточні умови та прогнози щодо економічних умов у майбутньому. Врахування цих факторів допомагає підприємству отримати вірогідну оцінку кредитного ризику і визначити адекватні резерви для покриття можливих кредитних збитків. Підприємство також повинно враховувати потоки від реалізації застави після завершення строку дії договору, щоб визначити повну вартість кредитного ризику. Оцінка очікуваних кредитних збитків проводиться на період, коли підприємство є вразливим до кредитного ризику, і не може бути зменшена шляхом управління ризиком. Згідно з МСФЗ, на кожну звітну дату підприємство повинно переглядати, чи зростав кредитний ризик значно після первісного визнання. Якщо договірні платежі стають простроченими більше ніж на 30 днів, припускається, що кредитний ризик зазнав значного зростання. Це дозволяє підприємству усвідомити потенційні ризики та вжити відповідних заходів для управління кредитним портфелем.

Оцінка і моделювання кредитного ризику відіграють суттєву роль у фінансовому управлінні, дозволяючи підприємству раціонально керувати своїм кредитним портфелем та забезпечувати адекватні резерви для покриття можливих кредитних збитків. Це сприяє покращенню якості фінансової звітності й впевненості у стабільності фінансових результатів підприємства. В Україні вся інформація стосовно кредитної історії й поточного фінансового положення клієнта надається через кредитне бюро. Це юридична особа, діяльність якої збір інформації про клієнтів банків, зберігання цієї інформації і її використання. В кредитній історії міститься інформація про ідентифікацію особи, відомості про минулі грошові зобов'язання суб'єкта кредитної історії, такі як суми кредиту, строк, порядок виконання, і поточний статус [8].

Методика п'ять "С" кредитного аналізу була сформована як результат розвитку та практики в галузі банківського кредитування. Історично точна дата й авторство формулювання цієї методики не є повністю визначеними, оскільки вона розвивалася поступово та широко використовувалася в практиці банківської та фінансової сфер. Вперше згадки про концепцію п'ять "С" були спостережені у публікаціях та розробках у галузі банківського кредитування на початку ХХ

століття. Але точний початок використання цієї методики не може бути чітко визначений через її плавний розвиток та широку адаптацію в різних контекстах. За десятиліття розвитку п'ять "С" стали широко використовуватися у банківській та фінансовій галузях як стандартна методологія для оцінки кредитної якості позичальників. Вона є неофіційною мнемонікою набору факторів ризику, що зазвичай вважаються впливовими на визначення кредитної здатності комерційних позичальників, включаючи кредитування малого та середнього бізнесу або кредитування корпоративних клієнтів. Попри те, що точні деталі формування цієї методики можуть бути неоднозначними, вона залишається цінним інструментом для оцінки ризику та прийняття рішень в кредитному процесі [3].

В розвитку методики п'ять "С" кредитного аналізу було визначено п'ять основних критеріїв, які широко використовуються для оцінки кредитної якості позичальників. Ці критерії включають ємність, капітал, характер, заставу та умови. На основі цих критеріїв банки та фінансові установи визначають здатність та готовність позичальника виконувати свої зобов'язання з погашення кредиту [9].

Характер (Character): Цей фактор оцінює моральні та етичні якості позичальника, його кредитну історію та відносини з попередніми кредиторами. Враховуються такі фактори, як платоспроможність, дотримання кредитних зобов'язань, репутація та здатність до управління фінансами. Кредитний звіт позичальника, що генерується кредитними бюро, містить докладну інформацію про його кредитну історію, включаючи кількість позик, що були взяті у минулому, та своєчасність їх погашення. Кредитний звіт також включає дані про інкасаційні рахунки та банкрутство, які зберігаються протягом тривалого періоду, зазвичай семи-десяти років. Кредитори використовують цю інформацію для оцінки кредитного ризику позичальника. Наприклад, модель кредитного скорину FICO використовує дані з кредитного звіту споживача для розрахунку кредитного рейтингу, який допомагає кредиторам передбачити ймовірність своєчасного погашення позичок. Багато кредиторів встановлюють мінімальний

кредитний рейтинг як вимогу для схвалення нових кредитів. Високий кредитний рейтинг позичальника збільшує ймовірність отримання схвалення на кредит. Крім того, кредитори використовують кредитний рейтинг для встановлення ставок і умов позичок, і зазвичай пропонують більш привабливі умови для позичальників з високою кредитною історією. Враховуючи важливість доброї кредитної оцінки та точних кредитних звітів для забезпечення позик, рекомендується скористатись послугами служб кредитного моніторингу для збереження цієї інформації. З метою покращення цього критерію, потенційні позичальники повинні впевнитися, що їх кредитна історія відображається коректно та точно у кредитних звітах. Важливо вчасно погашати регулярні рахунки, і для забезпечення цього можна розглянути використання автоматичних платежів. Регулярні погашення боргів і встановлення позитивної історії своєчасних платежів сприяють підвищенню кредитного рейтингу.

Здатність (Capacity): Цей фактор належить до здатності позичальника генерувати достатній дохід для погашення кредитного зобов'язання. Визначаються показники, такі як рентабельність, обсяги продажів, граничні платежі та інші економічні фактори, які дозволяють оцінити, наскільки позичальник може виконувати свої фінансові зобов'язання. Цей фактор залежить від внутрішніх і зовнішніх факторів.

Внутрішні фактори включають характеристики кредитора, такі як її здатність генерувати грошовий потік, розмір активів та зобов'язань, або розмір та строк позик на балансовому звіті.

Зовнішні фактори залежать від характеристик діяльності, а також від стану економіки [9].

Аналіз внутрішніх факторів означає дослідження фінансових показників клієнта, стабільності, капіталовкладень. Кожна з цих областей аналізується за допомогою кількох показників.

Аналіз фінансових показників і співвідношень включає:

Показники рентабельності та грошового потоку: аналіз зазвичай зосереджується на показниках, таких як маржі рентабельності, рентабельність

інвестованого капіталу та маржі грошового потоку. Здатність кредитора розрахуватися зі своїми зобов'язаннями залежить від його можливості генерувати достатньо фінансового потоку. Цей аналіз дозволяє аналітикам краще зрозуміти, чи вдасться компанії зібрати достатньо ресурсів для повернення позик.

Показники покриття: проблема не полягає тільки в тому, яку суму потрібно повернути як основну суму позики. Відсотки, що стягуються, неоднакові для кожної компанії і суттєво впливають на здатність позичальника. Тому аналіз показників покриття, таких як EBITDA (показник доходу після вирахування податків), є надзвичайно важливим.

Зовнішні фактори та здатність позичальника: деякі фактори, що впливають на здатність позичальника, є зовнішніми і, отже, мають мало відношення до конкретних характеристик. До таких зовнішніх факторів входять:

Сфера діяльності: сфера діяльності клієнта має бути проаналізована і взята в розрахунок. Компанія повинна класифікувати й визначити ризиковість цієї сфери діяльності, а також її стабільність. Клієнти які займаються ризиковою сферою діяльності мають більше ймовірність на порушення умов договору. В свою чергу клієнти з відсутністю стабільності можуть, так само, порушати умови договору не сплачуючи позику в відведений термін.

Економічна ситуація: треба брати до уваги загальну економічну ситуацію в країні, як вона впливає на дохід клієнта і наскільки він залежить від зовнішніх обставин.

Загалом, оцінка здатності позичальника вимагає комплексного підходу, який враховує як внутрішні, так і зовнішні фактори, що впливають на його фінансову стійкість та спроможність виконувати зобов'язання з позиками. Аналіз фінансових показників, стабільності, покриття та зовнішніх факторів допомагає зрозуміти ризики та можливості позичальника перед наданням позики.

Капітал (Capital): Цей фактор відноситься до фінансової стійкості позичальника і його власного капіталу. Оцінюється співвідношення власних коштів до всього капіталу, а також загальна фінансова структура компанії. Чим більший капітал у позичальника, тим більша його здатність покрити можливі

збитки і ризику. Велика кількість власного капіталу свідчить про рівень інвестицій позичальника, що може зробити кредиторів більш впевненими в продовженні кредиту.

Застава (Collateral): Цей фактор стосується активів або майна, яке може бути використане як гарантія в разі невиплати кредиту. Важливими елементами оцінки є тип та цінність заставного майна, його ліквідність і можливість його реалізації у разі необхідності. Застосування застави клієнта має значний вплив на його ризиковість і визначає рівень безпеки для кредитора. Застава допомагає забезпечити позичальника забезпеченням кредиту і надає кредитору гарантію, що в разі неплатоспроможності позичальника він може відновити частину втрачених коштів, реалізуючи заставлену власність. Зазвичай застава є самим об'єктом, на який позичаються гроші. Наприклад, автокредити забезпечуються автомобілями, а іпотечні кредити — житлом. Такі кредити, що забезпечуються заставою, часто називаються захищеними кредитами або захищеною заборгованістю. Вони вважаються менш ризикованими для кредиторів. У результаті, кредити, що забезпечені якоюсь формою застави, зазвичай надаються з нижчими процентними ставками та кращими умовами порівняно з іншими видами фінансування. Застосування застави допомагає знизити ризик для кредитора, оскільки він має право відновити втрачені кошти, якщо позичальник не виконує свої зобов'язання. Це створює додатковий стимул для позичальника дотримуватися умов позики та своєчасно сплачувати платежі. Коли позичальник своєчасно і повністю виконує свої зобов'язання, це сприяє його добрій кредитній історії та можливості отримати кредити на більш вигідних умовах у майбутньому. Важливо пам'ятати, що застава може мати певні ризики для позичальника. Якщо він не виконує свої зобов'язання і стає неплатоспроможним, кредитор має право конфіскувати заставлене майно. Тому перед використанням застави в якості забезпечення кредиту, позичальник повинен ретельно розглянути свою фінансову ситуацію та ризики, пов'язані зі зобов'язаннями. Важливо мати достатньо активів, які можна використовувати як заставу, та забезпечити своєчасну сплату зобов'язань для запобігання втраті заставленого майна. Таким

чином, застава клієнта впливає на його ризиковість, створюючи додаткову гарантію для кредитора та забезпечуючи можливість отримати кредит на кращих умовах. Однак, позичальник повинен бути обережним і усвідомлювати ризики, пов'язані зі зобов'язаннями та заставою, та своєчасно виконувати свої зобов'язання для забезпечення фінансової стабільності.

Умови (Conditions): Цей фактор враховує зовнішні умови, такі як стан ринку, економічна ситуація, законодавство та інші фактори, що можуть впливати на здатність позичальника погасити кредит. Визначається ступінь ризику, пов'язаного з зовнішніми факторами, які можуть впливати на позичальника. Умови кредитування клієнта мають значний вплив на його ризиковість і визначають рівень забезпечення для кредитора. Крім доходів позичальника, кредитори аналізують загальні умови, пов'язані з кредитом. Це може включати тривалість роботи позичальника на поточній роботі, стан його галузі, майбутню стабільність роботи. Умови кредиту, такі як відсоткова ставка і сума основного боргу, впливають на бажання кредитора фінансувати позичальника. Умови можуть відноситися до того, як позичальник збирається використовувати гроші. Бізнес-кредити, які можуть забезпечити майбутній грошовий потік, можуть мати кращі умови, ніж ремонт будинку в період затяжної ситуації на ринку нерухомості. Крім того, кредитори можуть брати до уваги умови, які знаходяться поза контролем позичальника, такі як стан економіки, тенденції у галузі або очікувані законодавчі зміни. Для компаній, які намагаються отримати кредит, ці необхідні умови можуть стосуватися перспектив ключових постачальників або фінансової стабільності клієнтів у найближчі роки.

Умови кредитування клієнта мають суттєвий вплив на його ризиковість, оскільки вони відображають стан і перспективи позичальника, а також зовнішні фактори, що можуть впливати на його здатність повернути кредит. Позичальникам варто ретельно розглянути умови кредитування, щоб забезпечити свою фінансову стабільність і мінімізувати ризики.

1.2 Застосування методів прогнозування для аналізу банківських ризиків

Прогнозування часових рядів є важливим інструментом в аналізі банківського ризику. В банківській сфері використовуються різні методи прогнозування, такі як авторегресійні моделі (AR), моделі ковзного середнього (MA), моделі авторегресії з ковзним середнім (ARMA), а також більш складні моделі, такі як авторегресійні інтегровані ковзні середні (ARIMA), моделі експоненційного згладжування (ESM), авторегресійні інтегровані ковзні середні зі згладжуванням сезонності (SARIMA) і моделі гармонічних регресійних моделей (HAR). Ці моделі дозволяють аналізувати та прогнозувати різні показники банківського ризику, такі як неплатоспроможність позичальників, коливання ринкових цін, портфельні ризики, кредитний ризик і ризик ліквідності. Прогнозування часових рядів дозволяє банкам оцінити майбутній ризик і прийняти відповідні стратегічні та оперативні рішення для управління ризиками. Для прогнозування часових рядів в аналізі банківського ризику часто використовуються статистичні методи, машинне навчання і інші аналітичні підходи. Ці методи базуються на аналізі минулих даних, виявленні трендів, сезонності, циклів та інших закономірностей в часових рядах, що дозволяє прогнозувати майбутні значення і ризики. Враховуючи широкий спектр даних, що використовуються в банківському секторі, прогнозування часових рядів стає корисним інструментом для виявлення тенденцій, виявлення високоризикових ситуацій та прийняття відповідних заходів для зниження ризиків.

В жовтні 2019 року вчені з Великобританії Joel Suss і Henry Treitel випустили дослідження присвячене розробці системи попередження про банківську небезпеку з використанням нових даних та методів машинного навчання. Автори використовують конфіденційні наглядові оцінки ризику фірм та дані регуляторних звітів для вибірки британських регульованих фірм з 2006 по 2012 роки. Дослідження має кілька ключових особливостей.

По-перше, використовується моделювання небезпеки фірм, а не прямого збанкрутування, що більш відповідає практичним потребам наглядових органів, оскільки дозволяє передбачити небезпеку задовго до випадку банкрутства. Дослідження також враховує різноманітні фактори ризику, включаючи менші нерозмічені організації, що часто виключаються з аналізу через обмежену доступність даних.

По-друге, автори використовують методи з машинного навчання поряд з традиційними підходами. Вони порівнюють шість різних моделей, включаючи логістичну регресію, випадковий ліс, алгоритм k найближчих сусідів (KNN), методи здвоєної класифікації дерев (random forest і boosting) та опорний векторний метод (SVM). Автори використовують різні метрики продуктивності, такі як площа під кривою ROC (AUC) та Brier Score, щоб оцінити ефективність моделей. Вони, що алгоритм випадкового лісу значно перевершує інші моделі за метриками продуктивності. Вони також аналізують вплив відносної вартості помилкових позитивів та помилкових негативів на рішення, виявляючи, що випадковий ліс виявляється найкращим в разі зростання вартості помилкових негативів порівняно з помилковими позитивами. Протягом дослідження автори також досліджують фактори, що призводять до банківської небезпеки, використовуючи методи машинного навчання для інтерпретації результатів. Вони визначають важливість різних макроекономічних змінних та фінансових показників на основі використання методу вектора Шеплі.

Результати цього дослідження свідчать про те, що випадковий ліс слід використовувати для побудови системи попередження про фінансову небезпеку. Щоб покращити прозорість алгоритму, автори досліджують фактори, що впливають на передбачені ймовірності моделі, використовуючи агрегацію значень Шеплі для кожного спостереження в тестовій вибірці та метод Шеплі регресії (Joseph 2019). Результати показують, що фактори, що впливають на випадковий ліс, якісно відрізняються від об'єднаної логістичної регресії, що пояснюється взаємодією між змінними. Шеплі регресія показує важливість макроекономічних змінних (особливо щорічної зміни середніх реальних доходів

у Великій Британії), чутливості фірми до ринкового ризику (відношення торгового портфеля до загальних активів), капітального буфера та чистої маржі відсотків. Крім того, автори також застосовуємо прості методи асемблювання для поєднання результатів всіх моделей, що приводить до істотних і статистично значущих покращень порівняно з випадковим лісом сам по собі [4].

Загалом, дане дослідження пропонує нові підходи до прогнозування банківської небезпеки з використанням нових даних та методів машинного навчання, і доводить наявність потенційного розвитку використання цих моделей. Хоча фінансові організації не можуть цілком опиратись на методи прогнозування в процесі аналізу ризиків, проте ці методи є ефективним інструментом для складання загальної оцінки. В свою чергу, включення результатів прогнозування в оцінку банківського ризику здатне значно покращити моделі оцінювання.

Також у 2022 році в міжнародному журналі менеджменту економіки і бізнесу вийшла наукова стаття написана Герчек Озпарлак (Gerçek Özparlak) на тему прогнозування банкрутства за допомогою методів машинного навчання. У статті здійснюється дослідження проблеми прогнозування банкрутства в секторі охорони здоров'я і порівнюються ефективність методів машинного навчання та класичних методів. Автори зазначають, що банкрутство є серйозним наслідком фінансових проблем підприємств і може викликати значні збитки для їхньої діяльності та зацікавлених сторін. Тому розробка моделей прогнозування банкрутства, які надають надійні результати, має велике значення для збереження підприємств та захисту рівня добробуту зацікавлених сторін. Одним зі стандартних методів прогнозування банкрутства є багатофакторний дискримінантний аналіз (MDA), який був запропонований Альтманом у 1968 році. Однак після цього було розроблено багато інших методів, включаючи логістичний аналіз, пробіт-аналіз та штучні нейронні мережі (ANN). Зокрема, ANN-моделі здобули широке застосування в прогнозуванні банкрутства, оскільки вони можуть успішно працювати з різними фінансовими проблемами і даними, включаючи часові ряди, пропуски і помилки. Для проведення

дослідження автори вибрали вибірку з 103 компаній у секторі охорони здоров'я, які оголосили про банкрутство між 2011 і 2021 роками, і 103 фінансово успішних компаній у тому ж періоді. Застосовуючи методи машинного навчання (SVM, ANN) та класичний метод MDA, дослідники намагалися побудувати моделі, які зможуть прогнозувати банкрутство в секторі охорони здоров'я. Результати дослідження показали, що методи машинного навчання (SVM, ANN) виявилися більш ефективними у прогнозуванні банкрутства, ніж класичний метод MDA. Використання SVM та ANN дозволило досягти високої точності та надійності прогнозів, здатних виявити ризики банкрутства заздалегідь. Такі результати свідчать про потенціал моделей машинного навчання для виявлення ризиків та забезпечення фінансової стабільності підприємств у секторі охорони здоров'я.

Згідно з дослідженням, використання методів машинного навчання для прогнозування банкрутства в секторі охорони здоров'я має переваги перед класичними методами, оскільки вони дозволяють здійснювати більш точний та надійний аналіз фінансових даних та ідентифікувати компанії з високим ризиком банкрутства. Це може сприяти покращенню стратегій управління ризиками та прийняттю своєчасних рішень для збереження стійкості та розвитку підприємств у цьому секторі.

Згідно з результатами цього дослідження штучні нейронні мережі є перспективними у сфері прогнозування банкрутства компаній завдяки їх високій успішності класифікації та простоті використання. Завдяки методу штучним нейронним мережам і застосуванням, отриманим у цьому дослідженні, дослідникам можна поради передбачити банкрутство компаній в інших секторах або країнах з високим ризиком банкрутства. За допомогою моделей машинного навчання також можна передбачити фінансові крахи підприємств у різних секторах. Наприклад, обробна промисловість, енергетика, автомобільна промисловість, транспорт і логістика, сфера послуг. Методи машинного навчання є перспективними завдяки своїм чудовим можливостям класифікації. Тому дослідникам у майбутніх дослідженнях рекомендується проводити дослідження,

у яких порівнюватиметься більше методів машинного навчання в різних секторах [5].

Проте використання лінійного дискримінантного аналізу теж має високу ефективність в порівнянні з методами нейронних мереж як було зазначено в роботі Giancarlo Marco і Franco Varetto. У своєму дослідженні вони порівнюють ефективність нейронних мереж і лінійного дискримінантного аналізу у бізнес-класифікації, зокрема в передбаченні фінансової небезпеки компаній. Дослідження доводить, що нейронні мережі мають великий потенціал і демонструють значну здатність розпізнавати здоров'я компаній. Результати, отримані нейронними мережами, у багатьох випадках близькі або перевершують результати дискримінантного аналізу. Однак, стаття також зазначає проблему неприйнятних типів поведінки в нейронних мережах, яка пов'язана з їх нелінійною природою та складністю архітектури. Ці типи поведінки можуть призвести до неточностей у прогнозуванні та ускладнити інтерпретацію результатів. Дослідники визнають, що лінійний дискримінантний аналіз, хоча вимагає більше часу на налаштування, все ж має переваги у швидкості оцінювання та інтерпретованості моделі. У цілому, стаття висвітлює переваги та обмеження нейронних мереж у порівнянні з лінійним дискримінантним аналізом в контексті бізнес-класифікації та передбачення фінансової небезпеки компаній [10].

1.3 Методи управління банківськими ризиками, застосування скорингової моделі для управління банківським ризиком

Скоринг кредитоспроможності є статистичним аналізом, який здійснюють кредитори та фінансові установи для визначення можливості позичання коштів особі або малому підприємству з власником-оператором. Скоринг кредитоспроможності використовується кредиторами для прийняття рішень

щодо надання або відмови у кредитуванні. Кредитний рейтинг може вплинути на можливість клієнта отримати фінансові продукти, такі як іпотека, автокредити, кредитні картки та приватні позики. Кредитні рейтинги визначають здатність особи позичати кошти для іпотек, автокредитів та особистих позик. Моделі скорингу кредитоспроможності можуть трохи відрізнятися в способі оцінки кредиту. Система скорингу кредитної історії Fair Isaac Corporation, відома як FICO-скоринг, є найбільш поширеною системою скорингу в фінансовій галузі, якою користуються понад 90% провідних кредиторів. Однак, ще одна популярна модель скорингу кредитоспроможності - VantageScore, яку створили три провідні кредитні агентства: TransUnion, Experian і Equifax. Кредитний рейтинг FICO є числом від 300 до 850, де 850 - найвищий можливий бал. Кредитні рейтинги для малих підприємств, такі як FICO Small Business Scoring Service (SBSS), варіюються від нуля до 300.

На кредитний рейтинг впливає на п'ять категорій:

- історія платежів (35%);
- суми заборгованості (30%);
- тривалість кредитної історії (15%);
- новий кредит (10%);
- склад кредиту (10%).

Кредитний рейтинг малого підприємства базується на інформації з його кредитного звіту, включаючи:

- інформація про компанію (включаючи кількість співробітників, оборот, власність та дочірні підприємства);
- історичні дані про діяльність підприємства;
- реєстраційні дані підприємства;
- огляд діяльності уряду;
- дані про операційну діяльність підприємства;
- класифікація та дані про галузь;
- публічні документи (застави, судові рішення та декларації з єдиного реєстру обороту).

Кредитори використовують скоринг кредитоспроможності для ризикоорієнтованого ціноутворення, при якому умови кредиту, включаючи відсоткову ставку, пропонуються позичальникам на основі ймовірності повернення коштів. Загалом, чим вище кредитний рейтинг, тим кращі умови пропонує фінансова установа. Хоча скоринг кредитоспроможності оцінює ризик позичальника, він не надає оцінки ймовірності неповернення позики. Він просто оцінює ризик позичальника від найвищого до найнижчого рівня. Отже, скоринг кредитоспроможності має обмеження, оскільки не може порівняти ризиковість окремих клієнтів між собою, оскільки оцінка винесена на основі окремих даних кожного клієнта і ці дані між собою не корелюються.

Ще одне обмеження скорингу кредитоспроможності полягає в тому, що він не може явно враховувати поточні економічні умови. Тобто, якщо у позичальника кредитний рейтинг 800, а економіка потрапляє в рецесію, то кредитний рейтинг не зміниться, якщо тільки поведінка клієнта або його фінансове становище не зміниться.

РОЗДІЛ 2

ОГЛЯД МЕТОДІВ АНАЛІЗУ І ПРОГНОЗУВАННЯ, ЩО ЗАСТОСОВУЮТЬСЯ ДЛЯ АНАЛІЗУ БАНКІВСЬКОГО РИЗИКУ

2.1 Оцінювання якості даних за допомогою коефіцієнту Херста, стандартного відхилення і коефіцієнту варіації

У фінансових даних, таких як дані по дохідності, часто доводиться оцінювати рівень ризику чи розмаху змін у часових рядах. Для цього часто використовують такі методи як коефіцієнт Херста, оскільки цей коефіцієнт може допомогти в оцінці майбутнього ризику, враховуючи властивості серії досягнутих значень. Стандартне відхилення є дуже розповсюдженою мірою варіації даних відносно середнього значення. Воно використовується для визначення рівня коливання або ризику у часовому ряді. Коефіцієнт варіації, в свою чергу, використовується для визначення відносної міри варіації у відсотках, для того щоб мати можливість порівнювати характеристики часових рядів між собою навіть якщо в них сильно відрізняються величини. Розрахунок цих статистичних мір допомагає оцінити і порівняти ризикові аспекти часових рядів, таких як дохідність.

Стандартне відхилення є важливим показником, що використовується для вимірювання розподілу даних і ризиків у фінансових аналізах і моделях. Воно вказує на розкид чи варіативність значень відносно середнього значення. Загальноприйнятою відповіддю на необхідність короткого виразу для дисперсії даних є зведення різниці кожного значення у квадрат із груповим середнім, отримуючи всі позитивні значення. Коли ці квадрати відхилень складають, а

потім ділять на кількість значень у групі, результатом є дисперсія. Дисперсія завжди є додатним числом, але в одиницях, що відрізняються від середнього. Щоб уникнути цієї незручності, можна використати квадратний корінь із дисперсії, який є стандартним відхиленням генеральної сукупності [11].

В контексті кредитного скорингу стандартне відхилення може мати декілька застосувань:

Вимірювання кредитного ризику: Стандартне відхилення може слугувати показником ризику, пов'язаного з кредитною ситуацією або фінансовим станом особи або підприємства. Велике стандартне відхилення може вказувати на більшу невпевненість або непередбачуваність у виплаті боргу, тоді як менше стандартне відхилення свідчить про менший ризик.

Порівняння з іншими кредитними суб'єктами: Стандартне відхилення дозволяє порівняти кредитний профіль однієї особи або підприємства з іншими. Чим менше стандартне відхилення, тим менші варіації і ризики пов'язані з цим кредитним профілем.

Визначення надійності моделей: Стандартне відхилення може використовуватись для оцінки надійності кредитних скорингових моделей. Якщо прогнознi значення моделі мають велике стандартне відхилення порівняно з фактичними даними, це може свідчити про меншу точність моделі і потребу у його вдосконаленні.

Стандартне відхилення є корисним інструментом для оцінки ризиків і варіацій в кредитному скорингу та фінансовому аналізі загалом. Воно допомагає краще розуміти розподіл даних та варіативність, що дозволяє приймати більш обґрунтовані рішення з урахуванням кредитного ризику. Стандартне відхилення набору спостережень часового ряду розраховується за формулою (див. формулу 2.1).

$$SD = \sqrt{\frac{\sum (r_i - r_{avg})^2}{n - 1}} \quad (2.1)$$

SD - це показник стандартного відхилення (від англ. Standard Deviation),

r_i – прибуток за одне конкретне спостереження

r_{avg} – середнє арифметичне всіх спостережень цього часового ряду

n – кількість спостережень в визначеному наборі даних

Середнє значення нормальної кривої (рис. 2.1) є серединою кривої (або піком дзвоника) з однаковою кількістю даних з обох сторін, тоді як стандартне відхилення кількісно визначає мінливість кривої (іншими словами, наскільки широка чи вузька крива).

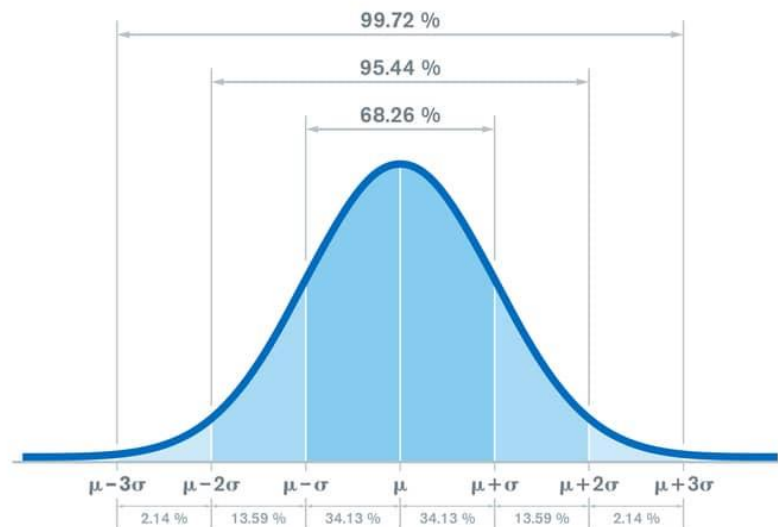


Рисунок 2.1 - Приклад нормального розподілу

Джерело: [12]

Припущення, яке можна зробити щодо даних, які слідує за нормальною кривою, полягає в тому, що площа під кривою є відносною того, на скільки стандартних відхилень спостереження віддалені від середнього. Область між плюс і мінус одним стандартним відхиленням від середнього містить 68% даних. Два стандартних відхилення містять 95% даних, а три стандартні відхилення містять 99,8% даних.

Для зручності порівняння результатів стандартного відхилення часто застосовують коефіцієнт варіації. Коефіцієнт варіації є статистичним показником, який використовується для вимірювання відносної дисперсії точок даних у ряді даних навколо їх середнього значення. Він обчислюється як відношення стандартного відхилення до середнього значення. Коефіцієнт варіації є корисним інструментом для порівняння ступеня варіації між різними

рядами даних, навіть якщо їх середні значення суттєво відрізняються. У фінансовій сфері коефіцієнт варіації допомагає інвесторам визначити рівень ризику порівняно з очікуваним прибутком від інвестицій. Чим нижче співвідношення стандартного відхилення до середнього значення прибутку, тим кращий компроміс між ризиком та прибутком. Коефіцієнт варіації в економіці використовується для вимірювання економічної нерівності. Крім того, його можна використовувати для перевірки точності певного процесу та досягнення ідеального балансу.

Основна перевага коефіцієнта варіації полягає в тому, що він дозволяє порівнювати набори даних, незалежно від одиниць вимірювання чи середніх значень. Це дає змогу зробити об'єктивні порівняння між різними наборами даних, оцінити ризик та винагороду. Коефіцієнт варіації є корисним інструментом для інвесторів, які шукають оптимальний баланс між ризиком і доходом від своїх інвестицій.

Коефіцієнт варіації є зручним інструментом, адже за його допомогою можна порівняти дисперсію відносно середнього доходу між різними клієнтами банку, що дасть зручний показник для аналізу ризиковості. При прогнозуванні можна звертати увагу на даний показник, що буде свідчити про загальну точність прогнозу.

Ще одним методом для оцінки якості часового ряду для подальшого прогнозування є коефіцієнт Херста. Коефіцієнт Херста, також відомий як коефіцієнт кореляції Херста або індекс кореляції розмахів, є статистичним показником, який використовується для вимірювання самоподібності або довготривалої залежності в послідовностях даних. Цей показник необхідний для визначення здатності часового ряду успадковувати тренд, що є важливим фактором при прогнозуванні часового ряду. Цей метод названий на честь голландського інженера Гарольда Херста, який першим використав його для вивчення коливань рівня води в Нілі. Коефіцієнт Херста вимірює ступінь кореляції між розмахами даних в різних масштабах часу. Він може приймати значення від 0 до 1, де 0 вказує на повну відсутність самоподібності, а 1 вказує

на досконалу самоподібність або довготривалу залежність даних. Значення біля 0,5 вважається ознакою слабкої самоподібності, тоді як значення більше 0,5 свідчить про наявність переважної самоподібності. Коефіцієнт Херста широко використовується в багатьох галузях, зокрема в фінансах, кліматології, аналізі часових рядів, телекомунікаціях та інших областях. У фінансовому аналізі він допомагає оцінити ризик та прогнозувати майбутні коливання цін активів. В кліматології використання коефіцієнта Херста допомагає вивчати довготривалі кліматичні зміни та передбачати майбутні погодні умови.

За допомогою коефіцієнта Херста ми можемо класифікувати часові ряди за типами і отримати деяке уявлення про їхню динаміку. Нижче наведені три типи класифікації часового ряду які ми можемо зробити на основі аналізу коефіцієнта Херста.

Часовий ряд Броунового руху (рис. 2.2): у часовому ряді Броунового руху немає кореляції між спостереженнями та майбутнім спостереженням. Ймовірність того, що наступне спостереження буде вищим або нижчим за поточне, однакова.

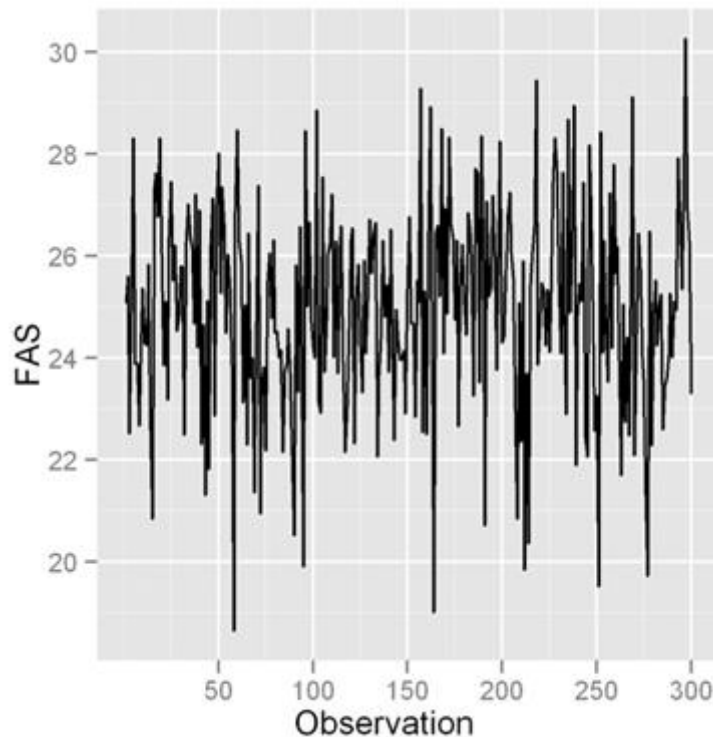


Рисунок 2.2 - Приклад Броунівського руху часового ряду

Джерело: [13]

На рисунку 2.2 наведено приклад часового ряду Броунівського руху. В даному прикладі коефіцієнта Херста для цих даних склав 0,53. Що близько до показника 0,5 і свідчить про класифікацію Броунівського руху.

Антипостійний часовий ряд: це тип часового ряду, який також називають середньорегресійним рядом. У такому ряді зміна значення зазвичай настає після зростання або навпаки, тобто можемо спостерігати, що значення мають тенденцію повертатися до середнього значення. Це означає, що майбутні значення мають тенденцію повертатися до довгострокового середнього значення. На рис. 2.3 продемонстровано приклад антипостійного часового ряду і його оцінений коефіцієнт Херста.

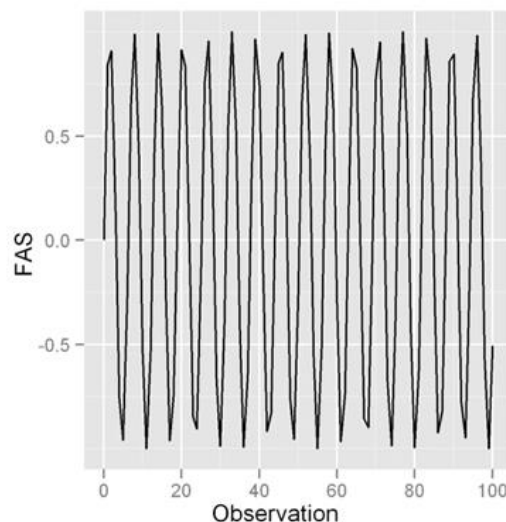


Рисунок 2.3 - Приклад антипостійного (середньорегресійного) часового ряду

Джерело: [13]

Оцінений коефіцієнт Херста для наведених вище даних становить 0,043. Значення коефіцієнта Херста в діапазоні від 0 до 0,5 свідчить про антипостійну поведінку, і чим ближче значення до 0, тим сильніше виявляється тенденція часового ряду повертатися до його довгострокового середнього значення.

Постійний часовий ряд — це тип часового ряду, у якому зростання значень зазвичай супроводжується подальшим зростанням у короткостроковій перспективі, а зменшення значень зазвичай супроводжується подальшим зменшенням у короткостроковій перспективі. На рис. 2.4 показано приклад постійного часового ряду.

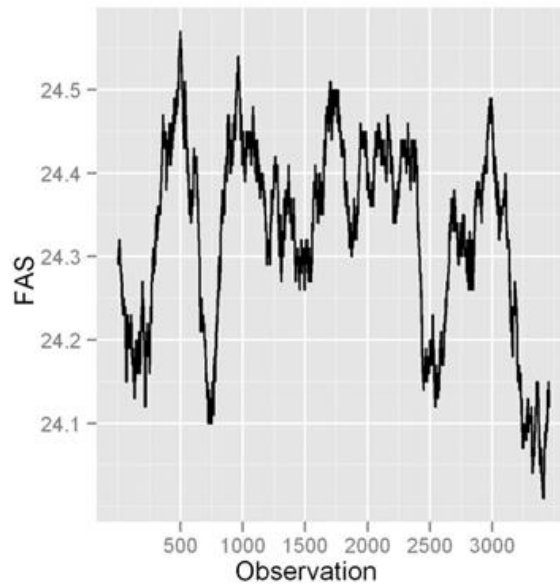


Рисунок 2.4 - Приклад постійного часового ряду

Джерело: [13]

На графіку показані дані де оцінений коефіцієнт Херста дорівнює 0,95, що свідчить про постійний часовий ряд. Значення коефіцієнта Херста в діапазоні від 0,5 до 1,0 свідчить про постійну поведінку; чим більше значення коефіцієнта, тим сильніше тренд.

2.2 Методологія прогнозування даних на основі ковзної середньої і простого експоненціального згладжування.

При обробці чи прогнозуванні часових рядів дуже важливим кроком є згладжування даних. Згладжування даних роблять для того щоб видалити шуми які здатні ускладнити аналіз прогнозування. Завдяки цьому при аналізі стає легше виявлення тренду, для визначення напрямку руху даних, чи сезонності, для того щоб зрозуміти повторювальні патерни і циклічність. Також методи згладжування необхідні для покращення прогнозування, оскільки виділення зайвих шумів дозволяє моделі прогнозу зосередитися на значущих змінах і тенденціях.

Ковзне середнє є одним з найбільш розповсюджених методів згладжування, на основі якого часто будують прогноз часових рядів.

Методи згладжування це класифікація методів попередньої обробки даних, основна мета даного методу є видалення небажаних шумів з даних. За рахунок чого можна побачити загальну тенденцію і як рухаються показники. При аналізі великої кількості даних, як правило надають перевагу згладженим даним, оскільки вони чіткіше ідентифікують зміни тренду, порівняно з не згладженими, де завелика кількість шумів. Ковзне середнє (англ. Moving Average – MA) – інструмент згладжування часових рядів, застосовуваний, головним чином, для відображення змін біржових котирувань акцій, цін на сировину і біржові товари. Ковзне середнє – один із найстаріших і найбільш поширених інструментів технічного аналізу. Він показує середнє значення ціни за певний період часу [14].

Розділяють два типи ковзних середніх:

1. Проста ковзна середня
2. Експоненційна ковзна середня

Ігнорування шумів і загальне згладжування даних відбувається за рахунок того що до розрахунку береться середнє значення за минулі періоди. При застосуванні простої ковзної середньої показник отримується шляхом додавання значень часового ряду за певний період і діленням суми на кількість періодів взятих в розрахунок. Тобто розраховується середнє значення за певну кількість минулих спостережень. Цей процес можна описати формулою (див. формулу 2.2):

$$SMA = \frac{C_1 + C_2 + \dots + C_n}{n} \quad (2.2)$$

n – кількість спостережень взятих в розрахунок, C_n – значення часового ряду в момент спостереження [15].

Експоненційна ковзна середня відрізняється від простої ковзної середньої тим що надає більше переваги останнім даним. Зважування, що застосовується до останніх цін, залежить від загальної кількості періодів. Ця різниця між простим та експоненційним середнім робить останню ефективнішою у

реагуванні на нову інформацію. У експоненційному середньому розрахунок залежить від розрахунків за всі попередні дні. Щоб обчислити експоненціальне ковзне середнє, вам потрібні дані принаймні за 10 спостережень. Експоненційна ковзна середня розраховується за формулою (див. формулу 2.3):

$$EMA(n) = k * P(n) + (1 - k) * EMA(n - 1) \quad (2.3)$$

$P(n)$ - ціна закриття поточного торгового періоду,

$EMA(n-1)$ - значення експоненціального змінного середнього, розрахованого для попереднього торгового періоду,

k - регулюючий коефіцієнт.

Початкове значення $EMA(1)$ приймається рівним ціні найпершого торгового періоду, який береться до розгляду, тобто $P(1)$. Чим більше значення регулюючого коефіцієнта k , тим краще крива експоненціального змінного середнього апроксимує графік, бо більше значення надається поточними цінами [16].

Фактор відставання є характеристикою затримки, яку мають рухомі середні через використання минулих даних. Чим довший період рухомої середньої, тим більша затримка відбувається. Крім того, тип рухомої середньої впливає на рівень затримки: експоненціально зважена рухома середня з більш актуальними даними має меншу затримку порівняно з простою рухомою середньою, яка надає однакову вагу даним з віддаленішого минулого. 10-денна рухома середня досить точно відображає ціни та змінюється незабаром після зміни цін. Короткострокові рухомі середні схожі на швидкі човни — вони здатні швидко реагувати на зміни. Натомість 100-денна рухома середня містить багато попередніх даних, що уповільнює її зміну. Для зміни курсу 100-денної рухомої середньої потрібно значно більше і триваліше руху цін порівняно з 10-денною рухомою середньою. Крім того, метод простого ковзного середнього зазвичай обчислюється швидший, ніж більш складні методи, такі як експоненціальне ковзне середнє. Він також вимагає менше параметрів і може використовуватися на коротших наборах даних. Також ефективність методу SMA доведена в

багатьох додатках, таких як аналіз фондового ринку, а також сезонне прогнозування.

Загалом метод ковзної середньої є ефективним інструментом для короткострокового прогнозування завдяки своїй гнучкості та простоті використання. Його здатність враховувати всі минулі значення під час створення прогнозів забезпечує точність, а його здатність визначати сезонні закономірності означає, що його можна також ефективно використовувати для довгострокового прогнозування. Крім того, його обчислювальна швидкість і мінімальні параметри роблять його популярним вибором для багатьох програм.

2.3 Методологія прогнозування методом Хольта, Хольта – Вінтерса і демпфованого тренду

В даному дослідженні дані знаходяться в форматі часового ряду, яким притаманні тренд і сезонність. При аналізі таких типів часових рядів можна використовувати метод Хольта, Хольта – Вінтерса, адитивний і мультиплікативні методи Хольта – Вінтерса, а також метод демпфованого тренду. Завдяки застосуванню цих методів є можливість урахувати сезонність, тренд і загальну тенденцію, а також визначити вклад в прогноз кожного з цих показників.

Метод прогнозування Хольта – один з методів експоненційного згладжування, який застосовується для прогнозування часових рядів. Розроблений американським професором Чарльзом Хольтом в 1957 році. Найчастіше даний метод використовується для прогнозування часових рядів без сезонності.

Популярності даний метод набув за рахунок своєї простоти і ефективності в прогнозуванні часових рядів з трендом і з відсутністю повторюваних циклічних варіацій.

Основна методологія методу Хольта полягає в тому, що аналіз, переважно, спирається на дві складові для прогнозування майбутніх часових рядів, це рівень згладжування і коефіцієнт тренду.

Ця модель враховує наявність тренду. Якщо присутня тенденція до зростання або спадання часового ряду, то разом з оцінкою поточного рівня ряду (як у простому експоненційному згладжуванні) потрібно виділити тренд. Для управління рівнем та нахилом у моделі Хольта вводиться два коефіцієнти згладжування – коефіцієнти згладжування ряду α та тренду β . Отже, щоб здійснити прогноз економічних показників на основі адаптивної моделі Хольта потрібно дотримуватися відповідного алгоритму: спочатку необхідно розрахувати експоненційно-згладжений ряд, далі визначається значення тренду, і насамкінець – робиться прогноз. Модель Хольта використовується лише для рядів, коли є дані за не повний цикл і сезонна компонента ще не виявлена. Цю модель розвинув вчений Вінтерс, який додав у модель Хольта сезонність [17].

Рівень згладжування це критерій який визначає наскільки швидко будуть впливати нові спостереження на прогнозований часовий ряд. Значення цього критерію між 0 і 1. Чим нижче даний критерій, тим більше буде згладжування. Тобто 0 означає повністю згладжений ряд, в той час, як 1 буде означати повне врахування останнього спостереження. Тобто чим більше значення рівня згладжування, тим більше ваги приділяється останньому спостереженню. Як наслідок в тренді буде відображатись більше змін. Зазвичай, якщо часовий ряд має меншу кількість шуму або змін, то більш вагомою вважаються найновіші спостереження, тому коефіцієнт визначають більше ніж 0.5. Якщо ж ряд має більше шуму або змін, то менша вага надається найновішим спостереженням, і тоді коефіцієнт менше ніж 0.5.

Коефіцієнт тренду це параметр, що визначає, наскільки швидко тренд може змінюватись. Він також приймає значення між 0 і 1, де 0 означає, що тренд не змінюється, а 1 - що тренд змінюється з максимальною швидкістю. Рівняння розрахунку рівня згладжування для моделі Хольта:

$$F_{t+1} = \alpha X_t + (1 - \alpha)(F_t + T_t) \quad (2.4)$$

Рівняння розрахунку коефіцієнта тренду для моделі Хольта:

$$T_{t+1} = \beta(F_{t+1} - F_t) + (1 - \beta)T_t \quad (2.5)$$

Формула прогнозування на m періодів в моделі Хольта:

$$H_{t+m} = F_{t+1} + mT_{t+1} \quad (2.6)$$

F_{t+1} - згладжене значення за період $t+1$

α - коефіцієнт згладжування рівня

X_t - наявне значення для періоду t

F_t - прогнозоване згладжене значення для часового ряду

T_{t+1} - оцінка тренду

β - коефіцієнт згладжування для оцінки тренду

bt - оцінка нахилу ряду для періоду t

m - кількість прогнозованих періодів

H_{t+m} - прогнозоване значення за методом Хольта для періоду m [18].

Враховуючи усі вище перераховані особливості і спираючись на метод застосування, стає зрозуміло чому метод Хольта використовується в багатьох галузях, де важливо прогнозувати майбутні значення часових рядів з лінійною тенденцією. Наприклад, в економічному аналізі його часто застосовують для прогнозування продажів товарів, попиту на послуги, руху цін на ринку або економічного зростання. Метод Хольта дозволяє отримати прогнозовані значення на основі історичних даних, з урахуванням лінійної тенденції та згладжування залежно від вибраних параметрів. Це дозволяє аналізувати та прогнозувати майбутні тренди, що є корисним інструментом для прийняття рішень у багатьох галузях діяльності.

У 2007 році вийшло дослідження американських спеціалістів Незіх Алтай, Френк Рудісілл і Льюїс А. Літераль в якому вони порівнювали прогнозування непостійного попиту за допомогою моделей Хольта і модифікованого методу Кростона [19].

Дослідження зосереджується на проблемі прогнозування попиту на запчастини, особливо у випадку періодичного попиту з наявності тенденції. У роботі порівнюються два методи прогнозування: модифікований метод Кростона та модифікований метод Хольта — Райта з урахуванням конкурентних пріоритетів фірми. Метод Хольта — Райта це модифікація методу Хольта, який використовує подвійне експоненційне згладжування для обчислення та оновлення лінії тренду. Цей метод вважається ефективною альтернативою при прогнозуванні попиту на запчастини з наявною тенденцією. Він враховує зміну лінійної тенденції та забезпечує більш точні прогнози в умовах взятих для дослідження. Результати дослідження показали, що фірми, які мають пріоритетом мінімізацію рівня запасів, краще використовувати модифікований метод Кростона. Водночас якщо високий рівень обслуговування клієнтів є пріоритетом, то модифікований метод Хольта — Райта є кращим варіантом. Також методи використані в цьому дослідженні можуть бути застосовані в аналізі банківського кредитного ризику.

Наприклад метод Кростона для прогнозування кредитного портфеля банку допомагаючи визначити майбутні зміни в ризиковості активів та визначити необхідні заходи управління ризиком.

Метод Хольта — Райта може бути застосований для аналізу кредитного ризику, дозволяючи врахувати сезонні коливання ризикових показників, а також прогнозувати майбутні зміни ризиковості на основі тренду.

Таким чином, модель прогнозування Хольта використовується в цьому дослідженні для прогнозування періодичного попиту на запчастини з урахуванням наявної тенденції. Вона дозволяє точніше прогнозувати майбутні тренди та забезпечує кращі результати в порівнянні з іншими методами прогнозування.

Проте в 1960 році модель Хольта була покращена Джоном Вінтерсом. Джон Вінтерс це статистик та дослідник, який вніс значний внесок до розвитку методології прогнозування, особливо в модель Хольта. Вінтерс спеціалізувався на прогнозуванні часових рядів та розробив покращену версію моделі Хольта,

відому як метод Хольта – Вінтерса, який став одним з найпопулярніших підходів до прогнозування сезонних часових рядів. Метод Хольта – Вінтерса включає в себе не лише експоненціальне згладжування тренду, як у базовій моделі Хольта, але також додає сезонну компоненту. Він враховує повторювані сезонні зміни в часовому ряді, що дозволяє забезпечити більш точні та надійні прогнози сезонних даних.

Метод Хольта – Вінтерса може бути застосований до ряду сезонних даних, які проявляють явну циклічність та мають тренд. Він дозволяє моделювати та прогнозувати як лінійні, так і не лінійні тренди, а також надає можливість враховувати кілька типів сезонності, таких як денна, тижнева, або щорічна. Завдяки внесеним Вінтерсом покращенням, метод Хольта – Вінтерса став популярним і широко використовується в різних галузях, де важливе прогнозування сезонних та трендових залежностей. Він забезпечує більш точні та надійні прогнози, що допомагають підприємствам приймати кращі рішення щодо планування виробництва, запасів, маркетингових стратегій та інших аспектів управління.

Сезонність – це характеристика часового ряду, в якому дані зазнають регулярних і передбачуваних змін, що повторюються кожного календарного року. Будь-які передбачувані коливання або закономірності, що повторюються або повторюються протягом одного року, називаються сезонними. Сезонні ефекти відрізняються від циклічних, оскільки сезонні цикли спостерігаються протягом одного календарного року, тоді як циклічні ефекти, такі як збільшення обсягів продажів через низький рівень безробіття, можуть охоплювати періоди, коротші або довші, ніж один календарний рік [20]. Сезонність в часовому ряду відноситься до повторюваних, регулярних та систематичних коливань, які спостерігаються в певний період часу, зазвичай від року до року. Ці коливання можуть бути спричинені як природно-кліматичними факторами, так і соціально-економічними впливами. Сезонність відображає нерівномірність використання ресурсів, робочої сили та попиту на ринку певних товарів або послуг, чи інших коливань економічних показників протягом року.

Аналіз сезонної компоненти може мати кілька цілей. В першому випадку, він допомагає встановити амплітуду сезонних коливань, визначити максимальні та мінімальні значення ряду, що може бути корисним для планування подальших дій компанії. В другому випадку, вилучення сезонної компоненти з часового ряду дозволяє досліджувати інші компоненти, такі як тренд. Це допомагає краще розуміти основні зміни та тенденції в ряду, окреслювати довгострокові зміни показників і ігнорувати коригування даних від сезонності. У третьому випадку, коригування поточних значень ряду на основі сезонної компоненти може бути важливим для врахування сезонних варіацій. Наприклад, це може бути використано для прогнозування безробіття в зимовий період, якщо враховувати вплив сезонності.

Загалом, аналіз сезонності часового ряду допомагає виявити повторювані шаблони та регулярні зміни, що може бути корисним при плануванні та прийнятті управлінських рішень [21].

Метод прогнозування Хольта – Вінтерса поділяють на два види, адитивний і мультиплікативний.

Різниця між адитивним і мультиплікативним методами Хольта – Вінтерса полягає у тому, як вони моделюють сезонну компоненту в часовому ряду. Адитивний метод Хольта – Вінтерса є одним з методів прогнозування часових рядів, особливістю якого є моделювання сезонної компоненти у вигляді адитивної моделі, тобто модель в яку фактори входять у вигляді алгебраїчної суми. Цей метод використовується для аналізу та прогнозування часових рядів, які мають виражену сезонність. Основна особливість адитивного методу полягає в тому, що він розглядає сезонну компоненту як сталий абсолютний приріст або зменшення в часовому ряді. Це означає, що амплітуда сезонності залишається постійною незалежно від рівня ряду.

При використанні адитивного методу Хольта – Вінтерса, прогнозоване значення для майбутнього періоду складається з трендової компоненти, яка відображає загальну тенденцію часового ряду, та сезонної компоненти, яка відображає періодичні коливання на певному інтервалі.

Застосування адитивного методу дозволяє аналізувати та передбачати вплив сезонності на зміни в часовому ряді. Наприклад, він може бути використаний для прийняття рішень щодо кількості готелів, які слід закрити після завершення сезону, з урахуванням очікуваної сезонної зміни.

Адитивна модель прогнозування Хольта – Вінтерса:

$$\begin{aligned}\hat{y}_{t+h|t} &= l_t + hb_t + s_{t+h-m(k+1)} \\ l_t &= \alpha(y_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}) \\ b_t &= \beta^*(l_t - l_{t-1}) + (1 - \beta^*)b_{t-1} \\ s_t &= \gamma(y_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}\end{aligned}\tag{2.7}$$

$\hat{y}_{t+h|t}$ – прогнозоване значення моделі

l_t – рівняння розрахунку рівня згладжування

b_t - рівняння розрахунку коефіцієнта тренду

s_t – рівняння розрахунку сезонної зміни

k є цілою частиною $(h-1)/m$, що гарантує, що оцінки сезонних індексів, які використовуються для прогнозування, походять з останнього часового ряду вибірки.

Рівняння рівня згладжування показує зважене середнє між сезонно скоригованими спостереженнями $(y_t - s_{t-m})$ і несезонний прогноз $(l_{t-1} + b_{t-1})$ на часі t . Рівняння тренду ідентично лінійному методу Хольта. Сезонне рівняння показує зважене середнє між поточним сезонним індексом, $(y_t - l_{t-1} - b_{t-1})$, і сезонний індекс того самого сезону минулого року (тобто m - періоди часу) [22].

Таким чином, адитивний метод Хольта – Вінтерса дозволяє враховувати сезонну компоненту як сталий абсолютний приріст або зменшення в часовому ряді, що дозволяє зробити більш точні прогнози та врахувати сезонні коливання при аналізі часових рядів.

На відміну від адитивного методу мультиплікативний метод Хольта – Вінтерса використовує мультиплікативну модель для сезонної компоненти. За цим методом, сезонна компонента множиться на прогнозований тренд, що означає, що вона розглядається як відсоткова зміна від рівня ряду. Це означає, що

амплітуда сезонності залежить від рівня ряду, і її значення змінюється пропорційно з рівнем ряду. Використання мультиплікативного методу дозволяє враховувати сезонну компоненту як відносну зміну, що пропорційно залежить від рівня ряду. Це дозволяє аналізувати та передбачати вплив сезонності на зміни в часовому ряді, зокрема у випадках, коли сезонні коливання мають виразні залежності від загального рівня ряду. Формула мультиплікативної моделі Хольта – Вінтерса:

$$\begin{aligned} y_t &= (l_{t-1} + b_{t-1})c_{t-m} + (l_{t-1} + b_{t-1})c_{t-m}\varepsilon_t \\ l_t &= l_{t-1} + b_{t-1} + \alpha_1(l_{t-1} + b_{t-1})\varepsilon_t \\ b_t &= b_{t-1} + \alpha_2(l_{t-1} + b_{t-1})\varepsilon_t \\ c_t &= c_{t-m} + \alpha_3c_{t-m}\varepsilon_t \end{aligned} \quad (2.8)$$

l_t - рівень згладжування для часового ряду,

b_t – коефіцієнт тренду

c_t – сезонна зміна

Рівняння спостереження показує зв'язок між часовим рядом y_t і базовими змінними стану в момент часу t . Рівняння стану показують перехід змінних стану, який можна інтерпретувати наступним чином у період часу t :

Таким чином мультиплікативний метод Хольта – Вінтерса дозволяє враховувати сезонну компоненту як відсоткову зміну, що залежить від рівня ряду, і дозволяє зробити більш точні прогнози та врахувати сезонні коливання при аналізі часових рядів.

Отже, різниця двох методів полягає у способі моделювання сезонної компоненти: адитивний метод додає сезонну компоненту до тренду, а мультиплікативний метод множить сезонну компоненту на тренд. Вибір між цими двома методами залежить від того, як сезонність впливає на зміни у часовому ряді та від характеру залежності амплітуди сезонної компоненти від рівня ряду.

Проте модель Хольта – Вінтерса можна покращити за рахунок додавання демпфованого тренду. Демпфований тренд є поняттям, яке використовується для

опису згладженого тренду в часовому ряді. Тренд представляє собою загальну довгострокову тенденцію зміни значень ряду в часі. В багатьох випадках тренд може бути підданий коливанням, які викликані сезонністю, циклічністю або іншими факторами, що плутає визначення основної тенденції тренду і погіршує точність подальшого прогнозування. Демпфований тренд є результатом застосування методів згладжування, які дозволяють виділити довгострокову зміну в часовому ряді, відокремивши її від короткострокових коливань. Згладжування тренду має на меті зменшити вплив випадкових коливань і зосередитися на основній тенденції ряду. Демпфований тренд відрізняється від оригінального тренду тим, що він менш піддається короткостроковим коливанням та має меншу амплітуду змін. Застосування методів згладжування може включати рухомі середні, експоненційне згладжування або інші підходи. Демпфований тренд є корисним інструментом для аналізу часових рядів, оскільки він дозволяє виділити основну тенденцію змін в довгостроковій перспективі, прибравши шуми та короткострокові зміни. Це може бути корисно для виявлення довгострокових змін в ряді, встановлення тренду та передбачення майбутніх значень.

В моделі Хольта – Вінтерса, демпфований тренд використовується для опису довгострокової зміни в часовому ряді, при якому тренд має тенденцію зменшувати свою амплітуду з часом. Це означає, що вплив тренду на значення часового ряду послаблюється з часом. Основна особливість демпфованого тренду в моделі Хольта – Вінтерса полягає у використанні двох параметрів: параметр тренду (b) і параметр демпфування (ϕ). Параметр тренду відповідає за швидкість зміни тренду, тобто за те, наскільки швидко тренд рухається вгору або вниз. Параметр демпфування визначає, наскільки швидко амплітуда тренду зменшується з часом. У моделі Хольта – Вінтерса з демпфованим трендом, оновлення значень тренду залежить від попереднього значення тренду та зваженої різниці між поточним спостереженням і передбаченим значенням на попередньому кроці часу.

Демпфування можливе як за допомогою адитивного, так і мультиплікативного методів Хольта – Вінтерса. Метод, який часто забезпечує точні та надійні прогнози для сезонних даних, – це метод Хольта – Вінтерса із затухаючим трендом і мультиплікативною сезонністю [24].

$$\begin{aligned}\hat{p}_{t+h|t} &= [l_t + (\phi + \phi^2 + \dots + \phi^h)b_t]c_{t+h-m(k+1)} \\ l_t &= \alpha \left(\frac{p_t}{c_{t-m}} \right) + (1 - \alpha)(l_{t-1} + \phi b_{t-1}) \\ b_t &= \beta^*(l_t - l_{t-1}) + (1 - \alpha)(l_{t-1} + \phi b_{t-1}) \\ c_t &= \gamma \frac{p_t}{l_{t-1} + \phi b_{t-1}} + (1 - \gamma)c_{t-m}\end{aligned}\tag{2.9}$$

Можемо побачити що до розрахунку додався параметр демпфування (ϕ), і за рахунок якого показник тренду буде слабкіше відповідати на викиди чи шуми, і як наслідок, давати чіткішу картину для прогнозування. Таким чином, параметр демпфування ϕ контролює, наскільки швидко амплітуда тренду зменшується з кожним кроком часу. Якщо значення ϕ близьке до 1, то тренд буде змінюватись повільно і матиме більшу амплітуду, якщо значення ϕ близьке до 0, то тренд буде згасати швидше і матиме зменшену амплітуду. Цей метод застосовується для прогнозування довгострокових змін в часових рядах, особливо тих, які відображають тренд, що має тенденцію зменшувати амплітуду з часом. Цей метод знаходить широке застосування в економічному прогнозуванні, фінансовому аналізі, управлінні запасами, маркетингових дослідженнях та інших галузях, де необхідно прогнозувати майбутні тенденції на основі історичних даних. Демпфований тренд в моделі Хольта – Вінтерса дозволяє враховувати зміну амплітуди тренду з часом, що є корисним при прогнозуванні рядів, де тренд піддається затуханню або зменшенню впливу з часом. Наприклад, в економічних рядів, таких як продажі товарів або фінансові показники, можуть спостерігатись довгострокові тенденції, які змінюються протягом часу, але амплітуда цих змін зменшується. Демпфований тренд дозволяє уникнути переоцінки майбутніх значень тренду, зменшуючи його

амплітуду з кожним кроком часу. Це допомагає забезпечити більш точні прогнози на більшу кількість періодів періоди.

На мою думку, при дослідженні банківського ризику наявна інформація характеризується сильною хаотичністю, сезонністю, тому доцільно демпфувати часові ряди, моделювати з урахуванням сезонності.

РОЗДІЛ 3

ПРОГНОЗУВАННЯ ДОХОДІВ КЛІЄНТІВ БАНКУ ДЛЯ ФОРМУВАННЯ ОЦІНКИ РИЗИКОВОСТІ З ВИКОРИСТАННЯМ СКОРИНГОВОЇ МОДЕЛІ

3.1 Опис початкових даних і технічної реалізації

Для фінансових компаній, які надають послугу кредитування, важливо бути впевненими в тому, що клієнт поверне запозичену суму з відсотками. Для цього клієнт, перш за все, повинен мати достатню кількість доходу, що б мати можливість сплачувати згідно з умовами кредиту в заданий строк. Головна мета цієї роботи, створити прогноз доходів клієнта на наступні три місяці, що б бути впевненим, що клієнт має можливість розрахуватись за надані послуги, або зменшити суму коштів які доступні клієнту для кредитування.

Дані які використовуються в роботі це реальні дані міжнародної фінансової компанії з надання послуги кредитування для виробників онлайн контенту з платформи YouTube. Датасет складається з:

- ID – код конкретного клієнта створений для полегшення процесу аналізу даних
- Country – країна в якій зареєстрований клієнт
- date – щомісячна дата спостережень для кожного клієнта
- value – дохід клієнта від платформи YouTube за конкретний місяць

Датасет включає в себе 298 клієнтів і щомісячні доходи за три роки, з 1 лютого 2020 року по 1 січня 2023 року. Деякі клієнти почали співпрацю пізніше, тому щомісячних спостережень в них менше.

Дослідження проводилось в середовищі RStudio мовою R, і за допомогою програми MS Excel.

“lubridate” – бібліотека для спрощення і покращення обробки даних пов’язаних з датою та часом [25];

“tidyr” – бібліотека функцій для форматування датасету в формат tidy(формат де кожен стовпець відповідає змінним, кожен рядок це наявні спостереження, кожна клітинка дорівнює відповідному значенню) [26];

“readr” – бібліотека, що надає швидкий спосіб читати табличні дані розділених табуляцією формату CVS чи TSV [27];

“forcats” – бібліотека для обробки категоріальних змінних, змінних, які мають фіксований набір можливих значень. forcats також корисний для зміни порядку векторів символів для покращення відображення [28];

“dplyr” – одна з найпоширеніших бібліотек з функціями для маніпулювання даними в датасеті (фільтрація, додавання нових змін та інші) [29];

“ggplot2” – бібліотека для побудови графіків і візуалізації даних [30];

“prasma” – бібліотек, що надає велику кількість функцій із числового аналізу та лінійної алгебри, чисельної оптимізації, диференціальних рівнянь, часових рядів, а також деякі добре відомі спеціальні математичні функції [31];

“zoo” – пакет, який надає клас S3 із методами для індексованих повністю впорядкованих спостережень, таких як дискретні нерегулярні часові ряди. Його ключовими цілями проєктування є незалежність від конкретного класу індексу/часу/дати та узгодженість з базовим R і класом "ts" для регулярних часових рядів [32];

“fabletools” – бібліотека, що надає інструменти для створення пакетів моделювання з акцентом на прогнозування часових рядів;

“feasts” – бібліотека надає широкий вибір інструментів для аналізу часових даних у форматі tsibble. Назва пакета є аббревіатурою його основної функції: вилучення функцій і статистика для часових рядів [33];

“forecast” – бібліотека, що надає методи та інструменти для відображення та аналізу прогнозів однофакторних часових рядів, включаючи експоненціальне згладжування за допомогою моделей простору станів і автоматичного моделювання ARIMA [34];

“readxl” – бібліотека, що дозволяє легко імпортувати дані з Excel формату “.xls” та “.xlsx” в R [35];

“gridExtra” – бібліотека, що надає функції для роботи з сітковою графікою, для впорядкування кількох графіків і графічного відображення таблиць [36];

“fable” – бібліотека, що надає колекцію широко використовуваних однофакторних і багатфакторних моделей прогнозування часових рядів, включаючи експоненціальне згладжування за допомогою моделей простору станів [37];

“patchwork” – бібліотека, що полегшує поєднання двох графіків прогнозування часових рядів в один використовуючи API [38];

“openxlsx” – бібліотека, що надає можливість створювати нові файли Excel і вивантажувати таким чином дані з R [39].

3.2 Підготовка і оцінювання якості часових рядів

Спершу необхідно було завантажити дані, які зберігались в Excel, в R. Для цього було використано `read_csv()` з бібліотеки `readr`. Отримано датасет який складається з 9877 спостережень і 4 змінних: ID, Country, date і value (рис. 3.1).

	ID	Country	date	value
	<chr>	<chr>	<date>	<dbl>
1	P-113572	Ukraine	2020-02-01	253.
2	P-113572	Ukraine	2020-03-01	539.
3	P-113572	Ukraine	2020-04-01	377.
4	P-113572	Ukraine	2020-05-01	503.
5	P-113572	Ukraine	2020-06-01	1178.
6	P-113572	Ukraine	2020-07-01	5523.

Рисунок 3.1 – Перші рядки датасету

Джерело: розроблено автором мовою R

Можемо побачити, що стовпчики ID і Country зберігаються у форматі chr – character, стовпчик date в форматі date і value (дохід клієнта за місяць) в форматі double, тобто числа з дробовою частиною.

Для того, що б покращити точність подальшого прогнозування і мати можливість розрахунку сезонності для усього датасету, було видалено ті ID по яким менше ніж 12 спостережень. За рахунок цього залишилися тільки ті клієнти по яким є інформація більше ніж 1 рік. Оскільки першочергові дані щомісячні і в них відсутні пропуски (тобто як тільки клієнт почав співпрацю надалі по цьому клієнту є дані за кожен місяць) можна залишити тих яких більше ніж 12 спостережень, а не конкретизувати 12 спостережень підряд. Тепер датасет скоротився до 9626 спостережень.

Тепер було згруповано датасет за ID за допомогою функції `group_by()` для того, що б подальші розрахунки виконувались окремо для кожного клієнта, а не по всьому датасету. За рахунок цього було пораховано середній дохід кожного клієнта за допомогою функції `mean()`, і стандартне відхилення за допомогою функції `sd()`. Результати розрахунку вивів в окремий датасет разом з ID (рис. 3.2). Отримав датасет з 298 спостережень і 3 змінних. Тобто можна зробити висновок, що наявні 298 клієнтів в яких спостережень більше ніж 1 рік.

	ID	mean_value	stand_dev
	<chr>	<dbl>	<dbl>
1	P-1006	1207.	480.
2	P-101836	643.	543.
3	P-102620	362.	115.
4	P-10484	2064.	1855.
5	P-106433	92279.	42067.
6	P-10709	142.	130.
7	P-110677	1353.	899.
8	P-110819	131.	70.5
9	P-111091	174.	196.
10	P-11121	159.	53.6

Рисунок 3.2 – середнє значення доходу і стандартне відхилення кожного клієнта

Джерело: Розроблено автором мовою R

Бачимо, що для багатьох клієнтів стандартне відхилення є досить великим, що свідчить про сильний розкид доходів. Це пов'язано з тим, що дохід від платформи YouTube повністю залежить від кількості переглядів за місяць, і рівню монетизованості клієнта (наскільки контент задовольняє умови платформи).

Тепер додаємо до цього датасету ще одну змінну де буде записано коефіцієнт варіації. Створено додатковий стовпчик де для кожного ID розраховується відношення стандартного відхилення до середнього доходу (рис. 3.3).

	ID	mean_value	stand_dev	CV
	<chr>	<dbl>	<dbl>	<dbl>
1	P-1006	1207.	480.	0.398
2	P-101836	643.	543.	0.845
3	P-102620	362.	115.	0.317
4	P-10484	2064.	1855.	0.899
5	P-106433	92279.	42067.	0.456
6	P-10709	142.	130.	0.914
7	P-110677	1353.	899.	0.664
8	P-110819	131.	70.5	0.539
9	P-111091	174.	196.	1.13
10	P-11121	159.	53.6	0.336

Рисунок 3.3 – Датасет з середнім доходом, стандартним відхиленням і коефіцієнтом варіації

Джерело: Розроблено автором мовою R

Коефіцієнт варіації для більшості клієнтів досить високий, що підтверджує, що дані не дуже постійні і наявний сильний розкид доходу для кожного окремого клієнта (за виключенням для деяких клієнтів де коефіцієнт варіації нижче).

Для полегшення сприйняття даних додано діаграму, де буде зображено середній дохід кожного клієнта (рис. 3.4).

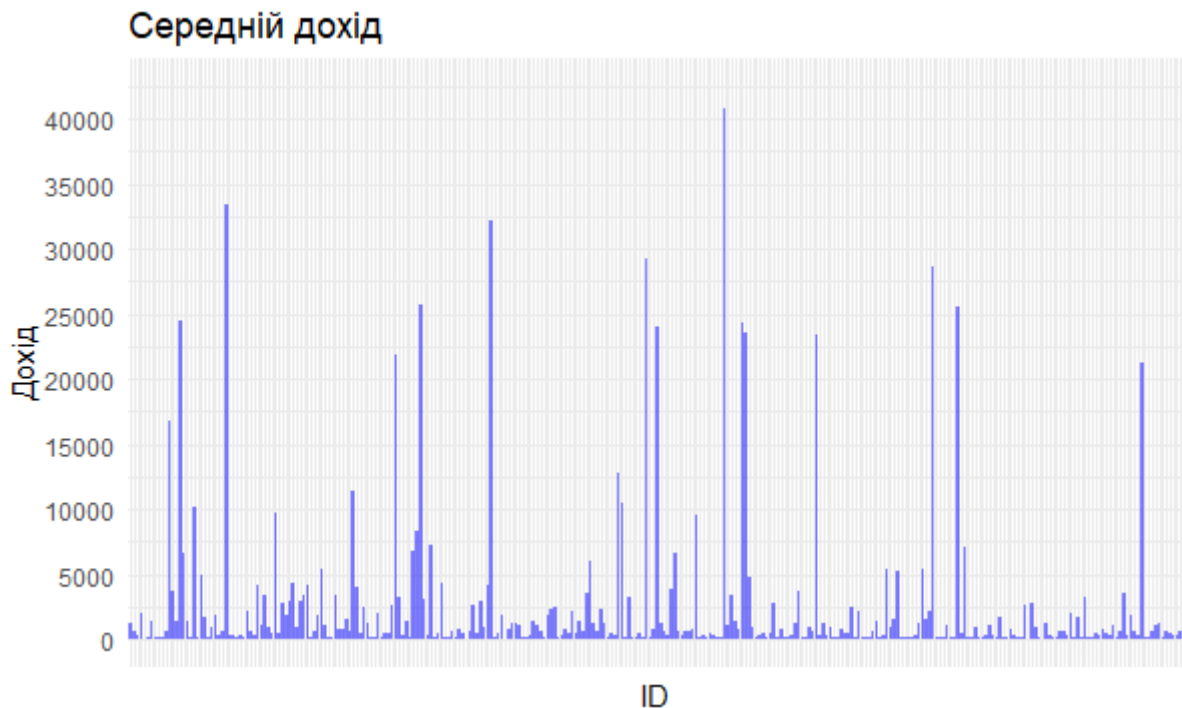


Рисунок 3.4 – Графік середнього доходу для всіх клієнтів

Джерело: Розроблено автором мовою R

На графіку по осі x розташовані ID клієнтів, по осі y кількість доходу. Можемо побачити, що середній дохід сильно відрізняється для кожного клієнта. Переважна більшість мають цей показник до 5000, проте є клієнти з середнім щомісячним доходом понад 20 000.

Для того, що б розрахувати коефіцієнт Херста необхідно перевести початкові дані в формат `tsbl`. Це формат в якому спостереження розташовані відносно часових міток. Тобто замість простого списку значень кожен рядок має часову мітку яка відповідає цьому значенню. Це необхідно для того, що б спростити роботу з часовими рядами і полегшення побудову прогнозування.

Для цього застосовуємо функцію `as_tsibble()` з бібліотеки “`tsibble`”. В якості змінної яка буде визначати індекс часу зазначив ID. Також позначив запис дати в щомісячні (рис. 3.5).

```
# A tsibble: 9,626 x 4 [1M]
# Key:      ID [298]
   ID      Country    date value
  <chr>    <chr>      <mt>  <dbl>
1 P-1006   Germany    2020 лют  1738.
2 P-1006   Germany    2020 бер  1737.
3 P-1006   Germany    2020 кві  1353.
4 P-1006   Germany    2020 тра   914.
5 P-1006   Germany    2020 чер   837.
6 P-1006   Germany    2020 лип  1175.
7 P-1006   Germany    2020 сер  1652.
8 P-1006   Germany    2020 вер  1421.
9 P-1006   Germany    2020 жов  1702.
10 P-1006  Germany    2020 лис  1717.
```

Рисунок 3.5 – Дані в форматі tsibble

Джерело: Розроблено автором мовою R

Оскільки тепер дані зберігаються в форматі tsibble і формати записів дати має формат mth я можу порахувати коефіцієнта Херста.

Створено окремий датасет `hurst_coef` в який було додано колонку ID і поряд результат обчислення коефіцієнта Херста. Розрахунок коефіцієнта Херста відбувається за рахунок функції `coef_hurst()` з бібліотеки “feasts”.

Цей показник необхідно розрахувати для визначення самоподібності даних і за рахунок цього розуміти властивість часових рядів до довготривалої пам’яті. Цей показник приймає значення від 0 до 1, де 0 означає повну відсутність самоподібності, а 1 говорить про досконалу самоподібність. Показники які набувають значення біля 0,5 говорить про часткову самоподібність.

Після проведення розрахунку коефіцієнта Херста отримано наступний результат (рис. 3.6).

```
# A tibble: 298 x 2
   ID      coef_hurst
  <chr>      <dbl>
1 P-1006      0.922
2 P-101836    0.909
3 P-102620    0.571
4 P-10484     0.986
5 P-106433    0.977
6 P-10709     0.956
7 P-110677    0.938
8 P-110819    0.961
9 P-111091    0.944
10 P-11121    0.945
```

Рисунок 3.6 – Результати коефіцієнта Херста

Джерело: Розроблено автором мовою R

Можемо побачити, що майже у всіх клієнтів цей показник є дуже високим, в середньому більше 0,9. Це свідчить, що дані схильні до самоподібності і що тимчасовий ряд має сильну тенденцію до збереження певних структур або закономірностей. Тобто подальше прогнозування буде мати досить високу точність.

3.3 Прогнозування доходів клієнтів

Перше прогнозування яке було зроблено це прогнозування на основі ковзної середньої. Для цього треба створити новий датасет який буде включати в себе всі вхідні змінні (ID, Country, date, value). За допомогою функції `mutate()` додати ще одну змінну в якій буде міститись показник ковзної середньої. Для цього групуємо дані по ID і за допомогою функції `rollmean()` з бібліотеки `zoo` розраховуємо ковзну середню для кожного клієнта окремо. Ширину рухомого вікна (кількість спостережень для розрахунку середнього) було обрано 4 місяці, оскільки усього спостережень від 12 до 36 місяців, і брати більше буде недоцільно (рис. 3.7).

	ID	Country	date	value	rolling_avg
	<chr>	<chr>	<mt>	<dbl>	<dbl>
1	P-1006	Germany	2020 лют	1738.	NA
2	P-1006	Germany	2020 бер	1737.	NA
3	P-1006	Germany	2020 кві	1353.	NA
4	P-1006	Germany	2020 тра	914.	1435.
5	P-1006	Germany	2020 чер	837.	1210.
6	P-1006	Germany	2020 лип	1175.	1070.
7	P-1006	Germany	2020 сер	1652.	1144.
8	P-1006	Germany	2020 вер	1421.	1271.
9	P-1006	Germany	2020 жов	1702.	1487.
10	P-1006	Germany	2020 лис	1717.	1623.

Рисунок 3.7 -Датасет з результатами згладжування ковзною середньою

Джерело: Розроблено автором мовою R

Тепер маємо датасет з додатковою змінною `rolling_avg`. Перші 3 спостереження порожні, оскільки для них відсутні 4 попередні часові ряди для розрахунку середнього.

Для зручності прогнозування було прибрано усі змінні окрім `ID` і `rolling_avg`, також видаляю перші три спостереження для кожного `ID`, оскільки значення ковзної середньої там порожні.

Для того, що б побудувати прогноз для кожного `ID` окремо мені необхідно застосовувати прогнозування окремо для кожного `ID`. Що б автоматизувати цей процес можна скористатися функцією `group_split()` з бібліотеки “`dplyr`”. Спершу я створив вектор `ID` який зберігається в такому ж порядку, що і дані ковзної середньої, зберігаю цей список окремо для зручності подальшої роботи. Тепер розділив датасет який складався з `ID` і ковзної середньої на окремі датасети кожен з яких відповідає кожному `ID` (рис. 3.8).

	<code>ID</code>	<code>rolling_avg</code>
	<code><chr></code>	<code><dbl></code>
1	P-97203	129.
2	P-97203	109.
3	P-97203	124.
4	P-97203	123.
5	P-97203	121.
6	P-97203	149.
7	P-97203	151.
8	P-97203	160.
9	P-97203	192.
10	P-97203	182.

Рисунок 3.8 – Один з розділених датасетів, що містить `ID` і результати ковзної середньої

Джерело: Розроблено автором мовою R

З цих датасетів тепер вилучаємо `ID` за допомогою функції `lapply()` для того, що б полегшити прогнозування.

Тепер за допомогою функції `forecast()` з бібліотеки `forecast` створено користувацьку функцію, що застосовує прогнозування на 3 періоди на кожен датасет окремо. Результати також зберігаються в списку датасетів в тому ж порядку в якому ці клієнти перебували в першочерговому датасеті, тому тепер, за допомогою функції `names()` надаємо кожному елементу цього списку, тобто кожному датасету результатів, ім'я, що дорівнює його значенням `ID`.

Отримав список, де кожен елемент називається по ID і містить прогнозування на 3 періоди за ковзною середньою (рис. 3.9).

mov_avg_forc	list [298]	List of length 298
P-1006	list [1]	List of length 1
rolling_avg	list [10] (S3: forecast)	List of length 10
model	list [19] (S3: ets)	List of length 19
mean	double [3] (S3: ts)	741 741 741
level	double [2]	80 95
x	double [32] (S3: ts)	1435 1210 1070 1144 1271 1487 ...
upper	double [3 x 2] (S3: mts, ts, matrix)	854 901 938 913 986 1042
lower	double [3 x 2] (S3: mts, ts, matrix)	629 582 545 569 497 441
fitted	double [32] (S3: ts)	1391 1435 1210 1070 1144 1271 ...
method	character [1]	'ETS(M,N,N)'
series	character [1]	'ts(x)'
residuals	double [32] (S3: ts)	0.0320 -0.1570 -0.1160 0.0698 0.1107 0.1702 ...
P-101836	list [1]	List of length 1
P-102620	list [1]	List of length 1
P-10484	list [1]	List of length 1

Рисунок 3.9 – Прогнозування для кожного клієнта в форматі списку

Джерело: Розроблено автором мовою R

Далі, за допомогою користувацької функції я зробив так, що б кожен запис прогнозу формувався в окремий датасет, до них додавалось 2 змінні – ID цього клієнта (посилаючись на назву спостереження в списку) і номер часового ряду. Ці датасети тепер я можу сформувати в один цілий (рис. 3.10).

rolling_avg.Point.Forecast	rolling_avg.Lo.80	rolling_avg.Hi.80	rolling_avg.Lo.95	rolling_avg.Hi.95	ID	period
741.437389	628.9388446	853.935934	569.3857125	913.48907	P-1006	33
741.437389	581.7918958	901.082882	497.2806766	985.59410	P-1006	34
741.437389	545.2301909	937.644587	441.3643742	1041.51040	P-1006	35
100.203085	75.5375045	124.868665	62.4803351	137.92583	P-101836	34
87.559403	51.0227359	124.096070	31.6813928	143.43741	P-101836	35
77.444454	30.9453591	123.943550	6.3302250	148.55868	P-101836	36
323.403500	282.7251911	364.081808	261.1913954	385.61560	P-102620	34
323.403500	269.0230096	377.763990	240.2357171	406.57128	P-102620	35
323.403500	258.0618371	388.745162	223.4720503	423.33495	P-102620	36
371.127510	273.4863921	468.768628	221.7983067	520.45671	P-10484	33
371.127510	231.6019179	510.653102	157.7415311	584.51349	P-10484	34
371.127510	198.4440925	543.810927	107.0310133	635.22401	P-10484	35

Рисунок 3.10 – Результати прогнозування в форматі датасету

Джерело: Розроблено автором мовою R

Можемо побачити, що тепер у нас є сформований датасет де сформовано прогноз для кожного клієнта на наступні 3 місяці, також наявні записи часового ряду, за рахунок чого прогноз буде легко суміщатися з вхідними даними.

Розглянемо побудування моделей прогнозування на прикладі одного клієнта – P-1006.

Побудуємо прогноз на основі експоненційного згладжування, модель ETS.

Перш за все беру вхідні дані по клієнту P-1006 і формую їх в окремий датасет. Далі за допомогою функції `ts()` формую значення доходу в формат часового ряду, вказавши частоту спостережень в 1 місяць, дату початку і кінця спостережень. Функцією `as.matrix()` форматував дані в матрицю для того, що б застосувати прогнозування моделі ETS(ANN). Це модель з адитивною похибкою, без адитивного тренду і без сезонності (рис. 3.11).

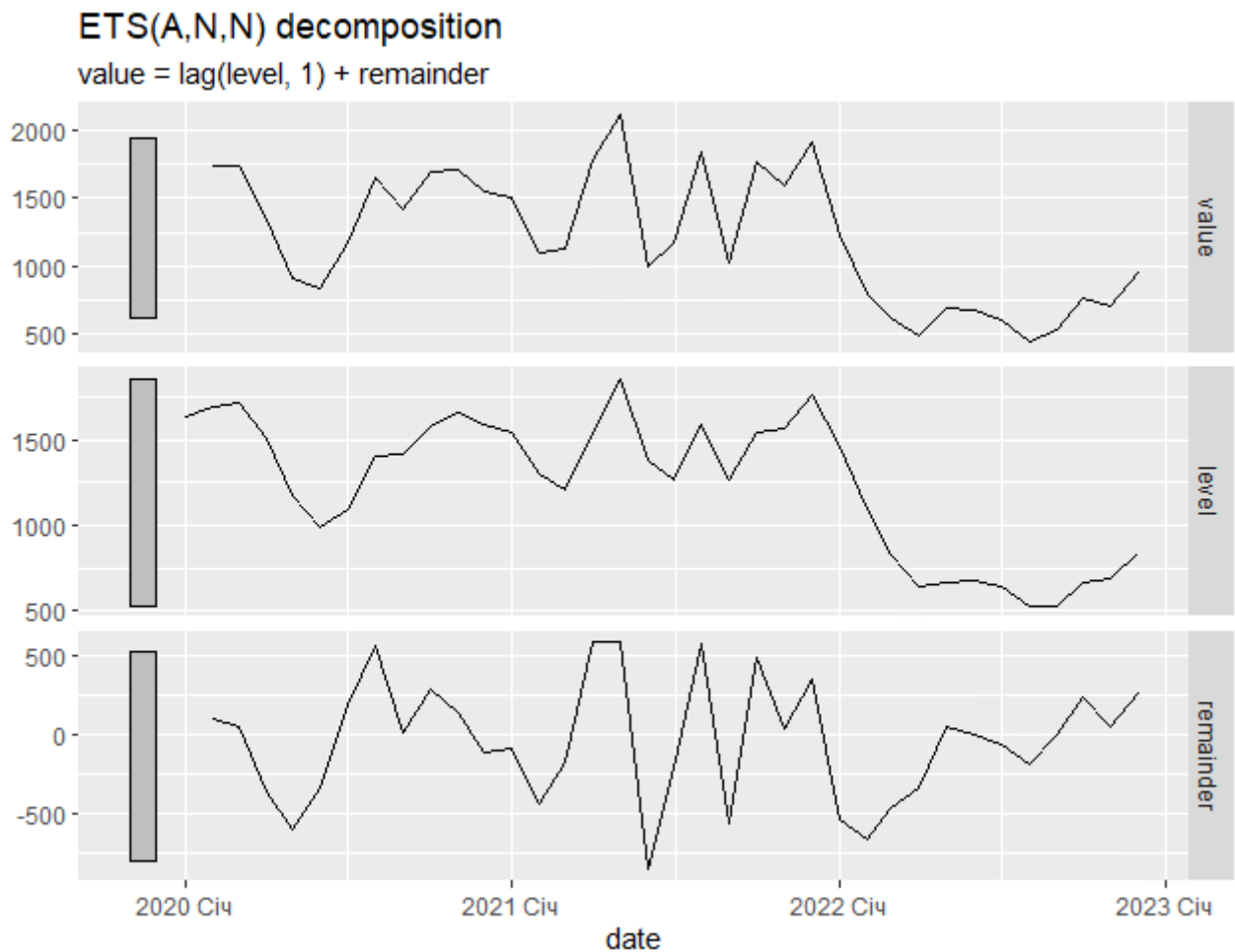


Рисунок 3.11 – Декомпозиція ETS моделі для клієнта P-1006

Джерело: Розроблено автором мовою R

Можемо побачити, що перший графік `value` відповідає реальним даним цього клієнта. Другий графік `level` демонструє загальну направленість, напрямок тренду, це є згладжений часовий ряд цього клієнта. Третій графік зображує викиди, тобто як згладжувався часовий ряд. По графіку можна відмітити значний спад в лютому 2021 року. Тепер можемо застосувати функцію `forecast()` до даної моделі і отримати прогнозування значень (рис. 3.12).

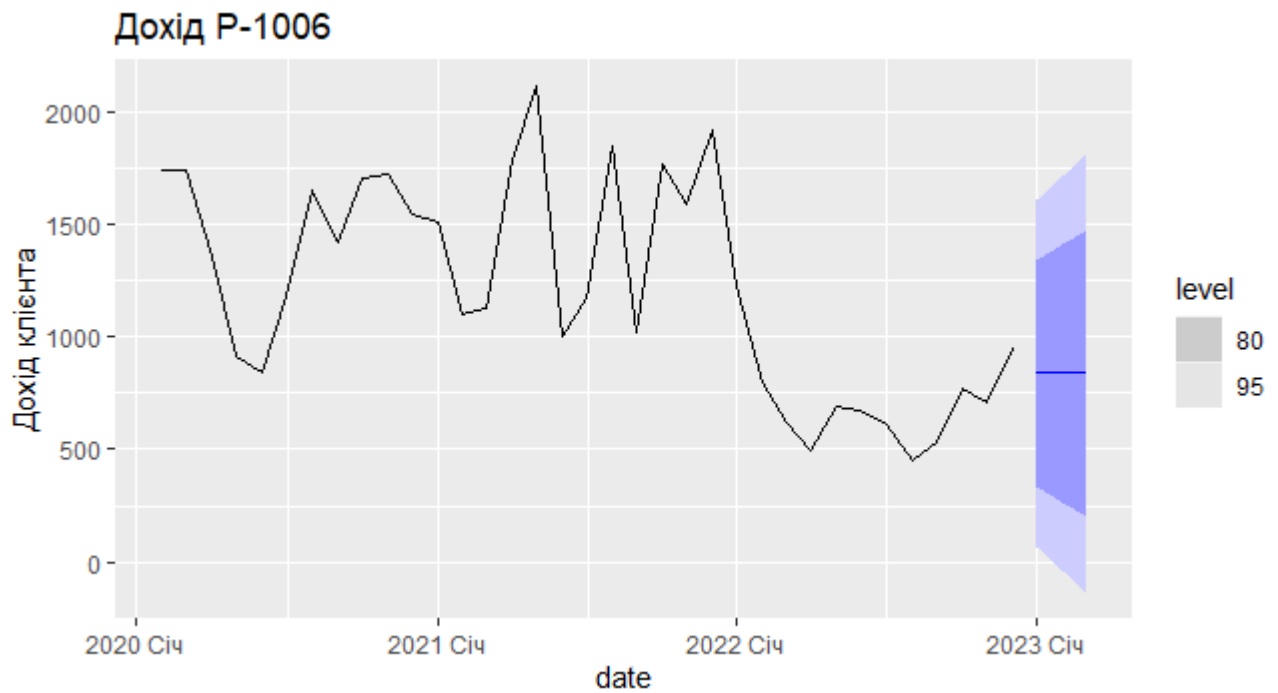


Рисунок 3.12 – Графік прогнозу доходу для клієнта Р-1006 за моделлю ETS(ANN)

Джерело: Розроблено автором мовою R

Можемо побачити прогнозоване значення нижче останнього спостереження. Також різними відтінками зображено ймовірність, що наступні спостереження будуть в цьому діапазоні. 80% - темніший і 95% - світліший.

Тепер зроблено прогнозування за моделлю Хольта з адитивною похибкою, і трендом, але без сезонності. Зроблено декомпозицію за цією моделлю (рис. 3.13).

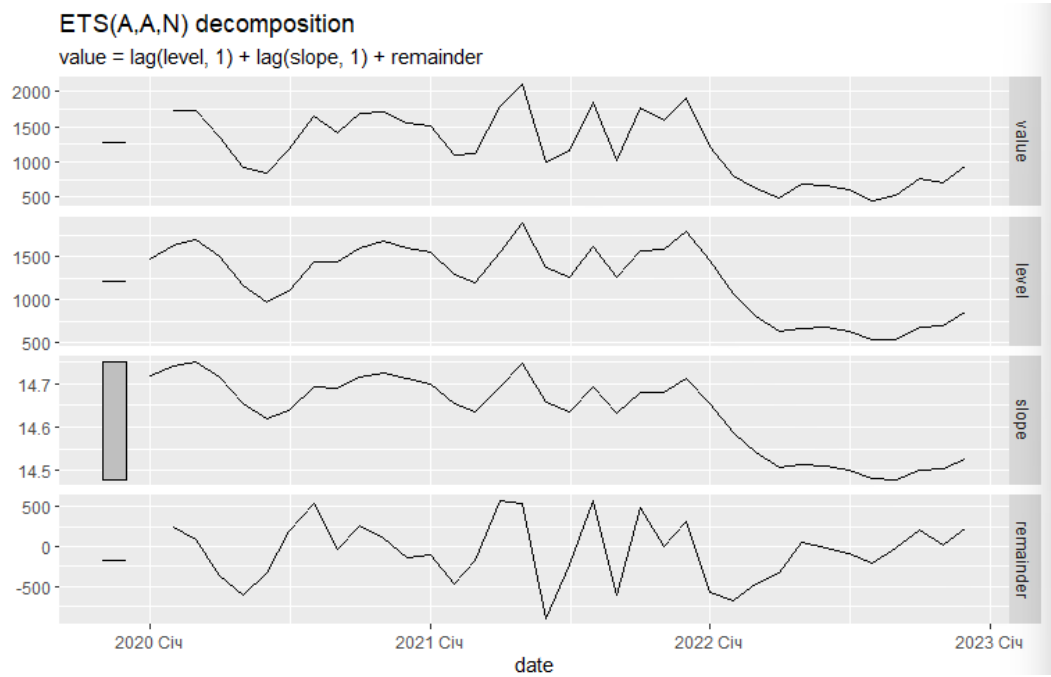


Рисунок 3.13 – Декомпозиція моделі Хольта

Джерело: Розроблено автором мовою R

Додався ще один графік *slope*, що демонструє коефіцієнт нахилу.

Побудовано прогнозування на основі цієї моделі за допомогою функції *forecast* (рис. 3.14).

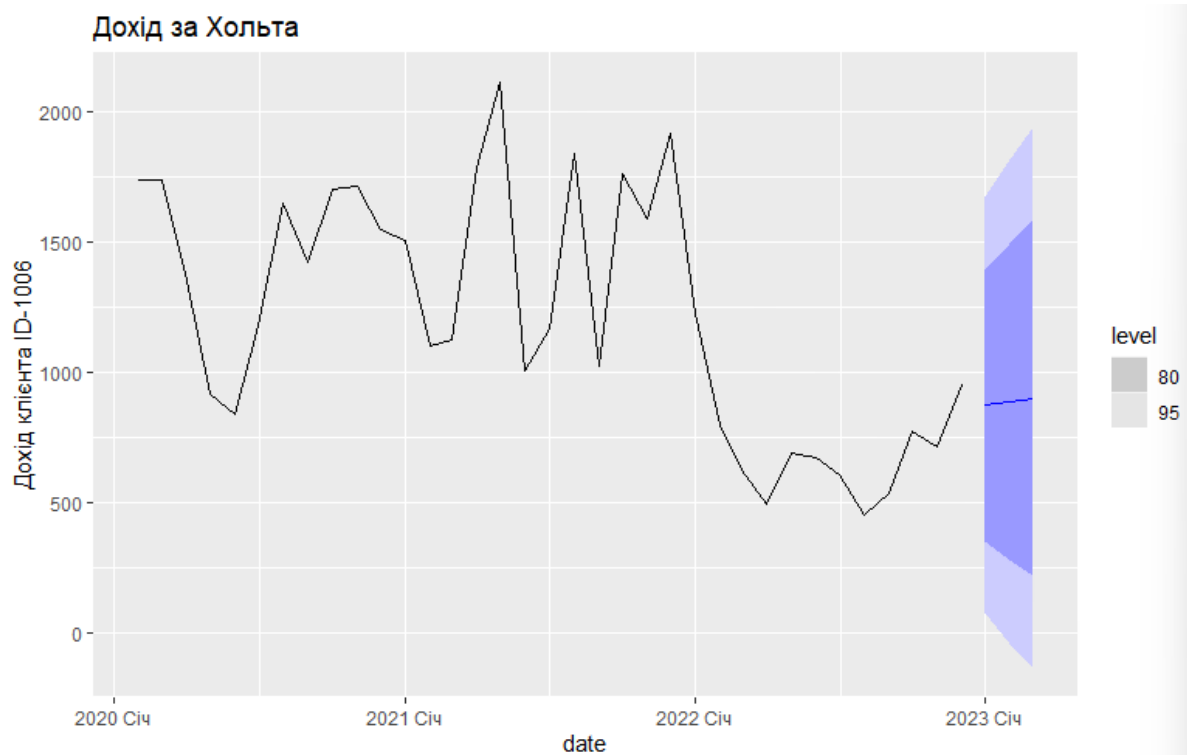


Рисунок 3.14 - Графік прогнозу доходу за моделлю Хольта

Джерело: Розроблено автором мовою R

Можемо побачити, що прогноз дуже схожий на прогноз за моделлю ANN, за виключенням того, що додався нахил прогнозу вгору, це означає що модель передбачає зростання значень. Оскільки початок прогнозування розташований нижче останнього спостереження можна зробити висновок, що модель передбачає зміну тренду або зростання темпу приросту значень.

Тепер слід побудувати модель Хольта – Вінтерса. Для адитивної моделі використовуються адитивні залишки, тренд і сезонність, а для мультиплікативної моделі – мультиплікативні залишки та сезонність і адитивний тренд.

Для цього створено дві ETS моделі, AAA — для адитивного методу і MAM для мультиплікативного. Побудовано по кожній моделі прогноз на 3 місяці (рис. 3.15).

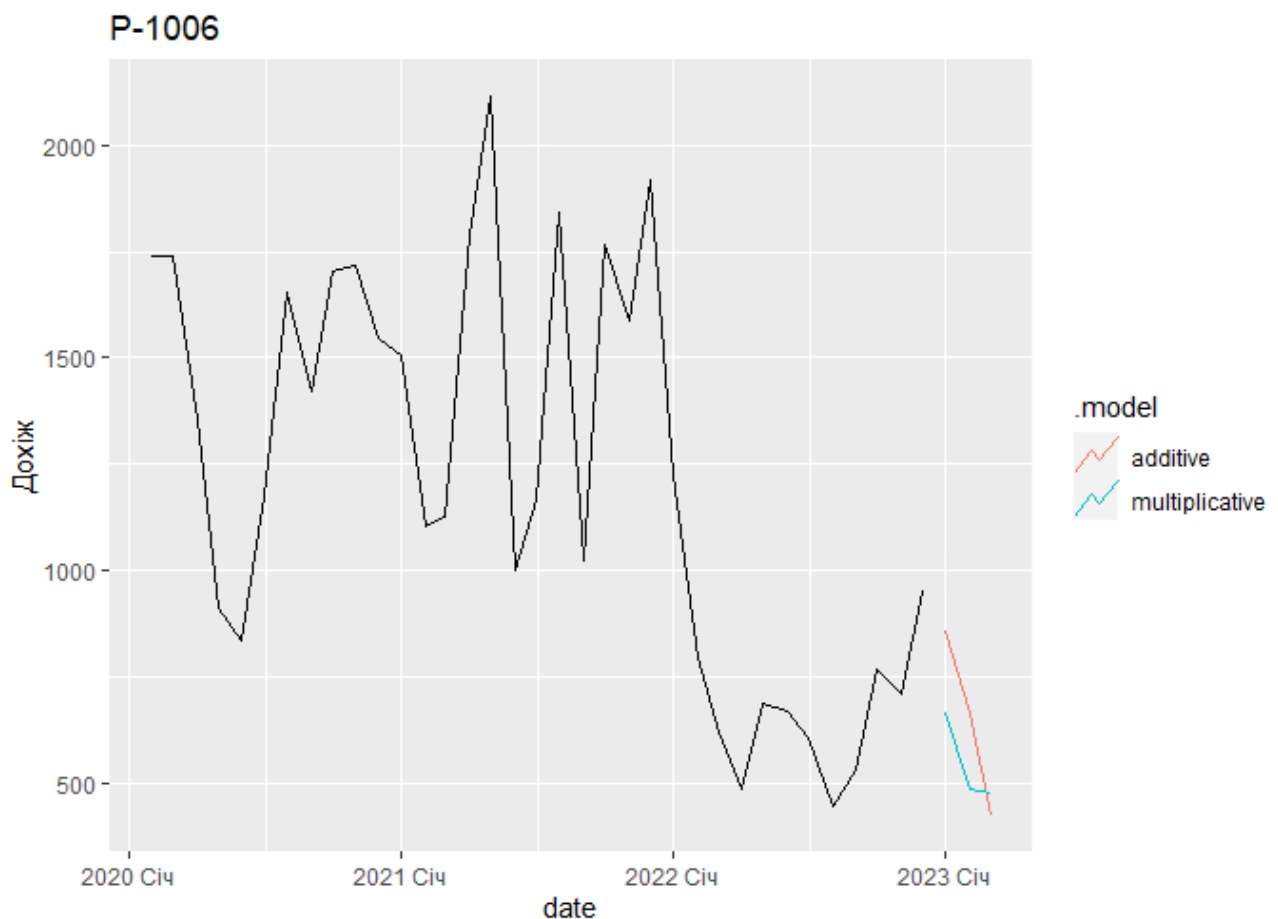


Рисунок 3.15 – Прогнозування за методом Хольта – Вінтерса

Джерело: Розроблено автором мовою R

Можемо побачити, що прогноз з урахуванням сезонності сильно відрізняється від попередніх моделей, тепер моделі прогнозують спадання

доходів, як адитивна, так і мультиплікативна. Для того, що б зрозуміти причину такого прогнозу зробив декомпозицію обох методів (рис. 3.16).

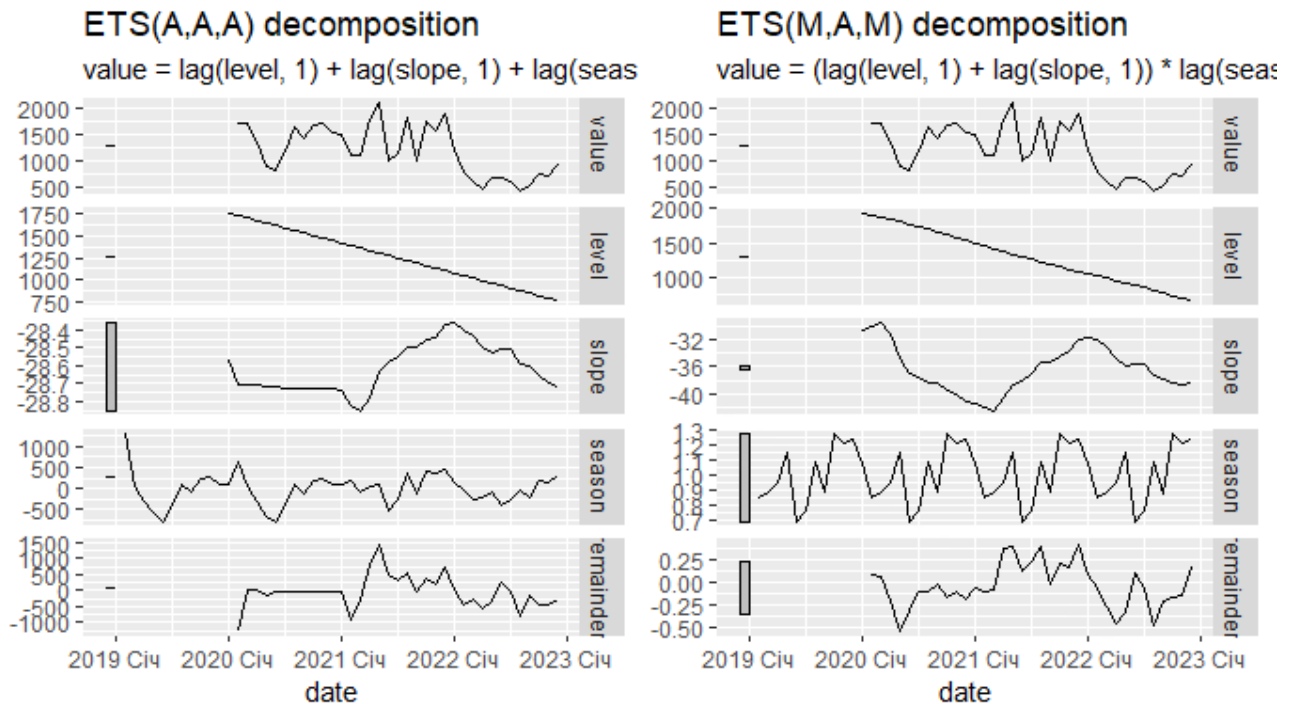


Рисунок 3.16 – Декомпозиція адитивного і мультиплікативного методу Хольта – Вінтерса

Джерело: Розроблено автором мовою R

В декомпозиції з'явився ще один графік, що відображає сезонність, тобто циклічні коливання впродовж року. З графіків можна помітити, що модель відмічає спад на початку майже кожного року. А оскільки реальні дані закінчуються в грудні 2022 року то модель передбачає спад впродовж наступних трьох місяців.

Тепер треба побудувати прогноз на основі демпфованого тренду за методом Хольта Вінтерса, для того, що б зменшити вплив короткострокових коливань і спрогнозувати дохід на три місяці за допомогою згладженого тренду (рис. 3.17).

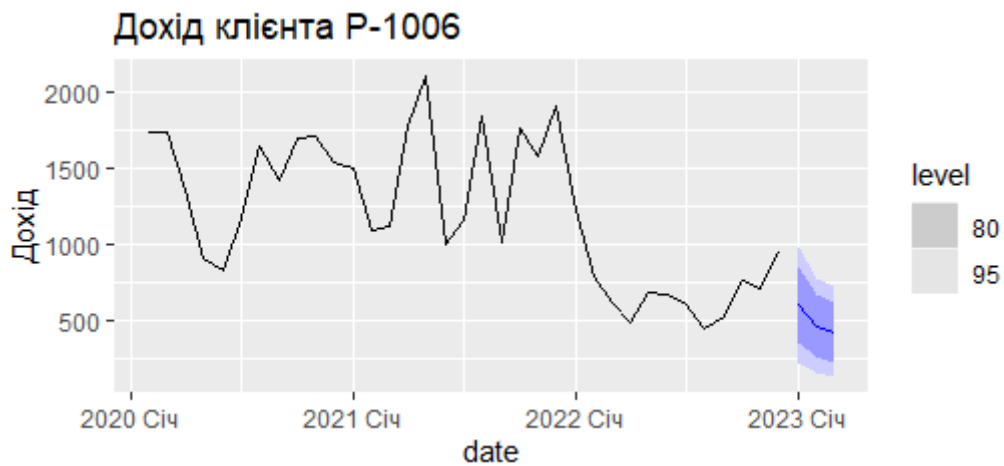


Рисунок 3.17 – Графік прогнозу демпфованого тренду Хольта – Вінтерса

Джерело: Розроблено автором мовою R

Можемо побачити що навіть при демпфуванні тренду, модель з сезонністю Хольта Вінтерса прогнозує спад доходів для наступних трьох місяців. Тобто цей спад обумовлений не короткостроковими коливаннями, а дійсною тенденцією тренду.

3.4 Формування скоринг моделі

Тепер коли всі моделі побудовані і є достатня кількість моделей для того, що б робити певні висновки можна сформуванати висновки для цього клієнту в форматі скорингової моделі.

В скорингову модель входять такі дані як коефіцієнт Херста, стандартне відхилення і коефіцієнт варіації, для оцінювання розсіювання і якості часових рядів. В якості моделей прогнозування — ковзна середня, експоненційне згладжування, модель Хольта, модель Хольта – Вінтерса, адитивний і мультиплікативний методи Хольта – Вінтерса, модель Хольта – Вінтерса з демпфованим трендом.

Для зручності скорингова модель була сформована в програмі Excel. Для цього мені необхідно перенести дані прогнозів з R в Excel. Функція `write.xlsx()` з бібліотеки `openxlsx` створює Excel файл в який експортує дані в форматі таблиці.

На окремому аркуші сформовано скоринг таблицю, в якій прописав результати всіх зроблених моделей прогнозування і критеріїв оцінки часового ряду, таких як стандартне відхилення і коефіцієнт варіації (рис. 3.18).

No.	Категорія / Category	Критерій оцінювання / Evaluation criteria	Бали / Points	Вага / Weight	Сума / Total	Count	
1	Коеф Херста	0,922003509	1000	8,0%	80	1000	80
			500		40	0	0
			0		0	0	0
2	Стандартне Відхилення	480,4103562	1000	5,0%	50	0	0
			500		25	500	25
			0		0	0	0
3	Коеф Варіації	0,397945762	1000	5,0%	50	0	0
			500		25	500	25
			0		0	0	0
4	Ковзаюча середня	741,4373891	1000	10,0%	100	1000	100
			500		50	0	0
			0		0	0	0
5	ETS	837,2953231	1000	15,0%	150	1000	150
			500		75	0	0
			0		0	0	0
6	Holt	900,4317667	1000	15,0%	150	1000	150
			500		75	0	0
			0		0	0	0
7	Holt Winters	427,8465091	1000	10,0%	100	0	0
			500		50	0	0
			0		0	0	0
8	Holt Winters drift(адативний)	425,1965243	1000	10,0%	100	0	0
			500		50	0	0
			0		0	0	0
9	Holt Winters drift(мультиплікативний)	478,2496149	1000	11,0%	110	0	0
			500		55	0	0
			0		0	0	0
10	Holt Winters drift(мультиплікативний)	427,8465091	1000	11,0%	110	0	0
			500		55	0	0
			0		0	0	0

Рисунок 3.18 – Скоринг таблиця

Джерело: Розроблено автором в середовищі Excel

В цій таблиці проставлена вага кожного критерію на кінцевий бал. Перші три категорії виділені окремим кольором, бо це не результати прогнозування, а критерії оцінки даних, наскільки велика дисперсія часового ряду і наскільки він схильний до збереження довготривалої пам'яті. Зеленим кольором виділяються моделі прогнозування і їх оцінка базується на тому прогнозує модель зростання чи спадання прогнозованого доходу. Для клієнта Р-1006 можемо побачити, що коефіцієнт варіації і стандартне відхилення свідчать про наявність коливань даних майже 40% від середнього доходу, що досить багато і каже про сильну дисперсію, проте коефіцієнт Херста показав результат 0,9, що є дуже високою оцінкою і каже про здатність часового ряду успадковувати тренд. За рахунок цього моделі прогнозування будуть мати більшу точність прогнозу.

Моделі ковзної середньої, ETS і Хольта прогнозують зростання доходів впродовж наступних трьох місяців. Але моделі з урахуванням сезонності прогнозують спадання доходів. Кінцевий бал оцінюється за системою, наведеною на рис. 3.19. Наведена таблиця, на якій зображено як формується кінцевий бал і який ризик він позначає.

MASTERSCALE	Score group	Client category	Points	Max points	1000	Points	530
	1	A - Very low risk	>700	Min points	0	Client category	MEDIUM RISK
	2	A - Low risk	601-700				
	3	B - Medium risk	501-600				
	4	B - Medium risk	401-500				
	5	C - High risk	351-400				
	6	C - High risk	301-350				
	7	D - Very high risk	201-300				
	8	E - Unacceptable risk	101-200				
	9	E - Unacceptable risk	51-100				
	10	E - Unacceptable risk	0-50				

Рисунок 3.19 – Оцінювання скоринг таблиці

Джерело: Розроблено автором в середовищі Excel

Сумарний бал скорингової моделі понад 700 буде означати дуже маленький ризик і що ймовірність того, що клієнт буде мати можливість сплачувати кредит протягом наступних трьох місяців, є дуже високою. Для клієнта Р-1006 результат 530, що означає середній ризик. Це обумовлено тим, що половина моделей показала зростання доходів, а інша половина відобразила спадання за рахунок врахування сезонності.

ВИСНОВКИ

Розрахунок кредитного ризику для фінансових організацій є дуже важливим і завжди актуальним завданням. Компанії постійно намагаються покращити точність розрахунку цього показника, що б зменшити збитки і максимізувати прибуток від кредитування. В сучасних умовах війни це завдання стало ще складнішим, оскільки багато даних, які використовуються в стандартних моделях управління ризиками — відсутні, а компанії потребують надійних методів розрахунку цих ризиків.

Впродовж цього дослідження було ретельно досліджено сучасні методи оцінки ризикованості клієнтів банків, використання прогнозування для розрахунку ризику і методології прогнозування часових рядів за різними моделями. Також були проаналізовані реальні дані кредитного портфеля фінансової міжнародної компанії з надання кредитних послуг, сформовані моделі прогнозування доходів клієнтів за методами згладжування мовою R і створена скорингова система в середовищі Excel.

Проведений аналіз показав, що дані часові ряди мають великий коефіцієнт варіації, в середньому більше 0,6 що свідчить про сильну дисперсію і відсутність постійної стабільності доходів. Це обумовлено специфікою партнерських доходів від платформи YouTube, де дохід за місяць формується виключно на кількість переглядів за цей місяць. Кількість переглядів в свою чергу залежить від якості контенту, кількості реклами, сезону (в середньому в теплі пори року переглядів менше ніж в холодні). Проте коефіцієнт Херста для більшості даних є дуже високим, що свідчить про сильну властивість даних успадковувати тренд.

Побудовані прогнози розділились на 2 групи. Моделі які не враховували сезонність прогнозували подальше зростання доходів протягом трьох місяців. Були використані моделі на основі ковзної середньої, ETS і модель Хольта. Всі вони спрогнозували приріст доходів. Проте моделі які враховують сезонність,

такі як моделі Хольта – Вінтерса, Хольта – Вінтерса адитивна та мультиплікативна моделі і модель Хольта Вінтерса з демпфованим трендом, показали прогноз на спад доходів. Що обумовлено тим, що модель визначила сезонність таким чином що на початку кожного року дохід спадає, а оскільки реальні дані для клієнта Р-1006 закінчуються в грудні 2022 то і тренд прогнозується на спад.

Всі ці результати прогнозів були експортовані в Excel де було створено зручну скоринг таблицю на основі якої фінансова організація з надання послуг кредитування може робити висновки щодо кредитування конкретного клієнта. Що потенційно здатно зменшити збитки компанії від ризикових клієнтів.

Цим дослідженням я довів важливість коректної оцінки банківського ризику, і необхідність застосування моделей прогнозування як один з методів управління цим ризиком. Також побудував алгоритм обробки даних і їх прогнозування, за рахунок чого створив скорингову модель для підрахунку ризикованості клієнтів банку.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Л Б Штефан. Проблеми управління кредитним ризиком в комерційних банках України. Ефективна економіка №3 2015. Журнал «Ефективна економіка» - наукове фахове видання з питань економіки. URL: <http://www.economy.nayka.com.ua/?op=1&z=3904> (дата звернення: 01.06.2023).
2. Counterparty credit risk in Basel III - Executive Summary. Bank for International Settlements. URL: https://www.bis.org/fsi/fsisummaries/ccr_in_b3.htm (date of access: 01.06.2023).
3. The 5 C's of Credit in the Lending Industry. Search eLibrary :: SSRN. URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1872804 (date of access: 01.06.2023).
4. Suss, Joel and Treitel, Henry, Predicting Bank Distress in the UK with Machine Learning (October 4, 2019). Bank of England Working Paper No. 831, October 2019, Available at SSRN: <https://ssrn.com/abstract=3465753> or <http://dx.doi.org/10.2139/ssrn.3465753>
5. Özparlak G., ÖZDEMİR DİLİDÜZGÜN M. CORPORATE BANKRUPTCY PREDICTION USING MACHINE LEARNING METHODS: THE CASE OF THE USA. International Journal of Management Economics and Business. 2022. URL: <https://doi.org/10.17130/ijmeb.1079688> (date of access: 01.06.2023).
6. Банківські ризики та їх види. Реферат. Освіта.UA. URL: <https://osvita.ua/vnz/reports/bank/19870/> (дата звернення: 01.06.2023).
7. Міністерство Фінансів України. URL: https://mof.gov.ua/storage/files/IFRS_9_Ukrainian-compressed.pdf (дата звернення: 01.06.2023).
8. Бюро кредитних історій. Забезпечення конституційних прав і свобод фізичних осіб при організації формування та обігу кредитних історій. Міністерство юстиції України.

9. CUA - The Five Cs of Credit. CUA - Personal Banking. URL: <https://cua.com/Home/Advice/CUAdvice/Five-Cs-of-Credit/> (date of access: 01.06.2023).
10. Welcome to Pages at the Stern School of Business, New York University. URL: <https://pages.stern.nyu.edu/~ealtman/CorporateDistressDiagnosis.pdf> (date of access: 01.06.2023).
11. Faculty of Medicine, McGill University. URL: <http://www.medicine.mcgill.ca/epidemiology/hanley/IntMedResidents/SD-SE.pdf> (date of access: 01.06.2023).
12. Standard Deviation: 6 Steps to Calculation.
URL: <https://www.sixsigmadaily.com/standard-deviation-6-steps-to-calculation/> (date of access: 01.06.2023).
13. Subir Mansukhani. The Hurst Exponent: Predictability of Time Series. Analytics.inform. URL: <https://pubsonline.informs.org/doi/10.1287/LYTX.2012.04.05/full/>.
14. Дяченко Ю. А. Науковий вісник Ужгородського національного університету. URL: <https://dspace.uzhnu.edu.ua/jspui/bitstream/lib/24214/1/РОЗВИТОК%20МЕТОДІВ%20ПРОГНОЗУВАННЯ%20ДИНАМІКИ%20БІРЖОВИХ%20ЦІН.pdf> (дата звернення: 01.06.2023).
15. Ashish Kumar Srivastav. Moving Average. WallStreetMojo. URL: <https://www.wallstreetmojo.com/moving-average-formula/>.
16. Експоненціальне ковзне середнє. Forex Broker InstaForex: trading on Forex market. URL: <https://www.instaforex.com/support/ua/article/121> (дата звернення: 01.06.2023).
17. Журнал «Ефективна економіка» - наукове фахове видання з питань економіки. URL: http://www.economy.nayka.com.ua/pdf/4_2019/42.pdf (дата звернення: 01.06.2023).

18. Joseph W. Holt's exponential smoothing - ppt video online download. SlidePlayer - Upload and Share your PowerPoint presentations. URL: <https://slideplayer.com/slide/7867571/> (date of access: 01.06.2023).
19. Altay, N., Rudisill, F., & Litteral, L. A. (2008). Adapting Wright's modification of Holt's method to forecasting intermittent demand. *International Journal of Production Economics*, 111(2), 389-408. <https://doi.org/10.1016/j.ijpe.2007.01.009>
20. Сезонність - Фінансова енциклопедія. Фінансова енциклопедія. URL: <https://ua.nesrakonk.ru/seasonality/> (дата звернення: 01.06.2023).
21. Українські підручники, посібники та статті - Бібліотека Posibniki. Електронна бібліотека підручників онлайн. Українські підручники, посібники та статті - Бібліотека Posibniki. Електронна бібліотека підручників онлайн. URL: <https://posibniki.com.ua/post-sezonna-dekompoziciya-chasovogo-ryad> (дата звернення: 01.06.2023).
22. Hyndman R. J., Athanasopoulos G. *Forecasting: Principles and Practice*. OTexts, 2021.
URL: <https://otexts.com/fpp2/holt-winters.html>
23. Monash University | Study at Monash Uni | Melbourne. URL: https://www.monash.edu/business/ebs/research/publications/ebs/forecasting_models_and_prediction_intervals_for_the_multiplicative_holt_winters_method.pdf (date of access: 01.06.2023).
24. Hyndman R. J., Athanasopoulos G. *Forecasting: Principles and Practice*. OTexts, 2021.
URL: <https://otexts.com/fpp2/holt-winters.html>
25. lubridate package - RDocumentation. Home - RDocumentation. URL: <https://www.rdocumentation.org/packages/lubridate/versions/1.9.2> (date of access: 01.06.2023).
26. tidyr package - RDocumentation. Home - RDocumentation. URL: <https://www.rdocumentation.org/packages/tidyr/versions/1.3.0> (date of access: 01.06.2023).

27. readr package - RDocumentation. Home - RDocumentation. URL: <https://www.rdocumentation.org/packages/readr/versions/2.1.4> (date of access: 01.06.2023).
28. forcats package - RDocumentation. Home - RDocumentation. URL: <https://www.rdocumentation.org/packages/forcats/versions/1.0.0> (date of access: 01.06.2023).
29. dplyr package - RDocumentation. Home - RDocumentation. URL: <https://www.rdocumentation.org/packages/dplyr/versions/1.0.10> (date of access: 01.06.2023).
30. ggplot2 package - RDocumentation. Home - RDocumentation. URL: <https://www.rdocumentation.org/packages/ggplot2/versions/3.4.2> (date of access: 01.06.2023).
31. pracma package - RDocumentation. Home - RDocumentation. URL: <https://www.rdocumentation.org/packages/pracma/versions/1.9.9> (date of access: 01.06.2023).
32. Achim Zeileis. zoo: An S3 Class and Methods for Indexed Totally Ordered Observations. CiteSeerX. URL: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=da6cede528db3a97150fa6716562ea5410033821> (date of access: 01.06.2023).
33. Introduction to feasts. The Comprehensive R Archive Network. URL: <https://cran.r-project.org/web/packages/feasts/vignettes/feasts.html> (date of access: 01.06.2023).
34. forecast package - RDocumentation. Home - RDocumentation. URL: <https://www.rdocumentation.org/packages/forecast/versions/8.21> (date of access: 01.06.2023).
35. readxl package - RDocumentation. Home - RDocumentation. URL: <https://www.rdocumentation.org/packages/readxl/versions/1.4.2> (date of access: 01.06.2023).

36. gridExtra package - RDocumentation. Home - RDocumentation. URL: <https://www.rdocumentation.org/packages/gridExtra/versions/2.3> (date of access: 01.06.2023).
37. fable package - RDocumentation. Home - RDocumentation. URL: <https://www.rdocumentation.org/packages/fable/versions/0.3.3> (date of access: 01.06.2023).
38. patchwork package - RDocumentation. Home - RDocumentation. URL: <https://www.rdocumentation.org/packages/patchwork/versions/1.1.2> (date of access: 01.06.2023).
39. openxlsx package - RDocumentation. Home - RDocumentation. URL: <https://www.rdocumentation.org/packages/openxlsx/versions/4.2.5.2> (date of access: 01.06.2023).

Ім'я користувача:
Математичного моделювання та статистики Даценк...

ID перевірки:
1015393411

Дата перевірки:
02.06.2023 13:56:16 EEST

Тип перевірки:
Doc vs Internet + Library

Дата звіту:
02.06.2023 13:58:19 EEST

ID користувача:
100005744

Назва документа: Василенко_Іщук_Плагіат

Кількість сторінок: 64 Кількість слів: 12221 Кількість символів: 95107 Розмір файлу: 847.26 KB ID файлу: 1015057818

4%

Схожість

Найбільша схожість: 0.88% з Інтернет-джерелом (http://www.economy.nayka.com.ua/pdf/4_2019/42.pdf)

2.8% Джерела з Інтернету

192

Сторінка 66

3.27% Джерела з Бібліотеки

346

Сторінка 67

0% Цитат

Вилучення цитат вимкнене

Вилучення списку бібліографічних посилань вимкнене

0%

Вилучень

Немає вилучених джерел