

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**Національний університет «Острозька академія»**  
**Економічний факультет**  
**Кафедра економіко-математичного моделювання та інформаційних**  
**технологій**

Кваліфікаційна робота

на здобуття освітнього ступеня магістра

на тему: **«Використання машинного навчання для**  
**прогнозування інвестицій»**

**Виконала:**

студентка 2 курсу, групи МЕК-61  
спеціальності 051 «Економіка»  
освітньо-професійної програми «Економічна кібернетика»  
другого (магістерського) рівня вищої освіти  
**Яцюк Дарія Сергіївна**

**Керівник:** старший викладач  
**Клебан Юрій Вікторович**

**Рецензент:**

Front-end Developer “DOODLE”, LLC  
**Місай Володимир Віталійович**

***"РОБОТА ДОПУЩЕНА ДО ЗАХИСТУ"***

**Завідувач кафедри економіко-математичного моделювання та**  
**інформаційних технологій \_\_\_\_\_ (проф. Ольга КРИВИЦЬКА)**

Протокол № 5 від 24 листопада 2022 р.

**Острог, 2022**

МІНІСТЕРСТВО ОСВІТИ І НАУКИ  
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «ОСТРОЗЬКА АКАДЕМІЯ»

Економічний факультет  
Кафедра економіко-математичного моделювання та інформаційних технологій  
Спеціальність 051 «Економіка»  
Освітньо-професійна програма «Економічна кібернетика»

ЗАТВЕРДЖУЮ

Завідувач кафедри  
економіко-математичного  
моделювання та інформаційних  
технологій

проф., д.е.н. Ольга КРИВИЦЬКА  
“ \_\_\_\_\_ ” \_\_\_\_\_ 202\_\_ року

**ЗАВДАННЯ  
НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТА**

***Яцюк Дарії Сергіївни***

1. Тема роботи: **Використання методів машинного навчання для прогнозування інвестицій**

керівник роботи: : *Клебан Юрій Вікторович, старший викладач*

затверджено наказом ректора Національного Університету «Острозька академія» від *29 жовтня 2021 р. №110*

2. Термін здачі студентом закінченої роботи на кафедрі: *09 грудня 2022 року*

3. Вихідні дані до роботи: *науково-публіцистичні праці з досліджуваної проблематики вітчизняних та зарубіжних вчених, мови програмування R, дані про обсяги капітальних інвестицій та макроекономічних показників України із сайту Державної статистики України, Міністерства фінансів України, ЗУ «Про інвестиційну діяльність», Постанови Кабінету міністрів України щодо регулювання інвестиційної діяльності.*

4. Зміст розрахунково-пояснювальної записки (перелік питань, які потрібно розробити): *1) підходи до аналізу та прогнозування капітальних інвестицій; 2) засади інвестиційної діяльності та аналіз основних факторів впливу на обсяг капітальних інвестицій 3) прикладні аспекти прогнозування капітальних інвестицій та порівняльний аналіз якості найбільш використовуваних методів та моделей.*

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень): *графічна візуалізація етапів побудови моделей та прогнозування обсягів капітальних інвестицій: дослідження часових рядів на стаціонарність та присутність у ньому трендової та сезонної складової, графічний вигляд побудованих методів регуляризації, та ансамблевих, результати прогнозування на основі побудованих моделей,*

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	Завдання Прийняв
Розділ 1	Клебан Ю. В., старший викладач кафедри економіко-математичного моделювання та інформаційних технологій	08.02.2022	08.02.2022
Розділ 2	Клебан Ю. В., старший викладач кафедри економіко-математичного моделювання та інформаційних технологій	16.05.2022	16.05.2022
Розділ 3	Клебан Ю. В., старший викладач кафедри економіко-математичного моделювання та інформаційних технологій	24.06.2022	24.06.2022

7. Дата видачі завдання: *01 листопада 2021 р.*

**КАЛЕНДАРНИЙ ПЛАН**

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів роботи	Примітка
1.	Вивчення літератури	до 07.02.2022	
2.	Розробка змісту (плану)	до 15.03.2022	
3.	Ознайомлення керівника з текстом кваліфікаційної роботи (чорновий варіант):		
3.1	Розділ 1	до 19.05.2022	
3.2	Розділ 2	до 09.06.2022	
3.3	Розділ 3	до 10.11.2022	
4.	Ознайомлення керівника з текстом кваліфікаційної роботи із врахуванням зауважень	до 20.11.2022	
5.	Попередній захист кваліфікаційної роботи	до 23.11.2022	
6.	Рецензування кваліфікаційної роботи	до 09.12.2022	
7.	Здача роботи на кафедрі Реєстрація на Moodle	до 09.12.2022	

Студент \_\_\_\_\_ Дарія ЯЦЮК

Керівник роботи \_\_\_\_\_ Юрій КЛЕБАН

**АНОТАЦІЯ**  
**кваліфікаційної роботи**  
**на здобуття освітнього ступеня магістра**

*Тема:* **«Використання методів машинного навчання для прогнозування інвестицій»**

*Автор:* Яцюк Дарія Сергіївна

*Науковий керівник:* старший викладач Клебан Юрій Вікторович

*Захищена «.....».....2022 року.*

**Короткий зміст праці:** Кваліфікаційна робота присвячена систематизації та оцінці порівняння якості методології розробки прогнозу інвестицій в основний капітал в рамках прогнозування соціально-економічного розвитку України.

У першому розділі наведено опис використовуваних методів машинного навчання (Ridge, LASSO, Post-Lasso, Adaptive LASSO, Elastic Net, Spike and Slab, випадковий ліс, бустинг) і альтернативних методів прогнозування, використовуваних для порівняння якості. Описані основні проблеми з якими можуть зіткнутись дослідники у процесі їх використання.

Другий розділ присвячений аналізу інвестиційної діяльності в Україні та основних чинників, що впливають на обсяг капітальних інвестицій. Описано сутність капітального інвестування в Україні. Проаналізовано основні тенденції розвитку капіталоінвестування. Визначено основні фактори впливу на обсяг капітальних інвестицій.

У третьому розділі описані використані дані, способи їх трансформації та методи побудови прогнозів, містяться емпіричні результати дослідження. Здійснено їх порівняльний аналіз. Визначено, що оцінки коефіцієнтів загальноприйнятих ключових факторів інвестицій, отримані за допомогою методів регуляризації, в цілому узгоджуються з економічною теорією. Описано основні недоліки використовуваних методів та моделей. Визначено основні проблеми при макроекономічному прогнозуванні.

**Ключові слова:** капітальні інвестиції, прогнозні моделі, порівняльний аналіз, машинне навчання, Ridge, LASSO, Post-Lasso, Adaptive LASSO, Elastic Net, Spike and Slab, випадковий ліс, бустинг.

\_\_\_\_\_ Дарія Яцюк

**ANNOTATION**  
**qualification work**  
**to obtain a master's degree**

*Topic:* «Use of machine learning methods to forecast investment»

*Author:* Dariia Yatsiuk

*Academic supervisor:* Senior lecturer Yuriy Kleban

*Protected by ".....".....2022.*

**Summary of the work:** the qualification work is devoted to systematization and evaluation of quality comparison of the methodology for developing a forecast of investments in fixed assets in the framework of forecasting socio-economic development of Ukraine.

The first section describes the machine learning methods used (Ridge, LASSO, Post-Lasso, Adaptive LASSO, Elastic Net, Spike and Slab, random forest, boosting) and alternative forecasting methods used to compare quality. The main problems that researchers may encounter in the process of using them are described.

The second section is devoted to the analysis of investment activity in Ukraine and the main factors affecting the volume of capital investments. The essence of capital investment in Ukraine is described. The main trends in the development of capital investment are analyzed. The main factors influencing the volume of capital investments are identified.

The third chapter describes the data used, methods of their transformation and methods of constructing forecasts, contains empirical results of the study and makes a comparative analysis of them. It is determined that the estimates of the coefficients of generally accepted key investment factors obtained using regularization methods are generally consistent with economic theory. The main disadvantages of the methods and models used are described. The main problems of macroeconomic forecasting are identified.

**Keywords:** capital investment, predictive models, comparative analysis, machine learning, Ridge, LASSO, Post-Lasso, Adaptive LASSO, Elastic Net, Spike and Slab, random forest, boosting.

\_\_\_\_\_ *Dariia Yatsiuk*

## ЗМІСТ

ВСТУП	3
РОЗДІЛ 1. ТЕОРЕТИКО-МЕТОДИЧНІ ОСНОВИ КОРОТКОСТРОКОВОЇ ОЦІНКИ ТА ПРОГНОЗУ МАКРОЕКОНОМІЧНИХ ПОКАЗНИКІВ	6
1.1 Аналіз актуальних публікацій щодо прогнозування інвестицій	6
1.2 Розгляд методів, що використовуються машинним навчанням та альтернативних методів для прогнозування інвестицій	9
1.2.1 Методи регуляризації	16
1.2.2 Ансамблеві методи	12
1.2.3 Альтернативні методи	16
1.3 Аналіз методів для порівняння якості моделей при прогнозуванні	22
РОЗДІЛ 2. АНАЛІЗ ІНВЕСТИЦІЙНОЇ ДІЯЛЬНОСТІ В УКРАЇНІ ТА ОСОБЛИВОСТІ МЕТОДІВ ДЛЯ ЗДІЙСНЕННЯ ПРОГНОЗУ ТЕМПІВ ЗРОСТАННЯ ІНВЕСТИЦІЙ	28
2.1 Нормативно-правове регламентування обліку капітальних інвестицій	28
2.2 Динаміка і структура інвестицій в основний капітал України	37
2.3 Основні чинники, що впливають на обсяг капітальних інвестицій в Україні	42
РОЗДІЛ 3. Моделювання та оцінка прогностичної моделі темпів зростання інвестицій в основний капітал України	49
3.1 Використання даних, способи їх трансформації та методи побудови прогнозу інвестицій	49
3.2 Порівняльна оцінка якості моделей	55
3.2 Проблеми короткострокового прогнозування макроекономічних показників	66
ВИСНОВОК	70
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ ТА ЛІТЕРАТУРИ:	73
ДОДАТКИ	86

## ВСТУП

**Актуальність теми дослідження.** Інвестиції в основний капітал необхідні для розвитку підприємств, галузей промисловості та соціальної інфраструктури. Збільшення інвестицій в основний капітал сприяє створенню та оновленню основних фондів, модернізації та розширенню виробництва. Тому важливо в цьому аспекті планувати інвестиції в основний капітал, що є одним з інструментів, що дозволяють сформулювати стратегію розвитку господарюючого суб'єкта. А тому, вибір найбільш оптимального методу прогнозування дозволить створити найбільш приблизний прогноз.

Завдання з розширення меж інвестиційної діяльності в Україні належить до першочергових, оскільки це забезпечить динамічне економічне зростання. Стратегією сталого розвитку «Україна — 2020», схваленою Указом Президента України від 12 січня 2015 року № 5/2015, Стратегією реформування системи управління державними фінансами на 2017—2020 роки, схваленою Розпорядженням КМУ від 8 лютого 2017 р. № 142-р, Програмою діяльності Кабінету Міністрів України, схваленою Постановою Верховної Ради України від 19 квітня 2016 року № 1099-19 та іншими національними програмно-стратегічними документами визначено необхідність збільшення інвестицій в реальний сектор економіки України [43]. Важливість цієї стратегічної мети в Україні обумовлена незадовільним станом капіталу, який є основою для зростання ВВП і підвищення ефективності економіки. Інформаційне видання Bloomberg у 2016 р. проаналізували ріст економіки України та дійшли висновку, що основним фактором росту ВВП став ріст капітальних інвестицій. Тому моніторинг стану елементів капіталу, виявлення проблем у сфері відтворення та обґрунтування напрямів розвитку інвестицій у капітал в Україні має важливе науково-практичне завдання. Забезпечення ефективного процесу формування та реалізації інвестиційної політики держави

вимагає відповідного професійного управління інвестиційними ресурсами. При розробці національної інвестиційної стратегії необхідно правильно прогнозувати кількість інвестиційних ресурсів, що надходять на інвестиційний ринок.

Поряд з поясненнями дослідників К. Паливода [35], О. Білоусова [3], А. Загородній та Ю. Стадницький [11], І. М. Крейдич [14], Й. Ридаровська [4] П. Кузьмович [25], Т. Остапчук [31], А. Пересада [4], Т. Радевич [46] про склад капітальних інвестицій, зазначення ознак їх класифікації і методів стимулювання, автори Ю.В. Подмешальська і С.В. Мельник [4], а також О.Ф. Ярмолюк та О.М. Дмитренко[32] підкреслюють необхідність уточнення значення категорії «капітальні інвестиції», щоб можна було поліпшити їх облік. Згодом такі автори як Т.В. Майорова, В.В. Крук та Я.В. Шевчук [4] представили нові методи визначення капітальних інвестицій, їх форм і проблем, пов'язаних з їх реалізацією, а також нові підходи до визначення їх змісту .

Предметом дослідження є економіко-математичні методи моделювання та прогнозування інвестицій.

Метою дослідження є систематизація та оцінка для порівняння якості методології розробки прогнозу інвестицій в основний капітал в рамках прогнозування соціально-економічного розвитку України.

Для досягнення поставленої мети потрібно виконати ряд завдань:

1. Здійснити аналіз літератури з метою узагальнення досвіду моделювання та прогнозування інвестицій в основний капітал і визначити найбільш придатний метод;
2. Порівняти різні методи машинного та альтернативного прогнозування.
3. Проаналізувати тенденцію і перспективу інвестування в основний капітал в Україні.



4. Зібрати макроекономічні дані для прогнозування капітальних інвестицій.
5. Здійснити оцінку прогностичної моделі темпів зростання капітальних інвестицій в Україні.

Інформаційною базою дослідження є наукові результати, опубліковані в монографіях, наукових журналах економічного і математичного напрямів, представлені в мережі Інтернет.

У процесі дослідження використані наступні наукові методи: аналіз літератури, методи статистичного прогнозування, графічний метод і порівняння. Для практичної реалізації прогнозу обрані наступні моделі: Ridge, LASSO, Post-Lasso, Adaptive LASSO, Elastic Net, Spike і Slab, AR, boost, часові ряди.

Практична значимість роботи полягає в можливості використання результатів дослідження, зокрема, запропонованого прогнозу для прогнозування інвестицій в основний капітал господарюючими суб'єктами.

## РОЗДІЛ 1

# ТЕОРЕТИКО-МЕТОДИЧНІ ОСНОВИ КОРОТКОСТРОКОВОЇ ОЦІНКИ ТА ПРОГНОЗУ МАКРОЕКОНОМІЧНИХ ПОКАЗНИКІВ

### 1.1 Аналіз актуальних публікацій щодо прогнозування інвестицій

Прогнозування обсягу інвестиції в основні засоби різних галузей промисловості протягом багатьох років є важливою темою досліджень. У науковому співтоваристві було створено і протестовано велику кількість статистичних моделей. Ернест Таннер представив результати прогнозування інвестицій у будівництво нежитлових приміщень у Канаді з використанням моделі розподіленої затримки [86]. Серед різних статистичних моделей були широко розроблені регресійні моделі для різних сценаріїв, включаючи попит на основні засоби для промислових цілей в США [94], попит на нерухомість в Таїланді [112] і інвестиції приватного сектора в британську промисловість [59]. Раймонд Ц. використовував дво-та тріступеневий метод найменших квадратів для вивчення впливу інвестицій та необхідності інвестицій у житлову нерухомість Гонконгу [114]. М. Скитмор, Кеунг Ф. В. розробили комбіновану генетичну модель лінійної регресії для попиту на нерухомість у Гонконзі [105]. Тео Хо Пін та Хо Пі Хуа використовували методи ARIMA для прогнозування попиту на будівництво у Сінгапурі [83]. Підходи структурного моделювання ARIMA також використовувались для прогнозування початку будівництва житла у Флориді, США [76].

Слід більш детально описати роботи авторів Шумін Ц., Чень Я., Цзінтао Г. та Чжанвен Д., [110] в якій розглядаються питання прогнозування у вугільній промисловості Китаю. Модель ARIMA була використана в дослідженні для прогнозування споживання вугілля, цін та інвестицій із 2016 по 2030 р.. На основі результатів прогнозування були зіставлені тенденції в еволюції

прогнозованих змінних. Автор віддає перевагу цьому методу з безлічі інших методів через доступність інформації про часові ряди [119]. При цьому перевага ARIMA перед іншими аналогічними методами полягає в можливості визначити найбільш придатну модель для відповідного часового ряду. Цей метод ефективний та надійний для прогнозування інвестицій, завдяки своїй гнучкості та впорядкованому пошуку відповідної моделі на кожному етапі .

У дослідженні А. Ш. Мохсена регресійні методи використовувалися для моделювання Йорданських приватних інвестицій в основний капітал будівельної галузі [60]. У статті розглядаються приватні інвестиції в будівництво в Йорданії, щоб сформулювати більш повне уявлення про структуру галузі, а також про її майбутній розвиток. Це дослідження також має на меті розглянути стаціонарність даних, що використовуються за допомогою процедур коінтеграції та диференціальної регресії, щоб гарантувати, що остаточно встановлена модель є статистично надійною та придатною для більш широкого застосування в аналогічних дослідженнях, а також для практичного використання.

Моделі множинної регресії також використовуються в більш недавньому дослідженні С. Томаса Нг. [104] при прогнозуванні приватних інвестицій в будівництво в Гонконзі. Крім того автор розглянув моделі VEC (векторне виправлення помилок), було поставлено завдання визначити найкращі моделі з точки зору їх точності і здатності обробляти дані нестационарних часових рядів. Автор дійшов висновку, що при обробці нестационарних даних модель VEC вважається більш точною та надійною. У той же час встановлено, що результати інвестиційного прогнозування чутливі до загальних економічних умов (валовому внутрішньому продукту, рівню безробіття тощо) та в умовах економічної нестабільності для більш точних прогнозів необхідно застосовувати моделі множинної регресії. Наприклад, валовий внутрішній продукт може відображати здатність інвесторів платити за будівельні проекти, тоді як рівень безробіття використовується як непрямий показник готовності кінцевих

користувачів купувати будівельні об'єкти [67]. Автор визнає, що прогнозування, у якому використовується регресійний аналіз, може дати більш точні результати. У той же час реалізація застосування такого підходу (особливо для розрахунку прогнозу на кілька періодів вперед) утруднена з практичної точки зору, оскільки для обчислення прогнозних значень залежної змінної часто необхідно спрогнозувати майбутні значення одного або декількох факторів.

Слід зазначити, що моделі VEC застосовувалися для прогнозування різних економічних показників, таких як безробіття [67], відсоткових ставок [112], обмінний курс [115], фондові ринки [100] і ціни на житло [66]. У той же час ці моделі рідко використовувалися для прогнозування інвестиції в основний капітал. У свою чергу, Тео Хо Пін та Хо Пі Хуа висловили думку, що метод VEC може стати перспективним інструментом для прогнозування інвестицій в нерухомість. Було запропоновано порівняти метод VEC з традиційними моделями, щоб встановити новий орієнтир для прогнозування попиту на будівництво. У дослідженні була використана модель і проведено порівняння VEC та традиційна модель множинної регресії з різними даними (для обробки нестационарних даних).

На думку деяких авторів, основним принципом прогнозування є простота його реалізації. У той же час простота прогнозування полягає в простоті використовуваних методів, наукових позицій, взаємозв'язків в моделях і між ними, прогнозів і рішень, отриманих для розуміння всіх етапів прогнозування особами, які приймають рішення. Однак, простота використання кількісних методів зменшується, а складність збільшує помилки прогнозування. Об'єднання прогнозів з використанням різних методів і моделей зменшує помилки прогнозу за межами вибірки, якщо методи і моделі надійні [84]. Наприклад, для прогнозування ринку інвестицій в житло Федеральна резервна система США часто використовує комбіновані прогнози, в тому числі на довгострокову перспективу, які показують зниження помилок прогнозування.

Для реалізації таких прогнозів використовуються щоквартальні середньомісячні дані, на основі яких формуються узгоджені (медіанні) прогнози зростання реальних інвестицій у житло з урахуванням результатів опитування (не менше 50 питань) по домогосподарствам (не менше 500 домогосподарств). Важливо відзначити, що прогнози Федеральної резервної системи США, а також приватні прогнози зростання інвестицій в житло мають високий ступінь точності [63].

Огляд показав, що в академічних колах використовується широкий спектр методів для прогнозування інвестицій в основний капітал. Однак, щоб мінімізувати розриви між моделюванням та реальністю, необхідно вивчити більш інтегровані моделі факторної оптимізації, які характеризують ширший діапазон невизначеностей. Зокрема, встановлено, що дослідження в цій галузі, як правило, передбачають використання спрощених моделей та припущень, які відокремлюють відповідні підходи від реальності і, таким чином, не розкривають реальних проблем перенесення моделювання в реальний світ [101].

## **1.2 Розгляд методів, що використовуються машинним навчанням та альтернативних методів для прогнозування інвестицій**

Машинне навчання важливе, оскільки воно дає підприємствам уявлення про тенденції поведінки клієнтів і бізнес-операційні моделі, а також підтримує розробку нових продуктів. Машинне навчання-це тип штучного інтелекту який надає програмам здатності вчитися та вирішувати проблеми без явного програмування. Далі піде мова про основні проблеми, які можна вирішити за допомогою машинного навчання при прогнозуванні, а також практичні рішення машинного навчання для їх вирішення.

Так само, як люди навчаються щодня, щоб дозволити машині вчитися, потрібно надати їй достатньо даних. Як тільки оброблюються дані, вона може

робити прогнози щодо майбутнього. Припускаючи, що ви хочете класифікувати електронні листи за тим, чи є вони спамом чи ні. Щоб вирішити цю проблему за допомогою машинного навчання, вам потрібно надати машині безліч позначених електронних листів, які вже класифіковані за правильними класами спаму та не спаму. Класифікатор проведе ітерацію за зразками і дізнається, які особливості визначають спам-лист. Якщо припустити, що модель машинного навчання була правильно навчена, то вона зможе з високою точністю передбачити, чи слід класифікувати майбутній електронний лист як спам чи ні. У багатьох випадках неможливо повністю зрозуміти, як модель передбачає клас.

Світ машинного навчання можна розділити на два типи проблем: контрольоване навчання та неконтрольоване навчання. У випадках, коли є немарковані дані, неконтрольоване навчання може бути правильним рішенням.

Під контрольованими навчальними проблемами є ще один поділ проблем регресії проти проблем класифікації. У задачах регресії значення, яке можна передбачити, є безперервним. Наприклад, вартість будинку. З іншого боку, у проблемах класифікації значення, є дискретним, як спам проти неспаму.

Дані, які потрібно надати для навчання моделі, залежать від проблеми та значення, яке потрібно передбачити. Припустимо, що потрібно передбачити ціну будинку на основі різних властивостей. Тож у цьому випадку кожен рядок у вашому наборі даних повинен (наприклад) складатися з:

- особливості: розмір будинку, кількість кімнат, поверх, наявність ліфта тощо.
- вартості: ціна будинку.

Вибір і збір характеристик, які найкраще описують будинок для прогнозування його ціни, може бути складним завданням. Це вимагає знання ринку, а також доступу до великих джерел даних. Характеристики - це ключі, на яких буде базуватися прогноз ціни на будинок.

Кожна проблема машинного навчання в основному є проблемою оптимізації . Тобто можна знайти або максимум, або мінімум певної функції. Функція, яку потрібно оптимізувати, зазвичай називається функцією втрат (або функцією витрат ). Функція втрат визначається для кожного алгоритму машинного навчання, який ви використовуєте, і це основний показник для оцінки точності вашої навченої моделі.

Існує безліч прикладів успішного використання машинного навчання при макроекономічному прогнозуванні. В області макроекономічного прогнозування дослідники Лі та Чен [91] досліджували можливості LASSO, Elastic Net і Group LASSO в порівнянні з динамічними факторними моделями . Дані у дослідженні включали 107 різних показників американської економіки з 1959 по 2008 р., і автори детально розглядали результати для 20 найважливіших показників. Було з'ясовано, що при прогнозах на один крок вперед методи регуляризації показували в середньому кращу якість ( Mean Squared Forecast Error), ніж DFM для 15, 18 і 19 змінних з 20 для LASSO, Elastic Net і Group LASSO. Крім того, комбінація методів регуляризації і DFM для кожного з 20 показників давала більш якісні прогнози, ніж DFM.

Для роботи з великою кількістю предикторів найбільш широко використовуються методи регуляризації. У дослідженнях і в літературі поширенні два види регуляризатора: LASSO або регуляризатор Р. Тібширані [85] і Ridge – регресія, також відома як регуляризатор А. Тихонова [112].

У роботі дослідників [58,85,91] була показана нестійкість LASSO у виборі змінних, яка обумовлена невизначеністю параметрів при оцінці коваріаційної матриці. Для вирішення цієї проблеми автори запропонували метод Elastic Net. Він узагальнює LASSO і Ridge. Головні відмінності цих двох методів полягають в наступному: LASSO дозволяє обнуляти коефіцієнти, а Ridge дає більш стабільні оцінки для висококорельованих змінних, в той час як оцінки LASSO можуть сильно помінятися при додаванні нових спостережень.

При застосуванні методу Elastic Net, з одного боку, можливе занулення коефіцієнтів (перевага LASSO), але з іншого боку – оцінки коефіцієнтів можуть виходити відносно стабільними (перевага Ridge).

Отже, проаналізувавши дослідження методів макроекономічного прогнозування можна дійти висновку, що найкращу якість показують ансамблеві методи- випадкового блукання і модель авторегресії порядку  $p$  AR ( $p$ ), що лідирують не тільки над простими еталонними моделями, але і над методами регуляризації. Відносно невисока якість прогнозів методів регуляризації. За результатами порівняння прогнозів можна побачити, що деякі з моделей машинного навчання дають значно якісніші передбачення короткострокової зміни інвестицій.

### 1.2.1 Методи регуляризації

Методи регуляризації за допомогою лінійної регресії, характеризується зокрема, двома моделями лінійної регресії — LASSO і Ridge [107].

Оскільки, використовується термін лінійний, тому спочатку потрібно уточнити, що лінійні моделі є одним із найпростіших способів прогнозування результатів за допомогою лінійної функції вхідних характеристик.

$$\hat{y} = w[0] \times x[0] + w[1] \times x[1] + \dots + w[n] \times x[n] + b \quad (1.1)$$

У наведеному вище рівнянні (1.1) ми показали лінійну модель на основі  $n$  числа ознак. Розглядаючи лише один об'єкт, оскільки  $w[0]$  буде нахилом, а  $b$  буде представляти перехоплення. Лінійна регресія [105] шукає шляхи оптимізації  $w$  і  $b$  таким чином, щоб мінімізувати функцію витрат. Рівняння функції витрат можна записати так:

$$\sum_{i=1}^M (y_i - \hat{y})^2 = \sum_{j=0}^p (w_j \times x_{ij})^2 \quad (1.2)$$



Як тільки ми використовуємо лінійну регресію для набору даних, розділеного на навчальний і тестовий набори, обчислення балів за того, хто навчає, і тестового наборів може дати нам приблизне уявлення про те, страждає модель від надмірного просіювання даних або недостатнього просіювання даних. Якщо дуже мало функцій у наборі даних, а оцінка низька як для навчального, так і для тестового набору, то це проблема недостатньої придатності. З іншого боку, якщо є велика кількість функцій, а оцінка тесту відносно нижча, ніж оцінка навчання, то це проблема надмірного узагальнення[91].

Ridge та LASSO -це методи зменшення складності моделі та запобігання надмірному просіюванні даних, яка може виникнути внаслідок простої лінійної регресії [58].

Ridge: метод являється удосконаленням лінійної регресії з підвищеною стійкістю до помилок, що накладає обмеження на коефіцієнти регресії для отримання куди більш наближеного до реальності результату. До того ж, цей результат набагато простіше інтерпретувати. Застосовується метод для боротьби з надмірністю даних, коли незалежні змінні корелюють один з одним (мультиколінеарність). В цілому метод забезпечує підвищення ефективності в задачах оцінки параметрів в обмін на допустиму кількість зміщення [91].

Регресія Ridge дозволяє зменшувати дисперсію оцінок коефіцієнтів, проте робить їх зміщеними. Формально формула виглядає наступним чином:

$$\beta^{Ridge} \in \arg \min_{\beta} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_2^2 \quad (1.3)$$

До стандартного функціоналу найменших квадратів додається функція, яка складається з-норми вектора  $\beta$ , помноженої на параметр  $\lambda$ . З ростом  $\lambda$  оцінки коефіцієнтів стають все ближче до 0, а при  $\lambda = 0$  завдання зводиться до звичайного методу найменших квадратів. Важливою властивістю такого підходу є наявність аналітичного рішення:

$$\beta^{Ridge} = (X'X + \lambda I_p)^{-1} X'Y \quad (1.4)$$

У разі якщо  $\text{rank}(X) < p$ , то метод найменших квадратів не має рішення, так як матриця  $X'X$  незворотна. Однак рішення існує для моделі Ridge у випадку, якщо  $\lambda \neq 0$ .

Недоліком цього методу є те, що він не може відбирати ненульові коефіцієнти моделі.

У регресії функція витрат змінюється додаванням штрафу, еквівалентного квадрату величини коефіцієнтів [87].

Це еквівалентно мінімізації функції витрат у рівнянні 1.2 за наведеної нижче умови:

$$\text{For some } c > 0, \sum_{j=0}^p w_j^2 < c \quad (1.5)$$

Таким чином, модель Ridge накладає обмеження на коефіцієнти ( $w$ ). Штрафний термін (лямбда) регулює коефіцієнти таким чином, що якщо коефіцієнти приймають великі значення, функція оптимізації штрафується. Отже, модель Ridge зменшує коефіцієнти, що допомагає зменшити складність моделі та мультиколінеарність. Повертаючись до екв. 1.3 можна побачити, що при  $\lambda \rightarrow 0$  функція вартості стає подібною до функції вартості лінійної регресії (рівняння 1.2). Отже, якщо зменшити обмеження (низький  $\lambda$ ) для функцій, модель буде нагадувати модель лінійної регресії [91].

LASSO: характеризує функцію вартості для регресії LASSO (найменшого абсолютного скорочення та оператора вибору) можна записати як:

$$\sum_{i=1}^M (y - \hat{y}_i)^2 = \sum_{j=0}^p (w_j \times x_{ij}) + \lambda \sum_{j=0}^p |w_j| \quad (1.6)$$

Так само, як функція вартості регресії, для лямбда = 0 наведене вище рівняння зводиться до рівняння 1.2. Єдина відмінність полягає в тому, що замість квадрата коефіцієнтів враховуються величини. Цей тип регуляризації

(L1) може призвести до нульових коефіцієнтів, тобто деякі функції повністю ігноруються для оцінки результатів. Таким чином, регресія LASSO не тільки допомагає зменшити перенавчання, але й може допомогти нам у виборі функцій. Подібно до Ridge, параметром регуляризації (лямбда) можна керувати

Метод LASSO був популяризований після роботи Р. Тібширані [85], однак і до цього зустрічався в літературі. Оцінка у дослідженні на основі LASSO виглядає наступним чином:

$$\beta^{Ridge} \in \arg \min_{\lambda} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1. \quad (1.7)$$

У статистиці та машинному навчанні LASSO - це метод регресійного аналізу, який виконує як вибір змін, так і регуляризацію з метою підвищення точності прогнозування та інтерпретації статистичної моделі, яку він виробляє. Використання LASSO на відміну від Ridge дозволяє вибирати із загального набору змінних лише кілька найбільш важливих і відкидати інші. При цьому аналітичного рішення моделі не існує. Методи регуляризації дозволяють в явному вигляді отримувати оцінки коефіцієнтів при предикторах.

Поки, що ми ознайомилися з основами регресії Ridge та LASSO та побачили кілька прикладів, щоб зрозуміти застосування. Тепер варто пояснити, чому регресія LASSO може призвести до вибору ознак, а регресія Ridge лише зменшує коефіцієнти, близькі до нуля, але не зменшує нуль [91]. Ілюстративний малюнок нижче допоможе краще зрозуміти, де ми припустимо гіпотетичний набір даних лише з двома характеристиками.

Використовуючи обмеження для коефіцієнтів регресії Ridge та LASSO (як показано вище у формулах), зображено для прикладу рисунок 3.1.

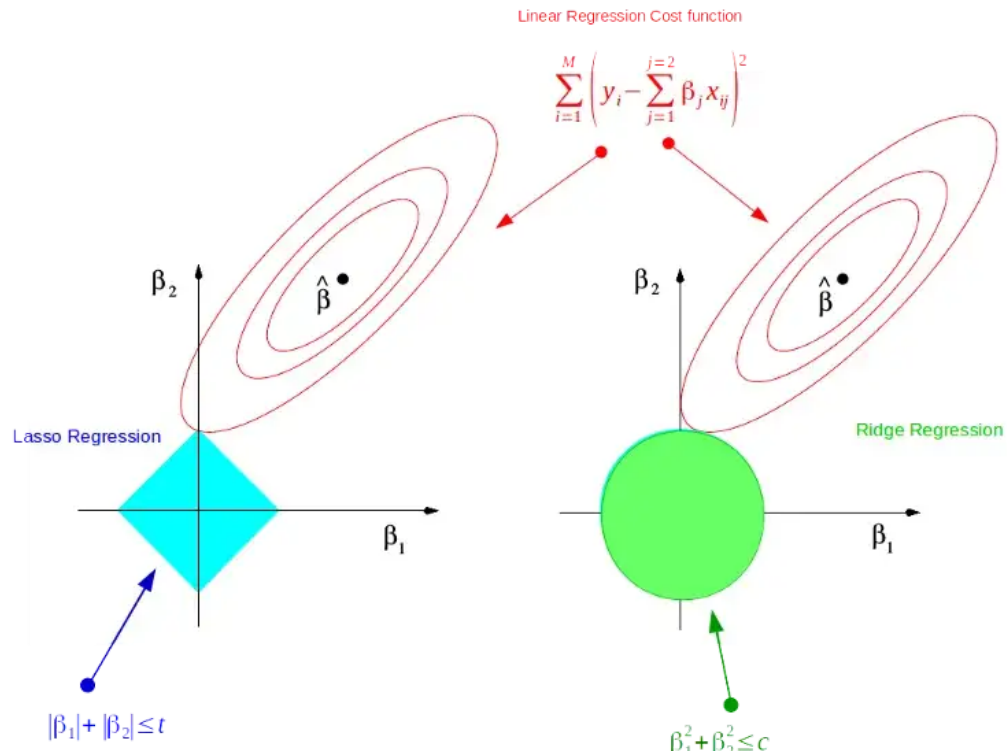


Рис. 1.1. Зменшення розмірності простору функцій

\*Джерело [88]

Для двовимірного простору об'єктів області обмежень нанесені для регресії LASSO і Ridge блакитним і зеленим кольорами. Еліптичні контури являють собою функцію витрат лінійної регресії (рівняння 1.2). Тепер, якщо є пом'якшені умови для коефіцієнтів, то обмежені області можуть збільшуватися, і в кінцевому підсумку вони потрапляють в центр еліпса. Це той випадок, коли регресія Ridge і LASSO нагадує результати лінійної регресії. В іншому випадку обидва методи визначають коефіцієнти шляхом знаходження першої точки, де еліптичні контури потрапляють в область обмежень. Ромб (LASSO) має кути на осях, і щоразу, коли еліптична область потрапляє в таку точку, одна з особливостей повністю зникає. Для просторових об'єктів більш високої розмірності на осі з регресією LASSO може бути багато рішень, і, таким чином, ми отримуємо тільки вибрані важливі об'єкти [58].

Комбінуючи моделі LASSO та Ridge, ми отримуємо регресію Elastic-Net . Elastic-Net Regression групує та звужує параметри, пов'язані з корельованими змінними, і залишає їх у рівнянні або видаляє їх усі одразу.

В силу неточності вимірювань даних або будь-яких ще помилок з метою побудови найкращої моделі вводять регуляризатор або кілька регуляризаторів. Тоді виходить наступне завдання оптимізації:

$$(1.8) \quad \|Xw - y\|^2 + z_1 \|w\| + z_2 \|w\|^2$$

Даний вид регресії створює умови для групового ефекту при високій кореляції змінних, а не обнуляє деякі з них, як метод LASSO.

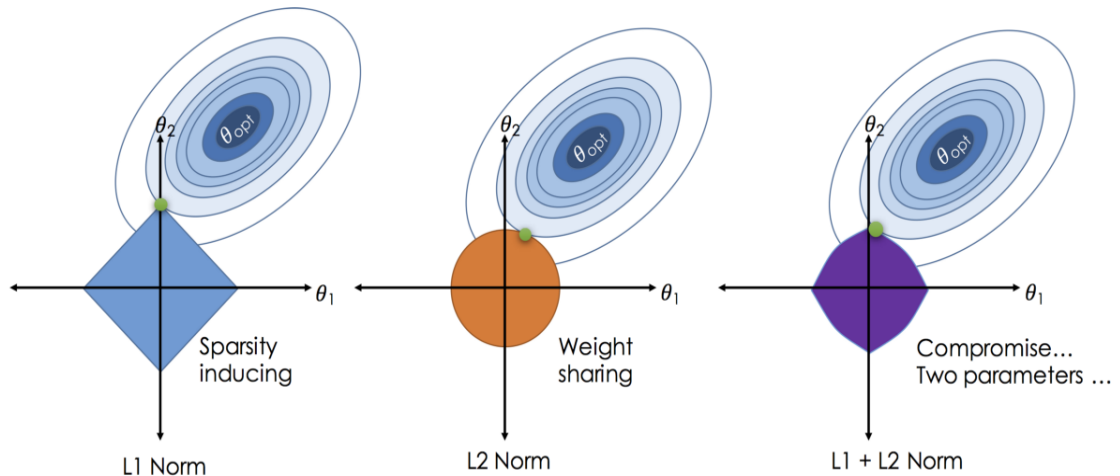


Рис. 1.2. Порівняння між регресією LASSO, Ridge та Elastic Net

\*Джерело [88]

Також немає обмежень за кількістю обраних змінних. На рис. 1.2 відображено порівняння LASSO, Ridge та Elastic Net.

### 1.2.1 Ансамблеві методи

На відміну від багатьох моделей машинного навчання, які зосереджені на високоякісному передбаченні, виконаному однією моделлю, алгоритми Boosting прагнуть покращити потужність передбачення шляхом послідовного навчання моделей (рисунк 1.3) , кожна з яких компенсує недоліки своїх попередників.

Щоб зрозуміти Boosting [105], важливо визнати, що Boosting — це загальний алгоритм, а не конкретна модель . Boosting вимагає, щоб вказали слабку модель (наприклад, Ridge, Lasso, тощо), а потім покращили її.

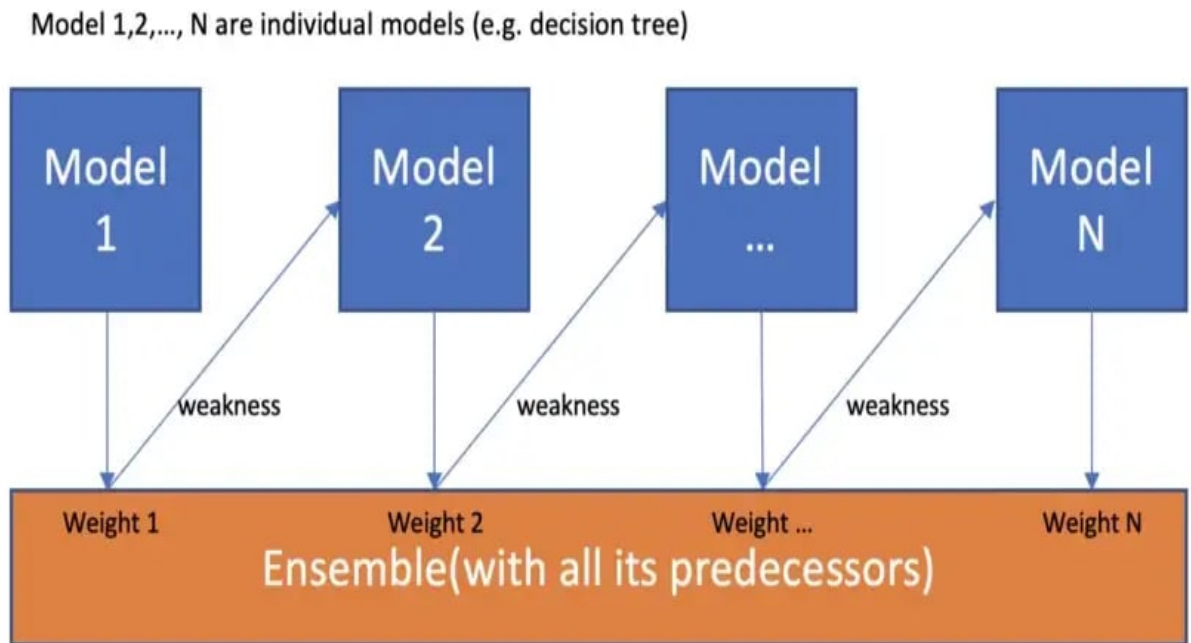


Рис.1.3. Послідовний процес навчання моделей

Тобто, бустинг насправді є загальним методом побудови методів, тому в якості базової можна використовувати практично будь-який тип моделі. Наприклад, випадковий ліс використовує дерево рішень як базову. Ми також використовуємо цей підхід у роботі. Навчання розширювальної моделі передбачає спочатку навчання базової моделі на всій навчальній вибірці, потім розгляд помилок моделі та навчання нової моделі на них. Формула для здійснення обрахунку має такий вигляд:

$$M = M + \eta m. \quad (1.9)$$

Параметр  $\eta$  контролює швидкість навчання, тобто коли швидкість навчання висока, модель має тенденцію до перенавчання, тоді як коли вона низька, для отримання хороших прогнозів потрібно більше ітерацій. Тому у результаті обрахунку змін для  $N$  отримаємо ось таку формулу для моделі:

$$Mn = M + \sum_{i=\eta i - m_i}^N \quad (1.10)$$

Використання методу boost у дослідженні складається із 4 етапів:

- Крок 1. Ініціалізування кореня.
- Крок 2. Обчислення залишкової величини, щоб вмістити у дерево рішень.
- Крок 3. Оновлення коефіцієнтів.

Щоб поєднати передбачення початкового кореня та дерева рішень, яке побудували, потрібно перетворити передбачені значення в дереві рішень таким чином, щоб вони відповідали логарифму (коефіцієнтам) у початковому корені.

- Крок 4. Повторення.

Використовуючи коефіцієнти, ми повторили крок 2 і крок 3. Отримали новий залишок і дерево рішень, створене для його прогнозування.

- Крок 5. Прогноз

Резюмуючи вищенаведене модель Boosting можна охарактеризувати так:

- легко інтерпретувати: boosting – це, по суті, ансамблева модель, тому її прогноз легко інтерпретувати
- сильна сила передбачення: зазвичай boosting > пакетування (випадкове передбачення) > дерево рішень.
- чутливий до викидів: оскільки кожен слабкий класифікатор присвячений виправленню недоліків своїх попередників, модель може приділяти занадто багато уваги викидам.
- важко розширити: оскільки кожен оцінювач побудований на своїх попередниках, процес важко розпаралелювати.

Велика частина машинного навчання — це класифікація — ми хочемо знати, до якого класу чи групи належить спостереження. Здатність точно класифікувати спостереження є надзвичайно цінною для різноманітних бізнес-застосунків, як-от передбачення того, чи придбає конкретний користувач продукт, або прогнозування того, чи буде непогашена дана позика чи ні.

Наука про дані надає безліч алгоритмів класифікації, таких як логістична регресія, метод опорних векторів, простий класифікатор Байєса та дерева рішень.

Випадковий ліс — це популярний алгоритм машинного навчання, який належить до методики навчання під наглядом. Його можна використовувати як для задач класифікації, так і для регресії в ML. Він заснований на концепції ансамблевого навчання, яке є процесом поєднання кількох класифікаторів для вирішення складної проблеми та покращення продуктивності моделі.

Як впливає з назви, « випадковий ліс» — це класифікатор, який містить кілька дерев рішень на різних підмножинах даного набору даних і бере середнє значення для підвищення точності прогнозування цього набору даних. Замість того, щоб покладатися на одне дерево рішень, випадковий ліс бере прогнози з кожного дерева прогнозів і прогнозує кінцевий результат.

Використання методу `randomforest` у дослідженні складається із 4 етапів:

- Крок 1. Створення початкових даних
- Крок 2. Побудування дерева рішень
- Крок 3. Повторне тестування.
- Крок 4. Отримання прогнозу, агрегування та оцінка

Під час навчання кожне дерево у випадковому лісі навчається з випадкової вибірки точок даних. Більша кількість дерев підвищує точність і запобігає проблемі перенавчання.

Зразки малюються із заміною , відомою як початкове завантаження, що означає, що деякі зразки використовуватимуться кілька разів в одному дереві. Ідея полягає в тому, що, навчаючи кожне дерево на різних зразках, хоча кожне



дерево може мати високу дисперсію щодо певного набору навчальних даних, загалом весь ліс матиме меншу дисперсію, але не за рахунок збільшення зміщення.

Отже, мета будь-якої проблеми машинного навчання полягає в тому, щоб знайти єдину модель, яка найкраще передбачить бажаний результат. Замість того, щоб створювати одну модель і сподіватися, що ця модель є найкращим/найточнішим прогнозом, який ми можемо зробити, ансамблеві методи враховують безліч моделей і усереднюють ці моделі для отримання однієї остаточної моделі.

### 1.2.3 Альтернативні методи

Також для прогнозування інвестицій використовують альтернативні моделі, які не відносяться до машинного навчання.

В якості базової моделі використовується проста модель випадкового блукання. Випадкове блукання - простий і зрозумілий індикатор якості, часто використовуваний в прикладних економічних дослідженнях. Формально прогноз на  $h$  кроків вперед записується наступним чином:

$$\hat{y}_{t+h} = y_t \quad (1.11)$$

Якщо існує деяка кореляція між значеннями в певному часовому ряді та тими, що з'являються до і після, базова модель, яка називається авторегресією, може оцінити майбутні показники на основі минулих показників (вперед і назад).

Регресія: метод прогнозування безперервної оцінки елемента на основі вхідних факторів.

Авто: прогнозує майбутні значення, використовуючи його історичну важливість.

Модель лінійної регресії, що включає змінні з відставанням як вхідні дані, є моделлю AR.

Рівняння моделі AR:

$$Y_t = C + b_1 Y_{t-1} + b_2 Y_{t-2} + \dots + b_p Y_{t-p} + E_{rt} \quad (1.12)$$

Авторегресія 1-го порядку:

$$Y_t = C_1 Y_{t-1} + C_2 \quad (1.13)$$

(Авторегресія 2-го порядку):

$$Y_t = C_1 + C_2 Y_{t-1} + C_3 Y_{t-2} \quad (1.14)$$

Ключові параметри:

- $p$  - минулі значення;
- $Y_t$  - функція різних минулих значень;
- $E_{rt}$  - помилки в часі;
- $C$  – перехоплення.

### 1.3 Аналіз методів для порівняння якості моделей при прогнозуванні

Важливим кроком у будь-якій моделі машинного навчання є оцінка точності моделі. Середньо-квадратична різниця, середня абсолютна помилка, корінь середньо-квадратичної різниці та R-квадрат або коефіцієнт детермінації використовуються для оцінки ефективності моделі [106].

Середня абсолютна похибка являє собою середнє значення абсолютної різниці між фактичними та прогнозованими значеннями в наборі даних. Вона вимірює середнє значення залишків у наборі даних.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}| \quad (1.14)$$

Середньо-квадратична різниця представляє середнє значення квадрата різниці між вихідним і прогнозованим значеннями в наборі даних. Він вимірює дисперсію залишків.

Корінь середньоквадратичної різниці— це квадратний корінь середньоквадратичної різниці. Він вимірює стандартне відхилення залишків.

Коефіцієнт детермінації або R-квадрат представляє частку дисперсії залежної змінної, яка пояснюється моделлю лінійної регресії. Це безмасштабна оцінка, тобто незалежно від того, малі чи великі значення, значення квадрата R буде менше одиниці [101].

Скоригований R-квадрат є модифікованою версією R-квадрата, і його скориговано на кількість незалежних змінних у моделі, і воно завжди буде менше або дорівнює R<sup>2</sup>.

Середньоквадратична різниця (MSE) і корінь середньоквадратичної (RMSE) різниці компенсують великі помилки передбачення порівняно з середньою абсолютною похибкою (MAE). Однак RMSE широко використовується, ніж MSE, для оцінки продуктивності регресійної моделі з іншими випадковими моделями, оскільки вона має ті самі одиниці, що й залежна змінна (вісь Y). Формула розрахунку RMSE має такий вигляд:

$$RSME = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2} \quad (1.15)$$

де,

$\sum$  - є сумою всіх значень;

f – прогнозоване значення;

o - спостерігається або фактичне значення;

$(\hat{f}_i - o_i)^2$  — це різниці між прогнозованими та спостережуваними значеннями та зведені в квадрат;

$N$  – загальний розмір вибірки.

MSE — це диференційована функція, яка полегшує виконання математичних операцій порівняно з недиференційованою функцією, як от MAE [71]. Тому в багатьох моделях RMSE використовується як метрика за замовчуванням для обчислення функції втрат, незважаючи на те, що її важче інтерпретувати, ніж MAE.

Менше значення MAE, MSE та RMSE означає вищу точність регресійної моделі. Однак більш високе значення квадрата  $R$  вважається бажаним.

$R$  Squared, Adjusted  $R$  Squared використовуються для пояснення того, наскільки добре незалежні змінні в моделі лінійної регресії пояснюють мінливість залежної змінної. Значення  $R$  Squared завжди збільшується з додаванням незалежних змінних, що може призвести до додавання надлишкових змінних у моделі. Однак скоригований  $R$ -квадрат може вирішити цю проблему.

Скоригований  $R$  у квадраті враховує кількість прогнозованих змінних і використовується для визначення кількості незалежних змінних у моделі. Значення скоригованого  $R$ -квадрату зменшується, якщо збільшення  $R$ -квадрата додатковою змінною недостатньо значне.

Для порівняння точності між різними моделями лінійної регресії RMSE є кращим вибором, ніж  $R$  Squared.

І RMSE, і  $R$ -квадрат кількісно визначають, наскільки модель лінійної регресії відповідає набору даних. RMSE показує, наскільки добре модель може передбачити значення змінної відповіді в абсолютних виразах, тоді як  $R$ -квадрат показує, наскільки добре змінні предиктора можуть пояснити варіацію змінної відповіді [85].

Часто використовувана модель для порівняння якості при прогнозуванні часових рядів це модель авторегресії порядку  $p$   $AR(p)$ . Вона передбачає, що

зміна процесу в момент часу  $t$  повністю залежать від  $p$  його попередніх значень і випадкової помилки  $\varepsilon_t$  :

$$y_t = a_0 + a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t \quad (1.16)$$

На практиці вибір аналізу кількості часових рядів зазвичай проводиться за допомогою формаційних критеріїв-AIC (Akaike Information Criterion) або BIC (Bayesian Information Criterion) [6].

AIC — це єдине число, яке можна використовувати, щоб визначити, яка з багатьох моделей, швидше за все, буде найкращою моделлю для даного набору даних. Вона оцінює моделі відносно , тобто оцінки AIC корисні лише в порівнянні з іншими оцінками AIC для того самого набору даних. Чим нижча оцінка AIC тим краще [99].

AIC найчастіше використовується в ситуаціях, коли неможливо легко перевірити продуктивність моделі на наборі тестів у стандартній практиці машинного навчання (невеликі дані або часові ряди) . Як наслідок, навчання на всіх даних і використання AIC може призвести до покращеного вибору моделі в порівнянні з традиційними методами вибору моделі навчання/перевірки/випробування.

AIC працює, оцінюючи відповідність моделі навчальним даним і додаючи штрафні терміни за складність моделі (основні принципи подібні до регуляризації ). Бажаний результат полягає в тому, щоб знайти найнижчу можливу AIC, яка вказує на найкращий баланс відповідності моделі та можливості узагальнення. Це служить кінцевій меті максимізації відповідності даних поза вибіркою. Рівняння AIC [98]:

$$AIC = -2 \ln(L) + 2k \quad (1.16)$$

де,

$L$  = правдоподібність,

$k$  = кількість параметрів.

AIC використовує оцінку максимальної правдоподібності моделі (логарифм правдоподібності) як міру відповідності. Логарифм правдоподібності є показником того, наскільки ймовірно хтось побачить свої спостережувані дані на основі моделі. Модель з максимальною ймовірністю – це та, яка найкраще «підходить» до даних. Натуральний логарифм правдоподібності використовується для зручності обчислень.

AIC зазвичай використовується, коли не має доступу до даних поза вибіркою і можна вибрати між кількома різними типами моделей або для зручності часу.

Часто велика вибірка є достатньою для приблизної оцінки, але оскільки використання AIC означає, що у вас є невеликий розмір вибірки, існує формула з коригуванням розміру вибірки під назвою AIC<sub>c</sub>, яка додає коригуючий термін, який збігається з відповіддю AIC для великих вибірок, але дає більш точну відповідь для менших зразків [88].

Тобто оцінки AIC насправді є імовірнісним ранжуванням моделей, які, ймовірно, мінімізують втрату інформації (найкраще відповідають даним).

Для більшості обчислень, де є достатньо даних, найкращий (і найпростіший) спосіб точно перевірити продуктивність моделі – це використовувати тренування, перевірку та тестовий набір у належній практиці машинного навчання. Але якщо виникають обставини, коли це неможливо (з невеликими даними або аналізом часових рядів), AIC може бути кращим тестом.

Також відомим та часто використовуваним методом для порівняння точності моделей прогнозування є тест Diebold Mariano [71], також відомий як тест DM.

Для прикладу цей статистичний інструмент використовується припустимо коли ми маємо дві моделі прогнозування, модель A та модель B, які використовуються для прогнозування. Ми хочемо визначити, яка з цих моделей точніша. Один із способів зробити це — порівняти помилки прогнозу двох

моделей. Якщо помилки моделі А менші, ніж помилки моделі В, можна зробити висновок, що модель А точніша.

Значення визначаються за допомогою класичного парного t-критерію, який порівнює середні значення двох парних вибірок. До прикладу це можуть бути вибірки для визначення помилки прогнозу двох моделей. Тест DM обчислює тестову статистику, яка називається статистикою DM, яка вимірює стандартизовану різницю втрат двох моделей. Якщо статистика DM істотно відрізняється від нуля, можна зробити висновок, що одна модель точніша за іншу. Формально,

$$DM_{AB} = \frac{\bar{d}}{\sigma_{\bar{d}_{AB}}} \sim N(0, 1)$$

(1.17)

Підсумовуючи, тест Дібольда-Маріано є корисним інструментом для порівняння точності двох моделей прогнозування. Це дозволяє нам визначити, яка модель краща, якщо така є, і може допомогти нам приймати більш обґрунтовані рішення при виборі методу прогнозування.

Отже, було проведено аналіз методів оцінки якості моделей. Виконавши даний аналіз, можна дійти до висновку, що в даній дослідницькій роботі необхідно використовувати саме такі методи оцінки якості моделей прогнозування: RMSE, AIC, тест Diebold Mariano. Завдяки чому можна буде правильно обробити дані та виконати оцінку моделей прогнозування.

## РОЗДІЛ 2

# АНАЛІЗ ІНВЕСТИЦІЙНОЇ ДІЯЛЬНОСТІ В УКРАЇНІ ТА ОСНОВНИХ ЧИННИКІВ, ЩО ВПЛИВАЮТЬ НА ОБСЯГ КАПІТАЛЬНИХ ІНВЕСТИЦІЙ

### 2.1 Нормативно-правове регламентування обліку капітальних інвестицій

Регулювання інвестицій здійснюється виконавчою гілкою влада, яку представляє Кабінет Міністрів, Міністерства та Відомства. Тому розглянемо основну характеристику та документи, якими виконавча гілка влади регулює облік і аналіз КІ:

- Постанова Кабінету міністрів України Про затвердження Національного стандарту № 3 «Оцінка цілісних майнових комплексів» була прийнята 2006 року та жодного разу не піддавалася зміні. Публікації документа велись в таких урядових виданнях, як Офіційний вісник України та Урядовий кур'єр. Містить дев'ять термінів серед яких у Ст. 2 є визначення капітальних інвестицій (далі – КІ)[30].
- Постанова Кабінету міністрів України Про затвердження Національного стандарту № 3 «Оцінка цілісних майнових комплексів» застосовується під час проведення оцінки цілісного майнового комплексу суб'єкта господарювання суб'єктами оціночної діяльності, а також особами, які відповідно до законодавства виконують рецензування звітів про оцінку цілісного майнового комплексу [42].
- Наказ Державного комітету статистики України «Про затвердження Методики розрахунку індексу капітальних» [38] (далі – Методика розрахунку індексу капітальних інвестицій) був прийнятий 25.12.2009 р.



Методика розрахунку індексу КІ розроблена на виконання заходів Стратегії розвитку державної статистики на період до 2012 року з метою вдосконалення статистичної методології з КІ. Завданням методики є створення інструментарію для оцінки динаміки інвестиційних процесів в Україні. Методика складається з передмови, чотирьох розділів, переліку використаної літератури й окреслює принцип формування та порядок розрахунку індексу капітальних інвестицій. Методика призначена для застосування фахівцями органів державної статистики та іншими зацікавленими користувачами статистичної інформації. Ця методика підготовлена відповідно до теоретичних принципів статистики та міжнародної практики обліку капітальних інвестицій з урахуванням властивостей національної економіки. Індокси КІ можуть бути застосованими 12 для аналізу динаміки інвестиційних процесів, при реалізації макроекономічних розрахунків, у прогнозуванні розвитку секторів економіки України. Положення методики будуть поправлятися з новим розвитком статистики КІ [38].

- Методичні рекомендації з бухгалтерського обліку основних засобів: Наказ Міністерства фінансів України від 30.09.2003р. №561 (зі змінами і доповненнями) були прийняті Наказом Міністерства фінансів України від 30.09.2003 р. № 561. Ці методичні рекомендації можуть бути застосовані підприємствами, спілками та іншими юридичними особами, їх філіями та іншими підрозділами незалежно від форм власності і організаційно-правових форм (крім банків і бюджетних установ та підприємств, які відповідно до законодавства складають фінансову звітність за міжнародними стандартами фінансової звітності). Бухгалтерський облік основних засобів зобов'язаний забезпечити: своєчасне відображення на рахунках бухгалтерського обліку повної інформації про оцінку, переоцінку і надходження основних засобів, їх

внутрішнє переміщення (зі складу в цех (відділ, дільницю), з цеху в цех тощо), вибуття (реалізація, ліквідація, безоплатна передача, нестача, псування тощо); відображення даних про нарахування амортизації основних засобів за звітний період та їх зносу з початку корисного використання; визначення витрат, пов'язаних з підтриманням основних засобів у робочому стані (технічні огляди, поточне обслуговування), їх ремонтом і поліпшенням; визначення фінансових результатів від реалізації та іншого вибуття основних засобів; інформацією для складання звітності про основні засоби і капітальні інвестиції [39].

- Методичні рекомендації щодо заповнення форм фінансової звітності прийняті Наказом Міністерства фінансів України від 28.03.2013 р. №433. У цих Методичних рекомендаціях аналізується питання розкриття інформації за статтями балансу (звіту про фінансовий стан), звіту про фінансові результати (звіту про сукупний дохід), звіту про рух грошових коштів, звіту про власний капітал. Мета складання балансу є надання користувачам повної, правдивої та неупередженої інформації про фінансовий стан підприємства на звітну дату [19].

Нормативно-правовою базою передбачено регулювання капітальних такими основними актами: Податковий Кодекс України (далі – ПКУ) та Закон України «Про інвестиційну діяльність» [13].

У Податковому кодексі України, інвестиції в основний капітал називають «капітальними інвестиціями». У категоріальному апараті національних нормативних документів, призначених для використання в різних сферах державного управління та економічної діяльності, економічна сутність основного капіталу позначається терміном "необоротні активи", де економічна сутність тлумачиться по-різному.

Визначення, наведене у Положенні (стандарті) бухгалтерського обліку №1 «Загальні вимоги до фінансової звітності» (П(с)БО) [41], затвердженому

Наказом Міністерства фінансів України від 7 лютого 2013 р. №73, найбільшою мірою корелює з категоріальним апаратом, що використовується в Системі національних рахунків, розроблених під егідою ООН, Єврокомісії, ОЕСР, МВФ, Групи Світового банку. У зазначеному П(с)БО основний капітал відноситься до активів, розташованих в першому розділі — необоротні активи. Сюди входять такі: нематеріальні активи, незавершені капітальні інвестиції, основні фонди, Інвестиційна нерухомість, довгострокові біологічні активи, довгострокові фінансові інвестиції, Довгострокова дебіторська заборгованість та відстрочені податки.

Відповідно до Закону України «Про інвестиційну діяльність» під інвестиціями розуміють всі види майнових та інтелектуальних цінностей, що вкладаються в об'єкти підприємницької та інших видів діяльності, в результаті якої створюється прибуток (дохід) та/або досягається соціальний та екологічний ефект [13].

Інвестиційна діяльність за своєю суттю спрямована на забезпечення безперебійного функціонування всіх без винятку господарюючих суб'єктів. Тобто, це включення набору реальних дій по реалізації інвестицій. У загальній структурі інвестиційних ресурсів підприємства займають капітальні вкладення.

Згідно з правилами (стандартами) бухгалтерського обліку, всі інвестиції можна розділити на внутрішні і зовнішні, тобто іноземні і внутрішні, а внутрішні, в свою чергу, діляться на реальні (виробничі) і фінансові. Вкладаючи особисті заощадження або особисте майно в статутний капітал підприємства, засновник - фізична особа - інвестує в його діяльність. Вкладаючи частину отриманого прибутку в будь - якій формі: грошової, матеріальної або нематеріальної, в статутний капітал підприємства, засновник-юридична особа інвестує в інше створене ним підприємство. Те ж саме можна сказати і про акціонерів. Все це, з точки зору підприємства, є зовнішньою інвестицією. До внутрішнього інвестування належить придбання (створення, будівництво)

об'єктів основних засобів та інших необоротних активів з метою розширення виробничих і невиробничих можливостей підприємства. Також внутрішнім інвестуванням є періодичне і безперервне вкладення коштів підприємства в господарський оборот, тобто звичайна і, зокрема, операційна діяльність будь-якого підприємства починається з інвестування цього процесу. Причому воно в цьому розумінні відбувається постійно і безперервно, адже для отримання економічної вигоди (прибутку) потрібні вкладення в оборот. Отже, всі свої ресурси, які підприємство вкладає «в себе», можна назвати внутрішнім інвестуванням (внутрішніми вкладеннями).

Реальні інвестиції розглядаються як капітальні витрати, які включають відтворення основних засобів, створення інноваційних нематеріальних активів або збільшення запасів товарно-матеріальних запасів та інших об'єктів інвестування, які пов'язані із здійсненням операційної діяльності підприємства. Реальні інвестиції є основною формою реалізації загальної стратегії економічного розвитку підприємства. Тому що саме така форма інвестування дозволяє підприємству успішно виходити на нові товарні та регіональні ринки, забезпечувати постійне зростання ринкової вартості підприємства. За допомогою реальних інвестицій також вирішують завдання по збільшенню обсягів виробництва і реалізації продукції, розширенню асортименту продукції, що випускається і підвищенню її якості [25].

При фінансових інвестиціях активи утримуються підприємствами з метою збільшення прибутку за рахунок відсотків, дивідендів і т.д. цінні папери, включаючи акції, облігації, депозитні сертифікати, казначейські облігації та інші цінні папери, є прикладами фінансових вкладень, які можуть збільшити вартість капіталу або отримати інші вигоди для інвесторів.

В рамках капітальних вкладень кошти використовуються для відтворення основних фондів, розширення, реконструкції та модернізації підприємств і структур, впровадження технологічних досягнень у всій економіці, будівництва

житла, шкіл, лікарень та інших соціальних і культурних об'єктів, досліджень і проектування. В результаті капітальні вкладення спрямовані на збільшення основних фондів, інших невідчужуваних основних засобів, нематеріальних активів, а також їх відновлення і модернізацію, що призводить до збільшення економічної вигоди в майбутньому [13].

В результаті капітальних вкладень економіка нашої країни значно розвивається, оскільки вони дозволяють підприємствам відтворювати і оновлювати свої основні фонди, підвищувати якість товарів і послуг, створювати робочі місця, захищати навколишнє середовище, будувати оборонні об'єкти, забезпечувати охорону здоров'я, освіту і соціальне забезпечення..

Бюджети капітальних витрат використовуються для оцінки капітальних вкладень підприємств або для розрахунку запланованих капітальних витрат.

В рамках бюджету капітальних витрат включається наступна інформація: первісна вартість всіх основних засобів; сума амортизаційних відрахувань, які будуть нараховані протягом планованого періоду; сума амортизаційних відрахувань; вартість основних засобів і сума амортизаційних відрахувань, які компанія матиме в кінці планового періоду [13].

Важливо розрізняти витрати на капітальний ремонт та на капітальні інвестиції, оскільки капітальні інвестиції відновлюють та розширюють основні засоби, тоді як капітальний ремонт лише підтримує їх працездатність.

Класифікація капітальних вкладень необхідна для забезпечення того, щоб фінансова звітність організації точно відображала інформацію про них, а розподіл класифікаційних ознак призначений для надання користувачам структурованої інформації, яка може вплинути на управлінські рішення і привести до позитивних результатів в бізнесі[35].

Інвестиції представляють собою цілеспрямоване використання капіталу, класифікація якого наведена на рис.2.1.



Рис. 2.1. Класифікація інвестицій

\*Джерело: [24]

За цілями застосування інвестицій їх поділяють на реінвестиції (заміна застарілих засобів) і нетто-інвестиції (придбання нових активів, розширення виробництва тощо). За характером участі в інвестуванні виділяють прямі і непрямі інвестиції. Під прямими інвестиціями розуміють особисту участь інвестора у виборі об'єктів інвестування і вкладенні коштів. Пряме інвестування

здійснюється обізнаними інвесторами, що мають досить точну інформацію про об'єкт інвестування і детально ознайомлені з механізмом інвестування.

Непрямі інвестиції пов'язані з інвестуванням, що опосередковується інвестиційними або іншими фінансовими посередниками. Не всі інвестори мають достатню кваліфікацію для ефективного вибору об'єктів інвестування і подальшого управління ними[13].

Непрямі інвестиції пов'язані з інвестуванням, що опосередковується інвестиційними або іншими фінансовими посередниками. Деякі інвестори не володіють досвідом і знаннями для ефективного вибору об'єктів інвестування та управління ними, тому вони інвестують в цінні папери, випущені інвестиційними або іншими фінансовими посередниками. Такими цінними паперами можуть бути інвестиційні сертифікати інвестиційних фондів або компаній. Акумуляовані таким чином кошти посередники розміщують на свій розсуд – обирають найефективніші об'єкти інвестування, беруть участь в управлінні ними, а отримані прибутки розподіляють серед своїх клієнтів [38].

За терміном інвестування розрізняють короткострокові та довгострокові інвестиції. Перші – це вкладення капіталу на період не більше одного року (короткострокові депозитні внески, купівля короткострокових ощадних сертифікатів тощо). Термін «довгострокові інвестиції» зазвичай використовується для позначення капітальних вкладень тривалістю більше одного року, хоча інвестиційні компанії описують довгострокові інвестиції такими способами: до 2 років; від 2 до 3 років; від 3 до 5 років; і більше 5 років.

Інвестиції за формою власності інвесторів поділяють на: приватні, державні, іноземні і спільні. В економічній теорії та практиці застосовується термін «приватні» інвестиції до вкладення коштів, що здійснюють фізичні особи, а також підприємства недержавної форми власності, насамперед – колективні. Під державними інвестиціями розуміють вкладення, що здійснюють центральні або місцеві органи влади за рахунок бюджетів, позабюджетних

фондів і позикових коштів, а також державні підприємства й установи за рахунок власних і позикових коштів. Іноземні інвестиції пов'язані з вкладеннями, що здійснюють іноземні громадяни, юридичні особи і держава. Під спільними інвестиціями розуміють вкладення, що здійснюють резиденти та нерезиденти.

Залежно від форми власності інвесторів, інвестиції поділяються на приватні, державні, іноземні та спільні.

В економічній теорії і практиці як правило, «приватні» інвестиції - це інвестиції, що здійснюються приватними особами, а також недержавними підприємствами, в першу чергу колективними. Державні інвестиції включають інвестиції, що здійснюються центральними та місцевими органами влади та державними підприємствами та установами за рахунок бюджетів, позабюджетних фондів та позикових коштів, а також інвестиції, що здійснюються державними підприємствами та установами за рахунок власних та позикових коштів. Термін «іноземні інвестиції» охоплює інвестиції, що здійснюються іноземними фізичними особами, корпораціями та урядами.

Активи інвесторів поділяються на чотири категорії: приватні, державні, іноземні та загальні. Спільні інвестиції-це інвестиції, що здійснюються резидентами та нерезидентами. В економічній теорії та практиці "приватні" інвестиції відносяться до інвестицій, що здійснюються приватними особами, а також до інвестицій, що належать недержавним корпораціям, які в першу чергу є колективами[24].

Капітал може бути основним і оборотним. До оборотного капіталу відносять матеріали, сировину, запаси вже готової продукції. Вони витрачаються за один цикл виробництва, а їх вартість повністю закладається у вартість готового продукту. У той час як основний капітал включає в себе кошти, які використовуються при виробництві протягом декількох років, а їх вартість



переноситься у вартість продукції частинами. До основного капіталу відносять будівлі, обладнання, машини, споруди та інше.

Таким чином, в загальноекономічному сенсі капітальні вкладення- це витрати на покупку або оновлення обладнання, будівель, споруд, транспорту. Такі витрати здатні забезпечити розвиток компанії і зростання прибутку в майбутньому. Якщо розглядати дану категорію капітальні інвестиції з точки зору підприємства, то її слід інтерпретувати як капіталізовану вартість ресурсів (матеріальних, фінансових, трудових і т.), спрямованих на відтворення основних активів. А капітал-це майно компанії, яке вона може використовувати для отримання прибутку.

Капітал може бути основним і оборотним. До оборотного капіталу відносять матеріали, сировину, запаси вже готової продукції. Вони витрачаються за один цикл виробництва, а їх вартість повністю закладається у вартість готового продукту. У той час як основний капітал включає в себе кошти, які використовуються при виробництві протягом декількох років, а їх вартість переноситься у вартість продукції частинами. До основного капіталу відносять будівлі, обладнання, машини, споруди та інше.

## **2.2 Динаміка і структура інвестицій в основний капітал України**

Економісти, зокрема Д. Йоргенсон, Ф. Голлоп і Б. Фраумені, [69] які проводили дослідження зростання економіки в багатьох країнах, дійшли висновку, що значна частина економічного зростання пояснюється накопиченням капіталу. Тобто, основним результатом висновку дослідження є те, що накопичення капіталу та інновації є важливими факторами довгострокового зростання. Більша кількість інновацій стимулює накопичення капіталу за рахунок збільшення граничного продукту капіталу. А більше

накопичення капіталу стимулює інновації, збільшуючи прибуток, нарощенню економічного потенціалу підприємств, підвищенню конкурентоспроможності продукції підприємств, створює умови для виробництва нових видів продукції та виходу на нові ринки збуту, а відтак є основою зростання обсягів формування національного багатства.

В Україні останнім часом спостерігається явна позитивна тенденція до зростання капіталовкладень. Так, в 2019 році капітальні вкладення склали 624 млрд., грн. проти 578,7 млрд., грн. в 2018 р., а в порівнянні з 2014 р. - збільшилася в 2,8 рази (рис. 2).

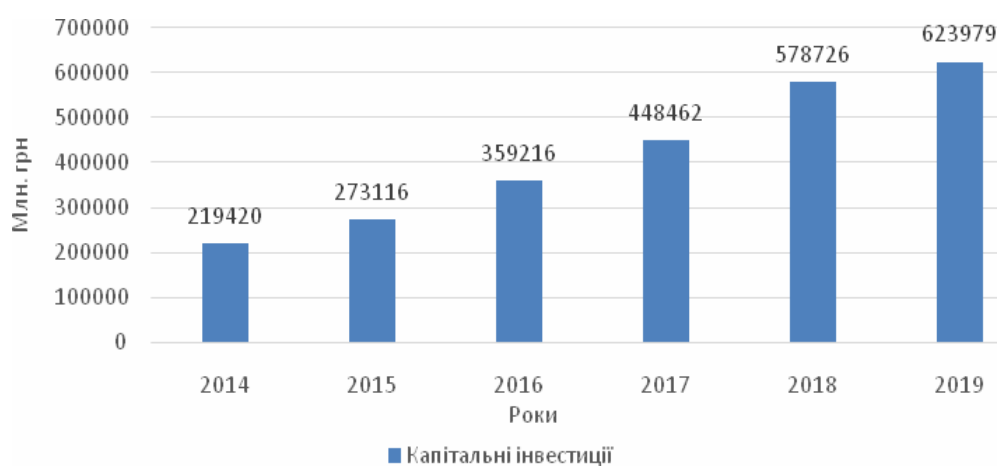


Рис. 2.2 Капітальні інвестиції в Україні в 2014-2019 роках

\*Джерело: [49,50, 51,52,53,54]

Для того щоб оцінити тенденцію в динаміці обсягів та ефективності капітального інвестування в Україні було проаналізовано співвідношення між сумою капітальних інвестицій та обсягами ВВП за 2010-2019рр. (табл. 1) Такий аналіз динаміки капітальних вкладень в Україні в 2010-2019 роках дозволяє зробити висновок, що до 2012 року їх обсяги росли, але темпи цього зростання знизилися. У 2013 році інвестиційні витрати скоротилися на 11,2%, що стало одним з факторів, що негативно вплинули на зміну ВВП в наступні роки. У 2014-2015 роках ВВП і капітальні вкладення скоротилися. Ця тенденція була в першу чергу обумовлена дестабілізацією політичної та військової ситуації на

сході України. Плюсом є те, що з 2016 року спостерігається певне зростання як ВВП, так і капітальних вкладень. Важливим показником ефективності капітальних вкладень є відношення їх суми до розміру валового внутрішнього продукту. У розвинених країнах значення цього показника в останні роки коливається від 20 до 46%.

**Таблиця 2.1**

Обсяги ВВП та капітальних інвестицій в Україні за 2010-2019 роки

Роки	Обсяги ВВП, млрд.дол	Обсяги капітальних інвестицій, млрд.дол.	Відношення обсягів капітальних інвестицій до ВВП, %	Темпи зростання обсягів ВВП, %	Темпи зростання обсягів капітальних інвестицій, %
2010	136	28,39	20,87	-	-
2011	163,2	36,61	22,44	20,0	29,0
2012	175,8	38,17	21,72	7,7	4,3
2013	183,3	33,89	18,49	4,3	-11,2
2014	133,5	17,88	13,40	-27,2	-47,2
2015	91	14,5	15,93	-31,8	-18,9
2016	93,3	20,28	21,72	2,5	39,9
2017	112,2	22,38	19,95	20,3	10,4
2018	130,8	24,56	18,77	16,6	9,7
2019	145,3	26,23	18,05	1,22	6,8

\*Джерело: складено авторами на основі [5, 24]

Аналізуючи інвестиційну активність, також було звернено увагу на структуру капітальних інвестицій. Найбільшу частку в структурі інвестицій в матеріальні активи зайняли інвестиції в машини, обладнання та витратні матеріали - 31,8%, далі йдуть інвестиції в інженерні споруди – 23,9%, нежитлові будівлі – 16,1%. Порівнюючи частку інвестицій в матеріальні основні засоби в структурі інвестицій в акціонерний капітал в 2020 році із 2019 роком можна сказати, що відбулося збільшення на 2,5 відсоткових пункти, а частка інвестицій в нематеріальні активи в питомому вираженні знизилася на 2,5

відсоткових пункти . Частка інвестицій в нежитлові будівлі в структурі капітальних вкладень за рік збільшилася на 0,7 відсоткових пункти, частка інвестицій в інженерні споруди збільшилася на 4,7 відсоткових пункти, частка інвестицій в транспортні засоби знизилася на 2,3 відсоткових пункти, інвестиції в права на товарні знаки, промислову власність, авторське право і суміжні права, патенти, ліцензії, концесії і т. д. протягом року скоротилися на 2,9 в. п. , таким чином слід звернути увагу на скорочення інвестицій в нематеріальні активи. Така динаміка дуже негативна, тому що в 21 столітті в розвинених країнах саме нематеріальні активи стають все більш привабливими об'єктами для інвестицій.

**Таблиця 2.2**

Динаміка капітальних інвестицій за видами економічної діяльності в Україні у 2014-2020 роках, млн. грн.

<b>Види економічної діяльності</b>	<b>2015</b>	<b>2016</b>	<b>2017</b>	<b>2018</b>	<b>2019</b>	<b>2020</b>
Сільське, лісове та рибне господарство, млн.грн	18796	30155	5084	64243	66104	59130
Промисловість, млн.грн	86242	87656	117754	14330	199896	254196
Будівництво, млн.грн	36057	43464	44444	52176	55994	62347
Оптова та роздрібна торгівля; ремонт автотранспортних засобів і мотоциклів, млн.грн	20716	20663	29957	33665	51818	44174
Транспорт, складське господарство, поштова та кур'єрська діяльність, млн.грн	15498	18704	25107	37944	50078	43793
Тимчасове розміщування й організація харчування, млн.грн	1482	1393	1478	2134	2675	2832
Інформація та телекомунікації, млн.грн	8175	22975	15651	18395	29885	21063
Фінансова та страхова діяльність, млн.грн	6214	6448	7679	8055	10652	11246
Операції з нерухомим майном, млн.грн	11230	11899	19665	22506	27557	28147
Професійна, наукова та технічна діяльність, млн.грн	2921	4065	6579	7965	10798	11912
Діяльність у сфері адміністративного та допоміжного обслуговування, млн.грн	3565	6535	10009	12747	11838	11741

Державне управління й оборона; обов'язкове соціальне страхування, млн.грн	5808	13942	22381	32844	44598	54306
Освіта, млн.грн	821	1540	2257	3493	4460	4789
Охорона здоров'я та надання соціальної допомоги, млн.грн	1224	2367	4479	6708	8139	9485
Мистецтво, спорт, розваги та відпочинок, млн.грн	509	1044	970	1649	3663	4146
Надання інших видів послуг, млн.грн	162	266	322	638	571	672
<b>Усього</b>	<b>219420</b>	<b>273116</b>	<b>359216</b>	<b>448462</b>	<b>578726</b>	<b>623979</b>

\*Джерело: складено авторами за даними [49,50, 51,52,53,54]

Джерелами капітальних інвестицій можуть бути кошти держбюджету, місцевих бюджетів, власні кошти підприємств та організацій, кредити банків та позики, кошти іноземних інвесторів, кошти населення та інші джерела фінансування.

Згідно з даними найбільшим джерелом фінансування капітальних інвестицій у 2020 році в Україні були власні кошти підприємств та організацій, їх частка у структурі капітальних інвестицій складала 65,4%, проти 70,8% у минулому році (рис. 2.3).

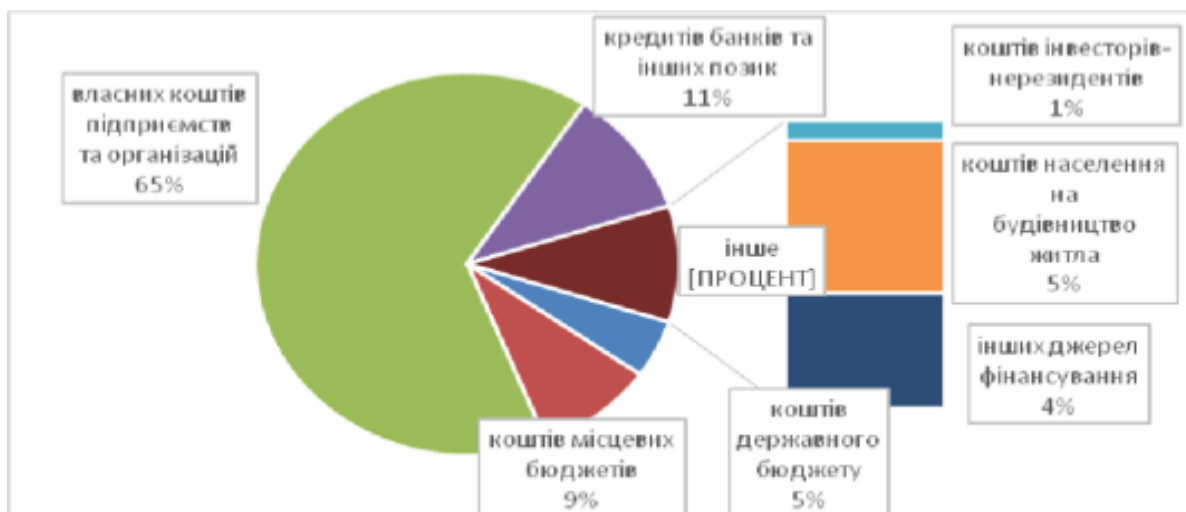


Рис.2.3. Джерела фінансування капітальних інвестицій у 2020 р.

\*Джерело [49,50, 51,52,53,54]

Важливе значення для забезпечення розвитку та підвищення ефективності як для економіки так і для окремих підприємств займають капітальні інвестиції

Проаналізувавши динаміку капітальних інвестицій в Україні за 2010-2019 років можна дійти висновку, що темпи зростання засвідчують тенденцію їхнього обсягу, однак темпи цього зростання в останні роки знижувалися.

### **2.3 Основні чинники, що впливають на обсяг капітальних інвестицій в Україні**

Обсяг інвестицій в основний капітал має вагомe значення як для компаній, так і для держави. Оскільки, інвестиції в основний капітал забезпечують розвиток виробництва: обладнання стає більш сучасним, виробляється більше, якісніше і швидше. Це призводить до зростання прибутку компанії. При цьому розвиток виробництва сприяє і зростанню економіки всієї країни. Причому на національній економіці позитивно відбивається як зростання обсягу виробництва, так і поліпшення якості виробництва. З аналізу статистичних даних ми можемо виявити, що у деякі періоди відбувається зростання або спад капітальних інвестицій, це вказує на те, що існують фактори які впливають на динаміку. Тому потрібно визначити і класифікувати ті фактори, які безпосередньо впливають на обсяги капітальних вкладень, вивчивши фактори, що формують інвестиційний клімат. Оскільки, капітальні інвестиції є економічним фактором, вони не можуть існувати незалежно від економічного середовища. Таким чином, економічні фактори є найбільш впливовими при

аналізі капітальних вкладень. Серед економістів фактори зазвичай поділяються [98]:

- Ті фактори, які в першу чергу обумовлені природними явищами або пов'язані з ними;
- ті, які суб'єктивно пов'язані з діяльністю окремої особи або організації, зокрема, в результаті їх управлінської або виробничої діяльності

Найбільш поширеним серед економістів є поділ факторів залежно від рівня їх сформованості.

Максимова В. Ф [100] класифікує фактори за їхніми сферами, які впливають на регулятивне, податкове, інформаційне забезпечення, економічні умови та екологію, ринкові умови, політичні, соціальні, соціокультурні, організаційні, правові, фінансові, непрозорість для інвесторів, заборгованість перед міжнародними економічними та соціальними організаціями фінансові організації. Таким чином, показники цієї конкретної системи класифікації використовуються для вивчення тісного взаємозв'язку між капітальними вкладеннями та їх наслідками.

За змістом досліджуваного явища автори В. М. Аскінадзі, Ю. Бугаєва, Л.[100] , та Білоусова О. С. [2] визначили, що на інвестиційний клімат країни впливають наступні фактори :

- адміністративно-правова система (система законодавства, судова система, дотримання гарантій прав власності, а також захисту прав учасників інвестиційного проекту, ступінь виконання законів і т. д.);
- економічна кон'юнктура (податкова система, потенціал ринку, середня прибутковість проектів з оновлення основного капіталу, значення основних макроекономічних показників (динаміка ВВП, темп інфляції, стійкість курсу валюти, стан конкурентного середовища, обсяг відтоку капіталу, ресурсний потенціал, частка заощаджень у ВВП , регіональний ВВП, частка заощаджень у ВВП в регіоні, кількість населення регіону,

обсяг чистого прибутку підприємств, відсоткова ставка за кредитами, обсяг реалізованої промислової продукції));

- політична ситуація (легітимність влади, розвиненість демократичних інститутів, громадянського суспільства, передбачуваність зовнішньої політики та ін.);
- стан і розвиненість інфраструктури (доступність державних послуг, доступність і об'єктивність інформації, розвиненість інфраструктури та ін.);
- розвиток промисловості (провідні галузі промисловості, рівень зносу основного капіталу промислових підприємств, ведення робіт в області створення високих технологій-наявність спеціалізованих науково-дослідних центрів, випробувальних лабораторій і т. д., рівень технічної оснащеності машинобудівних галузей, в першу чергу цивільного призначення);
- за формою впливу прямого впливу: обсяг прибутку підприємств, обсяг доходів населення, норма дохідності інвестицій; зворотного впливу: темп інфляції, ставка оподаткування, ставка банківського відсотка.

Також у дослідженнях все більше звертається уваги на основний чинник росту інвестицій – інноваційний розвиток, оскільки інвестиції в цю сферу переважають. Тому, що саме інновації здатні вивести економіку країни на новий рівень. Однією з причин, по якій може збільшитися інвестиційний попит, є технологічні нововведення.

Тому аби підтвердити даний взаємозв'язок між собою: зростання запасів капіталу і поліпшення технології, як вони впливають на обсяг виробництва, ми звертаємо увагу на пояснення моделі Солоу [98]. Попит і пропозиція на ринку виробленої продукції, які були ключовими елементами статичної моделі економіки, важливі і в моделі Солоу. Пропозиція визначає обсяг виробництва у



кожен момент часу, а попит визначає розподіл продуктів виробництва між різними напрямками використання і стійкий рівень капіталовкладень.

Представивши дві головних складових моделі Солоу - виробничу функцію і функцію споживання, можна проаналізувати, як накопичення капіталу забезпечує економічне зростання. Запаси капіталу можуть змінюватися з двох причин:

1. Інвестиції призводять до зростання запасів капіталу.
2. Частина капіталу зношується, тобто амортизується, що призводить до зменшення запасів капіталу.

Так, здійснивши аналіз основних факторів впливу на капітальні інвестиції, визначимо силу та напрям дії окремих факторів на обсяг капітальних інвестицій в Україні. За допомогою коефіцієнтів кореляції можна знайти щільність зв'язку між двома ознаками, при цьому нас цікавить форма зв'язку. Для аналізу було обрано декілька важливих макроекономічних показників, які визначаються незалежними змінними:

- обсяг ВВП ( $x_1$ ) [5],
- обсяг чистого прибутку підприємств ( $x_2$ )[33],
- обсяг реалізованої промислової продукції ( $x_3$ )[55],
- індекс цін виробників ( $x_4$ )[22] ,
- обсяг доходів населення ( $x_5$ )[1] ,
- обсяг прямих іноземних інвестицій ( $x_6$ )[55],
- відсоткова ставка за новими кредитами резидентам ( $x_7$ )[55].

Залежною змінною є обсяг обсяг капітальних інвестицій ( $y$ )[55].

Аналіз проводився за допомогою статистичною програми SPSS та набору даних.

**Model Summary<sup>b</sup>**

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,953 <sup>a</sup>	,927	,763	,00136
2	,513 <sup>b</sup>	,283	,576	,06088
3	,364 <sup>c</sup>	,142	,234	,25471
4	,568 <sup>d</sup>	,446	,565	,26217
5	,610 <sup>e</sup>	,346	,342	,21907
6	,984 <sup>f</sup>	,965	,433	,09149
7	,988 <sup>f</sup>	,969	,670	,77687

- a. Predictors: (Constant), обсяг ВВП,  
 b. Predictors: (Constant), обсяг прямих іноземних інвестицій  
 c. Predictors: (Constant), індекс цін виробників  
 d. Predictors: (Constant), обсяг чистого прибутку підприємств  
 e. Predictors: (Constant), обсяг за новим кредитами резидентами  
 f. Predictors: (Constant), обсяг реалізованої продукції  
 g. Predictors: (Constant), обсяг доходів населення  
 b. Dependent Variable: обсяг капітальних інвестицій

Рис.2.4. Результат кореляційно-регресійного аналізу

Спостереження, які представлені на рисунку 2.4 з проведеного дослідження показують, що на капітальні вкладення впливають коливання ВВП в Україні. Виходячи з коефіцієнта залежності між обсягом капітальних вкладень та обсягом ВВП - 0,953 вказує на те, що ці дві змінні сильно взаємопов'язані. Це пов'язано з тим, що ВВП є ключовим показником економічного розвитку та економічної безпеки країни і, як макроекономічний фактор, істотно впливає на всі показники. Примітно, що рівень капітальних вкладень в українську економіку значно змінився за останні роки, про що свідчить нестабільність ВВП.

Обсяг прямих іноземних інвестицій відіграє значну роль у капітальних вкладеннях, оскільки коефіцієнт зв'язку для прямих іноземних інвестицій становить 0,513, що вказує на середню щільність прямих інвестицій

Оскільки, коефіцієнт кореляції дорівнює 0,984 обсяг реалізованої промислової продукції істотно впливає на капітальні вкладення, збільшуючи прибуток підприємств, що відповідають за свої фінансові ресурси. Компанії, які

керують власними фінансами, збільшують свій прибуток у міру збільшення обсягу. Це збільшення не тільки збільшує прибуток, але й розширює інвестиційні можливості для компаній, які мають власні фінансові ресурси. Як результат, підприємства з власним фінансовим управлінням здатні отримувати більший прибуток, якщо у них є вільні гроші для інвестування. Крім того, вони покращують інвестиційні можливості для компаній, які керують власними фінансовими ресурсами, дозволяючи компанії інвестувати свої вільні гроші [118].

Зі збільшенням доходів збільшується обсяг інвестицій, спрямованих на максимізацію розподілу доходів населення на основі використання заощаджень. Коли кошти перерозподіляються банками, вони розміщують гроші на депозитних рахунках і стають ресурсами для кредитування юридичних осіб, тому, чим вище дохід, тим більше грошей економить населення, що полегшує банкам видачу кредитів. Оскільки банківські установи перерозподіляють грошові кошти, то наявні вклади на депозитних рахунках населення стають ресурсами для надання кредиту юридичним особам. Таким чином, чим вищі доходи населення, тим більше коштів вони заощаджують, що дозволяє банківським установам надавати більше позик. Про таку залежність свідчить коефіцієнт зв'язку доходів населення та капітальних інвестицій, величина якого складає 0,988.

Критичним аспектом нашої економічної моделі є величина чистого прибутку підприємств, яка значною мірою впливає на обсяги капітальних вкладень у сучасну економіку країни.

На капітальні вкладення індекс цін виробників не робить істотного впливу, як показує коефіцієнт 0,364. Таким чином, на обсяг інвестицій в капітальні витрати впливає нестабільність валютних курсів. Також більш імовірно, що капітальні інвестиції зменшаться, якщо очікуваний рівень інфляції зросте, що, в свою чергу, зменшить очікувану віддачу від інвестицій.

Таким чином було з'ясовано, що в економічній системі держави капітальні інвестиції дуже важливі. Зменшення рівня валових інвестицій в економіці країни призведе до того, що зменшиться кількість виробництва. Це все робить значний вплив на функціонування економіки держави і розвиток державних секторів. У позитивному ключі впливає на становлення економічної системи і розвиток її секторів акумуляцію інвестиційних фондів компанії, тобто збільшення можливостей для виробництва.

Кореляційно-регресійний аналіз показав, що в Україні найбільшою мірою на обсяг капітальних інвестицій впливають: обсяг ВВП, обсяг реалізованої промислової продукції і обсяг доходів населення, до факторів із середнім рівнем впливу на результативну ознаку належить: обсяг чистого прибутку підприємств відсоткова ставка за новими кредитами резидентам та обсяг прямих іноземних інвестицій; не має тісного зв'язку із результативною ознакою індекс цін виробників.

## РОЗДІЛ 3

### **Моделювання та оцінка прогностичної моделі темпів зростання інвестицій в основний капітал України**

#### **3.1 Використання даних, способи їх трансформації та методи побудови прогнозу інвестицій**

Наступні кроки формують основу для робочого процесу дослідження:

1. Сформулювання питання та визначення необхідних даних.
2. Підготування даних для моделі машинного навчання, тобто відбувається стандартизація даних.
3. Вибір моделей для прогнозування.
4. Навчання моделей на навчальних даних.
5. Здійснення прогнозу на основі даних тесту.
6. Порівняння прогнозів з відомими цільовими показниками тестового набору та обчислення показників ефективності.
7. Якщо продуктивність незадовільна, відбувається налаштування моделі, в такому випадку ми отримуємо більше даних або потрібно обирати іншу техніку моделювання.
8. Інтерпретування результату і здійснення опису про результати візуально та чисельно.

Прогнозованою змінною у дослідженні є темп зростання капітального інвестування. Ця змінна складається з трьох рядків зростання капітального інвестування, включаючи логарифмічну різницю з відповідним кварталом попереднього року. Тобто, були включені дані, які були отримані за три різні періоди: з 1 кварталу 1995 року по 4 квартал 2002 року, з 1 кварталу 2003 року по 4 квартал 2010 року і з 1 кварталу 2011 року по 1 квартал 2019 року. Такий

підхід дозволяє позбутися від сезонності з точки зору капітального інвестування. Крім того, такий підхід враховує можливу множинність та структурні зміни сезонності. Розширений тест Дікі-Фулера не показує розриву в перетвореному ряду. Після перетворення часовий ряд скорочується на перші 4 спостереження і починається з 1 кварталу 1996 року.

В якості залежних змінних використовувався набір з 31 макроекономічного показника української економіки: валове накопичення основного капіталу, ВВП в постійних цінах, ВВП, витрати зведеного бюджету, експорт, імпорт, індекс их (індекс української біржі) (на кінець кварталу), індекс реальної заробітної плати, індекс обсягу сільськогосподарського виробництва, індекс споживчих цін, індекс тарифів на вантажні перевезення, індекс цін виробників промислових товарів, індекс цін на будівельно-монтажні роботи, державний борг (в середньому за квартал), курс долара на міжбанку (на кінець кварталу), M0 (на кінець кварталу), M2 (на кінець кварталу), номінальний ефективний обмінний курс (на кінець кварталу), безробіття (в середньому за квартал), офіційний курс долара (на кінець кварталу), прибутковість 6-місячних державних облігацій (в середньому за квартал), прострочена дебіторська заборгованість (в середньому за квартал), прострочена кредиторська заборгованість (в середньому за квартал), ставка міжбанківського ринку, 1 день (в середньому за квартал), ставка міжбанківського ринку, 7 днів (в середньому за квартал), частка валового накопичення основного капіталу у ВВП. Повний список змінних, їх джерела наведені у додатку А.

Оскільки, багато з часто використовуваних показників макроекономічного прогнозування з'явилися в Україні відносно пізно, тому вихідні значення всіх рядів, використаних в роботі, починають свій відлік із 1 кварталу 1995 року.. Всі вони, при необхідності, були приведені до стандартизованого вигляду. Це робиться шляхом обчислення першої логарифмічної різниці за попередній квартал (для серії без сезонності) або за

відповідний квартал попереднього року (для серії з сезонністю). Після всіх перетворень лівою межею даних, використаних в роботі, став 1 квартал 1996 року.

На рисунку 3.1 показаний графік стандартизованих змінних, використаних у роботі, окремо виділяючи ряд прогнозованих змінних. Для оцінки моделей, відповідно до досліджень, стандартизація даних проводилася лише для методів регуляризації, оскільки для ансамблевих методів це робити не потрібно.

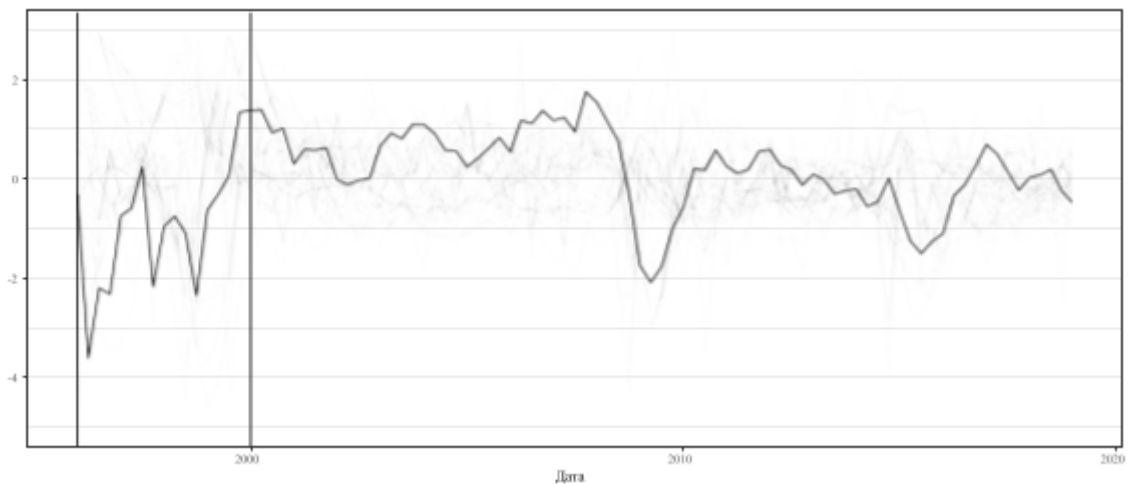


Рис. 3.1. Графік стандартизованих змінних  
(лініями визначена ліва межа навч. вибірок)

Наступним кроком у даній роботі є розгляд специфікацій для кожної моделі. Окремо розглядалися дві дати вибірки: перший квартал 1996 року та перший квартал 2000 року. Як вже зазначалось: перша дата була обрана тільки з тієї причини, що насправді багато макроекономічних показників до 1995 року взагалі недоступні, а після перетворень, проведених з даними, мінімальна доступна дата переноситься на перший квартал 1996 року. Тому можна припустити що додавання інформації за докризовий період 1998 року (під час першого кварталу 2000 року найраніша інформація, використана в навчанні, відноситься до першого кварталу 1999 року) не покращує якість моделей. Якщо

така гіпотеза опосередковано підтвердиться, вона може бути корисною в інших прикладних дослідженнях.

Для кожної зі специфікацій моделі регуляризації проводилось тестування з межі зліва направо, де права межа приймає значення з 1 кварталу 2012 року до 4 кварталу 2018 року, окремо також для прогнозування на горизонті  $h$  від 1 до 8 кварталів.

При цьому для моделей з  $h = 1$ , відповідно, максимальна дата закінчення навчального випробування — III квартал 2018 року, для моделей, прогнозованих на 2 квартали — II квартал 2018 року і так далі, для моделей, прогнозованих на 8 кварталів - IV квартал 2016 року.

Після навчання були побудовані прогнози поза вибіркою для кожного зразка навчання на горизонті  $h = 1, 2 \dots, 8$  кварталів для першого спостереження після тренувального випробування, а для нової ітерації межа тренувального зразка перемістила одне спостереження вперед. Отже, щоб отримати прогноз на квартал  $t + h$  використовувались лише значення змінних у кварталі  $t$ .

Параметри штрафу у відповідних моделях регуляризації були обрані шляхом перехресної перевірки у вікні навчання для кожної специфікації окремо. Перехресна перевірка в рамках тренувального випробування проводилася у фіксованому часовому інтервалі 40 кварталів і кроці 1.

Методологія прогнозування `randomforest` і моделей `boosting` не відрізнялася від методології прогнозування моделей регуляризації, за винятком того, що перехресна перевірка не використовувалася для вибору параметрів.

Вибір параметрів `randomforest` відбувався наступним чином: кількість змінних, доступних для побудови одного дерева рішень  $p$ , встановлюється на рівні, рекомендованому авторами цього методу Ліавом та Вінером [] для задач регресії  $p = p/3$ , де  $p$  - кількість змінних. Тобто  $p$  буде дорівнювати 12. Наступним етапом є вибір кількості дерев  $N$ . Згідно із дослідженнями у випадку дерев неможливо бути впевненим, що всі спостереження були використані в



моделі, однак із збільшенням кількості дерев збільшується обчислювальна складність. Тому у дослідженні використовуються чотири різні значення  $N$ : 100, 500, 1000, 2000. Оскільки, як показали результати для цих трьох моделей досить близькі за якістю, і жодна з них не лідирує за всіма специфікаціями, тоді як різниця між  $N = 1000$  і  $N = 2000$  майже непомітна, тому подальше збільшення  $N$  непотрібне. Також використовується стандартний критерій зупинки, рекомендований автором методу при побудові дерева: вершина розпізнається як кінцева, якщо подальше розщеплення призводить до менш ніж 5 спостережень в одній з дочірніх вершин.

Встановлення параметрів швидкості навчання виконується стандартним значенням  $\eta = 0,3$ . Хоча це значення зазвичай дозволяє отримати досить консервативну модель, специфікації додатково перевіряються на роботі  $\eta = 0,4$ ,  $\eta = 0,2$  і  $\eta = 0,1$ . У різних специфікаціях неможливо однозначно вибрати найкраще значення параметра. Параметр  $N$  відповідає за кількість ітерацій при їх збільшенні. Зазвичай при виборі  $N$  вони зупиняються на рівні, після чого прогнози моделі перестають змінюватися. У дослідженні достатньо було встановити  $N = 100$ .

Для побудови моделі авторегресії кількість використаних затримок спочатку вибиралася для кожної специфікації моделі за допомогою AIC (як правило, не перевищувала 2). Потім були оцінені коефіцієнти. Для цього був побудований рекурсивний прогноз на 8 кроків вперед, тобто послідовна оцінка всіх 8 прогнозних значень, замість окремої оцінки 8 рівнянь. Такий метод оцінки коефіцієнтів був використаний тому, що як показали дослідження Фауста і Райта [89], прогнози з цим підходом є більш точними.

Для того аби побудувати прогнози для всіх моделей було здійснено обчислення в R за допомогою наступних пакетів:

- gmnet для Ridge, LASSO, Post-Lasso, Adaptive LASSO, Elastic Net,
- spikeslab для Spike і Slab,

- randomForest для моделі випадкового лісу,
- forecast для моделі AR,
- xgboost для моделі boost.

Код із розрахунками представлений у додатку В. На рисунку 3.2 зображено результати для всіх відібраних моделей методів-візуально прогнози при різних значеннях параметрів дуже схожі. Можна бачити, що моделі boosting і randomforest виявляються досить близькими до реальних значень змін інвестицій.

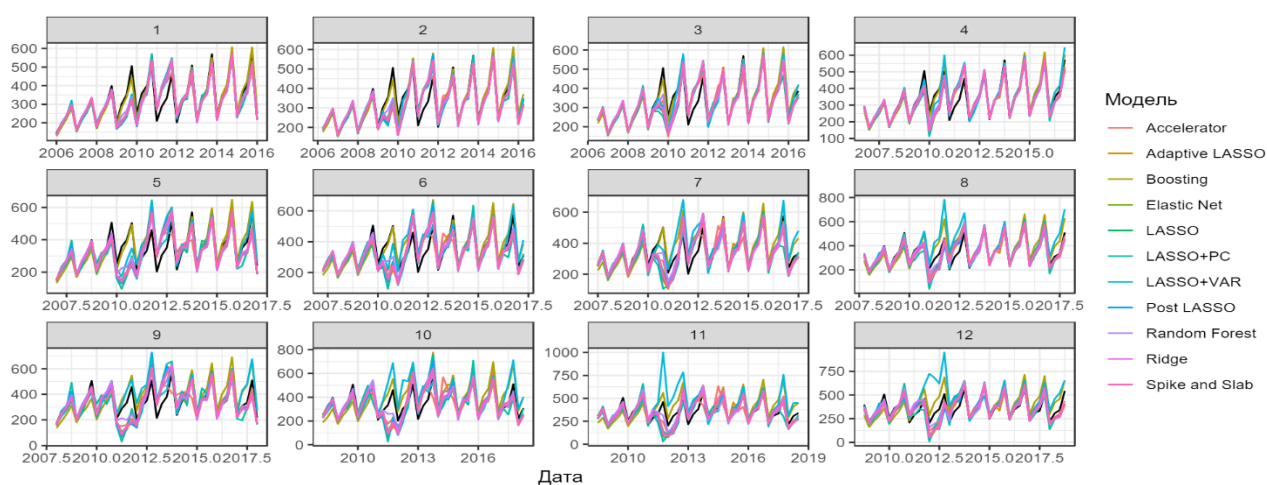


Рис. 3.2. Результати прогнозування

На рис. 3.3 зображено гістограму моделей і їх точності. Методи randomforest та boosting дали при перевірці найбільшу точність.

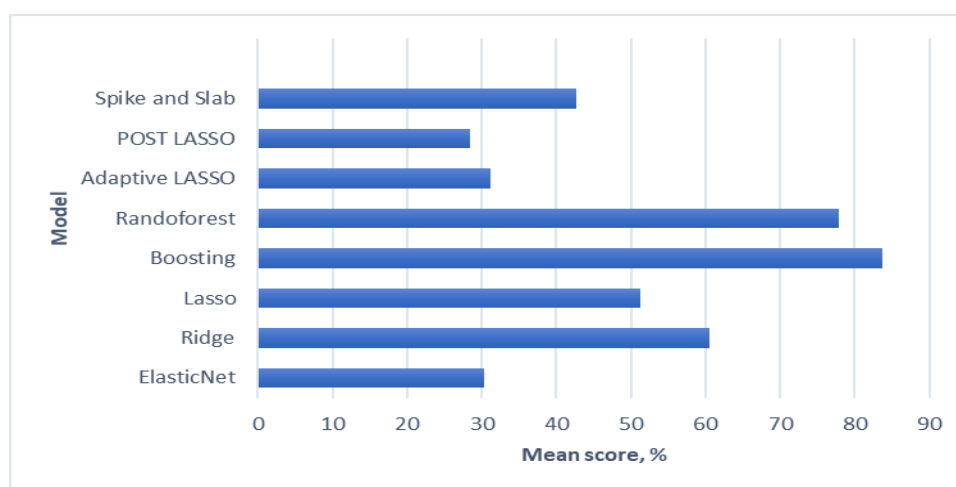


Рис.3.3. Точність кожної моделі

Найбільший відсоток прогностичної здатності мають моделі: Boosting та Randoforest.

```
#Results
Root Mean Squared Error (Lasso):      203421.22072610114
Root Mean Squared Error (ElasticNet):  203442.40673916895
Root Mean Squared Error (Ridge):      203411.816202574
```

Рис. 3.4. Сереньоквадратична помилка моделей

Сереньоквадратична помилка, що представлена на рисунку 3.4 показує безумовно, низьку продуктивність, що узгоджується із низьким відсотком прогностичної здатності, але певною мірою модель змогла спрогнозувати значення.

### 3.2 Порівняльна оцінка якості моделей

Після того, як дані були стандартизовані та прогнози були зроблені для кожної моделі, відбувся розрахунок RMSE на тренувальній вибірці. Метрика якості, що використовується в роботі, - RMSE (коренева середньоквадратична помилка прогнозу). Тобто, як вже зазначалось ця метрика застосовується тоді коли потрібно визначити квадратний корінь середньої квадратичної різниці між прогнозованими значеннями та фактичними значеннями в наборі даних. Чим нижче RMSE, тим краще модель відповідає набору даних. У дослідженні це значення розраховувалося по відношенню до моделі випадкового блукання для терміну прогнозування від нуля до восьми кварталів представлені в таблиці 3.1. Для лівої межі тренувальної вибірки в першому кварталі 1996 року і в таблиці 3.2 для лівої межі тренувальної вибірки в першому кварталі 2000 року.

Така метрика часто зустрічається при розгляді регресійних задач (вона використовується, наприклад, в роботах Фокіна і Полбіна, Байбузи [105]).

**Таблиця 3.1**

RMSE. Ліва межа тренувальної вибірки в першому кварталі 1996 року

Модель	1	2	3	4	5	6	7	8
Випадкове блукання	1,00	1,00	1,00	1,00	1,00	1,00	1,00	1,00
AR	0,92	0,85	0,79	0,73	0,68	0,67	0,68	0,61
Adaptive LASSO	0,99	0,91	0,79	0,84	0,75	0,77	0,61	0,5
Elastic Net	1,01	0,85	0,84	0,8	0,76	0,74	0,69	0,56
LASSO	1,01	0,89	0,83	0,81	0,82	0,72	0,71	0,55
Post-LASSO	1,05	0,95	0,83	0,93	0,79	0,78	0,72	0,58
Ridge	0,99	0,89	0,88	0,87	0,81	0,79	0,7	0,61
Spike and Slab	1,00	0,84	0,82	0,79	0,77	0,74	0,63	0,58
boosting ( $\eta = 0, 1$ )	0,92	0,74	0,81	0,66	0,67	0,53	0,62	0,62
boosting ( $\eta = 0, 2$ )	0,85	0,76	0,79	0,71	0,68	0,62	0,49	0,59
boosting ( $\eta = 0, 3$ )	0,97	0,79	0,7	0,71	0,73	0,52	0,53	0,6
boosting ( $\eta = 0, 4$ )	1,07	0,71	0,82	0,69	0,63	0,54	0,48	0,59
random forest ( $N = 100$ )	0,91	0,72	0,75	0,62	0,7	0,56	0,54	0,57
random forest ( $N = 500$ )	0,78	0,71	0,75	0,66	0,66	0,58	0,53	0,56
random forest ( $N = 1000$ )	0,80	0,71	0,76	0,66	0,66	0,58	0,55	0,55
random forest ( $N = 2000$ )	0,90	0,7	0,75	0,66	0,67	0,58	0,54	0,56

В результаті було встановлено, що для двох різних дат початку тренувального випробування найкраща якість майже завжди демонструється моделями random forest і boosting. У більшості дослідженнях такий феномен пояснюється тим, що часто погано інтерпретовані методи машинного навчання показують кращі результати прогнозування над методами регуляризації.

Таблиця 3.2

RMSE.Ліва межа тренувальної вибірки в першому кварталі 2000 року.

Модель	1	2	3	4	5	6	7	8
Випадкове блукання	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
AR	1.00	1,20	1,17	1,08	1,08	1,13	1,02	0.88
Adaptive LASSO	1.00	0.64	0.97	0.84	0.89	0.89	0.63	0.63
Elastic Net	0.78	0.62	0.96	0.76	0.81	0.82	0.67	0.59
LASSO	0.91	<b>0.58</b>	0.99	0.74	0.88	0.83	0.66	0.57
Post-LASSO	0.94	0.72	1,08	0.91	1,08	0.98	0.75	0.63
Ridge	0.88	0.64	0.79	0.78	0.76	0.79	0.68	0.66
Spike and Slab	0.85	0.64	0.80	0.80	0.91	0.83	0.66	0.60
boosting ( $\eta = 0, 1$ )	0.97	0.64	0.70	0.60	0.68	<b>0.59</b>	0.60	<b>0.44</b>
boosting ( $\eta = 0, 2$ )	0.90	0.63	<b>0.65</b>	0.58	0.69	0.63	0.63	0.57
boosting ( $\eta = 0, 3$ )	1,10	0.69	0.66	0.66	<b>0.63</b>	<b>0.59</b>	0.62	0.52
boosting ( $\eta = 0, 4$ )	1,08	0.66	0.72	<b>0.56</b>	0.67	0.66	<b>0.56</b>	0.50
random forest ( $N = 100$ )	<b>0.75</b>	0.62	0.73	0.62	0.67	0.63	0.62	0.56
random forest ( $N = 500$ )	0.81	0.61	0.73	0.64	0.67	0.63	0.61	0.58
random forest ( $N = 1000$ )	0.78	0.62	0.72	0.63	0.66	0.62	0.59	0.57
random forest ( $N = 2000$ )	0.78	0.62	0.73	0.64	0.66	0.62	0.59	0.58

Прикладами такого феномену є дослідження Байбуз О.[112], де метод random forest і boosting теж показували кращі результати прогнозування (майже у всіх специфікаціях під час прогнозування інфляції), або в роботі Квізгард І.[108], де методи регуляризації показували меншу якість прогностичної здатності під час прогнозування рівня ВВП та інфляції.

Наступним розрахунком у таблицях представленні дані моделей модифікацій LASSO. Результат показав, що в більшості специфікацій Adaptive LASSO і Post-LASSO не перевищують якість прогностичної здатності основного LASSO. Метод Elastic Net, в якому теоретично повинні використовуватися і Ridge і LASSO як виявляється, поступається у якості

прогностичної здатності принаймні одній з цих моделей. Однак усі вони досить близькі з точки зору помилок прогнозування

Наступними у таблиці представлені результати для моделі random forest. Для того щоб з'ясувати різницю між якостями прогнозів потрібно звернути увагу на показник параметру  $N$ . Значення параметрів  $N$  для моделі random forest, показав невелику різницю між значеннями. Тому це дає розуміння того, що подальше збільшення кількості дерев не призведе до поліпшення якості прогнозів. В цілому, всі протестовані моделі майже завжди показують якість не гірше, ніж випадкове блукання, але, прогнозуючи на квартал вперед, багато моделей виявляються гірше наївного прогнозу.

Наступним етапом проведення оцінки якості моделей ми доповнили тестом Дібольда-Маріано, описаним у Diebold and Mariano [71]. Оскільки, розмір тестових зразків невеликий-від 15 до 20 спостережень – у роботі використовувалася статистика тесту Дібольда-Маріано, адаптована до невеликих даних, запропонована Харві, Лейборном та Ньюболдом. Нульова гіпотеза ( $h_0$ ) для моделей полягає в тому, що обидві моделі мають однакову якість прогнозування. Альтернативна гіпотеза ( $h_1$ ) для кожної специфікації є односторонньою, що означає, що модель з найменшою вибіркою RMSE дає найкращі прогнози.

На рисунку 3.5 показані результати тестів Дібольда-Маріано для різних методів, навчених на одній вибірці та на одному горизонті прогнозування. Зелений колір на малюнках вказує на те, що модель зіставлення рядків була кращою, ніж модель зіставлення стовпців. Червоний колір вказує на перевагу моделі, що відповідає по стовпцю. Чим яскравіший колір, тим більше  $p$ -значення статистики тесту Дібольда-Маріано, а отже, різниця між двома прогнозами менша. Якщо значення  $p$  більше 0,1, різниця між двома моделями статистично незначна, і на малюнках це показано білим кольором.

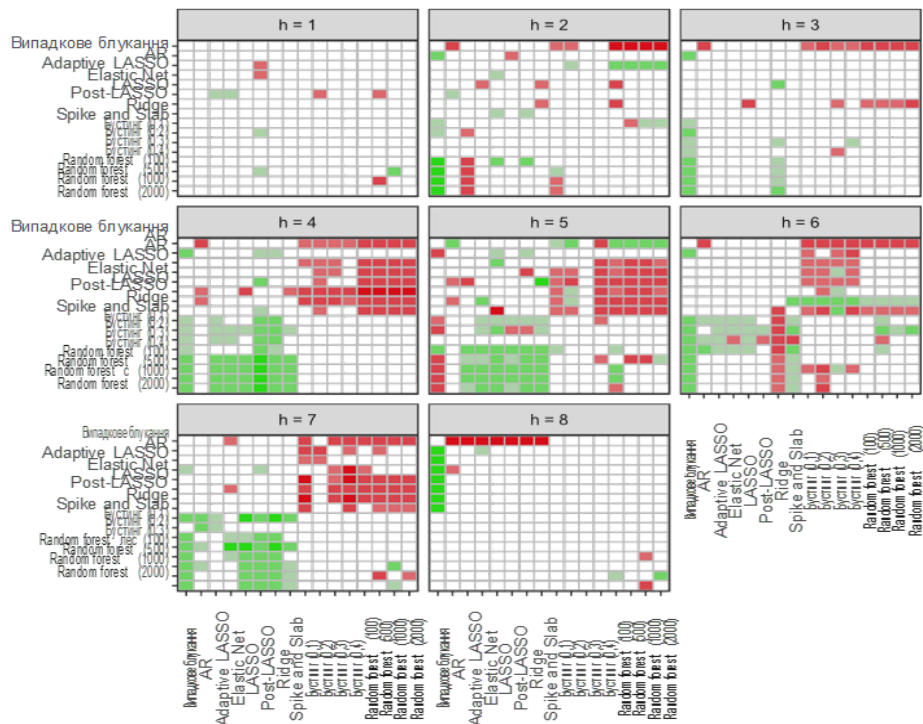


Рис.3.5. Тест Diebold-Mariano: якість моделей - ліва межа тренувальної вибірки 1-ий квартал 1996 р

Можна побачити, що в міру збільшення горизонту прогнозу (при  $\lambda > 4$ ) модельні прогнози починають все більше відрізнятися один від одного, і переважання ансамблевих методів над методами регуляризації проявляється все частіше. Однак навіть при  $\lambda = 8$  прогнози майже всіх моделей статистично дещо відрізняються один від одного.

Також тест Diebold-Mariano, що представлений на рисунку 3.5 вказує, як перехід до зменшеної вибірки впливає на різні методи: якісні моделі (boosting і random forest) стають ще кращими, тоді як середня якість методів регуляризації не збільшується або навіть зменшується. Єдиними винятком є модель AR, якість якої сильно знижується при переході на зменшену кількість спостережень, і модель випадкового блукання, яка має меншу якість прогностичної здатності за більшість моделей на більшості горизонтів прогнозування.

Можна також відмітити і те, що для усіх моделей машинного навчання, які були використанні у дослідженні, відносно моделі випадкового блукання зі збільшенням горизонту прогнозування, збільшується якість прогнозу, проте здатність витягувати корисну інформацію з даних, здається, зменшується. Також результат оцінки якості прогнозування показав, що ансамблеві методи статистично перевершують методи регуляризації для середніх горизонтів ( $\lambda = 4, \dots, 6$ ), проте ця перевага зникає при дворічному прогнозуванні. Насправді це означає, що в даних недостатньо інформації, щоб моделі могли робити піврічні прогнози.

На рисунку 3.6 показані прогнози для прогнозного горизонту до одного року ( $\leq 4$ ) з першого кварталу 2013 року до кварталу 2019 року для скороченого набору даних.

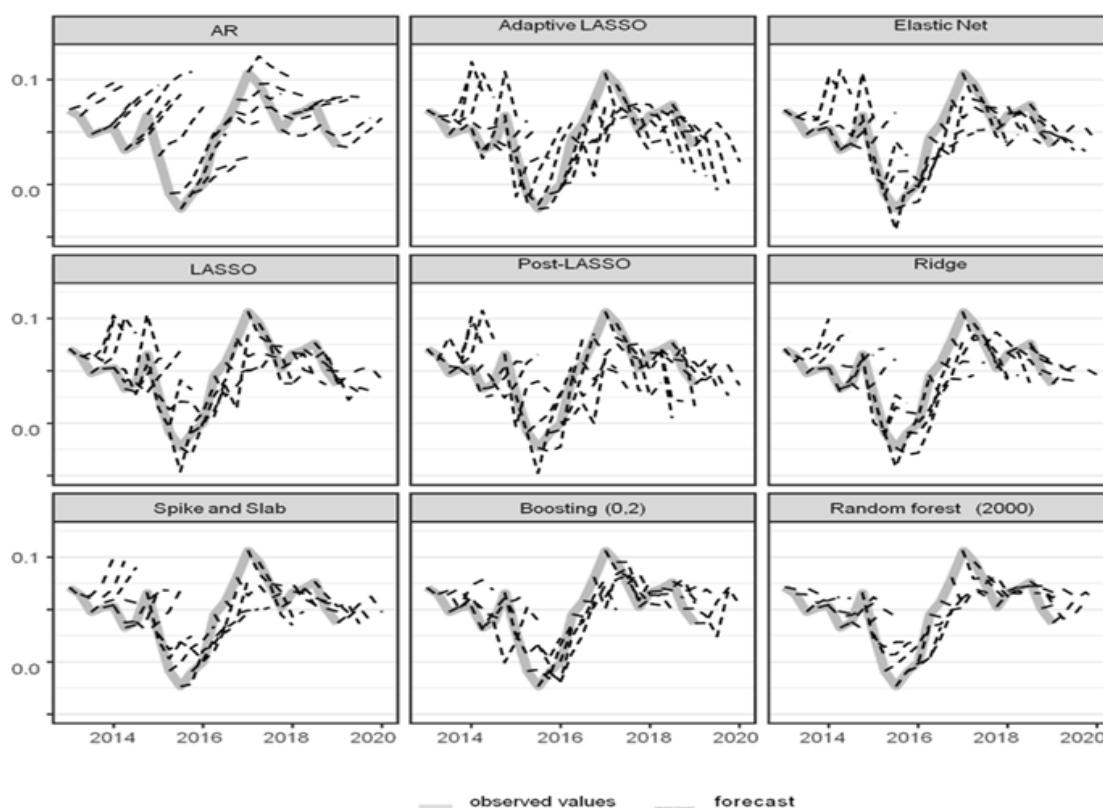


Рис. 3.6. Прогнозування капітальних інвестицій



Різниця в якості моделей, що була попередньо проведена оцінкою RMSE також підтверджується графіком: візуальним-але здається, що boosting і random forest краще повторюють рух прогнозованого ряду. На графіку показані прогнози boosting і random forest тільки для однієї специфікації тому, що візуально прогнози однакових моделей виявляються схожими при зміні параметрів. Прогнози моделі випадкового блукання, на графіку не показані тому що вони тривіальні і виходять шляхом зсуву спостережень вправо.

Коли ми отримали усі результати оцінок якості моделей методу RSME, то тоді наступним етапом є перевірка гіпотези ( $h_1$ ), про те, що модель з меншим показником RMSE дає найкращі прогнози. Перевірка гіпотези відбувалася із застосуванням тесту Дібольда-Маріано. У таблиці 3.4 показані зміни RMSE відносно моделі випадкового блудження при переході до зменшеної вибірки, а в дужках наведені значення статистики Дібольда-Маріано. Такий аналіз було важливо провести, оскільки із табл. було видно, що докризові дані майже ніколи не допомагають покращити прогнози. Це тому, що докризові спостереження дуже нестабільні, значення змінних незвичні для наступних змінних, і в результаті навчання на їх основі моделі починають гірше прогнозувати дані за останні періоди. На рисунку 3.2 показані значення всіх змінних, використаних у роботі. Можна відзначити зниження їх вартості з початку року (очікується, що в кризові 2008-2009 і 2014-2015 роки ряди знову стануть нестабільними).

Результати показують, що моделі машинного навчання, які були навчені на зменшених даних, дають прогнози не гірше, ніж аналогічні моделі на розширених даних в річному прогнозі ( $h \leq 4$ ). Єдиним винятком є модель підвищення, прогнози якої значно погіршуються при  $h = 1$ .

Коли справа доходить до прогнозування на два роки, якість декількох моделей також погіршується при  $h = 5,6$ , але в цілому жодна з двох специфікацій не перевершує іншу послідовно. Однак, беручи до уваги, що зі збільшенням  $h$ ,

тобто збільшенням розриву між датами, інтерпретація прогнозів повинна зменшуватися тому можна стверджувати, що видалення докризових даних з 1998 року, нетипових для наступні періоди, дозволяє отримати більш стабільні і відповідні результати, принаймні в прогнозі на майбутній рік.

З методів регуляризації модель Ridge виявилася найбільш стабільною щодо зменшення даних: або погіршення якості було дуже невеликим ( $h = 6,8$ ), або якість покращилася.

Як згадувалося вище, методи на основі LASSO дозволяють вибирати лише найважливіші змінні. Це дослідження демонструє, як, незважаючи на меншу прогностичну здатність LASSO, відповідні незалежні змінні все ще можуть бути обрані і інтерпретовані та узгоджуються з основними теоретичними уявленнями про інвестиційну функцію.

Тому важливо представити результати цієї моделі.

Варто почати з того, щоб показати, скільки змінних LASSO було вибрано в різних горизонтах прогнозування та навчальних зразках.. У таблиці показано кількість змінних, вибраних LASSO для двох різних лівих меж навчального зразка.

Кількість вибраних змінних є досить нестабільною, що зазвичай характерно для моделей LASSO, як показують Зоу Х. та Хасті Т. [73]. У той же час можна побачити, що при  $\lambda = 3,5$  кількість вибраних змінних стабільна для моделей, навчених за даними першого кварталу 1996 року (за розширеною вибіркою), і на цих горизонтах прогнозування LASSO показує погіршення якості зі зменшенням вибірки (див.додаток ). Навпаки, для  $\alpha = 2$  кількість коефіцієнтів у моделі, навчній на зменшеній вибірці, не змінюється так сильно, як у моделі, навчній на розширеній вибірці, що узгоджується зі статистично помітним покращенням якості моделі зі зменшенням вибірки.

Однак залишається питання, які конкретні змінні вибираються? Наступні таблиці демонструють середні стандартизовані коефіцієнти моделі LASSO для декількох горизонтів прогнозування за 1 квартал 2000 року.

**Таблиця 3.5**

Основні змінні в моделі LASSO для  $h = 1, \dots, 4$  (ліва межа тестової вибірки-1 квартал 2000 року)

1	h=1	h=2	h=3	h=4
1	ВВП в постійних цінах 0.047	М2 (НА КІНЕЦЬ КВАРТАЛУ)	Індекс цін виробників промислових товарів 0.0229	Індекс споживчих цін -0.293
2	Валове накопичення основного капіталу 0.035	Індекс реальної заробітної плати 0.025	Державний борг (в середньому за квартал) -0.060	Державний борг (в середньому за квартал) -0.084
3	Індекс реальної заробітної плати 0.025	Індекс реальної заробітної плати 0.024	М2 (НА КІНЕЦЬ КВАРТАЛУ) 0.055	Індекс реальної заробітної плати -0.067
4	М2 (НА КІНЕЦЬ КВАРТАЛУ) 0.01	Індекс тарифів на вантажні перевезення 0.015	Індекс реальної заробітної плати -0.043	Індекс реальної заробітної плати 0.056
5	Індекс цін на будівельно-вантажні роботи 0.008	Імпорт 0.017	КІ до ВВП -0.036	Індекс тарифів на вантажні перевезення -0.052

Ключові змінні для решти також можна знайти в таблиці 3.5 та 3.6.

Поточний ВВП і поточне відставання пояснюваної змінної є основними факторами при прогнозуванні на квартал вперед, що узгоджується з теорією Кларка і Гіттона [78], згідно з якою фірми прагнуть максимізувати прибуток, оптимально врівноважуючи капітал і випуск продукції. У цьому випадку

поточний рівень інвестицій повинен враховувати не тільки поточний випуск продукції, але і попередні інвестиції і попередній випуск продукції.

**Таблиця 3.6**

Основні змінні в моделі LASSO для  $h = 5, \dots, 8$  (ліва межа тестової вибірки-1 квартал 2000 року)

5	h=5	h=6	h=7	h=8
6	Індекс споживчих цін -0.023	Індекс споживчих цін -0.04	Індекс споживчих цін -0.020	Державний борг (в середньому за квартал) -0.013
7	Державний борг (в середньому за квартал) -0.08	Державний борг (в середньому за квартал) -0.02	Державний борг (в середньому за квартал) -0.011	Державний борг (в середньому за квартал) 0.010
8	Індекс реальної заробітної плати 0.04	Індекс тарифів на вантажні перевезення -0.004	Індекс цін на будівельно-вантажні роботи -0.006	Індекс споживчих цін 0.007
9	Індекс цін на будівельно-вантажні роботи -0.04	Державний борг (в середньому за квартал) 0.004	Державний борг (в середньому за квартал) 0.006	Індекс обсягу сільськогосподарського виробництва 0.005
10	Прострочена дебіторська заборгованість (в середньому за квартал) 0.01	Індекс цін виробників промислових товарів -0.003	КІ до ВВП -0.003	Індекс цін виробників промислових товарів -0.005

У таблицях варто звернути увагу на зміни частки накопичення валових основних засобів у ВВП, що важливі для декількох горизонтів з негативним знаком. Наявність цієї змінної фактично говорить про те, що існує довгострокове співвідношення інвестицій до обсягу виробництва, яке вказує на те, що коригування відбувається, якщо вона порушена. Тобто, при відносно надмірних інвестиціях, незмірно зростаючих до ВВП, через деякий час відбувається корекція, а темпи зростання інвестицій знижуються і навпаки.

Набір змінних у моделі LASSO не є стабільним, навіть якщо при цьому коефіцієнти та знаки не суперечать економічній теорії. Подібні умови виникають при моделюванні прогнозу модифікацій LASSO, що пояснює, чому прогнози мають нижчу якість, і поступаються за якістю ансамблевим методам.

### **3.2 Проблеми короткострокового прогнозування макроекономічних показників**

Макроекономісти займаються аналізом розвитку економіки як єдиного цілого. Вони збирають дані про доходи, ціни, безробіття і багато інших економічних показників, які стосуються різних країн і різних періодів. Потім вони намагаються сформулювати загальний теоретичний підхід до пояснення цих даних.

Для прогнозування макроекономічних змінних використовується метод відбору даних. Показники можуть мати значний вплив на точність прогнозу. Моделі прогнозування стикаються з двома проблемами: 1) вибір показників для аналізу (використання експертних знань є звичайним явищем, але не існує методології, яка могла б бути використана для чіткої демонстрації її обґрунтованості відбору показників для аналізу); 2) незбалансований набір показників історичних даних, що використовуються для побудови моделей. Органи державної статистики виконують свій обов'язок щодо забезпечення надійності та об'єктивності даних шляхом перегляду даних: з одного боку, вони несуть відповідальність за підтримку цілісності даних. З іншого боку, у них є зобов'язання щодо своєчасного оновлення даних. Орган державної статистики України переглядає статистичну інформацію відповідно до своєї політики щодо перегляду статистичних даних [12], яка визначає причини перегляду та класифікації змін. Перегляди можуть бути заплановані відповідно до політики наступним чином: регулярними, або спеціальним, а також – незапланованим.

Нерідко ключові статистичні дані публікуються із запізненням і часто переглядаються. Деякі короткострокові економічні показники можуть бути переглянуті протягом двох років після публікації, тоді як інші можуть бути переглянуті протягом першого місяця після публікації. Тому можна зробити висновок, що таким чином прогнози можуть викликати помилкову впевненість, оскільки найчастіше для побудови класичних прогнозів використовуються змінні, які не відповідають інформації про поточні економічні умови.

В даний час існує значна кількість наукових праць щодо застосування різних методів роботи з великими обсягами даних в макроекономіці. Першу модель макропрогнозування запропонував Ян Тімберген в 1936 році [74]. Із 1950-х рр. це поява моделі Клейна-Гольдбергера, яка була першою макроеконометричною моделлю змодельованою для створення макропрогнозів. Приблизно після середини 1950-х рр. ці так звані великомасштабні економетричні моделі швидко поширювалися, і вони домінували у виробленні політики та прогнозуванні в цій галузі. Починаючи з середини 1970-х рр., ці моделі прогнозування піддавалися критиці [97]. У 1976 р. Роберт Лукас сказав [5], що моделі, які оцінюються виключно на основі статистичних даних, можуть виявитися поганими орієнтирами для розробки політики. Статистичні дані, які традиційно або за певний період оцінені можуть не відображати зміни в економічній політиці. Крістофер С. [98] також критикував метод прогнозування, який називається векторною авторегресією, оскільки такий метод включає в себе менше припущень для складання прогнозів.

Однак з тих пір відбулися величезні інновації в доступності даних на рівні домогосподарств і фірм, в обчислювальних технологіях і в теоретичних міркуваннях. Тому у наш час важливість прогнозування продовжує зростати, тому що майже кожен бізнес повинен передбачати майбутнє, щоб приймати кращі рішення та ефективніше розподіляти ресурси. Щоб впоратися зі зростаючою різноманітністю та складністю проблем прогнозування, протягом

багатьох років було розроблено багато методів прогнозування. Кожне з них має своє особливе використання, тому потрібно бути обережним, щоб вибрати правильну техніку для конкретного застосування. Глибоке розуміння діапазону доступних методів прогнозування збільшує ймовірність того, що конкретне застосування принесе плоди. Вибір методу залежить від багатьох факторів — контексту прогнозу, актуальності та доступності даних, бажаного ступеня точності, періоду часу для прогнозування, вартості та переваг прогнозу та часу, доступного для проведення аналізу. Оскільки прогнозування макроекономічних показників є кількісним, мати дані про минуле є обов'язковим. Відповідний метод прогнозування значною мірою залежить від типу та кількості доступних даних[57].

Також важливо врахувати і те, що макроекономічне прогнозування зазвичай використовує для побудови прогнозу велику кількість даних, це може призвести до проблеми «прокляття розмірності». У нашому дослідженні щоб не зустрітись із цією проблемою, при виборі моделей для прогнозування ми також враховували цей фактор. Тому в результаті одним з підходів до роботи з великою кількістю змінних було обрано методи регуляризації. Який як вже було зазначено при оцінці параметрів використовує функцію штрафу за збільшення числа або розміру параметрів моделі. Труднощі, із великою розмірністю даних як показали інші дослідження можуть виникати також під час навчання моделей машинного навчання, аналізу чи візуалізації даних. Зі збільшенням розмірів дані стають розрідженими, а коли дані стають розрідженими, стає важко узагальнити модель[61].

Також важливим завданням дослідження із застосуванням макроекономічних показників є спрогнозування кризових ситуацій. Оскільки, як показало наше дослідження докризові спостереження дуже нестабільні, значення змінних незвичні для наступних змінних, і в результаті навчання на їх основі моделі починають гірше прогнозувати дані за останні періоди.

Під час час фінансової кризи, що розпочалась у 2008 р., відразу у засобах масової інформації розпочалася дискусія, про те що макроекономічні моделі не працюють тому, що не дозволяють спрогнозувати кризу. Основними аргументами було те, що макроекономічні моделі не враховують ефективність ринку, асиметричність інформації. Проте Нараян Кочерлакота [103] у своїй дослідницькій роботі проаналізував наукові роботи 44 сучасних макроекономістів і з'ясував, що аргументи щодо того, що усі макроекономічні моделі абстрактні і не враховують однорідність агентів, раціональність агентів, втручання держави і те, що не усі ринки неефективні - не відповідають дійсності. Оскільки, саме із цими термінами працюють макропрогнозисти. Однак, є деякі чіткі сфери у макроекономічному прогнозуванні, що викликають занепокоєння: фінанси та банківська справа (моделі не розкривають перерозподіл багатства на фінансових ринках), подрібнений підхід (жодна макромодель немає прогностичну здатність спрогнозувати стабільність на ринку праці)[97].

Тому видається невиправданим оцінювати прогнози тільки по кількісній точності. Очевидно, що надійність прогнозів визначається близькістю між їх фактичними та прогнозованими значеннями. Однак це може бути непередбачувано через інтерпретацію, здогадок експертів, неврахованих потрясінь і т. д. Тобто надійність прогнозування залежить від того, наскільки реалістичні і несуперечливі сценарії, а також від того, наскільки повно розглянуті і розраховані можливі комбінації зовнішніх і внутрішніх умов і наскільки достовірно визначені кількісні параметри факторів.

Отже, зважаючи на основні проблеми макропрогнозування ми вважаємо, що якісні макроекономічні прогнози повинні бути як кількісно точними, так і якісно точними, тобто здатністю прогнозувати і точно описувати та обґрунтовувати найбільш важливі процеси і тенденції економічної динаміки; прогнозувати зміни тенденцій і економічний розвиток, а також напрямок і інтенсивність структурних змін в економіці.



## ВИСНОВКИ

Для здійснення побудови прогнозів коефіцієнта накопичення капіталу в Україні було виконано ряд завдань. Спочатку ми здійснили аналіз літератури з метою узагальнення досвіду моделювання та прогнозування інвестицій в основний капітал та порівняли різні методи машинного навчання: регуляризації та ансамблеві. Також здійснили порівняльний аналіз методів для порівняння якості моделей при прогнозуванні. Так було визначено, що найкращу якість у прогнозуванні тенденції розвитку різних явищ і процесів показують ансамблеві методи- random forest і boost, і модель авторегресії порядку  $p$  AR ( $p$ ). Відносно невисока якість прогнозів методів регуляризації. У більшості дослідженнях такий феномен пояснюється тим, що часто погано інтерпретовані методи машинного навчання показують кращі результати прогнозування над методами регуляризації. Під час аналізу методів оцінки якості моделей прогнозування: RMSE, AIC, тест Diebold Mariano [72], були визначенні як найкращі методи, що дозволяють правильно обробити дані та виконати оцінку моделей прогнозування.

Значення терміну капітальні інвестиції в нормативно правових документах, визначається як витрати на покупку або оновлення обладнання, будівель, споруд, транспорту. Такі витрати здатні забезпечити розвиток компанії і зростання прибутку в майбутньому. Якщо розглядати дану категорію капітальні інвестиції з точки зору підприємства, то її слід інтерпретувати як капіталізовану вартість ресурсів (матеріальних, фінансових, трудових і т.), спрямованих на відтворення основних активів. А капітал-це майно компанії, яке вона може використовувати для отримання прибутку.

Проаналізувавши динаміку капітальних інвестицій в Україні за 2010-2019 років можна дійти висновку, що темпи зростання засвідчують тенденцію їхнього обсягу, однак темпи цього зростання в останні роки знижувалися. Кореляційно-регресійний аналіз показав, що в Україні найбільшою мірою на обсяг капітальних інвестицій впливають: обсяг ВВП, обсяг реалізованої промислової продукції і обсяг доходів населення, до факторів із середнім рівнем впливу на результативну ознаку належить: обсяг чистого прибутку підприємств відсоткова ставка за новими кредитами резидентам та обсяг прямих іноземних інвестицій; не має тісного зв'язку із результативною ознакою індекс цін виробників. Загальний аналіз літератури показав, що також однією з причин, по якій може збільшитися інвестиційний попит, є технологічні нововведення. Таким чином рівень інвестиційних вкладень значно впливає на модернізацію економіки, допомагає позбутися від довгострокових криз в системі. Капітальні інвестиції можуть вважатися одним з показників розвитку економіки держави. Вони показують потенційні тенденції зростання. Показник капітальних інвестицій на державному рівні дозволяє зробити висновок про затребуваність виробленої на території країни товарів та послуг, а також рівень ВВП, визначити привабливість для залучення зарубіжних інвестицій, оцінити рівень економічного розвитку в цілому.

Ми використовуємо великий набір залежних змінних і методи машинного навчання для побудови прогнозів коефіцієнта накопичення капіталу в Україні. За результатами кращу якість демонструють командні методи-random forest і boost, які перевершують не тільки прості еталонні моделі, але і методи регуляризації, що також узгоджується з літературою з макроекономічного прогнозування. Оскільки, аналіз дослідження методів макроекономічного прогнозування показав, що найкращу якість показують ансамблеві методи-random forest і boost, а також модель авторегресії порядку  $p$  AR ( $p$ ), Використання розширеного набору даних, що включає докризовий нагляд 1998

року, майже ніколи не покращує прогнози обмеженого збору даних. Внаслідок нестабільності їх специфікацій у різні моменти часу методи регуляризації передбачають відносно низьку якість. Тому, що докризові спостереження дуже нестабільні, значення змінних незвичні для наступних змінних, і в результаті навчання на їх основі моделі починають гірше прогнозувати дані за останні періоди.

Також було визначено, що моделі прогнозування стикаються з двома проблемами: 1) вибір показників для аналізу (використання експертних знань є звичайним явищем, але не існує методології, яка могла б бути використана для чіткої демонстрації її обґрунтованості відбору показників для аналізу); 2) незбалансований набір показників історичних даних, що використовуються для побудови моделей.

Тому розробка комплексних моделей прогнозування, а також вивчення інвестицій з використанням різнорідних статистичних даних є перспективною областю для подальших досліджень.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. Обсяг доходів населення  
[https://ukrstat.gov.ua/operativ/operativ2008/gdn/dvn\\_ric/dvn\\_ric\\_u/dn\\_reg2013\\_u.html](https://ukrstat.gov.ua/operativ/operativ2008/gdn/dvn_ric/dvn_ric_u/dn_reg2013_u.html) (дата звернення: 05.10.2022).
2. Безробіття. Державна служба статистики України/URL: [https://ukrstat.gov.ua/operativ/menu/menu\\_u/rp.htm](https://ukrstat.gov.ua/operativ/menu/menu_u/rp.htm) (дата звернення: 05.10.2022).
3. Білоусова О. С. Трансформація впливу держави на інвестиційні процеси. Інвестиції: практика та досвід. 2017. № 16. С. 32–37. 2. URL: [http://www.investplan.com.ua/pdf/16\\_2017/8.pdf](http://www.investplan.com.ua/pdf/16_2017/8.pdf) (дата звернення: 15.11.2022).
4. Богдан І. Прогнозування нагромадження основного капіталу К. : ДННУ «Акад. фін. Управління», 2011. 200 с. URL: [https://afu.kiev.ua/getfile.php?page\\_id=453&num=6](https://afu.kiev.ua/getfile.php?page_id=453&num=6) (дата звернення: 05.10.2022).
5. Валовий внутрішній продукт (ВВП). URL: <https://index.minfin.com.ua/ua/economy/gdp/> (дата звернення: 05.10.2022).
6. Васюник Т.І.. Аналіз інвестиційного забезпечення відтворення основного капіталу. Серія Економічні науки. URL: [http://www.ej.kherson.ua/journal/economic\\_22/1/12.pdf](http://www.ej.kherson.ua/journal/economic_22/1/12.pdf) (дата звернення: 15.11.2022).
7. Венгер В.В. Інвестиції в основний капітал як фактор економічного зростання металургійної галузі України. Міжнародні економічні відносини та світове господарство. URL: [http://www.irbis-nbuv.gov.ua/cgi-bin/irbis\\_nbuv/cgiirbis\\_64.exe?I21DBN=LINK&P21DBN=UJRN&Z21ID=&S21REF=10&S21CNR=20&S21STN=1&S21FMT=ASP\\_meta&C21COM=S&2\\_S21P03=FILA=&2\\_S21STR=Nvuumevcg\\_2016\\_8\(1\)\\_9](http://www.irbis-nbuv.gov.ua/cgi-bin/irbis_nbuv/cgiirbis_64.exe?I21DBN=LINK&P21DBN=UJRN&Z21ID=&S21REF=10&S21CNR=20&S21STN=1&S21FMT=ASP_meta&C21COM=S&2_S21P03=FILA=&2_S21STR=Nvuumevcg_2016_8(1)_9) (дата звернення: 26.10.2022).

8. Видатки зведеного бюджету України. URL : <https://index.minfin.com.ua/ua/finance/budget/cons/expense/> (дата звернення: 05.10.2022).
9. Державний борг (в середньому за квартал). URL: <https://index.minfin.com.ua/ua/finance/debtgov/> (дата звернення: 05.10.2022).
10. Експорт. URL: <https://index.minfin.com.ua/ua/economy/gdp/eximp/> (дата звернення: 05.10.2022).
11. Загородній А.Г. , Ю.І.Стадницький . Менеджмент реальних інвестицій: Навч. посіб. К.: Т-во «Знання», КОО, 2000. 209с.. URL: [http://znc.com.ua/ukr/publ/info/econ2003/zagorodniy\\_anatoliy\\_grygorovych.php](http://znc.com.ua/ukr/publ/info/econ2003/zagorodniy_anatoliy_grygorovych.php) (дата звернення: 05.10.2022).
12. Закон України «Про затвердження Політики перегляду статистичної інформації органами державної статистики України» URL: <https://zakon.rada.gov.ua/rada/show/v0283832-19> (дата звернення: 05.11.2022).
13. Закон України «Про інвестиційну діяльність». URL: <https://zakon.rada.gov.ua/laws/show/1560-12#Text>
14. І. М. Крейдич. Прогнозування прямих іноземних інвестицій та інвестицій в основний капітал в Україні. Теорія інвестицій URL: [http://www.investplan.com.ua/pdf/12\\_2009/3.pdf](http://www.investplan.com.ua/pdf/12_2009/3.pdf) (дата звернення: 03.10.2022).
15. Імпорт. URL: <https://index.minfin.com.ua/ua/economy/gdp/eximp/> (дата звернення: 05.10.2022).
16. Індекс UX (індекс Української біржі) (на кінець кварталу). URL: <https://index.minfin.com.ua/ua/markets/stock/ux/> (дата звернення: 05.10.2022).
17. Індекс обсягу сільськогосподарського виробництва. Держстат України. URL: [https://ukrstat.gov.ua/operativ/operativ2020/sg/sg\\_rik/arch\\_iosv.htm](https://ukrstat.gov.ua/operativ/operativ2020/sg/sg_rik/arch_iosv.htm) (дата звернення: 05.10.2022).
18. Індекс реальної заробітної плати. URL: <https://index.minfin.com.ua/ua/labour/salary/index/> (дата звернення: 05.10.2022).

19. Індекс споживчих цін.  
URL: <https://index.minfin.com.ua/ua/economy/index/inflation/> (дата звернення: 05.10.2022).
20. Індекс споживчих цін. Державна служба статистики України. URL: <https://ukrstat.gov.ua/> (дата звернення: 05.10.2022).
21. Індекс тарифів на залізничні перевезення. URL: <https://index.minfin.com.ua/ua/economy/index/railprice/> (дата звернення: 05.10.2022).
22. Індекс цін виробників промислових товарів. URL: <https://index.minfin.com.ua/ua/economy/index/prodprice/> (дата звернення: 05.10.2022).
23. Індекс цін на будівельно-монтажні роботи. URL: <https://index.minfin.com.ua/ua/economy/index/buildprice/> (дата звернення: 05.10.2022).
24. Капітальні інвестиції за джерелами фінансування. Державна служба статистики України. URL: [http://www.ukrstat.gov.ua/operativ/operativ2018/ibd/kindj/infin2018\\_u.htm](http://www.ukrstat.gov.ua/operativ/operativ2018/ibd/kindj/infin2018_u.htm) (дата звернення: 20.10.2022).
25. Кузьмович П. Економічна сутність капітальних інвестицій в контексті бухгалтерського обліку. URL: <http://dspace.tneu.edu.ua/bitstream/316497/12839/1/%D0%9A%D1%83%D0%B7%D1%8C%D0%BC%D0%BE%D0%B2%D0%B8%D1%87.pdf> . (дата звернення: 07.09.2020).
26. Курс долара на Міжбанку (на кінець кварталу). URL: <https://minfin.com.ua/ua/currency/mb/> (дата звернення: 05.10.2022).
27. M0. URL: <https://take-profit.org/statistics/money-supply-m1/ukraine/> (дата звернення: 05.10.2022).

28. M2 URL: <https://take-profit.org/statistics/money-supply-m1/ukraine/> (дата звернення: 05.10.2022).
29. Назаренко О. М. Імітація та прогнозування нестационарних фінансових часових рядів. Економічна кібернетика. 2011. № 4-6. С. 74-82
30. Національний стандарт № 3 «Оцінка цілісних майнових комплексів»: Постанова Кабінету Міністрів України від 29.11.2006 № 1655. URL: <https://zakon.rada.gov.ua/rada/show/1655-2006> (дата звернення: 28.11.2022).
31. Номінальний ефективний обмінний курс. URL: <https://minfin.com.ua/ua/2021/04/27/63979170/> (дата звернення: 05.10.2022).
32. О. Ф. Ярмолюк, О. М. Дмитренко. Особливості ідентифікації капітальних інвестицій в бухгалтерському обліку. URL: [http://www.economy.nayka.com.ua/pdf/9\\_2020/60.pdf](http://www.economy.nayka.com.ua/pdf/9_2020/60.pdf) (дата звернення: 06.10.2022).
33. Обсяг реалізованої промислової продукції <http://kh.ukrstat.gov.ua/obsiah-realizovanoi-produktsii-tovariv-posluh-subiektiv-hospodariuvannia-za-vydamy-ekonomichnoi-diialnosti> (дата звернення: 05.10.2022).
34. Офіційний курс долара (на кінець кварталу). URL: <https://minfin.com.ua/ua/currency/nbu/usd/> (дата звернення: 05.10.2022).
35. Паливода К.В. Капітальні інвестиції (на прикладі житлового будівництва в Україні): монографія. К.: «Знання», 2009. 711 с. (дата звернення: 07.10.2022).
36. Прибутковість 6-місячних державних облігацій. URL: <https://bank.gov.ua/ua/markets/ovdp-table> (дата звернення: 05.10.2022).
37. Присенко Г. В. Прогнозування соціально-економічних процесів. К.: КНЕУ, 2005. 378 с. URL: <http://ubooks.com.ua/books/000269/inx.php>
38. Про затвердження Методики розрахунку індексу капітальних інвестицій. URL: <https://zakon.rada.gov.ua/rada/show/v0303832-21#Text>

39. Про затвердження Методичних рекомендацій з бухгалтерського обліку для суб'єктів державного сектору. URL: <https://zakon.rada.gov.ua/rada/show/v0011201-15#Text>

40. Про затвердження Методичних рекомендацій щодо заповнення форм фінансової звітності. URL: <https://zakon.rada.gov.ua/rada/show/v0433201-13#Text>

41. Про затвердження Національного положення (стандарту) бухгалтерського обліку 1 «Загальні вимоги до фінансової звітності». URL: <https://zakon.rada.gov.ua/laws/show/z0336-13#Text>

42. Про затвердження Національного стандарту N 3 «Оцінка цілісних майнових комплексів». URL: <https://zakon.rada.gov.ua/laws/show/1655-2006-%D0%BF#Text>

43. Про схвалення Стратегії реформування системи управління державними фінансами на 2017-2020 роки. URL: <https://zakon.rada.gov.ua/laws/show/142-2017-%D1%80#Text>

44. Прострочена дебіторська заборгованість. URL: <https://aktiv.ua/ua/materials/articles/prosrochennaya-debitorskaya-zadolzhennost>

45. Прострочена кредиторська заборгованість. URL: <https://www.me.gov.ua/News/Detail?lang=uk-UA&id=026440bf-26a5-48f0-a656-25a8ba02d0c4&title=ZaborgovanistPoZarplatiVUkrainiZaPivtoraMisiatsiSkorotilasMaizheNa800-Mln-Grn-ZavdiakiPidprimstvamDerzhsektoru> (дата звернення: 05.10.2022).

46. Радевич Т. В. Нормативно-правове регулювання сутності капітальних інвестицій. Інвестиції : практика та досвід. К. : ТОВ «ДКС Центр», 2014. № 23. С. 64-68. URL: [http://www.investplan.com.ua/pdf/23\\_2014/17.pdf](http://www.investplan.com.ua/pdf/23_2014/17.pdf) (дата звернення: 07.09.2020).

47. Слово і Діло. «Динаміка притоку капітальних інвестицій в економіку України за останні шість років». 2020 URL:



<https://www.slovoidilo.ua/2020/10/01/infografika/ekonomika/dynamika-prytoku-kapitalnykh-investytsiy-ekonomiku-ukrayiny-ostanni-shist-rokiv> (дата звернення: 20.11.2022).

48. Ставка міжбанківського ринку, 1 день. URL: <https://index.minfin.com.ua/ua/banks/mb/index> (дата звернення: 05.10.2022).

49. Статистичний щорічник України 2014. Державна служба статистики України: веб-сайт. URL: <http://www.ukrstat.gov.ua> (дата звернення: 16.10.2022).

50. Статистичний щорічник України 2015. Державна служба статистики України: веб-сайт. URL: <http://www.ukrstat.gov.ua> (дата звернення: 16.10.2022).

51. Статистичний щорічник України 2016. Державна служба статистики України: веб-сайт. URL: <http://www.ukrstat.gov.ua> (дата звернення: 16.10.2022).

52. Статистичний щорічник України 2017. Державна служба статистики України: веб-сайт. URL: <http://www.ukrstat.gov.ua> (дата звернення: 16.10.2022).

53. Статистичний щорічник України 2018. Державна служба статистики України: веб-сайт. URL: <http://www.ukrstat.gov.ua> (дата звернення: 16.10.2022).

54. Статистичний щорічник України 2019. Державна служба статистики України: веб-сайт. URL: <http://www.ukrstat.gov.ua> (дата звернення: 16.10.2022).

55. Статистичний щорічник України 2020. Державна служба статистики України: веб-сайт. URL: <http://www.ukrstat.gov.ua> (дата звернення: 16.10.2022).

56. Удалих, О.О. Управління інвестиційною діяльністю промислового підприємства: Навч. посіб. К.: Центр навчальної літератури, 2006. URL: <https://readbookz.net/pbooks/book-28/ua/> (дата звернення: 05.10.2022).

57. Яцюк Д. С. Проблеми короткострокового прогнозування макроекономічних показників. X Міжнародна науково-практична конференція

58. Ahrens, A, Aitken, C, Ditzen, J, Ersoy, E, Kohns, D & Schaffer, ME 2021, A Theory-based Lasso for TimeSeries Data. in N Ngoc Thach, V Kreinovich & ND Trung (eds), Data Science for Financial Econometrics. Studies in Computational Intelligence, vol. 898, Springer, pp. 3-36, 3rd International Econometric Conference

of Vietnam 2020, Ho-Chi-Minh City, Viet Nam, 14/01/20. URL: [https://doi.org/10.1007/978-3-030-48853-6\\_1](https://doi.org/10.1007/978-3-030-48853-6_1) (дата звернення: 15.11.2022).

59. Akintoye A., Skitmore M. Models of UK private sector quarterly construction demand. *Construction Management and Economics*. 1994;12(1):3–13. URL: [https://www.researchgate.net/publication/27466040\\_Models\\_of\\_UK\\_private\\_sector\\_quarterly\\_construction\\_demand](https://www.researchgate.net/publication/27466040_Models_of_UK_private_sector_quarterly_construction_demand) (дата звернення: 15.11.2022).

60. Al-Abdulrazag B.A. The determinants of private construction investment in Jordan: an empirical study (1972–1996). *Damascus University Journal*. 2003;19:15–35. URL: <https://www.tandfonline.com/doi/abs/10.1080/01446193.2011.570356?needAccess=true&journalCode=rcme20> (дата звернення: 25.11.2022).

61. Fan, J. and Lv, J. (2009). A Selective Overview of Variable Selection in High Dimensional Feature Space (Invited Review Article). *Statistica Sinica*, 20 (1), pp.101–148.

62. Askinadzi V. M. V. F. Maksimova Investments: a textbook for bachelors Urayt Publishing House, 2014. - 422 p.

63. Baghestani H. Forecasts of growth in US residential investment: accuracy gains from consumer home-buying attitudes and expectations. *Applied Economics*. 2021;52:3744–3758. DOI: <https://doi.org/10.1080/00036846.2021.1885613> (дата звернення: 25.11.2022).

64. Bai J., Ng S. Forecasting Economic Time Series Using Targeted Predictors // *Journal of Econometrics*. 2008. № 146(2). pp. 304–317.

65. Board of governors of the federal reserve system URL: <https://www.federalreserve.gov/econres/us-models-about.htm>

66. Brissimis S.N., Vlassopoulos T. The interaction between mortgage financing and housing prices in Greece. *Journal of Real Estate Finance and Economics*. 2009;39(2):146–164. URL:

[https://econpapers.repec.org/article/kapjrefec/v\\_3a39\\_3ay\\_3a2009\\_3ai\\_3a2\\_3ap\\_3a146-164.htm](https://econpapers.repec.org/article/kapjrefec/v_3a39_3ay_3a2009_3ai_3a2_3ap_3a146-164.htm) (дата звернення: 25.11.2022).

67. Brüggemann R. Sources of German unemployment: a structural vector error correction analysis. *Empirical Economics*. 2006;31(2):409–431. URL: [https://www.researchgate.net/publication/24055116\\_Sources\\_of\\_German\\_unemployment\\_A\\_structural\\_vector\\_error\\_correction\\_analysis](https://www.researchgate.net/publication/24055116_Sources_of_German_unemployment_A_structural_vector_error_correction_analysis) (дата звернення: 05.11.2022).

68. Colin Lawrence and Aloysius Siow. Interest Rates and Investment Spending: Some Empirical Evidence for Postwar U.S. Producer Equipment, 1947-1980. *The Journal of Business*, Vol. 58, No. 4 (Oct., 1985), pp. 359-375 URL: <https://www.jstor.org/stable/2352831> (дата звернення: 08.10.2022).

69. Dale W. Jorgenson «Capital Theory and Investment Behavior,» Chapter 1 (1996a), pp. 1-16. URL: <http://gesd.free.fr/jorg63.pdf> (05.11.2022).

70. Description. `dm.test`: Diebold-Mariano test for predictive accuracy. RDocumentation. Search all packages and functions URL: <https://www.rdocumentation.org/packages/forecast/versions/8.5/topics/dm.test>

71. Diebold F. X., Mariano R. S. Comparing Predictive Accuracy // *Journal of Business and Economic Statistics*. – 1995. – № 13(3). – pp. 253–263. doi: 10.1080/07350015.1995.10524599

72. Diebold F. X., Mariano R. S. Comparing Predictive Accuracy // *Journal of Business and Economic Statistics*. – 2002. – № 20(1). – pp. 134–144. DOI: 10.1198/073500102753410444

73. Elastic Net Regression in R Programming. URL: <https://www.geeksforgeeks.org/elastic-net-regression-in-r-programming/>

74. Erwin Dekker Jan Tinbergen (1903–1994) and the Rise of Economic Expertise Cambridge University Press, 2021, pp. xxi + 463 URL: <https://www.tandfonline.com/doi/full/10.1080/10370196.2022.2035931?scroll=top&needAccess=true>

75. Fan, J. and Lv, J. (2009). A Selective Overview of Variable Selection in High Dimensional Feature Space (Invited Review Article). *Statistica Sinica*, 20 (1), pp.101–148. URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3092303/> (дата звернення: 05.11.2022).

76. Fullerton T.M., Laaksonen M.M., West C.T. Regional multi-family housing start forecast accuracy. *International Journal of Forecasting*. 2001;17(2):171–180. URL: [https://www.researchgate.net/publication/222336910\\_Regional\\_multi-family\\_housing\\_start\\_forecast\\_accuracy](https://www.researchgate.net/publication/222336910_Regional_multi-family_housing_start_forecast_accuracy) (08.11.2022).

77. Fumio Hayashi . «The Cost of Capital, Q, and the Theory of Investment Demand,» Chapter 4 in Lawrence J. Lau (2000a), pp. 55-84. URL: <https://scholar.harvard.edu/jorgenson/content/research-2> (03.11.2022).

78. Gaulti B. Eggertsson. What Fiscal Policy is Effective at Zero Interest Rates?. Staff Report 402. Federal Reserve Bank of New York 2009. URL: [https://www.newyorkfed.org/research/staff\\_reports/sr402.html](https://www.newyorkfed.org/research/staff_reports/sr402.html)

79. Genuer, R., Poggi, J.-M. and Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31 (14), pp.2225–2236. DOI: <http://dx.doi.org/10.1016/j.patrec.2010.03.014> (дата звернення: 11.11.2022).

80. Genuer, R., Poggi, J.-M. and Tuleau-Malot, C. (2010). Variable selection using random forests. *Pattern Recognition Letters*, 31 (14), pp.2225–2236.

81. Genuer, R., Poggi, J.-M. and Tuleau-Malot, C. (2015). VSURF: An R Package for Variable Selection Using Random Forests. *The R Journal*, 7 (2), pp.19–33. DOI: <https://doi:10.32614/RJ-2015-018> . (дата звернення: 05.11.2022).

82. Genuer, R., Poggi, J.-M. and Tuleau-Malot, C. (2015). VSURF: An R Package for Variable Selection Using Random Forests. *The R Journal*, 7 (2), pp.19–33..

83. Goh B.H., Teo H.P. Forecasting construction industry demand, price and productivity in Singapore: the Box-Jenkins approach. *Construction Management and*

Economics. 2000;18(5):607–618. URL: [https://www.researchgate.net/publication/24077454\\_Forecasting\\_construction\\_industry\\_demand\\_price\\_and\\_productivity\\_in\\_Singapore\\_The\\_Box-Jenkins\\_approach](https://www.researchgate.net/publication/24077454_Forecasting_construction_industry_demand_price_and_productivity_in_Singapore_The_Box-Jenkins_approach) (05.11.2022).

84. Green K.C., Armstrong J.S. Simple versus Complex Forecasting: The Evidence. Journal of Business Research. 2015;68(8):1678–1685. <http://dx.doi.org/10.1016/j.jbusres.2015.03.026> (04.11.2022).

85. Hastie, T., Tibshirani, R. and Tibshirani, R. (2017). Best Subset, Forward Stepwise or Lasso? Analysis and Recommendations Based on Extensive Comparisons. Statistical Science, 35 (4), pp.579–592.

86. Hodgins C.D., Tanner J.E. Forecasting nonresidential building construction. Canadian Journal of Economics. 1973;6(1):79–89. URL: <https://ideas.repec.org/a/cje/issued/v6y1973i1p79-89.html> (04.11.2022).

87. Hui Zou, Trevor Hastie . Regularization and variable selection via the elastic net. 2005. URL: <https://doi.org/10.1111/j.1467-9868.2005.00503> (дата звернення: 25.10.2022).

88. J. Friedman et.al., Springer “The element of statistical learning”;, Springer, pages- 79-91, 2008. URL: <https://link.springer.com/978-0-387-21606-5> (дата звернення: 25.11.2022).

89. James, G. et al. (2013). An introduction to statistical learning. Springer.URL: [https://www.academia.edu/36691506/An\\_Introduction\\_to\\_Statistical\\_Learning\\_Springer\\_Texts\\_in\\_Statistics\\_An\\_Introduction\\_to\\_Statistical\\_Learning](https://www.academia.edu/36691506/An_Introduction_to_Statistical_Learning_Springer_Texts_in_Statistics_An_Introduction_to_Statistical_Learning) (04.11.2022)

90. James, G. et al. (2013). An introduction to statistical learning. Springer.

91. Jiahan Li, Weiye Chen. Forecasting macroeconomic time series: LASSO-based approaches and their forecast combinations with dynamic factor models .Weiye Chen Department of Applied and Computational Mathematics and

Statistics, University of Notre Dame, 153 Hurley Hall, Notre Dame, IN 46556, USA.  
URL: <https://isiarticles.com/bundles/Article/pre/pdf/45737.pdf> (04.11.2022).

92. Jiang S., Yang C., Guo J., Ding Z. ARIMA forecasting of China's coal consumption, price and investment by 2030. *Energy Sources, Part B: Economics, Planning, and Policy*. 2018;13:190–195.  
<http://dx.doi.org/10.1080/15567249.2017.1423413> (04.11.2022).

93. Kenton W. Capital Investment. Investopedia. URL: <https://www.investopedia.com/terms/c/capital-investment.asp> (дата звернення: 25.11.2022)

94. Killingsworth R.A. Preliminary investigation into formulating a demand forecasting model for industrial construction. *Cost Engineering*. 1990;32(8):11–15.

95. Kitti Limskul. Economic Growth and Income Distribution: A Debate on Capital Accumulation in Thailand. Conference: Shakai Kagaku Ronshu 161 June 2020At: Saitama University Economic Conference. URL: [https://www.researchgate.net/publication/347932821\\_Economic\\_Growth\\_and\\_Income\\_Distribution\\_A\\_Debate\\_on\\_Capital\\_Accumulation\\_in\\_Thailand](https://www.researchgate.net/publication/347932821_Economic_Growth_and_Income_Distribution_A_Debate_on_Capital_Accumulation_in_Thailand) (04.11.2022).

96. Kocherlakota, Narayana R. Model fit and model selection. *Federal Reserve Bank of St. Louis Review* 89. 2007. URL: <https://files.stlouisfed.org/files/htdocs/publications/review/07/07/Kocherlakota.pdf>

97. Macroeconomic Models and Problems of Forecasting. *Econometrica*, pp. 31-33 (3 pages) Published By: The Econometric Society. URL: <https://www.jstor.org/stable/1910033>

98. Mankiw NG . *Macroeconomics*, 7th Edition. Worth Publishers;

99. Mankiw NG. The Macroeconomist as Scientist and Engineer. *Journal of Economic Perspectives*. 2006;20 (4) :29-46

100. Mayasami R.C., Koh T.S. A vector error correction model of the Singapore stock market. *International Review of Economics and Finance*.

2000;9(1):79–96.URL:

<https://econpapers.repec.org/RePEc:eee:revecov:9:y:2000:i:1:p:79-96> (06.11.2022).

101. Mohseni S., Brent A.C., Kelly S., Browne W.N. Demand response-integrated investment and operational planning of renewable and sustainable energy systems considering forecast uncertainties: A systematic review. *Renewable and Sustainable Energy Reviews*. 2022;158:112095. <https://doi.org/10.1016/j.rser.2022.112095> (06.11.2022).

102. Nadine McCloud. Does domestic investment respond to inflation targeting? A synthetic control investigation. *International Economics*. Volume 169, 2022, Pages 98-134. DOI: <https://doi.org/10.1016/j.inteco.2021.12.002> (дата звернення: 11.10.2022).

103. Narayana Kocherlakota; National Bureau of Economic Research. *Fragility of Purely Real Macroeconomic Models..* 2016. URL: <https://www.worldcat.org/title/fragility-of-purely-real-macroeconomic-models/oclc/936371796>

104. Ng S.T., Fan R.Y.C., Wong J.M.W. An econometric model for forecasting private construction investment in Hong Kong. *Construction Management and Economics*. 2011;29(5):519–534. <https://doi.org/10.1080/01446193.2011.570356> (06.11.2022).

105. Ng S.T., Skitmore M., Wong K.F. Genetic algorithm and linear regression for private housing demand forecast. *Building and Environment*. 2008;43(6):1171–1184. URL: [https://scholars.cityu.edu.hk/en/publications/using-genetic-algorithms-and-linear-regression-analysis-for-private-housing-demand-forecast\(acd35484-2977-422b-b3ed-7ea078e7c9f7\).html](https://scholars.cityu.edu.hk/en/publications/using-genetic-algorithms-and-linear-regression-analysis-for-private-housing-demand-forecast(acd35484-2977-422b-b3ed-7ea078e7c9f7).html) (06.11.2022).

106. Nijat Mehdiyev, David Enke. Evaluating Forecasting Methods by Considering Different Accuracy Measures *Procedia Computer Science* 95(4):264-271 DOI:10.1016/j.procs.2016.09.332

107. Prashanth Ashok. What is Ridge Regression?. 2022 URL: <https://www.mygreatlearning.com/blog/what-is-ridge-regression/> (дата звернення: 15.10.2022). (06.11.2022).

108. Sala-I-Martin, X. X. (1997). I Just Ran Two Million Regressions. *The American Economic Review*, 87 (2), pp.178–183. URL: [https://webdoc.agsci.colostate.edu/koontz/arec-econ535/papers/Sala-I-Martin%20\(AER%201997\).pdf](https://webdoc.agsci.colostate.edu/koontz/arec-econ535/papers/Sala-I-Martin%20(AER%201997).pdf) (06.11.2022).

109. Sala-I-Martin, X. X. (1997). I Just Ran Two Million Regressions. *The American Economic Review*, 87 (2), pp.178–183.

110. Sen P., Roy M., Pal P. Application of ARIMA for forecasting energy consumption and GHG emission: a case study of an Indian pig iron manufacturing organization. *Energy*. 2016;116:1031–1038.

111. Tan H.B., Baharumshah A.Z. Dynamic causal chain of money, output, interest rate and prices in Malaysia: evidence based on vector error-correction modelling analysis. *International Economic Journal*. 1999;13(1):103–120.

112. Tang J.C.S., Karasudhi P., Tachopiyagoon P. Thai construction industry: demand and projection. *Construction Management and Economics*. 1990;8(3):249–257.

113. The World Bank. Gross fixed capital formation. URL: <https://data.worldbank.org/indicator/NE.GDI.FTOT.ZS?end=2021&start=1960&view=chart> (дата звернення: 15.11.2022).

114. Tse R.Y.C., Ho C.W., Ganesan S. Matching housing supply and demand: an empirical study of Hong Kong's market. *Construction Management and Economics*. 1999;17(5):625–633. URL: <https://econpapers.repec.org/RePEc:taf:conmgt:v:17:y:1999:i:5:p:625-633> (06.11.2022).

115. Van Aarle B., Boss M., Hlouskova J. Forecasting the Euro exchange rate using vector correction models. *Review of World Economics*.



2000;136(2):232–258.URL: <https://link.springer.com/article/10.1007/BF02707687>  
(06.11.2022).

116. Wei Fan. Prediction of Monetary Fund Based on ARIMA Model.Procedia Computer Science.Volume 208,2022, Pages 277-285 DOI: <https://doi.org/10.1016/j.procs.2022.10.040> . (06.11.2022).

117. Yaniv Konchitchki, Panos N. Patatoukas. Accounting earnings and gross domestic product. Journal of Accounting and Economics, Volume 57, Issue 1, 2014, Pages 76-88. URL: <https://www.sciencedirect.com/science/article/pii/S016541011300058X> (дата звернення: 17.11.2022).

118. Yatsiuk D. «What are the main factors influencing the volume of capital investments in Ukraine?». Medium. 2022. URL: <https://medium.com/@dariia.yatsiuk/what-are-the-main-factors-influencing-the-volume-of-capital-investments-in-ukraine-d92f993c014d>

119. Yeboah S., Ohene M., Wereko T.B. Forecasting aggregate and disaggregate energy consumption using ARIMA models: a literature survey. Journal Statistical Economics Methods. 2012;1(2):71–79. URL: [https://www.researchgate.net/publication/286418448\\_Forecasting\\_aggregate\\_and\\_disaggregate\\_energy\\_consumption\\_using\\_arima\\_models\\_A\\_literature\\_survey](https://www.researchgate.net/publication/286418448_Forecasting_aggregate_and_disaggregate_energy_consumption_using_arima_models_A_literature_survey)  
(06.11.2022).

120. Anis Omri, Bassem Kahouli. The nexus among foreign investment, domestic capital and economic growth: Empirical evidence from the MENA region. Research in Economics, Volume 68, Issue 3, 2014, Pages 257-263. URL: <https://www.sciencedirect.com/science/article/pii/S1090944313000537> (дата звернення: 11.11.2022)

## ДОДАТКИ

## ДОДАТОК А

## Список змінних, їх джерела

Залежні змінні	Джерела
Валове накопичення основного капіталу	Держстат [23]
ВВП в постійних цінах	Мінфін[5]
ВВП	Мінфін [5]
Витрати зведеного бюджету	Мінфін [8]
Експорт	Мінфін [10]
Імпорт	Мінфін [15]
Індекс UX (індекс Української біржі) (на кінець кварталу)	Мінфін [16]
Індекс реальної заробітної плати	Мінфін [18]
Індекс обсягу сільськогосподарського виробництва	Держстат [17]
Індекс споживчих цін	Мінфін [19]
Індекс тарифів на вантажні перевезення	Мінфін [21]
Індекс цін виробників промислових товарів	Мінфін [21]
Індекс цін на будівельно-монтажні роботи	Мінфін [23]
Державний борг (в середньому за квартал)	Мінфін [9]
Курс долара на Міжбанку (на кінець кварталу)	Мінфін [26]
М0 (на кінець кварталу)	Мінфін [27]
М2 (на кінець кварталу)	Мінфін [28]
Номінальний ефективний обмінний курс (на кінець кварталу)	Мінфін [31]
Безробіття (в середньому за квартал)	Держстат [1]
Офіційний курс долара (на кінець кварталу)	Мінфін [34]
Прибутковість 6-місячних державних облігацій (в	Нацбанк,

середньому за квартал)	мінфін[35]
Прострочена дебіторська заборгованість (в середньому за квартал)	[42]
Прострочена кредиторська заборгованість (в середньому за квартал)	[43]
Ставка міжбанківського ринку, 1 день (в середньому за квартал)	Мінфін [46]
Ставка міжбанківського ринку, 7 днів (в середньому за квартал)	Мінфін[46]
Частка валового накопичення основного капіталу у ВВП	Розрахунки автора відповідно до джерел :[5],[24]

## ДОДАТОК Б

## Значення статистики Дібольда-Маріано

Модель	1	2	3	4	5	6	7	8
AR	-0.08 2 0.195	-0.24 3. (0.08 1)	-0.41 3* (0.01 8)	-0.441 ** (0.002 )	-0.477 ** (0.002 )	-0.458 ** (0.008 )	-0.337 * (0.021 )	-0,27 1 (0.02 1)
Adaptive LASSO	-0.00 4. (0.08 0)	0.264 * (0.02 5)	-0.18 1. (0.07 3)	-0.002 (0.186 )	-0.144 * (0.015 )	-0.121 (0.461 )	-0.013 (0.490 )	-0.12 8 (0.12 6)
Elastic Net	0.218 (0.41 2)	0.230 * (0.04 6)	-0.11 6 (0.12 8)	0.041 (0.453 )	-0.050 (0.421 )	-0.072 (0.379 )	0.027 (0.249 )	-0.03 7 (0.25 0)
LASSO	0.104 (0.47 2)	0.309 * (0.03 7)	-0.15 3 (0.11 3)	0.074 (0.343 )	-0.056 (0.403 )	-0.102 (0.421 )	0.055 (0.130 )	-0.02 1 (0.28 8)
Post-LASSO	0.105 (0.29 8)	0.229. (0.05 4)	-0.22 9 (0.12 6)	0.022 (0.176 )	-0.243 * (0.014 )	-0.201 (0.325 )	-0.025 (0.298 )	-0.09 5 (0.12 8)
Ridge	0.098	0.240 *	0.087	0.092	0.041	-0.005	0.025	-0.05 1

	(0.36 5)	(0.05 0)	(0.24 6)	(0.189 )	(0.162 )	(0.249 )	(0.484 )	(0.21 7)
Spike and Slab	0.142 (0.47 9)	0.207. (0.06 3)	0.022 (0.30 2)	-0.013 (0.418 )	-0.139 * (0.045 )	-0.083 (0.381 )	-0.024 (0.292 )	-0.02 1 (0.44 2)
boosting ( $\eta =$ 0, 1)	-0.05 8 (0.10 9)	0.100. (0.08 3)	0.113 (0.10 7)	0.066. (0.069 )	-0.011 (0.421 )	-0.054. (0.083 )	-0.147 ** (0.006 )	0.178 * (0.03 4)
boosting ( $\eta =$ 0, 2)	-0.03 3 (0.17 5)	0.131. (0.06 2)	0.139 * (0.02 3)	0.132 (0.128 )	-0.011 (0.493 )	-0.162 * (0.050 )	-0.136 ** (0.006 )	0.024 (0.13 4)
boosting ( $\eta =$ 0, 3)	-0.13 6* (0.04 2)	0.097 (0.22 0)	0.022 (0.22 4)	0.047. (0.055 )	0.099 (0.112 )	-0.064 (0.117 )	-0.091 (0.120 )	0.082 (0.11 4)
boosting ( $\eta =$ 0, 4)	-0.06 7 (0.10 2)	0.048 (0.20 7)	0.092 (0.17 0)	0.126. (0.089 )	-0.061. (0.079 )	-0.118 * (0.021 )	-0.081 (0.325 )	0.091 (0.18 5)
random forest ( $N =$ 100)	0.156 (0.26 9)	0.096 * (0.04 1)	0.024 (0.27 5)	0.026 (0.124 )	0.032 (0.203 )	-0.071 (0.269 )	-0.076 (0.219 )	0.009 (0.14 7)
random forest ( $N =$ 500)	0.074 (0.44 8)	0.096 * (0.03 6)	0.026 (0.35 3)	0.019. (0.093 )	-0.011 (0.392 )	-0.052 (0.200 )	-0.071. (0.079 )	-0.02 0 (0.35 6)
random forest ( $N =$ 1000)	0.114 (0.30 8)	0.075 * (0.04 0)	0.034 (0.27 9)	0.029. (0.093 )	-0.005 (0.486 )	-0.044 (0.224 )	-0.047 (0.256 )	-0.01 9 (0.36 9)
random forest ( $N =$ 2000)	0.124	0.083 *	0.027	0.027	0.008	-0.044	-0.040	-0.01 3

	(0.22 7)	(0.03 7)	(0.33 7)	(0.101 )	(0.385 )	(0.213 )	(0.199 )	(0.33 5)
--	-------------	-------------	-------------	-------------	-------------	-------------	-------------	-------------

## ДОДАТОК В

## Скрипт реалізації методів прогнозування

```
library(wbstats)
library(dplyr)
library(data.table)
library(rio)
library(stringr)
library(magrittr)
library(purrr)
library(ggplot2)
library(latex2exp)
library(stargazer)
library(glmnet)
library(corrplot)
library(dplyr)
library(lubridate)
library(rio)
library(stringr)
library(tibble)
```

```
library(magrittr)
library(purrr)
library(data.table)
library(tidyr)
library(forecast)
library(gridExtra)
library(grid)
library(Amelia)
library(fastDummies)
library(gnm)
library(parallel)
library(xgboost)
library(rbenchmark)
#library(plotly)

library(e1071)
library(rlang)
library(caret)
library(DiagrammeR)
library(sophisthse)
library(DiagrammeRsvg)
library(rsvg)
library(neuralnet)
library(ggribes)
library(tseries)
library(xts)
library(DtD)
library(readr)
```

library(TTR)  
library(zoo)  
library(seqHMM)  
library(hexbin)  
library(sophisthse)  
library(tseries)  
library(caret)  
library(sophisthse)  
library(tseries)  
library(MLmetrics)  
library(caret)  
library(xts)  
library(randomForest)  
library(spikeslab)  
library(BigVAR)  
library(xtable)  
library(readxl)  
library(httr)  
library(ggpubr)  
library(shiny)  
library(scales)  
library(DT)  
library(lemon)  
library(shinyWidgets)  
library(dplyr)  
library(multDM)  
library(tables)  
library(xtable)



```
library(animation)
library(gganimate)
library(multDM)
```

```
# import rawdata
load('raw.Rdata')
```

```
# # Індекс UX (індекс Української біржі) (на кінець кварталу), Ставка
міжбанківського ринку, 1 день (в середньому за квартал), прибутковість
6-місячних державних облігацій (в середньому за квартал),
```

```
# темпи росту номінального ефективного обмінного курсу (на кінець
кварталу), офіційного курс долара (на кінець кварталу),
```

```
# extended series ----
```

```
eseries <- c('investment', 'mkr_1d', 'mkr_7d', 'gov_6m', 'GKO',
            , 'neer', 'rts',
```

```
'GDPEA_C_Q', 'INVFC_Q', # службові для invest2gdp
```

```
'CPI_Q_CHI', # Індекс споживчих цін
```

```
'GDPEA_Q_DIRI', # ввп
```

```
'UNEMPL_Q_SH', # безробіття
```

```
'WAG_Q', # індекс реальної заробітньої плати
```

```
'CONI_Q_CHI', # Індекс цін на будівельно-монтажні роботи
```

```
'CTI_Q_CHI', # Індекс тарифів на вантажні перевезення
```

```
'AGR_Q_DIRI', # Індекс обсягу сільськогосподарського
```

```
виробництва
```

```
'TRD_Q_DIRI', # Індекс цін виробників промислових товарів
```

```

'HHI_Q_DIRI',# индекс реальных денежных доходов населения
'M0_Q', # M0
'M2_Q',# M2
'DEXRO_Q',# офіційний курс долара
'DEXRM_Q',# Курс долара на Міжбанку (на кінець кварталу)
'EX_T_Q',# экспорт
'IM_T_Q',# імпорт
'PPI_EA_Q' # (після 2004-01)
)

```

```

df <- rawdata[,eseries]
# Частка валового накопичення основного капіталу у ВВП
df$invest2gdp <- df$INVFC_Q/df$GDPEA_C_Q

# видалити лишнє
df$GDPEA_C_Q <- df$INVFC_Q <- NULL

# вставити gko в 6m gov bond yield
df$gov_6m[which(is.na(df$gov_6m))] <-
  df$GKO[which(is.na(df$gov_6m))]

df$GKO <- NULL

# нестационарні данні без сезонності ----
for(i in c('eer','neer','')){
  df[,i] %<>% diff.xts(log=TRUE)
}

```

```

# нестационарні данні із сезонністю ----
for(i in c('investment', 'CPI_Q_CHI',
          'invest2gdp',
          # 'deflator', із 1996
          'GDPEA_Q_DIRI',

          ,
          'UNEMPL_Q_SH',

          ##### 'TRP_Q_PASS_DIRI',
          'WAG_Q',
          'CONI_Q_CHI',
          'CTI_Q_CHI',
          'AGR_Q_DIRI',
          'TRD_Q_DIRI',
          'HHI_Q_DIRI',
          'M0_Q',
          'M2_Q',
          ##### 'IR_Q',
          ##### 'ICR_Q',
          'CBREV_Q',
          'FBREV_Q'
          'DEXRO_Q',# офіційний курс долару
          'DEXRM_Q',# курс доллара на ммвб
          'LIAB_T_Q',# Прострочена дебіторська заборгованість (в
середньому за квартал)

```

'LIAB\_UNP\_Q',# Прострочена кредиторська заборгованість (в середньому за квартал)

'EX\_T\_Q',# експорт

'IM\_T\_Q',# імпорт

'PPI\_EA\_Q' # (після 2004-01)

```
)) {
  df[,i] %<>% diff.xts(lag = 4, log=TRUE)
}
```

```
do.call.pipe <- function(args,what){
  do.call(what,args)
}
create_lag <- function(df, lag){
  out <- df
  if(lag > 0){
    for(i in 1:lag){
      x <- lag.xts(df, k = i)
      colnames(x) <- paste0(colnames(x), "_lag", i)
      out %<>% merge.zoo(x)
    }
  }

  return(out %>% as.xts)
```

```

}

train.model <- function(startdt= as.Date('1996-01-01'),
                        enddt = as.Date('2016-10-01'),
                        model,
                        # series parameter
                        # only for regularisation and machine learning models
                        series='e',
                        lag = 0L,
                        h = 1L,
                        target =
c('investment','TRD_Q_DIRI','GDPEA_Q_DIRI','UNEMPL_Q_SH', 'CPI_Q_CHI'),
                        # parameters for durable evaluations with function arguments
from expand.grid table
                        i = NULL,
                        N = NULL
){
  message(paste0(i, '/', N))
  # import df
  # seed
  set.seed(2019)

  if(series == ""){

    load('~/.investment_forecasting/data/stationary_data.RData')

  } else if(series == 'e') {

```

```

load('~investment_forecasting/data/stationary_data_ext.RData')

target <- match.arg(target)

if(!model %in% c('arima', 'rw')){
  if(!series %in% c("", 'e')){
    df %<>% df[, c(target,
                  series)]
  }
  # create lag
  if(lag != 0 |! model %in% c('arima', 'rw')){
    df %<>% create_lag(lag)
  }

  if(!is.integer(h)){
    message('h must be integer')
    return(NULL)
  }

  df$y <- lag.xts(df[,target], k = -h)

  if(h == 0){
    if(target == 'investment'){
      df$investment <-df$invest2gdp <- df$GDPEA_Q_DIRI <- NULL
    } else{
      df %<>% .[, colnames(.) != target]
    }
  }

```

```

} else{
  df$gdplag <- df$investmentlag <- df$invest2gdplag <- NULL
}

```

```
df %<>% .[rowSums(is.na(.[,colnames(.)!='y']))==0,]
```

```

if(startdt >= enddt){
  message('start must be greater then end')
  return(NULL)
}

```

```

startdt <- max(startdt %>% as.yearqtr,
               first(time(df)) %>% as.yearqtr) %>%
as.Date

```

```

enddt <- min(enddt %>% as.yearqtr,
             last(time(df)) %>% as.yearqtr) %>%
as.Date

```

```

if(df[paste0(startdt, "/", enddt)] %>% nrow < 48){
  message('train set length must be greater then 48 quarters')
  return(NULL)
}

```

```

        if(which((time(df) %>% as.Date())==(enddt%>% as.yearqtr%>%
as.Date))+1>length(df$y %>% na.omit)){
            #message('enddt must not be greater then last date when investment data is
available minus h quarters')
            return(NULL)
        }

```

```

train_n <- which(time(df) %>% as.Date())==(startdt %>% as.yearqtr %>%
as.Date)):

```

```

        which((time(df) %>% as.Date())==
            (enddt %>% as.yearqtr%>% as.Date))

```

```

test_n <- (which((time(df) %>% as.Date())==(enddt%>% as.yearqtr%>%
as.Date))+1):nrow(df)

```

```

df$y %<>% na.fill(0)

```

```

X.matrix <- model.matrix(y~0+., data = df)

```

```

X.train <- X.matrix[train_n,]

```

```

X.test <- X.matrix[test_n,]

```

```

y.train <- df$y[train_n] %>% as.numeric

```

```

y.test <- df$y[test_n] %>% as.numeric

```



```
# tc <- trainControl(method = 'cv', number = 10)
```

```
tc <- trainControl(method = "timeslice",  
  initialWindow = 40,  
  horizon = 1,  
  skip=0,  
  fixedWindow = TRUE)
```

```
lambda <- seq(-8,-1,length = 200) %>% exp()
```

```
if(model == 'lasso'){  
  train.out <- train(x=X.train,  
    y=y.train,  
    method = "glmnet",  
    metric = "RMSE",  
    trControl = tc,  
    tuneGrid =  
      expand.grid(.alpha = c(1),  
        .lambda = lambda))
```

```
model_fit <- glmnet(X.train,  
  y.train,  
  alpha = 1,
```

```
lambda = train.out$bestTune[1,2])

pred <- predict(model_fit,
  newx = rbind(X.train, X.test)) %>% as.numeric
}

else if(model == 'elnet'){
  train.out <- train(x=X.train,
    y=y.train,
    method = "glmnet",
    metric = "RMSE",
    trControl = tc,
    tuneGrid =
      expand.grid(.alpha = c(0.5),
        # new
        .lambda = lambda))

  model_fit <- glmnet(X.train,
    y.train,
    alpha = 0.5,
    lambda = train.out$bestTune[1,2])

  pred <- predict(model_fit,
    newx = rbind(X.train, X.test)) %>% as.numeric
}
```

```
else if(model == 'ridge'){
  train.out <- train(x=X.train,
                    y=y.train,
                    method = "glmnet",
                    metric = "RMSE",trControl = tc,
                    tuneGrid = expand.grid(.alpha = 0,
                                           .lambda = lambda))
  model_fit <- glmnet(X.train,
                     y.train,
                     alpha = 0,
                     lambda = train.out$bestTune[1,2])

  pred <- predict(model_fit,newx = rbind(X.train, X.test)) %>% as.numeric
}
else if (model == 'adalamo'){

  train.ridge <- train(x=X.train,
                      y=y.train,
                      method = "glmnet",
                      metric = "RMSE",trControl = tc,
                      tuneGrid = expand.grid(.alpha = 0,.lambda = lambda))

  m_ridge <- glmnet(X.train,
                   y.train,
                   alpha = 0,
                   lambda = train.ridge$bestTune[1,2])
```

```
w3 <- 1/abs(as.numeric(coef(m_ridge))
           [1:(ncol(X.train))] )^0.5 ## Using gamma = 1
w3[which(w3 == Inf)] <- 999999999 ## Replacing values estimated as
Infinite for 999999999
```

```
train.out <- train(x=X.train,
                  y=y.train,
                  penalty.factor = w3,
                  method = "glmnet",
                  metric = "RMSE",
                  trControl = tc,
                  tuneGrid = expand.grid(.alpha = 1,.lambda = lambda))
```

```
model_fit <- glmnet(X.train,
                   y.train,
                   alpha = 1,
                   lambda = train.out$bestTune[1,2],
                   penalty.factor = w3)
```

```
pred <- predict(model_fit,newx = rbind(X.train, X.test)) %>%
as.numeric
```

```
} else if (model == 'postlasso'){
```

```
train.out <- train(x=X.train,
```

```

y=y.train,
method = "glmnet",
metric = "RMSE",trControl = tc,
tuneGrid = expand.grid(.alpha = c(1),.lambda = lambda))

```

```

model_lasso <- glmnet(X.train,
  y.train,
  alpha = 1,
  lambda = train.out$bestTune[1,2])

```

```

nzvars <- predict(model_lasso, type = "nonzero") %>% .[[1]]

```

```

train_post <- as.data.frame(X.train) %>%
  select(nzvars) %>%
  mutate(y = y.train)

```

```

if(class(X.test)=='numeric'){
  X.test %<>% as.tibble %>% t %>% set_colnames(colnames(X.train))
}

```

```

test_post <- as.data.frame(X.test) %>%

```

```
select(nzvars) %>%
mutate(y = 0)

model_fit <- lm(y~., data = train_post)

pred <- predict(model_fit,newdata = rbind(train_post,test_post ))

} else if (model == 'rf'){

train.out <- NULL
model_fit <- randomForest(x = X.train,
                           y = y.train,
                           ntree = 100,
                           nodesize = 5)

pred <- predict(model_fit, newdata = rbind(X.train, X.test)) %>%
  as.numeric

} else if (model == 'ss'){

train.out <- NULL

model_fit <- spikeslab(x = X.train,
                       y = y.train,
                       n.iter2 = 500,
```

```
bigp.smalln = ncol(X.train)>=nrow(X.train),
intercept = TRUE)

pred <- predict(model_fit, newdata = rbind(X.train, X.test))$yhat.gnet

} else if (model == 'boost'){

tune_grid <- expand.grid(nrounds = 100,
                        max_depth = c(5),
                        eta = c(0.3),
                        gamma = 0,
                        colsample_bytree = 0.3,
                        min_child_weight = 1,
                        subsample = 1)

train.out <- NULL
model_fit <- train(x = X.train,
                  y = y.train,
                  method = "xgbTree",
                  metric = "RMSE",
                  tuneGrid = tune_grid)

pred <- predict(model_fit, newdata = rbind(X.train, X.test)) %>%
  as.numeric

}
} else
```

```
{

if(!is.integer(h)){
  message('h must be integer')
  return(NULL)
}

df %<>% .[, 'investment']

df$y <- df$investment

df %<>% na.omit

# проверка на start и end

if(startdt >= enddt){
  message('start must be greater than end')
  return(NULL)
}

startdt <- max(startdt %>% as.yearqtr,
               first(time(df)) %>% as.yearqtr) %>%
as.Date

enddt <- min(enddt %>% as.yearqtr,
             last(time(df)) %>% as.yearqtr) %>%
```



```
as.Date
```

```
if(df[paste0(startdt, "/", enddt)] %>% nrow < 48){
  message('train set length must be greater then 48 quarters')
  return(NULL)
}
```

```
train_n <- which(time(df) %>% as.Date()==(startdt %>% as.yearqtr %>%
as.Date)):
```

```
which((time(df) %>% as.Date())==
      (enddt %>% as.yearqtr %>% as.Date))
```

```
test_n <- (which((time(df) %>% as.Date())==(enddt %>% as.yearqtr %>%
as.Date))+1):nrow(df)
```

```
y.train <- df$y[train_n] %>% as.numeric
y.test <- df$y[test_n] %>% as.numeric
y.full <- c(y.train, y.test)
```

```
if (model == 'arima'){
  train.out <- NULL
  model_fit <- auto.arima(y.train,seasonal = FALSE, stationary = TRUE)
```

```

maxord <- arimaorder(model_fit) %>% sum

if(h == 0){
  pred <- Arima(y.full, model=model_fit) %>% fitted() %>% as.numeric()
} else {
  if(maxord == 1){
    pred <- c()
  } else {
    pred <- rep(NA, maxord-1)
  }
  for(i in maxord:length(y.full)){
    fit <- Arima(y.full[1:i],
                model=model_fit)
    pred <- c(pred, forecast(fit, h) %>% as.data.frame %>% .[h,1])
  }
}

} else if (model == 'rw'){
  if(h == 0){
    pred <- lag.xts(y.full)
  } else {
    pred <- y.full
  }
  train.out <- model_fit <- NULL
}
}

```

```

list(
  model = model,
  series = series,
  lag = lag,
  startdt = startdt,
  enddt = enddt,
  date = time(df)[c(train_n,test_n)] %>% as.Date,
  h = h,
  pred = pred,
  train.out = train.out,
  model_fit = model_fit
)
}

```

```

out <- expand.grid(startdt = c(as.Date('1996-01-01'), as.Date('2000-01-01')),
  enddt = seq(as.Date('2012-10-01'), as.Date('2030-10-01'), by =
'quarter'),
  lag = c(0L),
  h=c(1L:8L),
  model = c('arima',
            'boost',
            'adalamo',
            'postlasso',
            'lasso',

```

```

        'elnet',
        'rf',
        'rw',
        'ss',
        'ridge'
    )
) %T>%
(function(x) {assign('N', nrow(x), envir = globalenv())}) %>%
split(seq(1:nrow(.))) %>%

imap(function(x, i){
  train.model(startdt=x$startdt,
              enddt=x$enddt,
              model = x$model,
              lag=x$lag,
              h=x$h,
              i = i,
              N = N
  )
})

%>%
map_dfr(function(x){
  data.frame(model=x$model,
            lag = x$lag,
            startdt= x$startdt,
            enddt= x$enddt,
            h = x$h,

```

```

        date = x$date,
        pred=x$pred)

    })
save(out, file = 'out.RData')
ggplot(df %>% na.omit) +
  geom_line(aes(y = investment, x = time(df %>% na.omit)))+
  labs(title = "",
        y = "Капітальні інвестиції",
        x = "Дата") +
  theme_bw()
df %>%
  na.omit %>%
  as.data.frame %>%
  rownames_to_column('date') %>%
  melt(id.vars = 'date') %>%
  group_by(variable) %>%
  mutate(value = scale(value),
          date = as.Date(as.yearqtr(date))) %>%
  ggplot() +
  geom_line(aes(date, value, group=variable,
                alpha = ifelse(variable %in% c('investment'), 1, 0.3)), show.legend =
F)+
  labs(title = "",
        y = "",
        x = "") +
  scale_x_date(limits = c('1996-01-01', '2019-01-01') %>% as.Date)+

```

```

geom_vline(xintercept =c('1996-01-01','2000-01-01') %>% as.Date,
linetype='dashed')+
  scale_y_continuous(limits = c(-5, 3))+
  labs(x = 'Дата',
       y="")+
  theme_bw()

out %>% filter(enddt < as.Date(as.yearqtr(date)-h/4)) %>%
group_by(date, h, model, startdt) %>%
  filter(enddt == max(enddt)) %>%
  ungroup %>%
  mutate(forecastdate = as.Date(as.yearqtr(date) -h/4)) %>%
  mutate(pred = ifelse(h == 0, true, pred)) %>%
  mutate(gifttime =as.numeric(forecastdate)+0.2*((date -forecastdate) %>%
as.numeric())) %>%
  filter(forecastdate <='2019-01-01') %>% mutate(true = ifelse(date <=
forecastdate, true, NA))%>% filter(startdt ==
  '2000-01-01',
  h<5)%>%
  mutate(true_na = ifelse(date <= forecastdate, true, NA)) %>%
ggplot()+
  geom_path(data = fordata %>%
           mutate(true_na = ifelse(date <= forecastdate, true, NA)) %>%
na.omit,
           aes(date, true_na,
              alpha = 'Спостережені\пзначення',
              color = 'Спостережені\пзначення',
              size = 'Спостережені\пзначення',

```

```
linetype = 'Спостережені\значення'))+  
geom_line(aes(date, pred,  
              group = interaction(startdt,  
                                  forecastdate),  
              alpha = 'Прогноз',  
              color = 'Прогноз',  
              size = 'Прогноз',  
              linetype = 'Прогноз')  
)+  
facet_wrap(vars(model))+  
scale_y_continuous(limits = c(-0.2, 0.15))+  
labs(x = 'Дата',  
)
```