

Міністерство освіти і науки України

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА АКАДЕМІЯ»

Кафедра математики факультету інформатики

**ПЕРЕВІРКА ПРИПУЩЕННЯ ПРО НЕЗАЛЕЖНІСТЬ ТА
НОРМАЛЬНИЙ РОЗПОДІЛ ВИПАДКОВОЇ ВЕЛИЧИНИ В МОДЕЛІ
ЦІНОУТВОРЕННЯ НА ОСНОВНИЙ КАПІТАЛ**

**Текстова частина до курсової роботи
за спеціальністю „Прикладна математика” 113**

Керівник курсової роботи

к.ф.-м.н. Дрінь Світлана Сергіївна

(прізвище та ініціали)

(підпис)

“17” травня 2021 р.

Виконав студент Реніченко

Єгор Олегович

(прізвище та ініціали)

“17” травня 2021 р.

Зміст	Стор.
Анотація	3
ВСТУП	4
РОЗДІЛ 1: ТЕОРЕТИЧНІ ВІДМОСТІ	
1.1. Основні фінансові поняття.....	7
1.2. Статистичні тести	7
1.3. Узагальнений метод найменших квадратів.....	13
 РОЗДІЛ 2: ПРАКТИЧНА ЧАСТИНА	
2.1. Побудова однофакторних моделей.....	15
2.2. Побудова моделей залежності цін акцій компаній від комплексу змінних і їх оцінка за допомогою методу найменших квадратів.....	17
2.3. Побудова моделей залежності цін акцій компаній від комплексу змінних і їх оцінка за допомогою узагальненого методу найменших квадратів...	18
 Висновки по роботі	20
Список літератури	21
Додаток А: Скрипт побудови та оцінки однофакторних моделей.....	22
Додаток Б: Скрипт побудови та оцінки багатфакторних моделей методом найменших квадратів.....	24
Додаток В: Скрипт побудови та оцінки багатфакторних моделей методом узагальнених найменших квадратів.....	26

Анотація

Метою даної роботи є експеримент створення моделі залежності цін акцій провідних технологічних компаній від фондового індексу та цін на акції пошукової системи і соціальної мережі та її оцінка. Робота виконана за допомогою середовища R Studio, моделі оцінені за допомогою методу найменших квадратів та узагальненого методу найменших квадратів. Створені моделі було оцінено за допомогою методу найменших квадратів, узагальненого методу найменших квадратів та тестовано за чотирма критеріями до використання методу найменших квадратів.

ВСТУП

Сучасний світовий фондовий ринок є найважливішим інструментом фінансування і розвитку економіки як серед розвинених країн, так і в країнах, що розвиваються, де гравці фондового ринку активно розширюють свій вплив. Цей процес інформаційної доби грає особливу роль, тому що покликаний трансформувати заощадження домашніх господарств та підприємств у інвестиції. Основною метою інвесторів є вкладення у надійні та перспективні корпорації. Досягнення цієї мети можливе у тому числі за допомогою новітніх математичних рішень, зокрема, порівняно молодій науки економетрики.

У першій половині двадцятого сторіччя економетрика виникла як наука. Схильність до застосування спеціалістами точних інструментів для кращого розуміння економічних процесів, досягнення у вивченні економічної теорії та математичної статистики стали рушієм до досліджень у сфері новоствореної науки. У тридцятих роках 19 сторіччя за ініціативою Ірвінга Фішера, Рагнара Фріша, Яна Тінбергена, Йозефа Шумпетера, Оскара Андерсона та інших учених було створено економетричну спільноту. У 1933 році Рагнар Фріш опублікував перший випуск журналу «Економетрика», який до цього часу займає важливу роль у розвитку науки. Уже в 1941 році Ян Тінберген видає перший підручник з нової наукової дисципліни. У 1969 році Фріш і Тінберген стали першими, хто отримав Нобелівську премію з економіки «за створення і застосування динамічних моделей до аналізу економічних процесів». [1]

Створення економетричних моделей наразі є одним з кращих варіантів автоматизації та покращення прогнозування економічних процесів у бізнесі. Методи на базі цього розділу математики також є складовими частинами наукового інструментарію будь-яких техніко-економічних досліджень. Способом оцінки якості таких моделей є метод найменших квадратів, узагальнений метод найменших квадратів, і так далі.

Однак, використання будь-яких точних інструментів тягне за собою конкретні правила та критерії їх використання. Наприклад, метод найменших квадратів не може

бути використаний у випадку мультиколінеарності незалежних змінних у моделі, гетероскедастичності похибок, автокореляції або ненормальності розподілу похибок.

Виходячи з тенденцій та історії розвитку методів прогнозування за мету даної роботи був поставлений експеримент створення моделі залежності цін акцій провідних технологічних компаній від фондового індексу NASDAQ та цін на акції пошукової системи Google і соціальної мережі Facebook та її оцінка за допомогою середовища R Studio.

Цей експеримент може стати додатковим аналітичним інструментом при вивченні сучасних фондових ринків та тенденцій, які відбуваються із ними.

Робота складається з двох розділів.

Перший розділ присвячено теоретичній базі для створення та діагностування моделей залежності. Наведено критерії використання методу найменших квадратів та алгоритм використання узагальненого методу найменших квадратів.

В другому розділі наведено результати розробки скриптів у середовищі R Studio: однофакторні моделі залежності технологічних компаній від фондового індексу NASDAQ, багатфакторні моделі залежності технологічних компаній від фондового індексу NASDAQ, пошукової системи Google і соціальної мережі Facebook.

Створено програмний продукт, який призначений для експериментального дослідження при прогнозуванні цін акцій технологічних компаній та в навчальному процесі.

Постановка задачі

1. Систематизувати теоретичну базу про критерії використання методу найменших квадратів.
2. Навести алгоритм використання узагальненого методу найменших квадратів.
3. Дослідити залежності від фондового індексу NASDAQ.
4. Оцінити багатфакторні моделі залежності, створені відповідно до мети даної роботи, за допомогою методу найменших квадратів.

5. Перевірити моделі на наявність наступних явищ:

- ненормальність розподілу похибок;
- гетероскедастичність;
- автокореляція;
- мультиколінеарність.

6. Оцінити багатofакторні моделі залежності, створені відповідно до мети даної роботи, за допомогою узагальненого методу найменших квадратів.

Об'єктом дослідження є моделювання цін акцій на фондових ринках.

Методами дослідження є класифікація та систематизація теоретичної бази, формалізація фінансових процесів у першому розділі, а також моделювання у другому розділі.

Джерела дослідження: «Економетрика. Підручник» Єлісеєва І. – історичні відомості про виникнення та розвиток економетрики як науки; «Практика економетрії» Берндта Е. – основні фінансові поняття; «Економетрика» Лук'яненко І., «Базова Економетрика» Гуяраті Д., «Аналіз тесту на нормальність» Шاپіро С. Та Вілка М., «Мультиколінеарність у регресійному аналізі» Кумара К., «Вступ до вивчення статистики: за допомогою R» Гарета Дж. – відомості про статистичні тести критеріїв використання методу найменших квадратів; «Прикладні лінійні регресійні моделі» Катнера М. – використання узагальненого методу найменших квадратів; «Фактори Байєса» Касса Р. – застосування інформаційного критерія Байєса.

Наукова новизна одержаних результатів: створення експериментальної моделі залежності цін акцій деяких технологічних компаній від фондового індексу, пошукової системи та соціальної мережі.

РОЗДІЛ 1: ТЕОРЕТИЧНЕ ПІДҮНТЯ

1.1 Основні фінансові поняття

Припустимо, що коли інвестори вступають на ринок цінних паперів, їх поведінка є повністю раціональною у тому сенсі, що їх єдиною метою є отримати максимальні прибутки з мінімальними ризиками від власних капіталовкладень. Визначимо норму прибутку від капіталовкладень наступним чином:

$$r \equiv \frac{p_1 + d - p_0}{p_0}, \quad (1.1)$$

де p_1 – ціна акції у кінці часового періоду;

d – дивіденди (якщо вони наявні), які було сплачено за даний часовий період;

p_0 – ціна капіталу на початку часового періоду. [2]

1.2 Статистичні тести

Моделі, досліджувані у практичній частині, перевіримо за чотирма критеріями, які необхідні для отримання підстав до використання метода найменших квадратів. А саме: гомоскедастичність, мультиколінеарність, залежність факторів (автокореляцію) та нормальність похибок.

У наступних чотирьох підрозділах розміщено теоретичну базу цих явищ та тестів.

1.2.1 Автокореляція

Автокореляція — це кореляція випадкової величини з самою собою, яка зміщує певну величину незалежної змінної. Наявність автокореляції випадкових похибок у регресійній моделі призводить до погіршення якості оцінок параметрів регресії за

допомогою метода найменших квадратів (далі – МНК), а також до аномального перевищення тестових статистик, за якими оцінюється модель (створюється штучне поліпшення якості моделі врозріз із її дійсним рівнем точності). Тому тестування автокореляції випадкових похибок є необхідною процедурою побудови регресійної моделі.

Критерій Дарбіна-Уотсона - статистичний критерій, який використовується для тестування автокореляції першого порядку елементів досліджуваної послідовності. Нульова гіпотеза тесту - фактори є незалежними між собою (не присутня автокореляція), альтернативна – автокореляція існує.

Якщо e_t - похибка, має залежність $e_t = \rho e_{t-1} + v_t$, де v_t – випадкова величина, статистика Дарбіна-Уотсона стверджує, що нульова гіпотеза: $\rho = 0$, альтернативна гіпотеза $\rho \neq 0$, статистика тесту дорівнює:

$$d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2} \quad (1.2)$$

де T - кількість спостережень.

Значення d завжди знаходиться в межах від нуля до чотирьох. Якщо статистика Дарбіна-Уотсона істотно менше 2, є дані про позитивну серійну кореляцію. Як правило, якщо тест Дарбіна-Уотсона менше 1, це може свідчити про наявність автокореляції: невеликі значення d означають, що похибки позитивно корелюють. Якщо $d > 2$, похибки негативно корелюють. [3]

1.2.2 Гетероскедастичність

Послідовність випадкових величин можна назвати гетероскедастичною, якщо ця послідовність має різну дисперсію. Протилежне до гетероскедастичності поняття – гомоскедастичність, послідовність випадкових величин, яка має постійну

дисперсію. Із прикладами графічного зображення гетеро- та гомоскедастичності можна ознайомитись на рисунку 1.1.

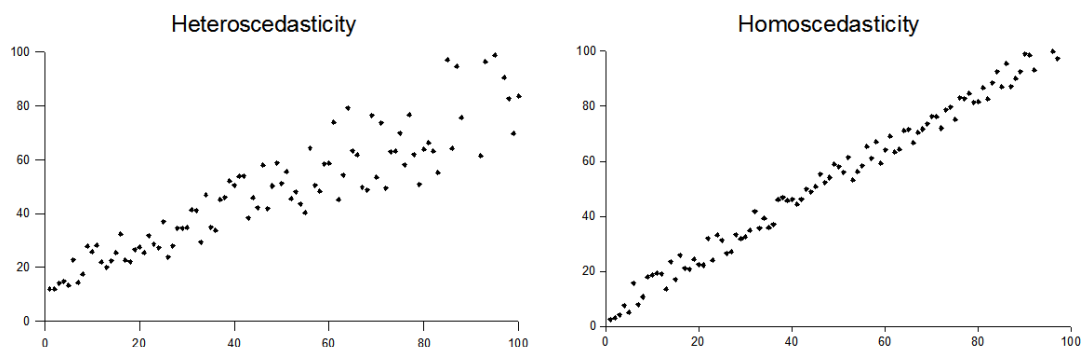


Рисунок 1.1 — Приклад графічного представлення гетеро- та гомоскедастичності.

Існування гетероскедастичності є серйозною проблемою в регресійному аналізі, оскільки за її наявності можна вважати недійсними статистичні тести значущості, які передбачають, що всі похибки мають однакову дисперсію. Оцінка моделі за допомогою метода найменших квадратів при наявності гетероскедастичності неефективна.

Тест Бройша-Пагана - один із статистичних тестів для перевірки наявності гетероскедастичності у похибках регресійної моделі. Нульовою гіпотезою тесту є гомоскедастичність моделі, альтернативною – наявність гетероскедастичності.

Тест перевіряє, чи оцінка дисперсії похибок регресії залежить від значень незалежних змінних. Оцінімо рівняння:

$$y = \beta_0 + \beta_1 x + e \quad (1.3)$$

Тепер ми можемо оцінити похибку e . Метод найменших квадратів обмежує e так, що його середнє дорівнює 0, звідси можемо обчислити дисперсію як середній квадрат значень. Простіше – протестувати регресію квадратів похибок e по кожній незалежній змінній, що і є тестом Бройша-Пагана:

$$\widehat{u}_t^2 = \beta_0 + \beta_1 x_t + v_t \quad (1.4)$$

Якщо F-тест підтверджує, що незалежні змінні разом значимі, то ми відхиляємо нульову гіпотезу про гомоскедастичність. [4]

1.2.3 Нормальність розподілу похибок

Нехай маємо вибірку з великою кількістю спостережень, за умови, що кожне було отримано незалежно від інших спостережень та випадковим. Розрахуємо для них середнє арифметичне. Як стверджує центральна гранична теорема, якщо цю процедуру повторити багато разів, то сукупність розрахованих середніх значень буде мати нормальний розподіл. Ілюстрація цієї процедури зображена на рисунку 1.2.

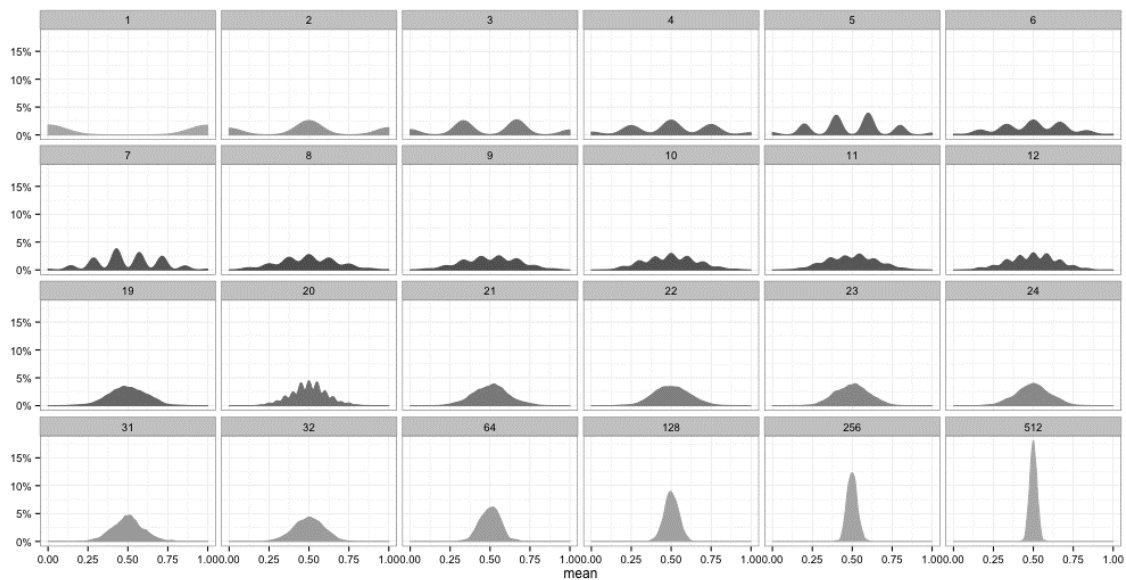


Рисунок 1.2 – Середнє значення біноміального розподілу моделювання різних розмірів вибірки, що показує, що при збільшенні можливих випадків розподіл концентрується навколо середнього значення. [5]

Можна констатувати, використовуючи центральну граничну теорему, що якщо статистичні дані випадкові та незалежні, то їх сукупність матиме нормальний розподіл. Це обґрунтовує вимогу до наявності нормального розподілу похибок у лінійних моделях.

Критерій Шапіро-Уїлка використовується для перевірки нульової гіпотези: «випадкова величина X розподілена нормально» і є одним найбільш ефективних критеріїв перевірки нормальності. За його допомогою можна перевірити, чи нормально розподілені похибки в моделі. Альтернативною гіпотезою тесту є ненормальність розподілу випадкової величини X .

Статистикою тесту є:

$$W = \frac{\left(\sum_{i=1}^n a_i x_{(i)}\right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (1.5)$$

де $x_{(i)}$ (не плутати з x_i) – це i -те найменше число у вибірці;

$\bar{x} = (x_1 + \dots + x_n) / n$ – середнє значення вибірки.

Коефіцієнти a_i задаються як:

$$(a_1, \dots, a_n) = \frac{n^\top V^{-1}}{C} \quad (1.6)$$

де $C = \|V^{-1}m\| = \sqrt{m^\top V^{-1}V^{-1}m}$

$m = (m_1, \dots, m_n)^\top$

m – складається з очікуваних значень статистики порядку незалежних та однаково розподілених випадкових величин, відібраних із стандартного нормального розподілу;

V - матриця коваріації цих статистичних даних нормального порядку. [6]

1.2.4 Мультиколінеарність

Мультиколінеарність – це явище, при якому одну незалежну змінну в моделі множинної регресії можна лінійно передбачити за допомогою інших з великою точністю. Мультиколінеарність не впливає на прогностні значення моделі або її надійність в цілому, але може спотворити результати знаходження необхідних та зайвих незалежних змінних у моделі.

Це явище можна оцінити за допомогою дослідження кореляції між обраними факторами. Якщо кореляція між факторами висока (по модулю більша за 0.7), то можна вважати, що між факторами присутня суттєва колінеарність.

Мультиколінеарність може бути точніше визначена з обчислення коефіцієнту “опуклої” дисперсії “здутості” або VIF (Variance inflation factor), який вимірює, наскільки завищена дисперсія коефіцієнта регресії через мультиколінеарність у моделі. Найменшим можливим значенням VIF є 1 (відсутність досліджуваного явища). Як правило, значення VIF, яке перевищує 5 або 10 (у залежності від точності дослідження), вказує на суттєву мультиколінеарність.

Ми можемо розрахувати k різних VIF (по одному для кожного X_i) за три кроки:

1. Спочатку ми створюємо звичайну регресію за допомогою метода найменших квадратів, яка має X_i як функцію всіх інших незалежних змінних у першому рівнянні. Наприклад, якщо $i = 1$, рівняння буде

$$X_1 = \alpha_0 + \alpha_2 X_2 + \dots + \alpha_k X_k + e \quad (1.7)$$

де α_0 - константа, а e - похибка.

2. Обчислити коефіцієнт VIF для $\hat{\alpha}_i$ за такою формулою:

$$VIF_i = \frac{1}{1 - R_i^2} \quad (1.8)$$

де R_i^2 - коефіцієнт визначення рівняння регресії на першому кроці з X_i ліворуч, а всі інші змінні фактору (усі інші змінні X) праворуч.

3. Далі, як вже було сказано, потрібно обрати величину A , з якою буде порівнюватись кожен VIF_i . Вибір, яку A використовувати - це питання особистих переваг, найчастіше використовується 5 та 10. Якщо VIF_i менший заданої величини A , то суттєвої колінеарності не знайдено, якщо більше – присутня суттєва колінеарність. [7-9]

1.3 Узагальнений метод найменших квадратів

Узагальнений метод найменших квадратів – метод оцінки параметрів регресійних моделей, узагальнення звичайного методу найменших квадратів. Узагальнений метод найменших квадратів використовується із мінімізацією узагальненої суми квадратів залишків регресії – $e^T W e$, де e – вектор похибок, W – позитивно визначена симетрична зважена матриця.

Узагальненим методом найменших квадратів характеризують як окремий випадок, коли в якості вагової матриці використовується матриця, яка є зворотною до коваріаційної матриці випадкових похибок моделі.

Припустимо, що розглядається модель виду:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + e_i \quad (1.9)$$

для якої дисперсія похибок величин виявилася пропорційна K_i^2 . K_i є коефіцієнтом пропорційності, який приймає різні значення для відповідних значень факторів x_1 і x_2 . Розглянемо цю модель, якщо похибки гетероскедастичні:

$$y_i = \beta_{0i} + \beta_{1i}x_1 + \beta_{2i}x_2 + K_i e \quad (1.10)$$

Для того, щоб отримати рівняння, де похибки стануть гомоскедастичними, перейдемо до нових змінних, розділивши всі члени рівняння (1.10) на коефіцієнт K_i :

$$\frac{y_i}{K_i} = \frac{\beta_{0i}}{K_i} + \frac{\beta_{1i}}{K_i}x_1 + \frac{\beta_{2i}}{K_i}x_2 + e_i \quad (1.11)$$

Параметри моделі залежать від того, що ми приймаємо за коефіцієнт пропорційності. Висунемо гіпотезу, що похибки пропорційні значенням фактору. Тоді якщо $K_i = x_1$, то узагальнений метод найменших квадратів передбачає оцінку параметрів наступного трансформованого з рівняння (1.11):

$$\frac{y_i}{x_1} = \frac{\beta_{0i}}{x_1} + \beta_{1i} + \frac{\beta_{2i}}{x_1}x_2 + \varepsilon \quad (1.12)$$

Застосування в цьому випадку узагальненого методу найменших квадратів призводить до того, що спостереження з меншими значеннями перетворених змінних мають при визначенні параметрів регресії більшу вагу, ніж з неперетвореними. Але слід мати на увазі, що новоутворена регресія має інший зміст, ніж регресія з початкових даних. [10]

РОЗДІЛ 2: ПРАКТИЧНА ЧАСТИНА

За допомогою сайту <https://finance.yahoo.com/> було отримано історичні дані цін на акції помісячно з 01.01.2009 по 31.12.2020 (за 12 років) компаній Uber, Google, Amazon, Tesla, Facebook, Apple та індексу NASDAQ.

NASDAQ – фондовий індекс, складений з звичайних акцій і подібних фінансових інструментів всіх компаній, що торгуються на біржі NASDAQ. Розраховується з 1971 року. Включає в себе більше 3000 американських і неамериканських корпорацій. NASDAQ-індекс вважається важливим показником динаміки курсів цінних паперів високотехнологічних і швидко зростаючих компаній. Індекс розраховується на основі середньозважених цін на момент закінчення торгів.

2.1 Побудова однофакторних моделей

Для початку було побудовано однофакторні моделі залежності окремо Uber, Google, Amazon, Tesla, Facebook, Apple від фондового індексу NASDAQ. На рисунку 2.1 можна ознайомитись із характером залежності у кожній моделі.

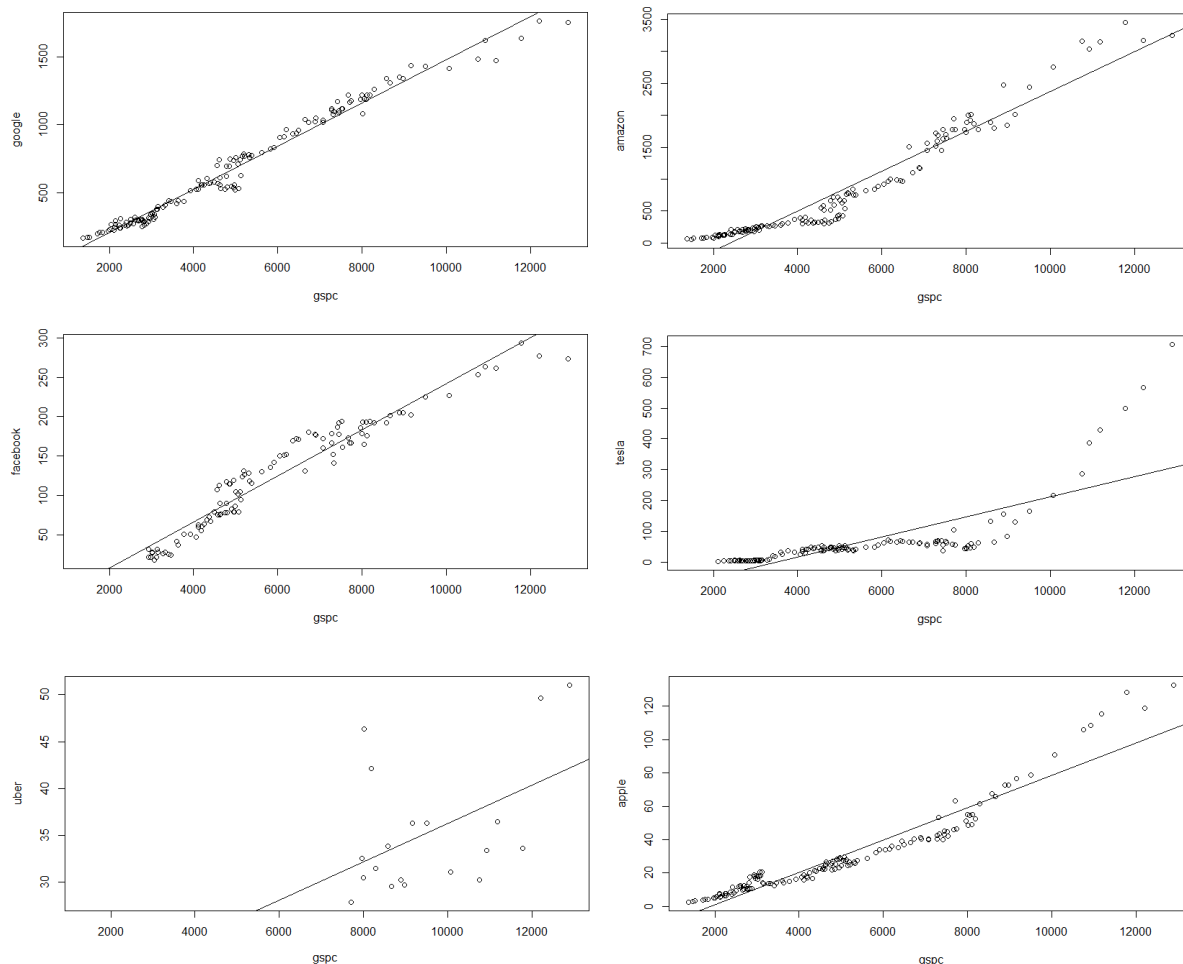


Рисунок 2.1 – Графіки залежності компаній Uber, Google, Amazon, Tesla, Facebook, Apple від фондового індексу NASDAQ.

З Рисунку 1 бачимо, що залежності Google та Amazon швидше лінійні, Facebook – логарифмічна, Tesla – експоненційна, Apple – квадратична, Uber має дуже нечітку залежність через малу кількість вимірювань (кількість рівна дев'ятнадцяти). Спираючись на візуальний аналіз, створюємо моделі та оцінюємо їх за допомогою метода найменших квадратів.

Результати можна побачити у таблиці 2.1.

№	Модель «Компанія ~ NASDAQ»	Коефіцієнт	Коефіцієнт детермінації	Статистика Фішера
1	Google	0.1586	0.9782	6384
2	Amazon	0.3133	0.9332	1984
3	Facebook	0.1828	0.9577	2288

4	Tesla	0.4386	0.7955	482
5	Uber	0.0020	0.2219	5
6	Apple	0.0008	0.9769	6009

Таблиця 2.1 – Моделі залежності компаній Uber, Google, Amazon, Tesla, Facebook, Apple від фондового індексу NASDAQ та ключові показники для оцінки моделей методом найменших квадратів: коефіцієнт бета, коефіцієнт детермінації та статистика Фішера.

Модель 1: google $\sim$ NASDAQ.

Модель 2: amazon $\sim$ NASDAQ.

Модель 3 facebook $\sim \ln(\text{NASDAQ})$.

Модель 4: $\ln(\text{tesla}) \sim \text{NASDAQ}$.

Модель 5: uber $\sim$ NASDAQ.

Модель 6: apple $\sim (\text{NASDAQ})^2$.

По значенням коефіцієнтів детермінації та статистики Фішера можна констатувати, що усі моделі надійні, окрім моделі залежності Uber $\sim$ NASDAQ.

2.2 Побудова моделей залежності цін акцій компаній від комплексу змінних і їх оцінка за допомогою методу найменших квадратів

Після побудови однофакторних моделей було досліджено залежність цін акцій компаній не тільки від індексу (умовно - ринку), а й від цін акцій пошукових систем та соціальних мереж (у нашому випадку - Google та Facebook).

Було побудовано моделі залежності ціни акцій окремо Uber, Tesla, Amazon та Apple від цін акцій разом NASDAQ, Google та Facebook (таблиця 2.2).

Модель «Компанія $\sim$ NASDAQ + Google + Facebook»	Коеф. NASDAQ	Коеф. Google	Коеф. Facebook	Коефіцієнт детермінації	Статистика Фішера

Amazon	0.3367	0.8621	-3.7694	0.9591	774
Tesla	0.4565	-2.1292	8.0044	0.7611	105
Uber	0.0088	-0.0243	-0.1748	0.3057	2
Apple	0.0009	-0.0197	0.0153	0.9789	1532

Таблиця 2.2 – Моделі залежності компаній Uber, Amazon, Tesla, Apple від фондового індексу NASDAQ, Facebook та Google, ключові показники для оцінки моделей методом найменших квадратів: коефіцієнти бета, коефіцієнт детермінації та статистика Фішера.

Модель 1: $\text{amazon} \sim \text{google} + \text{facebook} + \text{NASDAQ}$.

Модель 2: $\ln(\text{tesla}) \sim \text{google} + \text{facebook} + \text{NASDAQ}$.

Модель 3: $\text{uber} \sim \text{google} + \text{facebook} + \text{NASDAQ}$.

Модель 4: $\text{apple} \sim \text{google} + \text{facebook} + (\text{NASDAQ})^2$.

Також чотири багатофакторні моделі було перевірено на можливість використання метода найменших квадратів. Як можемо бачити з таблиці 2.3, жодна з моделей не може бути оцінена за допомогою метода найменших квадратів.

Модель «Компанія ~ NASDAQ + Google + Facebook»	Автокореляція	Гетероскедаст ичність	Ненормальність розподілу похибок	Мультиколінеа рність
Amazon	+	-	-	-
Tesla	-	-	-	-
Uber	+	+	-	-
Apple	+	-	-	-

Таблиця 2.3 – Моделі залежності, описані у Таблиці 2, критерії використання метода найменших квадратів.

2.3 Побудова моделей залежності цін акцій компаній від комплексу змінних і їх оцінка за допомогою узагальненого методу найменших квадратів

Оскільки критерії до використання класичного методу найменших квадратів не були виконані, ті самі моделі залежності було додатково оцінено за допомогою узагальненого методу найменших квадратів. Результати оцінки цим методом можна побачити у таблиці 2.4.

Модель «Компанія ~ NASDAQ + Google + Facebook»	Коеф. NASDAQ	Коеф. Google	Коеф. Facebook	BIC
Amazon	0.3367	0.8621	-3.7694	-38.03
Tesla	0.4565	-2.1292	8.0044	156.75
Uber	0.0088	-0.0243	-0.1748	-98.22
Apple	0.0009	-0.0197	0.0153	-785.15

Таблиця 2.4 – Моделі залежності компаній Uber, Amazon, Tesla, Apple від фондового індексу NASDAQ, Facebook та Google, ключові показники для оцінки моделей узагальненим методом найменших квадратів: коефіцієнти бета, інформаційний критерій Байєса (BIC). При порівнянні моделей менший BIC (Bayesian information criterion) позначає кращу модель. Чим більша різниця між двома моделями, що порівнюють, тим більш очевидно, що потрібно обрати модель з меншим показником (при різниці, більшій за 10 – очевидно краща модель з меншим показником).[11]

Модель 1: amazon ~ google + facebook + NASDAQ.

Модель 2: ln(tesla) ~ google + facebook + NASDAQ.

Модель 3 uber ~ google + facebook + NASDAQ.

Модель 4: apple ~ google + facebook + (NASDAQ)².

Висновки

У процесі роботи були зроблені наступні висновки.

Було систематизовано теоретичну базу про критерії використання методу найменших квадратів та наведено алгоритм використання узагальненого методу найменших квадратів.

Також було досліджено залежності Uber, Google, Facebook, Apple, Amazon та Tesla від фондового індексу NASDAQ, за допомогою яких створено й потім оцінено багатофакторні моделі залежності Uber, Apple, Amazon та Tesla від фондового індексу NASDAQ, пошукової системи Google і соціальної мережі Facebook за допомогою методу найменших квадратів.

Ці моделі було перевірено на наявність явищ, які унеможливають використання методу найменших квадратів. Результати тестування свідчать про неможливість використання методу найменших квадратів.

Оцінено створені вище багатофакторні моделі залежності за допомогою узагальненого методу найменших квадратів.

Список літератури

1. Эконометрика. Учебник / Под ред. Елисеевой И. И. — 2-е изд. — М.: Финансы и статистика, 2008. — 576 с.
2. The Practice of Econometrics: Classic and Contemporary. Ernst R. Berndt. Addison-Wesley Publishing Company, 1991 - Econometrics - 702 p.
3. Лук'яненко, Ірина. Економетрика : підручник / І. Лук'яненко, Л. Краснікова. - Київ : Знання, 1998. - 493 с.
4. Gujarati, Damodar N.; Porter, Dawn C. (2009). Basic Econometrics (вид. Fifth). New York: McGraw-Hill Irwin. с. 385–86.
5. <https://github.com/resendedaniel/math/tree/master/17-central-limit-theorem>
6. Shapiro, S. S.; Wilk, M. B. (1965). "An analysis of variance test for normality (complete samples)". *Biometrika*. 52 (3–4): 591–611.
7. Kumar, T. Krishna (1975). "Multicollinearity in Regression Analysis". *Review of Economics and Statistics*. 57 (3): 365–366.
8. James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2014. *An Introduction to Statistical Learning: With Applications in R*. Springer Publishing Company, Incorporated.
9. Kutner, M. H.; Nachtsheim, C. J.; Neter, J. (2004). *Applied Linear Regression Models* (4th ed.). McGraw-Hill Irwin.
10. Baltagi, B. H. (2008). *Econometrics* (4th ed.). New York: Springer.
11. Kass, Robert E.; Raftery, Adrian E. (1995), "Bayes Factors", *Journal of the American Statistical Association*, 90 (430): 773–795pp.

Додаток А

Скрипт побудови та оцінки однофакторних моделей.

Модель 1: google ~ NASDAQ.

Модель 2: amazon ~ NASDAQ.

Модель 3 facebook ~ ln(NASDAQ).

Модель 4: ln(tesla) ~ NASDAQ.

Модель 5: uber ~ NASDAQ.

Модель 6: apple ~ (NASDAQ)².

```
library(readxl)
data <- read_excel("data.xlsx")
attach(data)
plot(nasdaq, google)
model_goog_nasdaq <- lm(google ~ nasdaq, data = data)
print(summary(model_goog_nasdaq))
abline(model_goog_nasdaq)
# model_goog_nasdaq Multiple R-squared:  0.9782
# F-statistic: 6384 on 1 and 142 DF, p-value: < 2.2e-16
plot(nasdaq, amazon)
model_amzn_nasdaq <- lm(amazon ~ nasdaq, data = data)
print(summary(model_amzn_nasdaq))
abline(model_amzn_nasdaq)
# model_amzn_nasdaq Multiple R-squared:  0.9332
# F-statistic: 1984 on 1 and 142 DF, p-value: < 2.2e-16
plot(log(nasdaq), facebook)
model_fcbk_nasdaq <- lm(facebook ~ l(log(nasdaq)), data = data)
print(summary(model_fcbk_nasdaq))
```

```
abline(model_fcbk_nasdaq)
# model_fcbk_nasdaq Multiple R-squared:  Multiple R-squared: 0.9577
# F-statistic: 2288 on 1 and 101 DF, p-value: < 2.2e-16
plot(nasdaq, uber)
model_uber_nasdaq <- lm(uber ~ nasdaq, data = data)
print(summary(model_uber_nasdaq))
abline(model_uber_nasdaq)
# model_uber_nasdaq Multiple R-squared: 0.2219
# F-statistic: 4.849 on 1 and 17 DF, p-value: 0.04176
plot(nasdaq, log(tesla))
model_tsla_nasdaq <- lm(l(log(tesla)) ~ nasdaq, data = data)
print(summary(model_tsla_nasdaq))
abline(model_tsla_nasdaq)
# model_tsla_nasdaq Multiple R-squared: 0.7955
# F-statistic: 482.4 on 1 and 124 DF, p-value: < 2.2e-16
plot(nasdaq^2, apple)
model_appl_nasdaq <- lm(apple ~ l(nasdaq^2), data = data)
print(summary(model_appl_nasdaq))
abline(model_appl_nasdaq)
# model_appl_nasdaq Multiple R-squared: 0.9121
# F-statistic: 1474 on 1 and 142 DF, p-value: < 2.2e-16
```

Додаток Б

Скрипт побудови та оцінки багатфакторних моделей методом найменших квадратів.

Модель 1: $\text{amazon} \sim \text{google} + \text{facebook} + \text{NASDAQ}$.

Модель 2: $\ln(\text{tesla}) \sim \text{google} + \text{facebook} + \text{NASDAQ}$.

Модель 3 $\text{uber} \sim \text{google} + \text{facebook} + \text{NASDAQ}$.

Модель 4: $\text{apple} \sim \text{google} + \text{facebook} + (\text{NASDAQ})^2$.

```
library(readxl)
library(car)
data <- read_excel("data.xlsx")
attach(data)
# ----- Amazon -----
model_amzn_GoFbNasdaq <- lm(amazon ~ google + facebook + nasdaq, data = data)
print(summary(model_amzn_GoFbNasdaq))
# Multiple R-squared:  0.9591,    Adjusted R-squared:  0.9579
# F-statistic: 774.7 on 3 and 99 DF,  p-value: < 2.2e-16
print(shapiro.test(model_amzn_GoFbNasdaq$residuals)) # +
print(car::ncvTest(model_amzn_GoFbNasdaq)) # -
print(car::durbinWatsonTest(model_amzn_GoFbNasdaq)) # -
print(car::vif(model_amzn_GoFbNasdaq)) # -
# ----- Uber -----
model_uber_GoFbNasdaq <- lm(uber ~ google + facebook + nasdaq, data = data)
print(summary(model_uber_GoFbNasdaq))
# Multiple R-squared:  0.3057,    Adjusted R-squared:  0.1668
# F-statistic: 2.201 on 3 and 15 DF,  p-value: 0.1302
print(shapiro.test(model_uber_GoFbNasdaq$residuals)) # +
print(car::ncvTest(model_uber_GoFbNasdaq)) # +
print(car::durbinWatsonTest(model_uber_GoFbNasdaq)) # -
print(car::vif(model_uber_GoFbNasdaq)) # -
# ----- Tesla -----
model_tsla_GoFbNasdaq <- lm(l(log(tesla)) ~ google + facebook + nasdaq, data = data)
```

```

print(summary(model_tsla_GoFbNasdaq))
# Multiple R-squared:  0.7611,    Adjusted R-squared:  0.7538
# F-statistic: 105.1 on 3 and 99 DF,  p-value: < 2.2e-16
print(shapiro.test(model_tsla_GoFbNasdaq$residuals)) # -
print(car::ncvTest(model_tsla_GoFbNasdaq)) # -
print(car::durbinWatsonTest(model_tsla_GoFbNasdaq)) # -
print(car::vif(model_tsla_GoFbNasdaq)) # -
# ----- Apple -----
model_appl_GoFbNasdaq <- lm(apple ~ google + facebook + I(nasdaq^2), data = data)
print(summary(model_appl_GoFbNasdaq))
# Multiple R-squared:  0.9789,    Adjusted R-squared:  0.9783
# F-statistic: 1532 on 3 and 99 DF,  p-value: < 2.2e-16
print(shapiro.test(model_appl_GoFbNasdaq$residuals)) # +
print(car::ncvTest(model_appl_GoFbNasdaq)) # -
print(car::durbinWatsonTest(model_appl_GoFbNasdaq)) # -
print(car::vif(model_appl_GoFbNasdaq)) # -

```

Додаток В

Скрипт побудови та оцінки багатофакторних моделей методом узагальнених найменших квадратів.

Модель 1: $\text{amazon} \sim \text{google} + \text{facebook} + \text{NASDAQ}$.

Модель 2: $\ln(\text{tesla}) \sim \text{google} + \text{facebook} + \text{NASDAQ}$.

Модель 3: $\text{uber} \sim \text{google} + \text{facebook} + \text{NASDAQ}$.

Модель 4: $\text{apple} \sim \text{google} + \text{facebook} + (\text{NASDAQ})^2$.

```
library(readxl)
library(nlme)
library(MASS)
data <- read_excel("data.xlsx")
attach(data)

# ----- Amazon -----
model_amzn_GoFbNasdaq <- lm.gls(amazon ~ google + facebook + nasdaq,
                                na.action=na.exclude, data = data)
summary(model_amzn_GoFbNasdaq)

# ----- Uber -----
model_uber_GoFbNasdaq <- gls(uber ~ google + facebook + nasdaq,
                             na.action=na.exclude, data = data)
print(summary(model_uber_GoFbNasdaq))

# ----- Tesla -----
model_tsla_GoFbNasdaq <- gls(l(log(tesla)) ~ google + facebook + nasdaq,
                             na.action=na.exclude, data = data)
print(summary(model_tsla_GoFbNasdaq))

# ----- Apple -----
model_appl_GoFbNasdaq <- gls(apple ~ google + facebook + I(nasdaq^2),
```

```
na.action=na.exclude, data = data)  
print(summary(model_appl_GoFbNasdaq))
```