

Міністерство освіти і науки України

Національний університет «Києво-Могилянська академія»

Факультет інформатики

Кафедра математики

## **Кваліфікаційна робота**

освітній ступінь – бакалавр

на тему: **«МОДЕЛЬ СИСТЕМИ ОДНОЧАСНИХ РІВНЯНЬ З  
ЛАГОВИМ ЕФЕКТОМ ДЛЯ ПРОГНОЗУВАННЯ ПРОДАЖІВ»**

Виконав: студент 4-го року навчання  
освітньої програми «Прикладна  
математика»,  
спеціальності 113 Прикладна  
математика

Резніченко Єгор Олегович

Керівник: Дрінь С.С.,  
кандидат фіз.-мат. наук, ст. викладач

Рецензент \_\_\_\_\_  
(прізвище та ініціали)

Кваліфікаційна робота захищена  
з оцінкою \_\_\_\_\_

Секретар ЕК \_\_\_\_\_

« \_\_\_\_ » \_\_\_\_\_ 20 \_\_\_\_ р.

Київ – 2022

Міністерство освіти і науки України  
 НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ «КИЄВО-МОГИЛЯНСЬКА  
 АКАДЕМІЯ»  
 Кафедра математики факультету інформатики

ЗАТВЕРДЖУЮ  
 Зав.кафедри математики,  
 проф., д.ф-м.н.  
 \_\_\_\_\_ Б. В. Олійник  
 (підпис)  
 „\_\_\_\_” \_\_\_\_\_ 2021 р.

**ІНДИВІДУАЛЬНЕ ЗАВДАННЯ**  
 на дипломну роботу

студенту 4-го курсу бакалаврської програми, факультету інформатики  
Єгора Резніченка

Розробити математичну модель системи прогнозування об'єму продажів в  
умовах специфікації та ідентифікації одночасних рівнянь з  
розподіленим лагом та розробити комп'ютерну програму.

Вихідні дані:

- об'єм продажів;
- похибка вимірювання – 5 %;

Зміст ТЧ до магістерської роботи:

Зміст

Анотація

Вступ

1. Огляд існуючих теорій з цього питання;
2. Розробка структурної моделі залежності між продажами, рекламою;
3. Розробка алгоритму програми;
4. Програма для прогнозування розміру бюджету;
5. Результати, які будуть отримані при дослідженні реальних української даних;

Висновки

Список літератури

Додатки

Дата видачі „\_\_\_\_” \_\_\_\_\_ 2021 р. Керівник \_\_\_\_\_  
 (підпис)

Завдання отримав \_\_\_\_\_  
 (підпис)

**Тема: Модель системи одночасних рівнянь з лаговим ефектом для прогнозування продажів**

**Календарний план виконання роботи:**

№ п/п	Назва етапу дипломного проекту (роботи)	Термін виконання етапу	Примітка
1.	Отримання завдання на дипломну роботу.	23.11.2021	
2.	Огляд технічної літератури за темою роботи.	25.11.2021	
3.	Виконати аналіз сучасних методів	27.11.2021	
3.	Розробка алгоритму	15.12.2021	
4.	Розробка алгоритму	15.01.2022	
5.	Програмування розробленого алгоритму	15.02.2022	
6.	Виконання порівняльного аналізу результатів, для різних секторів в оплаті праці	15.03.2022	
7.	Написання пояснювальної роботи.	10.04.2022	
8.	Створення слайдів для доповіді та написання доповіді.	20.04.2022	
8.	Аналіз отриманих результатів з керівником, написання доповіді та попередній захист магістерської роботи.	30.04.2022	
10.	Корегування роботи за результатами попереднього захисту.	25.05.2022	
11.	Остаточне оформлення пояснювальної роботи та слайдів.	30.05.2022	
12.	Захист магістерської роботи (проекту)	04.06.2022	

Студент: Резніченко Є.О.

Керівник: Дрінь С.С.

“23” жовтня 2021 р.

Зміст	Стор.
Анотація.....	4
ВСТУП.....	5
<b>РОЗДІЛ 1: ТЕОРЕТИЧНЕ ПІДРУНТЯ.....</b>	<b>8</b>
1.1. Реклама як явище.....	8
1.1.1. Види реклами.....	9
1.1.1.1. Реклама на телебаченні.....	9
1.1.1.2. Інтернет-реклама .....	11
1.1.1.3. Інша реклама.....	12
1.1.2. Конкуренція та медійний тиск .....	13
1.1.3. Відкладений ефект – adstock .....	13
1.2. Економетричний підхід у рекламі.....	17
1.2.1. Тест Хаусмана.....	18
1.2.2. Тест Гренжера.....	22
1.2.3. Метод двокрокового МНК.....	24
1.3. Модель залежності продажів від реклами.....	25
<b>РОЗДІЛ 2: ПРАКТИЧНА ЧАСТИНА.....</b>	<b>28</b>
2.1. План та опис даних.....	28
2.2. Застосування тесту Хаусмана.....	30
2.3. Застосування тесту Гренжера.....	33
2.4. Формування, опис моделі та її оцінка за допомогою 2МНК.....	40
<b>Висновки по роботі та рекомендації для подальших досліджень.....</b>	<b>47</b>
Список літератури.....	49
Перелік прийнятих скорочень.....	50
Додаток А.....	51
Додаток Б.....	52
Додаток В.....	53

## Перелік прийнятих скорочень

СУМ-11	Словник української мови в 11 томах;
МНК	метод найменших квадратів;
2МНК	двокроковий метод найменших квадратів;
SKU	Stock Keeping Unit, ідентифікатор товарної позиції, одиниця обліку товарів;
GRP	Gross Rating Point, сумарний рейтинг;
ROI	Return on investment, величина, яка ілюструє рівень успішності (прибутковості) рекламних кампаній для продажів;
RSS	Residuals sum of squares, сума квадратів залишків у моделі.

## Анотація

Робота присвячена створенню узагальненої системи одночасних рівнянь, застосування до неї тестів Хаусмана та Гренжера задля виявлення методу оцінки, та оцінка моделі залежності обсягів продажів від різних факторів, у тому числі реклами та її лагів. Робота виконана для реальних даних за допомогою середовища RStudio та Google Colab, модель оцінена за допомогою двокрокового методу найменших квадратів. Також до системи одночасних рівнянь було застосовано тести Хаусмана та Гренжера. Модель протестовано та зроблені висновки щодо даних.

## ВСТУП

Реклама як один із важливих рушіїв та посилювачів попиту – сфера, яка потребує якісних досліджень із застосуванням математичних методів. Разом із появою реклами на ринку сформувався запит на збір даних та оцінка проведених кампаній, проведення аналітики, що дозволяла би планувати рекламні кампанії так, аби отримати найбільше продажів: обирати рекламні канали, на яких та чи інша компанія могла би краще залучати свою цільову аудиторію, виявляти оптимальні дні та час для розміщення рекламних роликів, рівень та архітектуру рекламного тиску, довжину ролику, характер передачі інформації і т.д. У точному аналізі й полягає актуальність цієї роботи.

Робота складається з двох розділів.

Перший розділ присвячено теоретичній базі з реклами, предметної області роботи, а також з економетричного підходу до аналізу в предметній області. Розглянуто види реклами та їх особливості, наведено принципи застосування реклами в умовах конкуренції і відкладений ефект реклами, наведено критерії використання специфікаційного тесту Хаусмана та тесту причинності за Гренджером, розглянуто алгоритм застосування двокрокового методу найменших квадратів. Також була розроблена узагальнена система одночасних рівнянь обсягів продажів товару від різних факторів, у тому числі від рекламних кампаній.

У другому розділі було описано дані деякого бренду фармацевтичної компанії, побудовано систему одночасних рівнянь та застосовано до них специфікаційний тест Хаусмана та тест причинності за Гренджером у RStudio та Google Colab. Отримавши висновки з тестів, модель обсягу продажів оцінено за допомогою 2МНК у RStudio. Оцінка за допомогою 2МНК проаналізована та результати рекламних кампаній розглянуто як ROI. Створено програмний продукт, який призначений для

експериментального дослідження впливу реклами при прогнозуванні продажу продукту та в навчальному процесі.

**Мета** даної **роботи** полягає у створенні узагальненої системи одночасних рівнянь, застосуванні до неї тестів Хаусмана та Гренжера задля виявлення методу оцінки, та оцінка моделі залежності обсягів продажів від різних факторів, у тому числі реклами та її лагів, кількість яких буде визначена тестом Гренжера, методом, який буде визначено тестом Хаусмана.

**Постановка завдань роботи:**

1. Систематизувати теоретичну базу про види реклами, оглянути особливості рекламних кампаній, що вимагають аналітичного підходу.
2. Навести алгоритм використання методів Хаусмана та Гренжера.
3. Навести алгоритм застосування двокрокового методу найменших квадратів.
4. Розробити узагальнену систему одночасних рівнянь – залежності продажів від різних факторів, у тому числі від реклами, а також залежності реклами від обсягу продажів та ціни на рекламу.

У рамках практичної частини:

5. Описати дані та фактори, які впливають на продажі бренда.
6. Застосувати специфікаційний Хаусмана для визначення методу оцінки моделі продажів.
7. Застосувати метод причинності за Гренжером аби виявити причинні зв'язки та глибину лагів реклами.
8. Побудувати модель виходячи з розробленого загального рівняння, попередньо розділивши дані на навчальну та тестову вибірки.
9. Оцінити коефіцієнти моделі за допомогою двокрокового методу найменших квадратів, проаналізувати отримані значення. Оцінити значення для тестової вибірки та порівняти з реальними.
10. Порахувати ROI для рекламних кампаній та проаналізувати їх.

**Об'єктом дослідження** є моделювання і прогноз продажів брендів, а також оцінка якості рекламних кампаній.

**Методами дослідження** є класифікація та систематизація теоретичної бази, формалізація видів реклами та створення системи одночасних рівнянь у першому розділі, а також моделювання та аналіз у другому розділі.

**Джерела дослідження:** "Практика економетрики: классика и современность: Учебник для студентов вузов, обучающихся по специальностям 060000 экономики и управления" Берндта Е., "Словник української мови» Паламарчук Л., "Медиапланирование на 100%" Назайкін А., "One Way TV Advertisements Work" Broadbent, S., "Generalizing what is known about temporal aggregation and advert-over" Leone, R., "Економетрика" Лук'яненко І.

**Наукова новизна** одержаних результатів: створення експериментальної системи одночасних рівнянь, серед яких присутня модель залежності продажу товару від цінового індексу, дистрибуції, реклами з відкладеним ефектом а також інших факторів.

## РОЗДІЛ 1: Теоретичне підґрунтя

### 1.1 Реклама як явище

Латинське слово “reclamare” може перекладатися на українську як “повторно вигукувати” або “гукати раз-у-раз”. Вважається, що це слово є предком українського слова “реклама”. Відповідно до “Словнику української мови в 11 томах” (СУМ-11) можна виокремити два окремих види реклами: інформативну (також - пряму) та непрямую. За СУМ-11 ці визначення є:

*“Інформативна (пряма) реклама — реклама, яка створює первинний попит та інформує споживача про існування товару, його властивості.*

*Непряма реклама — неявна реклама, рекламування товарів та послуг через інтерв'ю, публікації, радіо і телепередачі, в яких згадується рекламований товар.” [3]*

Ці два види дуже важливо розрізняти у контексті досліджень – у нашому випадку економетричних – оскільки вплив та функції прямої реклами відрізняються від непрямой. У книзі “Медіапланування на 100%” доктор філологічних наук Олександр Назайкін переконує, що “потрібно розуміти, що люди дивляться саме телепрограми, а не рекламу. Тому аудиторія рекламних пауз відрізняється від аудиторії телепрограм” [4]. Вплив реклами на різні аудиторії слід розглядати як окремі чинники для формування подальшого попиту на товар.

### 1.1.1 Види реклами

#### 1.1.1.1 Реклама на телебаченні

У 1941 році у США вперше використали рекламу у телебаченні – глядачам було запропоновано купувати годинники під час трансляції баскетбольного матчу. Ще через 13 років реклама на телебаченні посіла перше місце у США як найбільш активно використовуваний засіб реклами. До України в складі СРСР реклама прийшла у 70-х, але тоді вона ще не була комерційною. В Україні комерційна реклама по телебаченню набула популярності на розі 80-х та 90-х. Наразі телебачення досі є одним з найбільш впливових майданчиків і формує найбільший ринок реклами в Україні.

Разом із появою реклами сформувався запит на збір кількісних даних про перегляди та проведення аналітики, що дозволяла би планувати рекламні кампанії так, аби отримати найбільше продажів: обирати телевізійні канали, на яких та чи інша компанія могла би краще залучати свою цільову аудиторію, виявляти оптимальні дні тижня та час для розташування рекламних роликів, рівень та архітектуру рекламного тиску, довжину ролику, характер передачі інформації (емоційна та раціональна складові).

Збір такої інформації спочатку відбувався за допомогою спеціальних щоденників, де фокус-групи буквально заповнювали розклад свого теледивлення. Зважаючи на “людський” фактор можна припустити, що цей метод був не достатньо точним. З часом було винайдено піплметрію – метод використання програмного забезпечення, що встановлюється на телевізори кожного домогосподарства з фокус-груп. Доволі точно описує цей метод Олександр Назайкін: “ці прилади [піплметри] автоматично реєструють канали та час перегляду. За кожним членом сім’ї закріплено окрему кнопку реєстрації [на спеціальному пульті]. Також передбачена гостьова кнопка.

Вся інформація надходить у центральний комп'ютер, який зрештою і видає результати перегляду з точністю до секунди” [4]. Метод піпметрії значно спростив та покращив якість збору інформації, мінімізувавши кількість похибок “людського” фактору. Так, коли ресерчерські компанії у сфері телебачення Естонії перейшли від щоденникової моделі до вимірювання переглядів за допомогою піпметрів, показники обсягу середньодобового перегляду закономірно зменшились більш ніж на 30 хвилин. [4]

Поділ на пряму та непрямую рекламу значною мірою відображається у телебаченні. Пряма реклама випускається у спеціальних рекламних блоках, які розташовані в перерві між телепрограмами; непрямая застосовується у вигляді продакт-плейсменту (речові реквізити у фільмах, телепрограмах і т.д., що мають реальний комерційний аналог) у телепрограмах або фільмах. Як було зазначено вище, слід розрізняти вплив реклами на глядача у цьому вимірі. Так само важливо розрізняти її за довжиною, особливо це стосується прямої реклами. Емпіричним шляхом було досліджено, що існує кореляція між довжиною ролика та тим, як добре його запам'ятовують. Короткі ролики найчастіше застосовуються для підтримки знання та лояльності до вже відомих брендів, а довші – для формування думки про бренд, розкриття його особливостей за допомогою інформації емоційного та/або раціонального характеру.

Вимірювання кількісних показників рекламного розміщення розраховується за допомогою рейтингів. Рейтинг – це відсоток аудиторії, яка переглянула рекламне повідомлення, серед всієї аудиторії, яка мала можливість бачити повідомлення.

$$Rating = (n / N) \times 100, \quad (1.1)$$

де  $n$  – число глядачів, які переглянули рекламне повідомлення,

$N$  – уся аудиторія телебачення на момент показу рекламного повідомлення.

Шляхом сумування рейтингів показів, отриманих під час розміщення реклами протягом всієї рекламної кампанії ми отримуємо сукупний або сумарний рейтинг, який зазвичай називають GRP (gross rating point). Як правило, GRP виражається у відсотках, при цьому сам знак % опускається. У зв'язку з тим, що сукупний рейтинг розраховується простою сумою, його величина може перевищувати 100%.

Наприклад, під час кампанії реклама була розміщена 4 рази в телепрограмах, що мають рейтинг 15%, 3 рази в телепрограмах з рейтингом 20% і 5 разів в телепрограмах з рейтингом 25%. GRP складе 245 пунктів (%):

$$GRP = (4 * 15 + 3 * 20 + 5 * 25) = 245. [4]$$

Під час досліджень нас цікавить відсоток саме цільової аудиторії, яка переглянула рекламне повідомлення. Тому з GRP можна вивести поняття TRP (target rating point) як GRP по цільовій аудиторії.

#### 1.1.1.2 Інтернет-реклама

Реклама в інтернеті почала відігравати істотну роль з 1994 року, коли журналом Wired був створений HotWired – перший веб-сайт, що існував за рахунок реклами. На цьому сайті були розміщені банери таких компаній, як AT&T, IBM та Zima. [4]

Інтернет, на відміну від телебачення, пропонує величезний вибір майданчиків для комунікації з потенційним клієнтом: пошукові системи, тематичні сайти, новинні та інформаційні ресурси, соціальні мережі. Однією з найбільш помітних особливостей реклами в інтернеті є доступний та

точковий таргетинг (спеціальний рекламний механізм, який дозволяє виділити з усієї аудиторії лише цільову й показати рекламу саме їй). Якщо у телебаченні необхідно проводити дослідження цільових аудиторій на каналах, виявляти час, коли та чи інша аудиторія присутня в більшій мірі, а також володіти великим бюджетом на рекламу, то мережа Інтернет дозволяє дуже точно задіяти рекламу на необхідну цільову аудиторію, а також зробити це у тому числі й за менший бюджет, що відкриває можливості для стартапів, малих брендів та бізнесу.

Інтернет реклама також поділяється на види: банерну, відео (наприклад, у роликах YouTube), у соціальних мережах, контекстну (у пошукових запитах). Було би помилкою не врахувати особливості цих видів при аналізі, так само як з видами телевізійної реклами.

Вимірюється інтернет реклама найчастіше у impressions, тобто кількості реакцій (переглядів).

### 1.1.1.3 Інша реклама

Інші канали реклами, що часто застосовують, це зовнішня, на радіо та у пресі. Особливістю та перевагою зовнішньої реклами є можливість дуже точно налаштувати цільову аудиторію географічно, з чим не змогла би впоратись реклама на телебаченні та лише частково – у Інтернеті. Реклама на радіо також має ряд своїх переваг та недоліків: її неможливо побачити (втрачається вплив візуальних характеристик, що часто підсилюють ефект), можна тільки уявити за допомогою голосу диктора. Тим не менше, для досягнення деяких цільових аудиторій вона – незамінний інструмент. Складно не погодитись з Назайкіним, який пише: “цей засіб масової інформації охоплює такі категорії людей, які недоступні для телебачення і преси, які знаходяться поза домом, у дорозі, які слухають інформацію під

час якогось заняття” [4]. Те ж саме можна сказати про пресу, хоча з розвитком інтернету цей канал страждає та скорочується найбільше.

### 1.1.2 Конкуренція та медійний тиск

Складність охоплення споживачів за допомогою телебачення полягає насамперед у тому, що воно не має постійної аудиторії через дуже високий рівень ротації глядачів. Люди постійно перемикаються з каналу на канал, переходять від однієї телепрограми до іншої, уникають рекламних блоків. Різні канали і телепрограми можуть дивитися майже ті самі люди. Через це реклама на телебаченні потребує значної кількості повторів. Телебачення досить зашумлене рекламою, і вигідно позиціонувати себе серед конкурентів досить складно.

Тим не менше, медійний тиск та паритет з головними конкурентами – це одна з частих задач для команди просування бренда. Відповідність або паритетність з рекламними витратами конкурентів реалізується за допомогою відстеження їх рекламних дій. На основі даних за обсягами реклами та величини їх бюджетів розраховується власний бюджет – на тому ж рівні або більший. Це допомагає тиснути на потенційних покупців принаймні не менш часто, ніж конкуренти, що у свою чергу змушує покупця вибирати.

### 1.1.3 Відкладений ефект – adstock

Припустимо, що деяка компанія починає розміщувати рекламу свого бренда на телебаченні. Разом із saleshouse (агентством, що спеціалізується на продажі реклами у інтернеті, телебаченні та пресі) було домовлено про те, що ролик з рекламою цього бренда будуть виставляти упродовж  $k$  (нехай  $k = 10$ ) тижнів з повтором через тиждень (класична стратегія week on – week off) з середнім  $GRP_{1...k}$ .

$$GRP_{1...10} = \{n, 0, n, 0, n, 0, n, 0, n, 0\} \quad (1.2)$$

Представимо на графіку, як буде виглядати архітектура рекламної кампанії (див. Рис. 1.1).

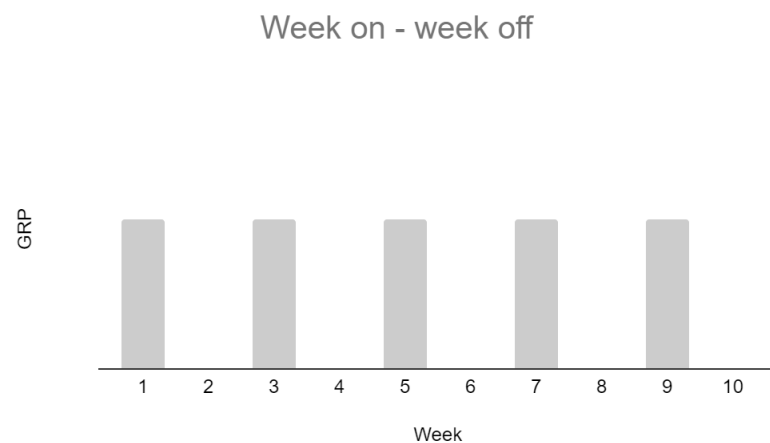


Рисунок 1.1 – приклад графічного представлення архітектури рекламної кампанії, виміряної у GRP

Можна припустити, що створюючи модель, ми могли би використати вимірювання такої кампанії як фактор: вплив телевізійної реклами підвищує обізнаність у свідомості споживачів, впливаючи на їхнє рішення про покупку в 1, 3, 5, 7 та 9 тижнях. Тим не менше, не можна не погодитись з економістом Бродбентом Саймоном, котрий вважає, що “кожне нове охоплення (перегляд) рекламою підвищує обізнаність, і поінформованість (awareness) буде вищою, якщо нещодавно охоплення були, і нижчою, якщо їх не було. За відсутності подальшого впливу рекламний ефект зрештою зменшується до незначних рівнів” [5]. Отже, було би помилково вважати, що ефект від реклами діє тільки у період (у нашому випадку – тиждень)

рекламного розміщення. Кожне нове розміщення ролику збільшує обізнаність про товар, і ефект від цього може продовжуватись більше, ніж період показу реклами. Цей ефект називають Advertising Adstock.

Компонент подовженого чи відкладеного ефекту Advertising Adstock (надалі – adstock) можна математично змодельовати і він зазвичай виражається як “період розпаду” (лаговий ефект) рекламної кампанії та моделюється у нашому випадку за допомогою GRP. Наприклад, “двотижневий період розпаду” з коефіцієнтом adstock  $\alpha = 0,5$  означає, що потрібно два тижні для того, щоб обізнаність про рекламне розміщення зменшилося до половини свого початкового рівня. У деяких академічних дослідженнях припускається, що “період розпаду” становить близько 7-12 тижнів. [6]

Для попереднього прикладу (Мал. 1.1) adstock з коефіцієнтом  $\gamma$  для  $k = 10$  періодів (наприклад – тижнів) можна розрахувати за наступною формулою:

$$\begin{aligned} AdstockGRP_1 &= GRP_1, \\ AdstockGRP_i &= AdstockGRP_{i-1} * \gamma + GRP_i, i = \{2, \dots, k\}, \end{aligned} \quad (1.3)$$

Нижче наводимо графік, який ілюструє модель adstock для прикладу вище (див. Рис. 1.2).

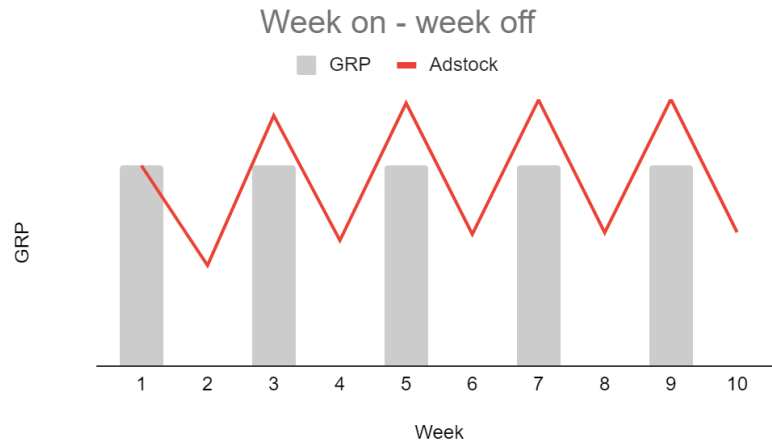


Рисунок 1.2 – приклад графічного представлення архітектури рекламного розміщення, виміряного у GRP та adstock

Також на Рисунках 1.3 та 1.4 можна побачити графіки для інших архітектур рекламних розміщень.

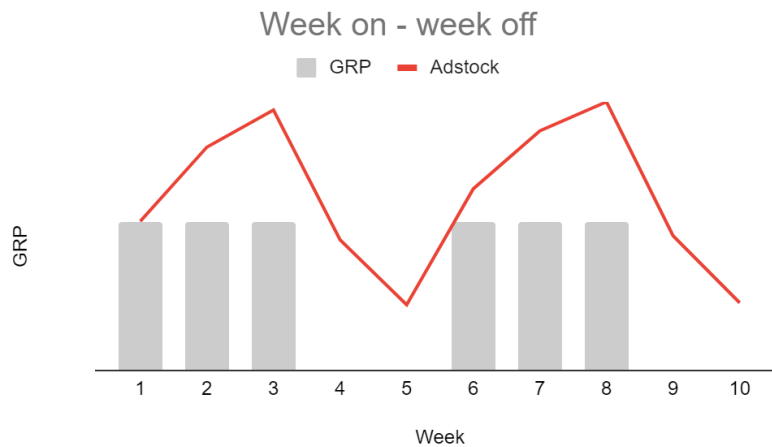


Рисунок 1.3 – приклад графічного представлення архітектури рекламного розміщення, виміряного у GRP та adstock

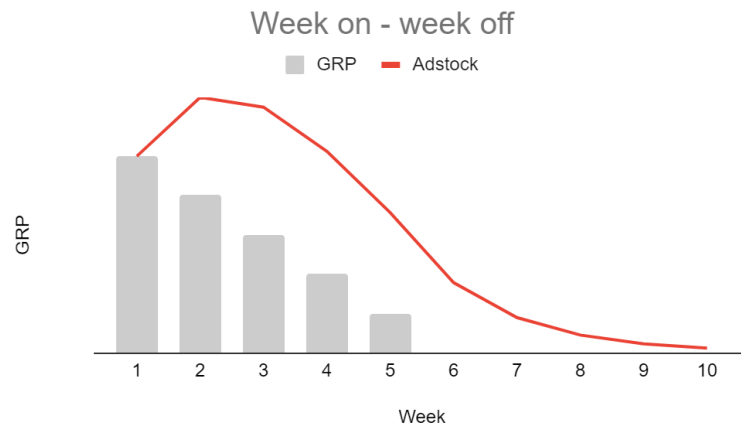


Рисунок 1.4 – приклад графічного представлення архітектури рекламного розміщення, виміряного у GRP та adstock

Adstock слід розглядати у контексті усіх медійних каналів. Реклама у мережі Інтернет, радіо, пресі, зовнішня реклама – ефект від розміщень на усіх цих каналах також має відкладений ефект.

Тим не менше, adstock можна замінити лагами, але тоді необхідно пильно підбирати метод оцінки моделі, аби уникнути автокореляції.

## 1.2 Економетричний підхід у рекламі

Реклама є однією зі сфер, де необхідно застосовувати математичні методи. Великі бюджети, які нині витрачають компанії для вдалого просування свого продукту вимагають планування рекламних стратегій із заданою точністю. Рекламні стратегії формуються на основі досліджень предметної області, вивчення попереднього досвіду своїх продуктів та продуктів конкурентів.

Економетричні поняття, які ми застосуємо у роботі, включають у себе проблеми вимірювання обсягу та ціни реклами, процедури застосування тесту специфікації Хаусмана (Hausman) та перевірки причинності по

Гренжеру (Granger), правильного обчислення  $adstock$  для вимірювання кумулятивного ефекту реклами за умови розподілених лагів.

Введемо деякі позначення. Нехай  $M$  означатиме кількість рекламних повідомлень,  $T$  - ціну одного повідомлення, також нехай функція попиту на продукт матиме вид  $Q = Q(M, P)$ , де  $Q$  – кількість товару на виході,  $P$  – ціна.

Можемо визначити прибуток як загальний дохід віднявши витрати на виробництво та рекламу:

$$\pi = P * Q(M, P) - C[Q(M, P)] - M * T, \quad (1.4)$$

### 1.2.1 Тест Хаусмана

Розглянемо наступну систему рівнянь взаємної ендогенності об'єму продажів та реклами:

$$S_t = a + bM_t + cP_{s,t} + u_t; \quad (1.5,a)$$

$$M_t = d + eS_t + fP_{m,t} + w_t; \quad (1.5,b)$$

де  $S$  - об'єм продажів у кількісному вимірі,

$M$  - кількість рекламних повідомлень,

$P_s$  та  $P_m$  – ціна відповідно товару та реклами,

$u$  та  $w$  – випадкові величини (викиди),

$t$  – період часу, якому належать усі показники вище.

Припустимо, що  $u$  та  $w$  незалежні та однаково нормально розподілені, і що  $u$  розподілені незалежно від  $P_s$ , а  $w$  – незалежно від  $P_m$ , а також, що  $u$  та  $w$ , узяті в один той самий момент можуть корелювати. Також ми очікуємо, що  $b$  та  $e$  є додатними, а  $c$  та  $f$  – від’ємні.

Вирішимо рівняння (1.5,б) спочатку відносно змінної  $S_t$ , прирівнявши її до (1.5,а), а отримане відношення, що вже не міститиме  $S_t$ , вирішимо відносно  $M_t$ :

$$\begin{aligned} M_t &= d + e * (a + bM_t + cP_{s,t} + u_t) + fP_{m,t} + w_t ; \\ M_t &= d + ea + beM_t + ecP_{s,t} + eu_t + fP_{m,t} + w_t ; \\ (1 - be)M_t &= d + ea + ecP_{s,t} + eu_t + fP_{m,t} + w_t . \end{aligned} \tag{1.6,a}$$

Для всіх коефіцієнтів (1.6,а) зробимо заміну:

$$\begin{aligned} \alpha_0 &= \frac{d+ea}{1-be}, \alpha_1 = \frac{f}{1-be}, \alpha_2 = \frac{ec}{1-be}, v_t = \frac{eu_t+w_t}{1-be}, \\ M_t &= \alpha_0 + \alpha_1 P_{m,t} + \alpha_2 P_{s,t} + v_t . \end{aligned} \tag{1.6,б}$$

Те саме зробимо для  $M_t$ :

$$\begin{aligned} S_t &= a + b * (d + eS_t + fP_{m,t} + w_t) + cP_{s,t} + u_t ; \\ S_t &= a + bd + beS_t + bfP_{m,t} + bw_t + cP_{s,t} + u_t ; \\ (1 - be)S_t &= a + bd + bfP_{m,t} + bw_t + cP_{s,t} + u_t ; \end{aligned}$$

$$\text{Заміна: } \beta_0 = \frac{a+bd}{1-be}, \beta_1 = \frac{bf}{1-be}, \beta_2 = \frac{c}{1-be}, z_t = \frac{bw_t+u_t}{1-be}, \quad (1.7,a)$$

$$S_t = \beta_0 + \beta_1 P_{m,t} + \beta_2 P_{s,t} + z_t. \quad (1.7,b)$$

Тепер маємо наступну систему:

$$M_t = \alpha_0 + \alpha_1 P_{s,t} + \alpha_2 P_{m,t} + v_t; \quad (1.8,a)$$

$$S_t = \beta_0 + \beta_1 P_{s,t} + \beta_2 P_{m,t} + z_t, \quad (1.8,b)$$

де випадкові компоненти  $v$  та  $z$  розподілені незалежно від регресорів  $P_{s,t}$  та  $P_{m,t}$ .

З цього слідує, що оцінка рівнянь у зведеній формі (1.8a,б) за допомогою МНК дає найкращі лінійні незміщені оцінки параметрів зведеної форми.

Одна з проблем, що впливає зі зведеної форми, це вибір застосування статистичних критеріїв оцінки моделей продажів або реклами.

Гіпотеза полягає у тому, що чинники моделі екзогенні і ми можемо застосувати МНК, альтернативна – що деякі з них ендогенні, і тоді потрібно застосувати метод двокрокового МНК (надлі – 2МНК). У обох випадках 2МНК дає конзистентні оцінки. Метод найменших квадратів дає конзистентні оцінки тільки при екзогенності факторів. Вибір між методами дозволяє зробити процедура, яка називається специфікаційним тестом Хаусмана.

Нульова гіпотеза теста Хаусмана, що впливає із гіпотези вище:  $M_t$  та  $u_t$  рівняння (1.5,a) не корельовані при великих об'ємах вибірки (і можна застосувати МНК). Альтернативна гіпотеза –  $M_t$  та  $u_t$  корельовані та

застосування МНК необхідно замінити використанням 2МНК. Нульова гіпотеза перевіряється за допомогою порівняння оцінок, які вираховуються за допомогою МНК та 2МНК.

Щоб застосувати тест, спочатку оцінимо рівняння (1.8,а) за допомогою МНК та обчислимо  $\widehat{M}_t$  (регресійне значення). Після цього застосуємо МНК-процедуру оцінки до розширеного рівняння регресії (1.5а):

$$S_t = a + bM_t + cP_{s,t} + h\widehat{M}_t + u_t, \quad (1.9)$$

де  $\widehat{M}_t$  – додатковий регресор.

Нульова гіпотеза про некорельованість  $M_t$  та  $u_t$  має підтвердитися і зводиться до перевірки гіпотези, що  $h = 0$ . Перевірку можна зробити шляхом вирахування відношення оцінки параметра  $h$  до його стандартної похибки та потім порівнявши значення з табличним критичним значенням нормального розподілу.

Звідси наша нульова гіпотеза  $M(h) = 0$ . Очевидно, що при справедливості нульової гіпотези, маємо  $M(\underline{h}) = 0$ . Використовуючи незміщену оцінку дисперсії

$$s_h^2 = \frac{\sum_{t=1}^n (h_t - \underline{h})^2}{n - 1} \quad (1.10)$$

Маємо наступну  $t$  статистику для перевірки  $h = 0$ :

$$t = \frac{(\underline{h} - 0)\sqrt{n}}{S_h} \quad (1.11)$$

При нульовій гіпотезі розподіл цієї статистики  $t(n - 1)$ . Отже, при перевищенні значення статистики за абсолютною величиною критичного значення даного розподілу (за заданим рівнем значущості) нульова гіпотеза відкидається.

### 1.2.2 Тест Гренжера

Звична концепція про те, що  $Y$  залежить від  $X$  має бути розглянута з обережністю, оскільки вона здатна відкрити приховані та складні проблеми, які виходять за рамки економетрики. Хоча й не існує загальноприйнятого універсального визначення причинної залежності, існує одна процедура, яка необхідна для виявлення причинної залежності в аналізі даних. Вона має назву тесту причинності за Гренжером. По суті, тест причинності за Гренжером включає у себе використання тесту Фішера, який застосовується для перевірки того, чи має лагова інформація про змінну, наприклад  $X$ , статистично значущий вплив при поясненні  $Y_t$  з урахуванням у якості змінних, що її пояснюють, і лагових значень  $Y$ . Якщо при наявності лагових значень  $Y$  лагові  $X$  не вносять статистично значимого впливу в пояснення  $Y_t$ , то кажуть, що “ $X$  не є причиною змін у  $Y$  за Гренжером”.

У ситуації з об’ємом реклами та продажами причинність за Гренжером може бути досліджена наступним чином:

$$M_t = \alpha + \sum_{i=1}^I \beta_i M_{t-i} + \sum_{j=1}^J \gamma_j S_{t-j} + \varepsilon_t, \quad (1.12)$$

де  $\varepsilon_t$  – випадкова похибка (є білим шумом) з нульовим середнім та дисперсією  $\sigma^2$ ,

$J$  та  $I$  – достатньо великі, щоб можливо було дослідити варіювання “глибини” автокореляції.

Рівняння регресії (1.12) оцінюється при включенні змінних  $S_{t-j}$  та без них, а потім виконується тест Фішера для перевірки нульової гіпотези про те, що  $\gamma_j = 0, j = 1, \dots, J$ . Отже, маємо два рівняння з індексами  $S$  і  $L$ , які відносяться відповідно до короткої та довгої моделі (моделі з обмеженнями та моделі без обмежень):

$$L: M_t = \alpha + \sum_{i=1}^I \beta_i M_{t-i} + \sum_{j=1}^J \gamma_j S_{t-j} + \varepsilon_t, \quad (1.13,а)$$

$$S: M_t = \alpha + \sum_{i=1}^I \beta_i M_{t-i} + \varepsilon_t, \quad (1.13,б)$$

Статистика тесту Фішера для перевірки лінійних обмежень по параметрах класичної нормальної лінійної регресії визначається за такою формулою:

$$F = \frac{(RSS_S - RSS_L)(n - k_L)}{RSS_L q}, \quad (1.14)$$

де  $q = \text{count\_paramters}(L) - \text{count\_paramters}(S)$  – кількість обмежень (різниця між кількістю параметрів у  $L$  та  $S$  моделях),

$n$  – об'єм вибірки

$k$  – кількість параметрів моделі,

$RSS$  – сума квадратів залишків моделей.

Якщо величина отриманої F-статистики більше критичного значення (нульова гіпотеза відкидається), то кажуть, що об'єм продажів є причиною за Гренжером для реклами. У протилежному випадку про причинні зв'язки за Гренжером між рекламою та об'ємом продажів сказати нічого не можна.

Навпаки відбувається для визначення причинності за Гренжером чи є реклама причиною для об'єму продажів.

### 1.2.3 Метод двокрокового МНК

У системах одночасних рівнянь двокроковий МНК (далі – 2МНК) застосовується для оцінки параметрів структурних рівнянь, оскільки в останніх як фактори беруть участь ендогенні змінні моделі і застосування звичайного МНК призводить до зміщених і неспроможних оцінок.

2МНК тісно пов'язаний з методом інструментальних змінних. Іноді його і називають узагальненим чи просто методом інструментальних змінних. Оцінюючи одиночні рівняння використовуються додаткові (інструментальні) змінні, які безпосередньо у модель не включені. Їх використання обумовлене тим, що частина факторів моделі можуть не задовольняти вимогу екзогенності.

Суть 2МНК полягає в наступному:

Крок 1. За допомогою МНК оцінюється залежність ендогенних змінних від усіх екзогенних.

Крок 2. За допомогою МНК оцінюється структурна форма моделі, де замість ендогенних змінних використовуються їх оцінки, отримані на першому кроці. [1]

### 1.3 Система одночасних рівнянь обсягів продажів та реклами

Як ми вже зазначали вище, якісний аналіз у сфері реклами неможливий без математичного, а у нашому випадку економетричного, підходу. Регресійна модель є одним з простих, але досить надійних методів для визначення впливових факторів та прогнозування.

У нашій моделі мусимо визначити ендогенну та екзогенні змінні. Ендогенна змінна формується внаслідок комбінації екзогенних (на які впливають зовнішні фактори) та ендогенних змінних. Виокремимо ендогенні змінні – змінні, які формуються всередині моделі і є результатом комбінації моделлю екзогенних показників. Ендогенними змінними нашої моделі є продажі (у грошовому еквіваленті, у літрах, кілограмах, пачках, тощо), а точніше доля ринку продажів серед конкурентів, та обсяги реклами. Під ринком мається на увазі група конкуруючих брендів, доля на ринку – це відсоткове чи абсолютне значення частки продажів конкретного бренда від продажів ринку в цілому.

$$S_i = \frac{s_i}{S_A}, \quad (1.15)$$

де  $S_i$  – сума продажів  $i$ -го бренда за даний період,

$S_A$  – сума продажів усіх брендів категорії за даний період.

Використання долі ринку на відміну від звичайного використання обсягів продажів є необхідним, аби зменшити вплив сезонних чинників. Особливо це важливо при створенні регресійної моделі у сфері реклами: рекламні кампанії часто проводять тоді, коли є більший попит на категорію бренда. Наприклад, бренд морозива розташовує рекламу найбільше влітку, аби залучити покупців у найближчій перспективі. Якщо ж ми братимемо під час побудови моделі абсолютні значення продажів морозива у цей період, модель скоріше за все виявить величезний вплив реклами на продажі там, де це був простий сезонний підвищений попит. Доля ринку ж показує, наскільки наш бренд змінюється відносно інших брендів, що значно згладжує вплив сезонності (але не завжди).

Основні екзогенні чинники, які можемо навести для моделі долі ринку:

- **Дистрибуція** є одним з ключових факторів продажів бренда. Під цим терміном маємо на увазі відсоток магазинів, торгових точок, у яких можна придбати товар певного бренда (за винятком інтернет-магазинів та майданчиків). Чим більша дистрибуція бренда, тим більше ймовірність покупки товару. Зазвичай розглядається окрема дистрибуція для кожного SKU (одиниця обліку товару).
- Для основної групи товарів маємо очевидний факт: більша ціна породжує менший попит. У роботі буде розглядатись саме **ціновий індекс** – відношення ціни бренда до середньої ціни на ринку. Такий підхід є важливим для нівелювання впливу інфляції та сезонних факторів: таким чином ми розглядаємо не просто ціну, а зміни цінової політики бренда відносно інших брендів.
- **Реклама** як фактор з відкладеним ефектом (adstock). Зазвичай для великих рекламних кампаній окрім поділу по каналах відбувається поділ на види реклами.

- Також за потреби додаються **інші фактори**, як-от активність конкурентів, святкові дні, температура повітря, тощо.

Основні чинники для моделі залежності обсягу реклами:

- Доля ринку.
- Ціна на рекламу.

Отже, наша модель буде виглядати наступним чином:

$$\left\{ \begin{array}{l} S_t = \alpha_0 + \sum_{i=1}^I \alpha_i d_{i,t} + \alpha_j P_{S,t} + \sum_{k=1}^K \alpha_k M_{k,t} + \\ \quad \sum_{l=1}^L \alpha_l s_{l,t} + e_{1,t}, \\ M_{1,t} = \beta_0 S_t + \beta_1 P_{M,t} + e_{2,t}, \\ \quad \dots \\ M_{k,t} = \gamma_0 S_t + \gamma_1 P_{M,t} + e_{k+2,t} \end{array} \right. \quad (1.16)$$

де  $S_t$  - доля ринку бренда,

$d_{i,t}, i = [1, \dots, I]$  – дистрибуції усіх SKU,

$P_{S,t}$  – ціновий індекс бренда,

$M_{k,t}, k = [1, \dots, K]$  – рекламні розміщення по всіх каналах та їх лаги,

$s_{l,t}, l = [1, \dots, L]$  – інші фактори,

$e_t$  – похибка.

Оцінивши оцінку такої моделі можна провести шляхом застосуванням 2МНК або МНК в залежності від теста Хаусмана, а застосувати модель можна для прогнозування у короткостроковій перспективі та для вирахування ROI.

## РОЗДІЛ 2: Практична частина

### 2.1 План та опис даних

Хід практичної частини:

1. Описати дані та фактори, які впливають на продажі бренда.
2. Застосувати специфікаційний Хаусмана для визначення методу оцінки моделі продажів.
3. Застосувати метод причинності за Гренжером аби виявити причинні зв'язки та глибину лагів реклами.
4. Побудувати модель виходячи з розробленого загального рівняння, попередньо розділивши дані на навчальну та тестову вибірки.
5. Оцінити коефіцієнти та коефіцієнти відкладеного ефекту за допомогою двокрокового методу найменших квадратів, проаналізувати отримані значення. Оцінити значення для тестової вибірки та порівняти з реальними.
6. Порахувати ROI для рекламних кампаній та проаналізувати їх.

Дані беремо по деякому бренду А з фармацевтичної категорії. , вони представлені потижнево у часових рамках з першого тижня 2019 року по вісімнадцятий тиждень 2021 року. Таблиця містить наступні колонки даних:

- year – рік,
- week – тиждень,
- category\_sales\_in\_units – обсяги продажів по всьому ринку,
- sales\_in\_currency – обсяги продажів бренду А у грошовому еквіваленті,
- sales\_in\_units – обсяги продажів бренду А в упаковках,
- SOM – share of market, доля ринку бренду А,
- distr\_1 – дистрибуція першого SKU бренду А,

- distr\_2 – дистрибуція першого SKU бренда А,
- price\_index – ціновий індекс бренда А,
- adv\_tv – реклама по телебаченню бренда А у TRP,
- adv\_tv\_comps – сумарна реклама по телебаченню конкурентів бренда А у TRP,
- adv\_int – реклама у інтернеті бренда А у TRP,
- adv\_int\_comps – реклама у інтернеті конкурентів бренда А у TRP,
- adv\_price\_tv – ціна на рекламні кампанії по телебаченню,
- adv\_price\_int – ціна на рекламні кампанії по телебаченню.

Упродовж 2019-2021 років доля ринку бренда А змінюється від 4% до 7,5% (див. Рисунок 2.1). Можемо припустити виходячи з візуальної оцінки, що присутня залежність між “піками” росту долі ринку та рекламними розміщеннями.

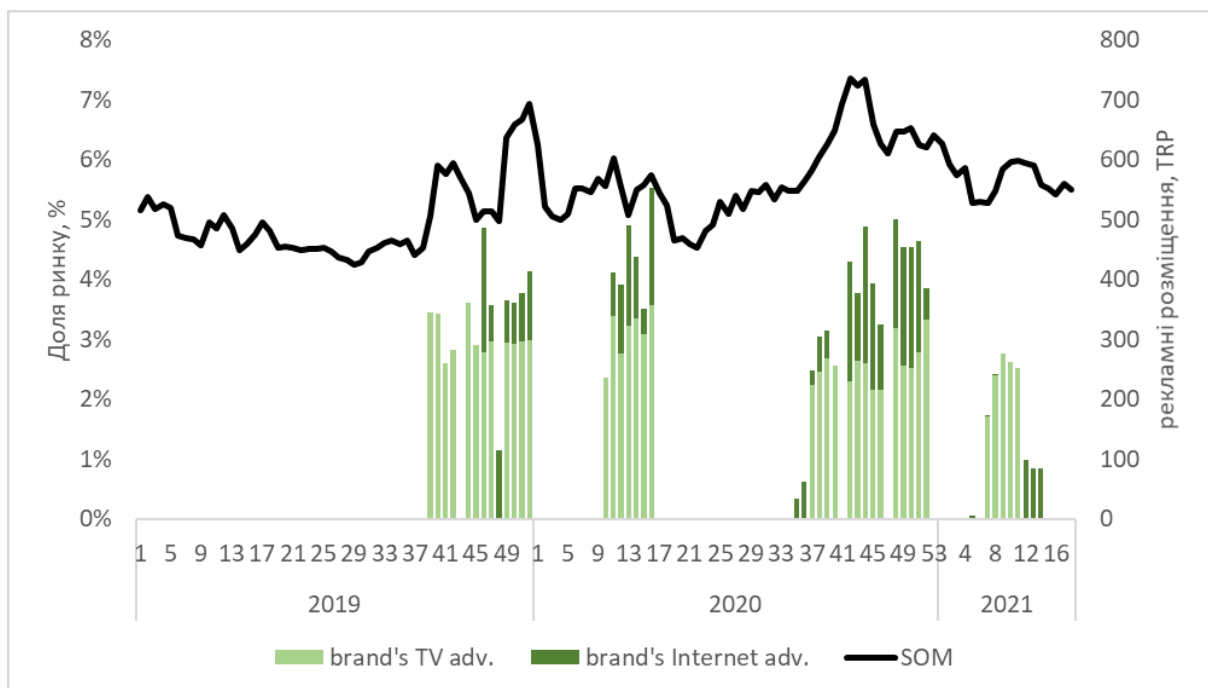


Рисунок 2.1 – доля ринку бренда А та його рекламні розміщення, виміряні у TRP

## 2.2 Застосування тесту Хаусмана

За допомогою специфікаційного тесту Хаусмана перевіримо, чи є  $M_t$  ендогенною чи екзогенною змінною, звідси з'ясуємо, який метод краще використати – метод двокрокового МНК (надлі – 2МНК) чи звичайний МНК. Для створення моделей з даних створимо три моделі відповідно до формули загальної системи одночасних рівнянь (1.16).

$$S_t = \alpha_0 + \alpha_1 d_{1,t} + \alpha_2 d_{2,t} + \alpha_3 P_t + \alpha_4 M_{TV,t} + \alpha_5 M_{int,t} + \quad (2.1a)$$

$$+ \alpha_6 s_{TV,t} + \alpha_7 s_{int,t} + e_{1,t} \quad (2.1b)$$

$$M_{TV,t} = \beta_0 S_t + \beta_1 P_{TV,t} + e_{2,t}$$

$$M_{int,t} = \gamma_0 S_t + \gamma_1 P_{int,t} + e_{3,t} \quad (2.1b)$$

$$\begin{aligned} S_t &= \alpha_0 + \alpha_1 d_{1,t} + \alpha_2 d_{2,t} + \alpha_3 P_t + \\ &+ \alpha_4 \beta_0 S_t + \alpha_4 \beta_1 P_{TV,t} + \alpha_4 e_{2,t} + \alpha_5 \gamma_0 S_t + \alpha_5 \gamma_1 P_{int,t} + \alpha_5 e_{3,t} + \\ &+ \alpha_6 s_{TV,t} + \alpha_7 s_{int,t} + e_{1,t}, \\ (1 - \alpha_5 \gamma_0 - \alpha_4 \beta_0) S_t &= \alpha_0 + \alpha_1 d_{1,t} + \alpha_2 d_{2,t} + \alpha_3 P_t + \\ + \alpha_4 \beta_1 P_{TV,t} + \alpha_4 e_{2,t} + \alpha_5 \gamma_1 P_{int,t} + \alpha_5 e_{3,t} + \alpha_6 s_{TV,t} + \alpha_7 s_{int,t} + e_{1,t}, \end{aligned} \quad (2.2)$$

Поділимо та виконаємо заміну коефіцієнтів у (2.2):

$$\begin{aligned} S_t &= \eta_0 + \eta_1 d_{1,t} + \eta_2 d_{2,t} + \eta_3 P_t + \\ &+ \eta_4 \beta_1 P_{TV,t} + \eta_5 \gamma_1 P_{int,t} + \eta_6 s_{TV,t} + \eta_7 s_{int,t} + e_{4,t}, \end{aligned} \quad (2.3)$$

Підставимо  $S_t$ , отримане в (2.2), у (2.1б,в):

$$M_{TV,t} = \beta_0(\eta_0 + \eta_1 d_{1,t} + \eta_2 d_{2,t} + \eta_3 P_t + \eta_4 \beta_1 P_{TV,t} + \eta_5 \gamma_1 P_{int,t} + \eta_6 S_{TV,t} + \eta_7 S_{int,t} + e_{1,t}) + \beta_1 P_{TV,t} + e_{2,t} \quad (2.4a)$$

$$M_{int,t} = \beta_0(\eta_0 + \eta_1 d_{1,t} + \eta_2 d_{2,t} + \eta_3 P_t + \eta_4 \beta_1 P_{TV,t} + \eta_5 \gamma_1 P_{int,t} + \eta_6 S_{TV,t} + \eta_7 S_{int,t} + e_{1,t}) + \gamma_1 P_{int,t} + e_{3,t} \quad (2.4б)$$

Замінімо коефіцієнти (2.4а,б) та отримаємо наступні два рівняння для  $M_{TV,t}$  та  $M_{int,t}$ , виражених через екзогенні змінні.

$$M_{TV,t} = \theta_0 + \theta_1 d_{1,t} + \theta_2 d_{2,t} + \theta_3 P_t + \theta_4 \beta_1 P_{TV,t} + \theta_5 \gamma_1 P_{int,t} + \theta_6 S_{TV,t} + \theta_7 S_{int,t} + e_{5,t} \quad (2.5a)$$

$$M_{int,t} = \vartheta_0 + \vartheta_1 d_{1,t} + \vartheta_2 d_{2,t} + \vartheta_3 P_t + \vartheta_4 \beta_1 P_{TV,t} + \vartheta_5 \gamma_1 P_{int,t} + \vartheta_6 S_{TV,t} + \vartheta_7 S_{int,t} + e_{6,t} \quad (2.6б)$$

У середовищі RStudio створимо та оцінимо регресійні значення  $M_{TV,t}$  та  $M_{int,t}$  (2.5а,б), за допомогою МНК. МНК ми можемо застосувати, оскільки у  $M_{TV,t}$  та  $M_{int,t}$  виражені тільки через екзогенні змінні.

```

Call:
lm(formula = adv_tv ~ distr_1 + distr_2 + price_index + adv_price_tv +
    adv_price_int + adv_tv_comps + adv_int_comps, data = df_H)

Residuals:
    Min       1Q   Median       3Q      Max
-108.140  -5.911  -2.804   5.799  104.727

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  20.705326  10.949156   1.891  0.0611 .
distr_1      -4.095968   2.853514  -1.435  0.1539
distr_2      -2.439481   6.682352  -0.365  0.7157
price_index  -33.548695  39.090899  -0.858  0.3926
adv_price_tv   0.196784   0.005294  37.169 <2e-16 ***
adv_price_int -0.017555   0.016292  -1.078  0.2835
adv_tv_comps  -0.048005   0.020552  -2.336  0.0212 *
adv_int_comps  0.014758   0.020716   0.712  0.4777
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27.16 on 115 degrees of freedom
Multiple R-squared:  0.9601,    Adjusted R-squared:  0.9577
F-statistic: 395.5 on 7 and 115 DF,  p-value: < 2.2e-16

Call:
lm(formula = adv_int ~ distr_1 + distr_2 + price_index + adv_price_tv +
    adv_price_int + adv_tv_comps + adv_int_comps, data = df_H)

Residuals:
    Min       1Q   Median       3Q      Max
-100.763  -7.163   0.110   6.029  117.194

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -1.323662  14.181631  -0.093  0.92580
distr_1      -2.760963   3.695946  -0.747  0.45657
distr_2       9.029916   8.655155   1.043  0.29900
price_index  -50.051944  50.631545  -0.989  0.32496
adv_price_tv   0.018380   0.006857   2.680  0.00844 **
adv_price_int  0.190111   0.021102   9.009 5.27e-15 ***
adv_tv_comps  -0.066522   0.026620  -2.499  0.01387 *
adv_int_comps  0.037255   0.026832   1.388  0.16769
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 35.18 on 115 degrees of freedom
Multiple R-squared:  0.6595,    Adjusted R-squared:  0.6387
F-statistic: 31.81 on 7 and 115 DF,  p-value: < 2.2e-16

```

Рисунок 2.2 – скріншот оцінки у RStudio моделей методом МНК

Далі підставляємо їх як  $\hat{M}_{TV}$  та  $\hat{M}_{int}$  у модель (2.1a) (див. Рис. 2.3).

$$\begin{aligned}
 S_t = & \alpha_0 + \alpha_1 d_{1,t} + \alpha_2 d_{2,t} + \alpha_3 P_t + \alpha_4 M_{TV,t} + \alpha_5 M_{int,t} + \\
 & + \alpha_6 s_{TV,t} + \alpha_7 s_{int,t} + \alpha_8 \hat{M}_{TV,t} + \alpha_9 \hat{M}_{int} + e_{1,t}
 \end{aligned} \tag{2.7}$$

```

Call:
lm(formula = SOM ~ adv_tv_fitted + adv_int_fitted + adv_tv +
    adv_int + distr_1 + distr_2 + price_index + adv_tv_comps +
    adv_int_comps, data = df_H)

Residuals:
    Min       1Q   Median       3Q      Max
-0.014018 -0.002821  0.000097  0.002788  0.011716

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.586e-02  1.889e-03  29.572 < 2e-16 ***
adv_tv_fitted  3.659e-05  1.684e-05   2.172  0.031944 *
adv_int_fitted  2.193e-05  1.892e-05   1.159  0.031783 *
adv_tv        -2.397e-05  1.597e-05  -1.500  0.136296
adv_int        2.489e-05  1.233e-05   2.018  0.045992 *
distr_1        2.336e-03  4.867e-04   4.799  4.91e-06 ***
distr_2       -4.428e-03  1.131e-03  -3.915  0.000155 ***
price_index   -8.457e-03  6.642e-03  -1.273  0.205546
adv_tv_comps   6.383e-06  3.591e-06   1.777  0.078230 .
adv_int_comps -7.918e-06  3.642e-06  -2.174  0.031783 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.004648 on 113 degrees of freedom
Multiple R-squared:  0.6109,    Adjusted R-squared:  0.5799
F-statistic: 19.71 on 9 and 113 DF,  p-value: < 2.2e-16

```

Рисунок 2.3 – скріншот оцінки у RStudio моделі (2.2) за допомогою МНК

Обчисливши t-статистику отримуємо, що при рівні значимості 5% p-value t-статистики по тобто каналах реклами менші за 0.05, що означає, що коефіцієнти при інструментальних змінних  $\hat{M}_{TV}$  та  $\hat{M}_{int}$  не можна відкинути. У такому випадку, регресори можна вважати ендогенними та краще застосувати 2МНК.

Висновок по результатам тесту Хаусмана: при виборі між використанням звичайного МНК та 2МНК у нашому прикладі віддаємо перевагу використанню 2МНК, оскільки всі деякі змінні моделі виявились ендогенними для нашого прикладу.

### 2.3 Застосування тесту причинності по Гренжеру

Застосуємо тест про причинність за Гренжером. За допомогою нього визначимо, чи є реклама причиною за Гренжером для продажів та чи є продажі причиною за Гренжером для реклами. Також у разі наявності

причинних зв'язків визначимо “глибину” лагів, яку нам варто взяти для моделі. Для створення моделей з даних виберемо наступні стовпчики:

- $M_{TV}$ , обсяг TRP рекламних кампаній по телебаченню,
- $M_{int}$ , обсяг TRP рекламних кампаній в інтернеті,
- $S_t$ , обсяг продажів бренду.

Для досліджень застосуємо наступний алгоритм. Створимо моделі залежності обсягу реклами від обсягу продажів з від 1 до  $K$  лагових значень.

Формула (1.13) для нашого прикладу виглядатиме наступним чином:

$$M_{1,t} = \alpha + \sum_{i=1}^1 \beta_i M_{t-i} + \sum_{j=1}^1 \gamma_j S_{t-j} + \varepsilon_t, \quad (2.8)$$

...

$$M_{10,t} = \alpha + \sum_{i=1}^K \beta_i M_{t-i} + \sum_{j=1}^K \gamma_j S_{t-j} + \varepsilon_t,$$

Сформуємо моделі з (2.8), короткі (S – short) та довгі (L – long):

$$L: M_{1,t} = \alpha + \sum_{i=1}^1 \beta_i M_{t-i} + \sum_{j=1}^1 \gamma_j S_{t-j} + \varepsilon_t, \quad (2.9,a)$$

...

$$M_{1,t} = \alpha + \sum_{i=1}^K \beta_i M_{t-i} + \sum_{j=1}^K \gamma_j S_{t-j} + \varepsilon_t$$

$$S: M_{1,t} = \alpha + \sum_{i=1}^1 \beta_i M_{t-i} + \varepsilon_t, \quad (2.9,b)$$

...

$$M_{1,t} = \alpha + \sum_{i=1}^K \beta_i M_{t-i} + \varepsilon_t$$

Для моделей порахуємо критерій Фішера за загальною формулою (1.14). Після цього порівняємо табличне критичне значення Фішера з таблиці для 5% та критерій Фішера. Якщо перший буде більший за другий, то ми можемо прийняти нульову гіпотезу про те, що всі коефіцієнти при  $S_t$  та його лагових значеннях рівні нулю. Також з цього робимо висновок, що для наших даних незалежна змінна не є причиною для залежної за Гренжером. У протилежному випадку приймаємо, що незалежна змінна є причиною для залежної.

Для виконання цього алгоритму застосуємо середовище Google Colab (функцію `grangercausalitytests`). Спочатку перевіримо, чи є по Гренжеру обсяг продажів причиною для реклами.

Застосовуємо `grangercausalitytests` до кожної з 5 моделей, у кожному з яких включаємо від 1 до 5 лагів змінної  $S_t$ , ми отримали F-статистики для кожної моделі. З Рисунку 2.4 бачимо, що починаючи з першої лагової змінної F-статистика та її p-value свідчать про те, що обсяг продажів не є

причиною для реклами по Гренжеру, оскільки при рівні значимості у 1%, 5% та 10% p-value їх значно перевищує.

```

Granger Causality
number of lags (no zero) 1
ssr based F test:      F=0.2642 , p=0.6082 , df_denom=119, df_num=1
ssr based chi2 test:  chi2=0.2709 , p=0.6027 , df=1
likelihood ratio test: chi2=0.2706 , p=0.6029 , df=1
parameter F test:     F=0.2642 , p=0.6082 , df_denom=119, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:      F=1.9882 , p=0.1416 , df_denom=116, df_num=2
ssr based chi2 test:  chi2=4.1477 , p=0.1257 , df=2
likelihood ratio test: chi2=4.0782 , p=0.1301 , df=2
parameter F test:     F=1.9882 , p=0.1416 , df_denom=116, df_num=2

Granger Causality
number of lags (no zero) 3
ssr based F test:      F=1.7768 , p=0.1556 , df_denom=113, df_num=3
ssr based chi2 test:  chi2=5.6607 , p=0.1293 , df=3
likelihood ratio test: chi2=5.5312 , p=0.1368 , df=3
parameter F test:     F=1.7768 , p=0.1556 , df_denom=113, df_num=3

Granger Causality
number of lags (no zero) 4
ssr based F test:      F=0.9513 , p=0.4374 , df_denom=110, df_num=4
ssr based chi2 test:  chi2=4.1163 , p=0.3905 , df=4
likelihood ratio test: chi2=4.0467 , p=0.3997 , df=4
parameter F test:     F=0.9513 , p=0.4374 , df_denom=110, df_num=4

Granger Causality
number of lags (no zero) 5
ssr based F test:      F=0.6696 , p=0.6473 , df_denom=107, df_num=5
ssr based chi2 test:  chi2=3.6922 , p=0.5945 , df=5
likelihood ratio test: chi2=3.6356 , p=0.6030 , df=5
parameter F test:     F=0.6696 , p=0.6473 , df_denom=107, df_num=5

```

Рисунок 2.4 – screenshot виконання тесту причинності обсягу реклами від обсягу продажів по Гренжеру з Google Colab

Тепер перевіримо зворотне твердження – чи є обсяг реклами причиною по Гренжеру для обсягу продажів. Дослідимо два випадки: для рекламних кампаній по телебаченню та в інтернеті.

F-статистики та їх p-value для моделей з лагами реклами по телебаченню (див. Рисунок 2.5) свідчать про те, що при рівні значущості 5% ми би могли вважати впливовими перші 7 лагів, при рівні 1% – 3. Також

вони підтверджують, що обсяг реклами по телебаченню є причиною по Гренжеру для обсягів продажів.

```

Granger Causality
number of lags (no zero) 1
ssr based F test:      F=11.0538 , p=0.0012 , df_denom=119, df_num=1

number of lags (no zero) 2
ssr based F test:      F=5.0486  , p=0.0079  , df_denom=116, df_num=2

number of lags (no zero) 3
ssr based F test:      F=4.1126  , p=0.0082  , df_denom=113, df_num=3

number of lags (no zero) 4
ssr based F test:      F=2.8533  , p=0.0271  , df_denom=110, df_num=4

number of lags (no zero) 5
ssr based F test:      F=3.4158  , p=0.0067  , df_denom=107, df_num=5

number of lags (no zero) 6
ssr based F test:      F=2.5566  , p=0.0238  , df_denom=104, df_num=6

number of lags (no zero) 7
ssr based F test:      F=2.2124  , p=0.0392  , df_denom=101, df_num=7

number of lags (no zero) 8
ssr based F test:      F=1.8666  , p=0.0739  , df_denom=98,  df_num=8

number of lags (no zero) 9
ssr based F test:      F=1.5804  , p=0.1322  , df_denom=95,  df_num=9

number of lags (no zero) 10
ssr based F test:      F=1.3783  , p=0.2027  , df_denom=92,  df_num=10

```

Рисунок 2.5 – screenshot виконання тесту причинності по Гренжеру обсягу продажів від обсягу реклами по телебаченню з Google Colab (вивід скорочено)

F-статистики та їх p-value для моделей з лагами реклами по інтернету (див. Рисунок 2.6) свідчать про те, що при рівні значущості 5% ми би могли вважати впливовими перші 14 лагів, при рівні 1% – 8. Також вони підтверджують, що обсяг реклами в інтернеті є причиною по Гренжеру для обсягів продажів.

```

Granger Causality
number of lags (no zero) 1
ssr based F test:      F=9.8501 , p=0.0021 , df_denom=119, df_num=1

number of lags (no zero) 2
ssr based F test:      F=4.6060 , p=0.0119 , df_denom=116, df_num=2

number of lags (no zero) 3
ssr based F test:      F=3.3964 , p=0.0204 , df_denom=113, df_num=3

number of lags (no zero) 4
ssr based F test:      F=3.9625 , p=0.0048 , df_denom=110, df_num=4

number of lags (no zero) 5
ssr based F test:      F=3.4820 , p=0.0059 , df_denom=107, df_num=5

number of lags (no zero) 6
ssr based F test:      F=3.5426 , p=0.0031 , df_denom=104, df_num=6

number of lags (no zero) 7
ssr based F test:      F=2.8799 , p=0.0088 , df_denom=101, df_num=7

number of lags (no zero) 8
ssr based F test:      F=2.7757 , p=0.0083 , df_denom=98, df_num=8

number of lags (no zero) 9
ssr based F test:      F=2.4839 , p=0.0136 , df_denom=95, df_num=9

number of lags (no zero) 10
ssr based F test:      F=2.3179 , p=0.0176 , df_denom=92, df_num=10

number of lags (no zero) 11
ssr based F test:      F=2.2667 , p=0.0174 , df_denom=89, df_num=11

number of lags (no zero) 12
ssr based F test:      F=2.0721 , p=0.0272 , df_denom=86, df_num=12

number of lags (no zero) 13
ssr based F test:      F=2.0987 , p=0.0224 , df_denom=83, df_num=13

number of lags (no zero) 14
ssr based F test:      F=2.0029 , p=0.0277 , df_denom=80, df_num=14
|
number of lags (no zero) 15
ssr based F test:      F=1.7901 , p=0.0513 , df_denom=77, df_num=15

```

Рисунок 2.6 – screenshot виконання тесту причинності по Гренжеру обсягу продажів від обсягу реклами в інтернеті з Google Colab (вивід скорочено)

Отже, обидві змінні (рекламні кампанії по телебаченню та в інтернеті) є причиною для обсягів продажу по Гренжеру, причому при рівні значущості 1% маємо 3 лаги для телебачення та 8 – для інтернету.

Тим не менше додатково застосуємо аналіз отриманих статистик Фішера, щоб краще дослідити лагову залежність. Для цього побудуємо два графіки, на яких відкладемо залежність F-статистики моделі від кількості використаних лагів (див. Рисунок 2.7). Можемо зазначити, що оскільки після першого лагу відбувається зменшення F-статистики більше ніж у два рази для обох каналів реклами, а після цього присутні набагато менші зміни (до 30%), можемо взяти тільки перші лаги обох змінних.



Рисунок 2.7 – графіки значень F-статистик тесту причинності по Гренжеру в залежності від кількості лагів у моделі

Отже, робимо висновок, що обсяг реклами є причиною для обсягу продажів за Гренжером, але цього не можна сказати про зворотне припущення – причинність обсягу реклами від продажів за Гренжером. У цьому випадку слід розглядати тільки модель залежності обсягів продажів від обсягів реклами. Також ми визначили “глибину” лагів, яку нам варто взяти для моделі: по одному лагу для обох каналів реклами.

## 2.4 Формування, опис моделі та її оцінка за допомогою 2МНК

Позначимо фактори, які будемо використовувати у моделі. Інформацію про них наведемо нижче у Таблиці 2.1.

Таблиця 2.1 – характеристика змінних у моделі

Змінна у моделі	Назва	Тип даних	Опис
$S_t$	Доля ринку (share of market)	$R$ $\in [0,1]$	Залежна змінна. Абсолютне значення частки продажів конкретного бренда від продажів ринку в цілому
$d_i$	Дистрибуція, розповсюдженість	$R$ $\in [0,1]$	Абсолютне значення частки магазинів, торгових точок, у яких можна придбати товар бренда, від загальної кількості перших. Це значення взято як різниця поточного та мінімального у вибірці – ми досліджуємо зміну дистрибуції. У моделі представлено дві змінні дистрибуції двох SKU: $d_1$ та $d_2$
$P_t$	Ціновий індекс	$R$	Відношення ціни бренда до середньої ціни на ринку
$M_{TV}$	Перегляд реклами бренда А по ТВ	$R, TRP$	Рейтинги перегляду реклами бренда А по телебаченню
$M_{int}$	Перегляд реклами бренда А в інтернеті	$R, TRP$	Рейтинги перегляду реклами бренда А в інтернеті
$s'_{TV}$	Перегляд реклами	$R, TRP$	Сумарні рейтинги перегляду реклами конкурентів бренда А по

	конкурентів по ТБ		телебаченню
$s'_{int}$	Перегляд реклами конкурентів в інтернеті	$R, TRP$	Сумарні рейтинги перегляду реклами конкурентів бренда А в інтернеті
$P_{TV}$	Ціна на рекламу по ТБ	$R$	Ціна реклами бренда А по телебаченню
$P_{int}$	Ціна на рекламу в інтернеті	$R$	Ціна реклами бренда А в інтернеті

Узагальнену систему рівнянь (1.17) запишемо для нашого прикладу наступним чином, додавши перші лаги рекламних кампаній бренда А:

$$\begin{cases} S_t = \alpha_0 + \alpha_1 d_{1,t} + \alpha_2 d_{2,t} + \alpha_3 P_t + \alpha_4 M_{TV,t} + \alpha_5 M_{int,t} + \\ + \alpha_6 M_{TV,t-1} + \alpha_7 M_{int,t-1} + \alpha_8 S_{TV,t} + \alpha_9 S_{int,t} + e_{1,t}, & (2.10a) \\ M_{TV,t} = \beta_0 S_t + \beta_1 P_{TV,t} + e_{2,t}, & (2.10б) \\ M_{int,t} = \gamma_0 S_t + \gamma_1 P_{int,t} + e_{3,t} & (2.10в) \end{cases}$$

Поділимо дані на навчальну та тестову вибірки, навчальна вибірка – 2019-2020 роки, тестова – 2021 рік 1-18 тижні.

Застосуємо метод 2МНК до системи одночасних рівнянь (1.18а,б,в) та оцінимо коефіцієнти  $(\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9)$ . Для цього у середовищі RStudio спочатку застосуємо МНК до необмеженої зведеної форми (2.5а,б) взявши тільки навчальну вибірку.

```

Call:
lm(formula = adv_tv ~ distr_1 + distr_2 + price_index + adv_price_tv +
    adv_price_int + adv_tv_comps + adv_int_comps, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-73.776  -4.476  -2.437   3.339   90.327

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.391250  11.544734   0.554  0.5811
distr_1     -3.178970   2.980924  -1.066  0.2888
distr_2      0.976944   6.495143   0.150  0.8807
price_index -10.386318  37.244152  -0.279  0.7809
adv_price_tv  0.210318   0.006524  32.240 <2e-16 ***
adv_price_int -0.046474  0.020504  -2.267  0.0256 *
adv_tv_comps -0.036290  0.032065  -1.132  0.2605
adv_int_comps 0.024255  0.019949   1.216  0.2270
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 25.19 on 98 degrees of freedom
Multiple R-squared:  0.9675,    Adjusted R-squared:  0.9651
F-statistic: 416.2 on 7 and 98 DF,  p-value: < 2.2e-16

Call:
lm(formula = adv_int ~ distr_1 + distr_2 + price_index + adv_price_tv +
    adv_price_int + adv_tv_comps + adv_int_comps, data = df)

Residuals:
    Min       1Q   Median       3Q      Max
-99.430  -8.153   2.154   8.583  95.531

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -36.87988  14.38792  -2.563  0.0119 *
distr_1      0.69137   3.71505   0.186  0.8528
distr_2     21.20028   8.09474   2.619  0.0102 *
price_index -37.08340  46.41648  -0.799  0.4263
adv_price_tv  0.00891   0.00813   1.096  0.2758
adv_price_int  0.23862  0.02555   9.338 3.31e-15 ***
adv_tv_comps  0.06684  0.03996   1.673  0.0976 .
adv_int_comps 0.05767  0.02486   2.320  0.0224 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 31.4 on 98 degrees of freedom
Multiple R-squared:  0.7556,    Adjusted R-squared:  0.7381
F-statistic: 43.28 on 7 and 98 DF,  p-value: < 2.2e-16

```

Рисунок 2.8 – скріншот оцінки у RStudio моделі (2.5а,б) МНК для навчальної вибірки

Тепер підставимо інструментальні змінні замість ендогенних змінних у рівняння (1.18а):

$$\begin{aligned}
 S_t = & \alpha_0 + \alpha_1 d_{1,t} + \alpha_2 d_{2,t} + \alpha_3 P_t + \alpha_4 \widehat{M}_{TV,t} + \alpha_5 \widehat{M}_{int,t} + \\
 & + \alpha_6 \widehat{M}_{TV,t-1} + \alpha_7 \widehat{M}_{int,t-1} + \alpha_8 s_{TV,t} + \alpha_9 s_{int,t} + e_{1,t}
 \end{aligned} \tag{2.11}$$

На Рисунку 2.9 представлено виконання RStudio оцінки моделі (2.11) за допомогою МНК. Коефіцієнт детермінації був оцінений у 65,6%, adjusted – у 62,3%. середня квадратична похибка – 0,004652. F-статистика підтверджує, що усі коефіцієнти моделі не можна прирівняти до нуля. Отже, модель адекватна.

```
Call:
lm(formula = df$SOM ~ estim_tv + estim_int + estim_tv_lag1 +
    estim_int_lag1 + df$distr_1 + df$distr_2 + df$price_index +
    df$adv_tv_comps + df$adv_int_comps)

Residuals:
    Min       1Q   Median       3Q      Max
-0.0127279 -0.0027157 -0.0005301  0.0025912  0.0103032

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.649e-02  2.208e-03  25.576 < 2e-16 ***
estim_tv     6.888e-06  6.263e-06   1.100  0.274188
estim_int    3.556e-05  1.688e-05   2.108  0.037678 *
estim_tv_lag1 2.654e-06  6.269e-06   0.423  0.672980
estim_int_lag1 3.870e-05  1.654e-05   2.340  0.021340 *
df$distr_1   1.892e-03  5.587e-04   3.387  0.001027 **
df$distr_2  -4.600e-03  1.182e-03  -3.890  0.000185 ***
df$price_index -7.382e-03  6.884e-03  -1.072  0.286269
df$adv_tv_comps 4.699e-06  5.864e-06   0.801  0.424888
df$adv_int_comps -1.087e-05  3.900e-06  -2.787  0.006405 **
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.004652 on 96 degrees of freedom
Multiple R-squared:  0.6556,    Adjusted R-squared:  0.6233
F-statistic: 20.3 on 9 and 96 DF,  p-value: < 2.2e-16
```

Рисунок 2.9 – скріншот оцінки у RStudio моделі (2.11) МНК для навчальної вибірки

Нижче представлено графік, на якому можна візуально оцінити акуратність змодельованої долі ринку бренда А (estimated SOM).

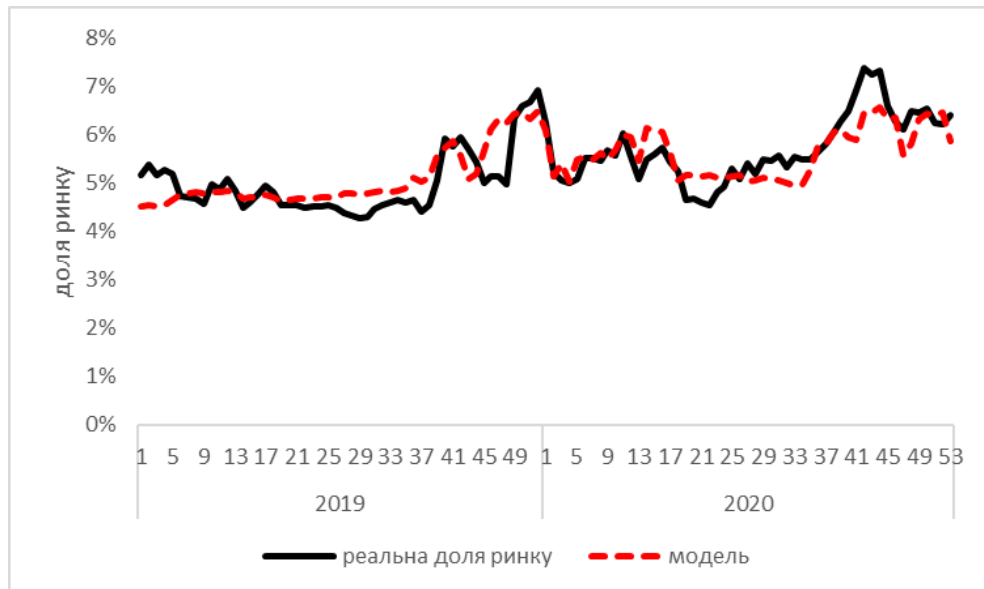


Рисунок 2.10 – Реальна та оцінена доля ринку бренда А (відповідно до значень, оцінених за допомогою 2МНК)

Опишемо оцінені коефіцієнти моделі з Рисунка 2.9.

Intercept, вільний коефіцієнт  $\alpha_0$ , свідчить про те, що за відсутності всіх інших факторів чи їх рівності нулю доля ринку бренда А дорівнювала би приблизно 5,6%.

Сумарна дистрибуція для двох SKU бренда є фактором позитивним, але оцінена моделлю дещо вироджено. Оскільки, по-перше, дистрибуція у даному контексті – це різниця від поточним значенням та мінімальним значенням, по-друге, дистрибуція обох SKU впродовж досліджуваного періоду змінювалась лише у межах 3%, тому вплив цього фактора не є вагомим у даній моделі і він міг бути оціненим неточно.

Реклама та перший лаг бренда А по телебаченню не внесла вагомий позитивний вплив на долю ринку, про відносно значимий позитивний внесок можна сказати про рекламу у інтернеті та її перший лаг. Реклама у інтернеті конкурентів внесла відчутний негативний вплив, телебачення – не внесла вагомого впливу.

Також було поставлено коефіцієнти моделі у тестову вибірку. Як бачимо з графіка (див. Рисунок 2.4), модель достатньо точно описує поведінку реальної долі ринку бренда А. Середня квадратична похибка становить 0,000002 (трохи менше, ніж у моделі – 0,000014). Можемо вважати, що модель є доволі точною.

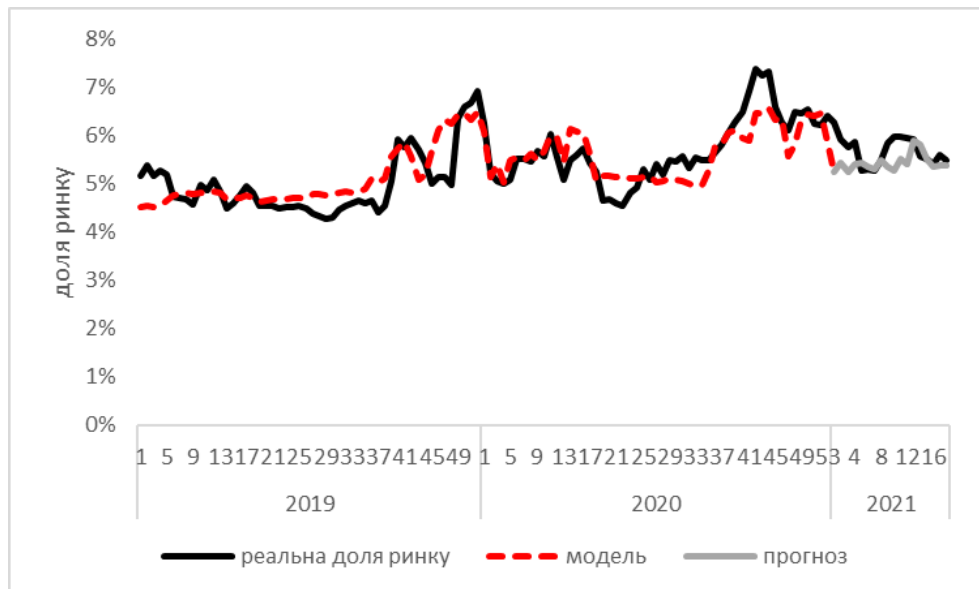


Рисунок 2.11 – Реальна, оцінена та прогнозована на новий період доля ринку бренда А (відповідно до значень, оцінених за допомогою 2МНК)

Також для рекламних кампаній по телебаченню та у інтернеті було пораховані ROI. Значення рекламних кампаній, що були оцінені як внесок у долю ринку бренда А, було помножено на загальну кількість проданих у категорії товарів (цим самим ми підраховували, скільки товару було продано за допомогою рекламних кампаній), а потім також помножено на ціну товару, аби отримати виручку, яку продажі бренда А принесли компанії. Отримані значення було поділено на рекламні бюджети за кожен рік та в цілому. Отримали наступну таблицю ROI:

Таблиця 2.2 – ROI рекламних кампаній бренда А по телебаченню та в інтернеті

Період	ROI ТБ	ROI Інтернет
2019	2,04	1,34
2020	1,64	1,33
2021	1,58	0,42
загалом	1,76	1,14

З Таблиці 2.2 можемо зробити наступні висновки:

- 1) майже всі рекламні канали є прибутковими й генерують більше доходу, ніж у них закладається бюджету
- 2) реклама у інтернеті гірше працює, ніж реклама по телебаченню
- 3) найкраще рекламні кампанії відпрацювали у 2019 році.

## Висновки

У процесі роботи були зроблені наступні висновки.

Було систематизовано теоретичну базу про види реклами, оглянуто особливості рекламних кампаній, що вимагають аналітичного підходу, наведено алгоритм використання методів Хаусмана та Гренжера, алгоритм застосування двокрокового методу найменших квадратів.

Розроблено узагальнену систему одночасних рівнянь – залежності продажів від різних факторів, у тому числі від реклами, а також залежності реклами від обсягу продажів та ціни на рекламу.

У рамках практичної частини було:

Описано дані та фактори, які впливають на продажі бренда деякої фармацевтичної компанії, застосовано специфікаційний Хаусмана для визначення методу оцінки моделі продажів. Як висновок, рекламні кампаніями виявились ендогенними змінними, тому надалі як метод оцінки розглядався 2МНК.

Застосовано тест причинності за Гренжером аби виявити причинні зв'язки та глибину лагів реклами. Виявлено, що для практичних даних обсяг реклами є причиною для обсягу продажів за Гренжером, але цього не можна сказати про зворотне припущення – причинність обсягу реклами від продажів за Гренжером. Звідси також дійшли висновку, що слід розглядати тільки модель залежності обсягів продажів від обсягів реклами, і визначено “глибину” лагів, яку було варто взяти для моделі: по одному лагу для обох каналів реклами.

Було побудовано модель виходячи з розробленого загального рівняння та результатів тестів Хаусмана та Гренжера, попередньо розділено дані на навчальну та тестову вибірки.

Було оцінено коефіцієнти моделі за допомогою двокрокового методу найменших квадратів та оцінено значення для тестової вибірки та порівняти

з реальними. Усі показники свідчили про адекватність моделі. Також було оцінено ROI рекламних кампаній, і виявлено, що реклама є прибутковим інструментом для розглянутого бренду.

## Список літератури

1. Лук'яненко, Ірина. Економетрика : підручник / І. Лук'яненко, Л. Краснікова. - Київ : Знання, 1998. - 493 с.
2. Берндт, Эрнст Роберт. Практика економетрики: классика и современность: Учебник для студентов вузов, обучающихся по специальностям 060000 экономики и управления / Пер. с англ. под ред. проф. С.А. Айвазяна / Э.Р. Берндт. – М.: ЮНИТИ-ДАНА, 2005. – 863 с. (Серия “Зарубежный учебник”).
3. Паламарчук Л. С. Словник української мови (СУМ) // Українська мова : енциклопедія / НАН України, Інститут мовознавства ім. О. О. Потебні, Інститут української мови ; ред. В. М. Русанівський [та ін.]. — К. : Українська енциклопедія, 2000. — ISBN 966-7492-07-9.
4. Назайкин А. Н. Медиапланирование на 100% / А. Н. Назайкин — «Автор», — (Бизнес на 100%).
5. Broadbent, S. (1979) "One Way TV Advertisements Work", Journal of the Market Research Society Vol. 23 no.3.
6. Leone, R.P. (1995) "Generalizing what is known about temporal aggregation and advert-over", Marketing Science, 14, G141-G150.

## Додаток А

Скрипт побудови та тестування моделей на причинність за Гренжером у середовищі Google Colab (мова – Python).

```
import pandas as pd
from statsmodels.tsa.stattools import grangercausalitytests

data = pd.read_excel('data.xlsx', sheet_name = "model_data") \
    [['adv_tv','adv_int','sales_in_units','distr_competitors','price_index']]
data['adv'] = data['adv_tv'] + data['adv_int']
data['sales'] = data['sales_in_units']
data = data[['adv','adv_tv','adv_int','sales','distr_competitors','price_index']]
data
grangercausalitytests(data[['sales','adv']], maxlag=5)
grangercausalitytests(data[['adv_tv','sales']], maxlag=10)
grangercausalitytests(data[['adv_int','sales']], maxlag=15)
```

## Додаток Б

Скрипт використання тесту Хаусмана у середовищі RStudio (мова – R).

```
library(readxl)
library(qpcR)
library(openxlsx)

# data reading
df_H <- read_excel("../data.xlsx", sheet = "model_data")
print(names(df_H))

# Hausman test
model_adv_tv <- lm(adv_tv ~ distr_1+distr_2+price_index+adv_price_tv+
  adv_price_int+adv_tv_comps+adv_int_comps, data = df_H)
summary(model_adv_tv)

# Internet
model_adv_int <- lm(adv_int ~ distr_1+distr_2+price_index+adv_price_tv+
  adv_price_int+adv_tv_comps+adv_int_comps, data = df_H)
summary(model_adv_int)

df_H["adv_tv_fitted"] <- fitted(model_adv_tv)
df_H["adv_int_fitted"] <- fitted(model_adv_int)

model_SOM <- lm(SOM ~ adv_tv_fitted+adv_int_fitted+adv_tv+adv_int+
  distr_1+distr_2+price_index+adv_tv_comps+adv_int_comps,
  data = df_H)
summary(model_SOM)
```

## Додаток В

Скрипт побудови моделі обсягу продажів від різних факторів та оцінки методом двокрокового методу найменших квадратів у середовищі RStudio (мова – R).

```

library(readxl)
library(nlme)
library(openxlsx)
library(tidyverse)

# data reading
df <- read_excel("../data.xlsx", sheet = "model_data")[c(1:106),]
print(names(df))

# TV
model_adv_tv_ <- lm(adv_tv ~ distr_1+distr_2+price_index+adv_price_tv+
  adv_price_int+adv_tv_comps+adv_int_comps, data = df)
summary(model_adv_tv_)

# Internet
model_adv_int_ <- lm(adv_int ~ distr_1+distr_2+price_index+adv_price_tv+
  adv_price_int+adv_tv_comps+adv_int_comps, data = df)
summary(model_adv_int_)

# getting fitted y from models
estim_tv = fitted(model_adv_tv_)
estim_int = fitted(model_adv_int_)
estim_tv_lag1 = unlist(sapply(1:length(estim_tv), function(x) estim_tv[x-1]))
estim_tv_lag1 <- c(0,estim_tv_lag1)
estim_int_lag1 = unlist(sapply(1:length(estim_int), function(x) estim_int[x-1]))
estim_int_lag1 <- c(0,estim_int_lag1)

```

```

model <- lm(SOM ~ estim_tv+estim_int+estim_tv_lag1+estim_int_lag1+
           distr_1+distr_2+price_index+
           adv_tv_comps+adv_int_comps, data=df)
summary(model)
# x <- data.frame(df$SOM, fitted(model))
# write.xlsx(x, 'output.xlsx')
df_test <- read_excel("../data.xlsx", sheet = "model_data")[c(107:123),]
# TV
model_adv_tv_test <- lm(adv_tv ~ distr_1+distr_2+price_index+adv_price_tv+
                       adv_price_int+adv_tv_comps+adv_int_comps, data = df_test)
summary(model_adv_tv_test)
# Internet
model_adv_int_test <- lm(adv_int ~
                       distr_1+distr_2+price_index+adv_price_tv+
                       adv_price_int+adv_tv_comps+adv_int_comps, data = df_test)
summary(model_adv_int_test)
# getting fitted y from models
estim_tv_test = fitted(model_adv_tv_test)
estim_int_test = fitted(model_adv_int_test)
estim_tv_lag1_test = unlist(sapply(1:length(estim_tv_test), function(x)
estim_tv_test[x-1]))
estim_tv_lag1_test <- c(0,estim_tv_lag1_test)
estim_int_lag1_test = unlist(sapply(1:length(estim_int_test), function(x)
estim_int_test[x-1]))
estim_int_lag1_test <- c(0,estim_int_lag1_test)
df_test["estim_tv"] <- estim_tv_test
df_test["estim_int"] <- estim_int_test
df_test["estim_tv_lag1"] <- estim_tv_lag1_test
df_test["estim_int_lag1"] <- estim_int_lag1_test
df_for_test = df_test[,c("estim_tv","estim_int", "estim_tv_lag1",

```

```
"estim_int_lag1", "distr_1", "distr_2", "price_index", "adv_tv_comps",  
"adv_int_comps"]  
fitted_test = predict(model, newdata = df_for_test, interval = "prediction")[,1]  
x <- data.frame(c(df$SOM, df_test$SOM), c(fitted(model), fitted_test))  
write.xlsx(x, 'output.xlsx')
```