МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ

О.Т. Іващук, Н. В. Дзюбановська

Економетрика

Методичні рекомендації для підготовки до практичних занять

Тернопіль – 2017

ТНЕУ

УДК 330.43

ББК 65в641

Розглянуто і рекомендовано до друку на засіданні кафедри економіко-математичних методів (протокол №1 від 28.08.2017 р.).

Іващук О. Т., Дзюбановська Н. В. Методичні рекомендації для підготовки до практичних занять з дисципліни «Економетрика». – Тернопіль: ТНЕУ, 2017. – 164 с.

У посібнику наведено короткі теоретичні відомості та приклади розв’язування задач з дисципліни «Економетрика». Для студентів денної форми навчання всіх спеціальностей ТНЕУ.

УДК 330.43

ББК 65в641

Зміст

Вступ…………………………………………………………… 4

Завдання 1. Однофакторна економетрична модель……….. 12

Завдання 2. Багатофакторна економетрична модель……… 30

Завдання 3. Дослідження наявності мультиколінеарності.. 53

Завдання 4. Гетероскедастичність………………………….. 64

Завдання 5. Економетричні моделі динаміки……………… 75

Завдання 6. Природа автокореляції та методи її усунення.. 116

Завдання 7. Економетричні моделі на основі системи одночасних рівнянь……………………………………….. 135

Література……………………………………………………… 161

Вступ

Процес прийняття науково обґрунтованих прогнозних рішень в економічній системі базується на тісному поєднанні кількісних методів математики, статистики та економетрики.

В економічну науку термін “економетрика” був введений норвезьким вченим Р. Фрішем в його доповіді, опублікованій у 1926 р., яка стала основою для визначення самостійної галузі наукових досліджень. Основна мета цих досліджень полягала у встановленні конкретних кількісних закономірностей взаємозалежності економічних процесів з допомогою математико-статистичних методів. Слово “економетрика” поєднує у собі комбінацію двох слів: “економіка” і “метрика” (від грецького слова “метрон”). Отже, термін “економетрика” означає вимірювання в економіці, і вимірювання справді є важливою складовою економетрики. Але не всі вимірювання можна розглядати як складові економетрики.

Перед тим, як визначити формально предмет економетрики, приведемо її визначення різними вченими. Так, польський економіст і статистик О. Ланге писав, що економетрика займається вивченням існуючих в економічному житті конкретних закономірностей, застосовуючи для цієї мети статистичні методи. Французький вчений Е. Маленво розглядав економетрику як потужний інструмент для виявлення найбільш стійких характеристик у поведінці реальних економічних одиниць.

За визначенням Т. Шаттеліса, економетрика – термін, котрий надається поняттю, сфера якого відноситься до дуже складної наукової дисципліни, яка народилася у галузі інтерференції математики, статистики та економіки. У наших дослідженнях ми будемо цікавитися не тільки виявленням існуючих економетричних законів і зв’язків між економічними показниками, але і підходами до формалізації цих процесів. Об’єктом економетрики є економічні системи та процеси різного рівня складності та ієрархій. Предметом – методи побудови та дослідження математико-статистичних моделей економічних процесів та явищ, оцінка та прогнозування їх розвитку.

Метою – кількісний аналіз функціонування економічних систем і процесів з допомогою економічних методів і моделей, їх застосування при прийнятті вигідних рішень. Завданням – оцінка параметрів моделей, перевірка адекватності та сценарії розвитку економічних процесів.

Необхідно розуміти відмінність економетрики від застосування математичної статистики до прикладних досліджень. Економетрика займається кількісною ідентифікацією моделей теоретичної економіки. Якщо ви обробляєте економічну інформацію за допомогою математичної статистики, але не виходите із певної економічної моделі, то ви не займаєтесь економетрикою.

Підсумовуючи сказане, необхідно відзначити, що немає такого, одночасно достатньо лаконічного та точного визначення, яке б дозволило завжди однозначно вирішувати: належить або не належить дане дослідження до економетрики.

На наш погляд, економетрика – прикладна економіко-математична дисципліна, яка синтезує економіку, математику та статистику і не може обмежуватись лише використанням регресійного аналізу в економіці, а повинна використовувати весь наявний математичний інструментарій прикладної математики для проведення економічних досліджень. Окрім цього, економетрика виступає базовою основою для визначення альтернативних варіантів у системах прийняття управлінських рішень. В умовах конкуренції прогнозні рішення повинні прийматися на основі модельних альтернативних варіантів в тісному поєднанні з наявною інформаційною базою та сучасними програмними продуктами.

Тому знання основ прикладної математики та методів економетричного аналізу даних є невід’ємною складовою професійних навиків майбутніх спеціалістів усіх освітніх рівнів, а відповідні програмні продукти – робочим інструментом кожного спеціаліста-аналітика.

Кваліфікований аналітик, який використовує економетричні методи у повсякденній практиці, в певній мірі повинен бути:

а) економістом – щоб використовувати економічну теорію для аналізу емпіричних даних;

б) математиком – щоб формулювати економічну теорію на математичній мові, зробивши її придатною для перевірки статистичних гіпотез;

в) спеціалістом у економічній статистиці – щоб володіти процесами формування інформаційної бази даних і вміти ставити у відповідності до змінних економічної теорії, реально виміряні макро- та мікроекономічні емпіричні показники;

г) спеціалістом у математичній статистиці – щоб використовувати для аналізу емпіричних даних кількісні методи.

До даного переліку слід додати необхідність обов'язкового знання та володіння комп'ютерною технікою, освоїти хоч б один статистичний пакет програмних продуктів, без використання яких немислимий системний аналіз сьогодні.

Із визначення економетрики слідує, що її походження і головне призначення – це соціально-економічні застосування та модельний опис існуючих кількісних взаємозв'язків і взаємозалежностей між аналізуючими показниками економічних процесів.

Наприклад, з теорії відомо, що споживання (y) залежить від величини доходу (x). Тобто має місце залежність: $y=f(x)$. У даному випадку економетрика дає відповідь на наступні питання: 1) Чи підтверджується це на практиці? 2) Як кількісно виражається цей вплив?

Економетричні моделі – важливий клас моделей, які пропонує математика аналітику. З допомогою цих моделей описуються явища, в яких присутні статистичні фактори, які важко пояснювати їх в чисто детермінованому підході. Типові моделі такого роду мають трендциклічну компоненту та випадкову складову. Хоче того чи ні, аналітик не може виключати випадкову складову і повинен будувати свої прогнозні висновки, враховуючи їх наявність.

Введемо випадкову змінну u, що включає в себе частину варіації результативного показника у, яка не пояснюється незалежними змінними. Тобто між розрахованими за моделлю значеннями $\hat{y}_i$ та фактичними y_i будуть спостерігатися відхилення u_i:

$$u_i = y_i - \hat{y} \text{ або } y_i = \hat{y}_i + u_i \qquad (1)$$

де i – індекс спостереження, $i = \overline{1,n}$.

Випадкову величину u назвемо збуренням (залишком, відхиленням). Її значення можуть змінюватися від одного спотереження до іншого. Наприклад, при вивченні залежності національного доходу від капітальних вкладень збуруюча змінна включала би в себе вплив на національний дохід таких факторів, як число працюючих у сфері виробництва, продуктивність праці, використання основних фондів і т. д., а також інші випадкові чинники.

Таким чином, найпростішу економетричну модель можна представити наступним чином:

$$y = \hat{y} + u \text{, тобто } y = f(x) + u \text{ або } y = f(x_1, .., x_n) + u. \quad (2)$$

Цей вид запису дозволяє інтерпретувати випадкову величину u як змінну, що враховує неправильну специфікацію функції, тобто неправильний вибір форми рівняння, яке описує залежність.

Завдяки введенню випадкової величини u змінна y також стає випадковою, оскільки при заданих значеннях пояснюючих змінних $x_1, .., x_n$ змінній y не можна приписувати або ставити у відповідність тільки одне певне значення. Якщо, наприклад, ми вивчаємо залежність обсягу податкових надходжень від величини ставки податку, то задавши значення ставки податку, можна вказати інтервал, в якому можуть знаходитися відповідні обсяги надходжень.

Весь процес економетричного дослідження певних залежностей можна розкласти на ряд основних етапів. Ці етапи описуються у відповідності із хронологією їх реалізації, проте декотрі з них, крім цього, знаходяться у співвідношенні ітераційної взаємодії: результати реалізації більш пізніх етапів можуть містити висновки про необхідність повторного розгляду вже пройдених (рис. 1). Дана схема в основному використовується для дослідження залежностей між кількісними змінними.

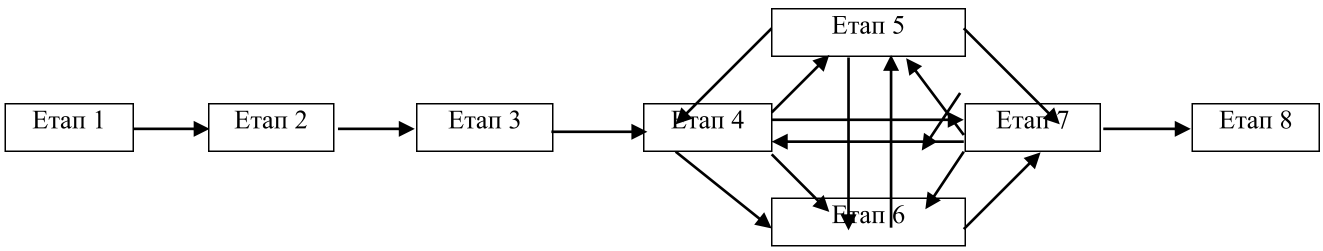


Рис. 1. Схема хронологічно-ітераційних взаємозв'язків основних етапів економетричного дослідження залежностей

Розглянемо основні етапи процесу економетричного моделювання.

Перший етап (постановчий) – визначення елементарної одиниці або об'єкта дослідження кінцевої мети моделювання, сукупності факторів і показників, які найбільш повно характеризують модель, їх роль та місце в дослідженні економічних процесів.

Другий етап (формалізаційний) – прийняття гіпотези взаємозв'язку, вибір загального виду моделі, специфікація економетричних зв'язків у рівняннях.

Третій етап (інформаційний) – ретроспективний аналіз економічної сутності досліджуваного процесу, формулювання та формалізація інформаційної бази моделі.

Четвертий етап (кореляційний) – спрямований на вирішення основної задачі кореляційного аналізу, дає можливість відповісти на питання, чи є взагалі який небудь зв'язок між досліджуваними змінними, яка структура та форма цих зв'язків і як виміряти їх щільність? Даний етап досить повно забезпечений необхідним математичним апаратом і відповідними програмними продуктами.

П'ятий етап (параметричний) – визначення загального виду, структури шуканих зв'язків між Y та X і класу допустимих розв'язків у вигляді певної параметричної сукупності функції.

Шостий етап (проблемний) – аналіз мультиколінеарності прогнозних змінних, її усунення та вибір найбільш інформативних із них.

Сьомий етап (ідентифікаційний) – обчислення оцінок невідомих параметрів, що входять в економетричну модель, їх достовірності.

Восьмий етап (верифікація моделі) - співставлення реальних і розрахункових даних, перевірка адекватності моделі, оцінка точності та стійкості отриманих рівнянь зв'язку, побудова прогнозів та сценаріїв розвитку.

При побудові економетричних моделей, які являють собою систему взаємозв'язаних рівнянь регресії, прийнятий у регресійному аналізі розподіл змінних на пояснюючі та залежні втрачає змістовність, оскільки одна й та сама змінна входить в одне з рівнянь як залежна, а в друге – як пояснююча. Тому необхідна інша класифікація змінних, яка найбільш повно відповідала б їх змістовності в економетричній моделі, а також відображала їх роль та характер в процесі моделювання.

У доповнення до етапів побудови економетричних моделей розглянемо деякі загальні принципи складання рівнянь, які входять у модель:

1) вибирають цільові функції, тобто результативні ознаки, які становлять особливий інтерес;

2) для кожної цільової функції будується рівняння, в якому її змінні пояснюються з допомогою інших змінних;

3) для пояснюючих змінних в свою чергу складаються власні рівняння і цей процес продовжується доти, поки непояснювальними залишаться тільки ті змінні, які неможливо або не припускається пояснити в рамках даної моделі;

4) усі параметри отриманих рівнянь повинні бути оцінені статистичними методами на основі даних у формі часових рядів. Первинними даними для оцінювання коефіцієнтів є послідовність багатомірних векторів, які містять значення ендогенних і екзогенних змінних моделі, які відповідають деякій послідовності моментів часу;

5) рівняння, які отримуються за допомогою екстраполяції, і результати прогнозу співставляються з фактичними даними як за межами, так і в межах базисного періоду.

Враховуючи приведені вище міркування, в економетричних моделях будемо розрізняти наступні змінні.

1. *Ендогенні змінні.* Ендогенними (внутрішньосистемними) назвемо ті змінні, які пояснюються даною моделлю при побудові причинно-наслідкових відносин. Значення ендогенних змінних формується у результаті одночасної взаємодії змінних, які входять у модель. Ендогенні змінні залежать від екзогенних та збурюючих змінних. Ендогенні змінні є залежними.

2. *Екзогенні змінні.* Значення екзогенних змінних в кожний період часу t визначаються поза моделлю. Екзогенними (позасистемними) назвемо такі вхідні змінні при побудові причинно-наслідкових відносин, які не пояснюються в рамках даної моделі. Вони являються зовнішніми наперед заданими економічними величинами і пояснюються не моделлю, а економічними факторами та закономірностями, що містяться за межами даної моделі. Екзогенні змінні визначають ендогенні, але самі не знаходяться під їх впливом. Таким чином, між ендогенними та екзогенним змінними існують тільки односторонні стохастичні причинні відносини.

3. *Наперед визначені змінні*. Ендогенні та екзогенні змінні можуть бути також лаговими (запізнюючими). Під лаговою розуміємо змінну, значення якої відстає на один або кілька періодів. Якщо $x_{t,k}$ – значення звичайної змінної x_k, зафіксовані в даний момент часу t, то $x_{t-1,k}$ – її лагові значення, зміщені на один період. При наявності в моделі лагових ендогенних та екзогенних змінних значення ендогенної змінної в період часу t залежить як від своїх власних значень в попередні періоди, так і від значень екзогенних змінних в ті ж періоди. Кожна із лагових екзогенних та ендогенних змінних при цьому розглядаються як самостійна змінна. Оскільки лагові змінні в період часу t так само не пояснюються економетричною моделлю, ми можемо віднести їх до наперёдзаданих екзогенних. У зв'язку з цим до класу наперед визначених змінних ми відносимо:

а) звичайні лагові змінні; вони завчасно визначені, оскільки пояснюються не економетричною моделлю, а факторами, які містяться поза моделлю;

б) лагові екзогенні змінні; вони завчасно визначені, оскільки їх значення належить попереднім періодам і пояснюється поза моделлю;

в) лагові ендогенні змінні; їх наперeдвизначеність випливає із попереднього пояснення в економетричній моделі.

4. *Спільнозалежні змінні*. Спільнозалежними змінними називаються звичайні ендогенні змінні, які пояснюються економетричною моделлю в момент часу t. Вони спільно залежні тому, що між ними існують багатосторонні зв'язки, і визначаються не одним рівнянням, а одночасними рівняннями моделі. В зв'язку з цим економетричну модель можна розглядати як спосіб визначення спільнозалежних змінних через наперeдвизначені змінні та збурення.

5. *Збурення або латентні змінні*. Збурення – це економічні величини, які не входять в рівняння економетричних моделей, але впливають на спільнозалежні змінні. Вони теж формуються за рахунок випадкових впливів, помилок і припущень. Наприклад, при використанні типу функцій, які не адекватно відображають досліджуване явище чи неправильному виборі способу оцінювання. Збурення являються стохастичними змінними. На противагу спільнозалежним та наперeдвизначеним змінним емпіричні значення збурених змінних невідомі. Їх значення знаходять як залишки за окремими рівняннями після оцінки невідомих параметрів моделі. Із

приведеного вище зрозуміло, що змістовна інтерпритація збурених змінних в економетричній моделі така сама, як і у випадку одного рівняння регресії.

Завдання 1. Однофакторна економетрична модель.

Однофакторною економетричною моделлю називається рівняння зв'язку між двома змінними *y* та *x* виду:

$$y = f(x) + u,$$

де *y* – залежна змінна; *x* – незалежна змінна; *u* – випадкова змінна.

Розрізняють два види залежностей: лінійна та нелінійна регресії.

Найбільш поширені моделі:

- лінійна – $\hat{y} = a + bx$;
- гіперболічна – $\hat{y} = a + \frac{b}{x}$, заміна $\frac{1}{x} = x^*$;
- експоненціальна – $\hat{y} = e^{a+bx}$; $\ln \hat{y} = a + bx$, заміна $y^* = \ln \hat{y}$;
- параболічна – $\hat{y} = a + bx + cx^2$, заміна $x^2 = x^*$;
- показникова – $\hat{y} = a \cdot b^x$, $\ln \hat{y} = \ln a + x \ln b$, заміна $y^* = \ln \hat{y}$, $a^* = \ln a$, $b^* = \ln b$;
- степенева – $\hat{y} = a \cdot x^b$, $\ln \hat{y} = \ln a + b \ln x$, заміна $y^* = \ln \hat{y}$, $a^* = \ln a$, $x^* = \ln x$.

Провівши відповідні заміни, ми прийдемо до лінійного виду.

Приклад 1.1. Побудувати економетричну модель впливу вартості основних виробничих фондів на обсяг отриманого прибутку деяким умовним підприємством регіону та виконати процедури оцінки достовірності отриманих результатів. Статистичні дані для розрахунку необхідних параметрів приведені в табл. 1.1.

Таблиця 1.1

Вплив вартості основних виробничих фондів на прибуток підприємства

№ підприємства	Прибуток, млн. грн. y_i	Основні фонди, млн. грн. x_i	x_i^2	$x_i y_i$
1	***2***	***3***	***4***	***5***
1	1.2	2.5	6.25	3.0
2	1.5	2.8	7.84	4.2
3	1.9	3.0	9.0	5.7
4	2.2	3.6	12.96	7.92
5	2.8	3.9	15.21	10.22

Продовження таблиці 1.1

1	***2***	***3***	***4***	***5***
6	3.1	4.2	17.64	13.02
7	3.4	4.5	20.25	15.3
8	4.5	5.0	25.0	22.5
9	4.8	5.6	31.36	26.88
10	5.4	6.0	36.0	32.4
Всього	30.8	41.1	181.51	141.84

Розв'язування. Розв'язок даної задачі представимо з допомогою наступних кроків.

І. Ідентифікація змінних.

Позначимо:

y – залежна змінна, прибуток підприємства, млн. грн.;

x – незалежна змінна, вартість основних виробничих фондів, млн. грн.

ІІ. Специфікація моделі.

Проведемо специфікацію моделі з допомогою діаграми розсіювання (рис. 1.1).

Для цього побудуємо діаграму розсіювання залежності обсягу прибутку (y) від вартості основних виробничих фондів підприємства (x).

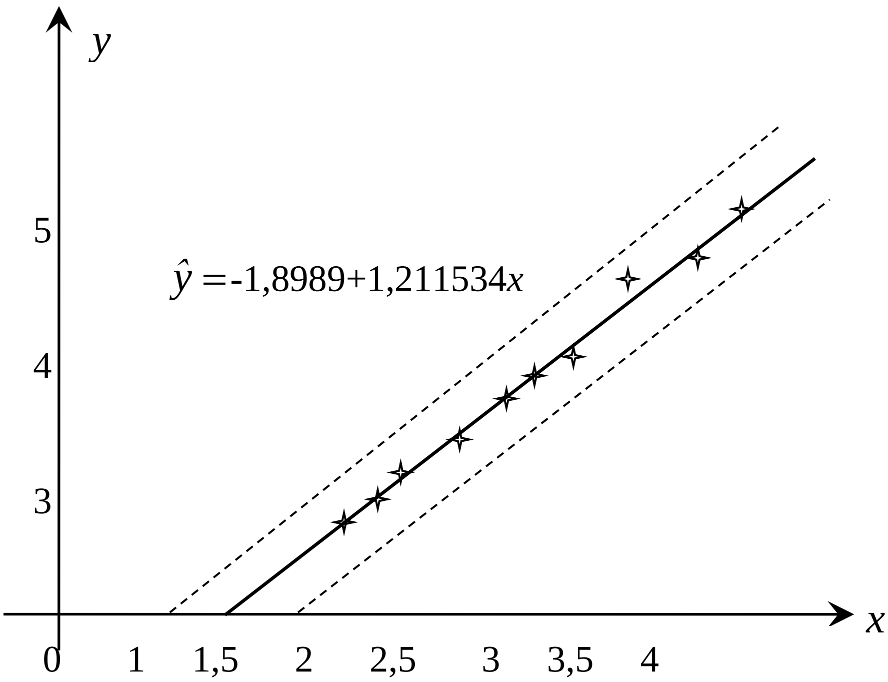


Рис. 1.1. Діаграма розсіювання та регресійна пряма, що відображає залежність прибутку від основних фондів

Розміщення точок на діаграмі розсіювання (рис. 1.1) дає можливість зробити припущення про існування лінійної форми зв'язку у вигляді функції:

$$\hat{y} = a + bx, \tag{1.1}$$

де $\hat{y}$ – розрахунковий обсяг прибутку, млн. грн.; x – вартість основних виробничих фондів, млн. грн.; a та b відповідні оцінки параметрів α та β економетричної моделі:

$$y = \alpha + \beta x + u. \tag{1.2}$$

ІІІ. Знаходження оціночних параметрів.

Розрахувати оцінки a та b можна з допомогою наступних методів:

- МНК (за системою нормальних рівнянь);
- МНК (через відхилення від середніх);
- матричним методом;
- стандартними пакетами прикладних програм (STADIA, MS EXCEL та ін.).

Метод найменших квадратів (МНК).

В основу методу найменших квадратів покладена вимога мінімізації суми квадратів відхилень емпіричних даних y_i від розрахункових значень регресії $\hat{y}_i$.

Необхідною умовою цього є перетворення в нуль перших частинних похідних цієї функції по кожній змінній a та b. Оскільки

$$F(a,b) = \sum_{i=1}^{n} e_1^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - a - bx_i)^2, \tag{1.3}$$

то

$$\begin{cases} \dfrac{\partial F}{\partial a} = -2\sum_{i=1}^{n} (y_i - a - bx_i) = 0 \\ \\ \dfrac{\partial F}{\partial b} = -2\sum_{i=1}^{n} (y_i - a - bx_i)x_i = 0 \end{cases} \tag{1.4}$$

Результатом виконання відповідних перетворень є наступна система нормальних рівнянь відносно невідомих a та b:

$$\begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases} \tag{1.5}$$

Далі знаходимо $\sum_{i=1}^{n} x_i$, $\sum_{i=1}^{n} x_i^2$ та $\sum_{i=1}^{n} x_i y_i$ (табл. 1.1).

Враховуючи, що n=10, наша система нормальних рівнянь (1.5) прийме наступний вид:

$$\begin{cases} 10a + 41{,}1b = 30{,}8, \\ 41{,}1a + 181{,}51b = 141{,}84; \end{cases} \tag{1.6}$$

Отже, для знаходження значень a та b нам необхідно розв'язати систему рівнянь (1.6).

<u>Метод Крамера.</u>

$$\Delta = \begin{vmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{vmatrix} = n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2 = 10 \cdot 181{,}51 - (41{,}1)^2 = 125{,}89;$$

$$\Delta_a = \begin{vmatrix} \sum_{i=1}^{n} y_i & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i^2 \end{vmatrix} = \sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} x_i y_i = 30{,}8 \cdot 181{,}51 - 141{,}84 \cdot 41{,}1 =$$

$$= -239{,}116;$$

$$\Delta_b = \begin{vmatrix} n & \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i y_i \end{vmatrix} = n \cdot \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_i = 10 \cdot 141{,}84 - 41{,}1 \cdot 30{,}8 = 152{,}52.$$

Тоді $a = \dfrac{\Delta_a}{\Delta} = \dfrac{-239{,}116}{125{,}89} = -1{,}8994$, $b = \dfrac{\Delta_b}{\Delta} = \dfrac{152{,}52}{125{,}89} = 1{,}211534$.

Для знаходження значень визначника можна використати МОПРЕД (MS Excel).

<u>Метод відхилення від середніх.</u>

Основу даного методу складають властивості оцінок, знайдених МНК, які полягають в тому, що лінія регресії обов'язково проходить через точку середніх значень $\bar{x}, \bar{y}$.

Виписуємо перше рівняння системи (1.4) і ділимо його на величину n:

$$na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \; : n$$

У результаті ділення отримаємо

$$\bar{y} = a + b\bar{x}. \tag{1.7}$$

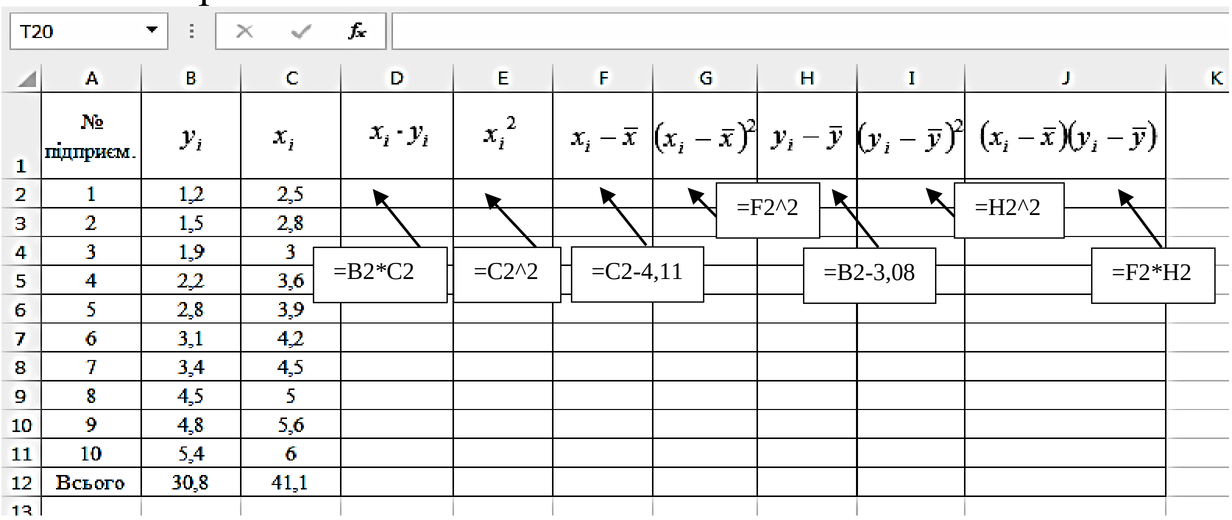
Можна довести, що

$$b = \frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}. \qquad (1.8)$$

Враховуючи (1.7), знайдемо параметр

$$a = \bar{y} - b\bar{x}. \qquad (1.9)$$

Знайшовши $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{30{,}8}{10} = 3{,}08$ та $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{41{,}1}{10} = 4{,}11$ доповнимо нашу розрахункову таблицю значеннями $x_i - \bar{x}$, $y_i - \bar{y}$, $(x_i - \bar{x})(y_i - \bar{y})$, $(y_i - \bar{y})^2$ та $(x_i - \bar{x})^2$ і знайдемо їх відповідні суми. Для проведення такого типу обчислень скористаємось табличним процесором MS Excel. Сформуємо в MS Excel таблицю і заповнимо другий рядок (клітинки D2, E2, F2, G2, H2, I2 та J2), використовуючи формули як показано на рис. 1.2.



	A	B	C	D	E	F	G	H	I	J	K
1	№ підприєм.	y_i	x_i	$x_i \cdot y_i$	x_i^2	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$	
2	1	1,2	2,5								
3	2	1,5	2,8								
4	3	1,9	3								
5	4	2,2	3,6								
6	5	2,8	3,9								
7	6	3,1	4,2								
8	7	3,4	4,5								
9	8	4,5	5								
10	9	4,8	5,6								
11	10	5,4	6								
12	Всього	30,8	41,1								

Рис. 1.2. Фрагмент документу MS Excel, в якому описано формули для здійснення подальших розрахунків

Для того, щоб заповнити решта клітинок (від третього до одинадцятого рядка, тобто діапазон D3:J11), потрібно спочатку виділити клітинки діапазону D2:J2 (як показано на рис. 1.3), потім підвести курсор до нижнього правого кута виділеного фрагменту, щоб з’явився чорний хрестик і, затиснувши лівою клавішею миші, протягнути вниз до одинадцятого рядка. Таким чином, ми скопіюємо формули із клітинок другого рядка у решта незаповнених клітинок.

	A	B	C	D	E	F	G	H	I	J
1	№ підприєм.	y_i	x_i	$x_i \cdot y_i$	x_i^2	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
2	1	1,2	2,5	3	6,25	-1,61	2,5921	-1,88	3,5344	3,0268
3	2	1,5	2,8							
4	3	1,9	3							
5	4	2,2	3,6							
6	5	2,8	3,9							
7	6	3,1	4,2							
8	7	3,4	4,5							
9	8	4,5	5							
10	9	4,8	5,6							
11	10	5,4	6							
12	Всього	30,8	41,1							

Рис. 1.3. Фрагмент документу MS Excel, в якому виділено діапазон клітинок для подальшого копіювання формул

Після того, як ми заповнили майже всі клітинки нашої таблиці, залишається обчислити відповідні суми розрахованих значень. Заповнимо дванадцятий рядок (клітинки D12, E12, G12, I12 та J12), використовуючи формули як зображено на рис. 1.4.

	A	B	C	D	E	F	G	H	I	J
1	№ підприєм.	y_i	x_i	$x_i \cdot y_i$	x_i^2	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
2	1	1,2	2,5	3	6,25	-1,61	2,5921	-1,88	3,5344	3,0268
3	2	1,5	2,8	4,2	7,84	-1,31	1,7161	-1,58	2,4964	2,0698
4	3	1,9	3	5,7	9	-1,11	1,2321	-1,18	1,3924	1,3098
5	4	2,2	3,6	7,92	12,96	-0,51	0,2601	-0,88	0,7744	0,4488
6	5	2,8	3,9	10,92	15,21	-0,21	0,0441	-0,28	0,0784	0,0588
7	6	3,1	4,2	13,02	17,64	0,09	0,0081	0,02	0,0004	0,0018
8	7	3,4	4,5	15,3	[illegible]	[illegible]	0,1521	0,32	0,1024	[illegible]
9	8	4,5	5	22,5	[illegible]	[illegible]	0,7921	1,42	2,0164	[illegible]
10	9	4,8	5,6	26,88	31,36	1,49	2,2201	1,72	2,9584	2,5628
11	10	5,4	6	32,4	36	1,89	3,5721	2,32	5,3824	4,3848
12	Всього	30,8	41,1	=СУММ(D2:D11)	=СУММ(E2:E11)		=СУММ(G2:G11)		=СУММ(I2:I11)	=СУММ(J2:J11)
13										
14										

Рис. 1.4. Фрагмент документу MS Excel, в якому позначені формули для знаходження сум відповідних показників

У кінцевому результаті після всіх обчислень одержимо таблицю із потрібними нам значенями (рис. 1.5).

№ підприєм.	y_i	x_i	$x_i \cdot y_i$	$x_i^{\,2}$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$(x_i - \bar{x})(y_i - \bar{y})$
1	1,2	2,5	3	6,25	-1,61	2,5921	-1,88	3,5344	3,0268
2	1,5	2,8	4,2	7,84	-1,31	1,7161	-1,58	2,4964	2,0698
3	1,9	3	5,7	9	-1,11	1,2321	-1,18	1,3924	1,3098
4	2,2	3,6	7,92	12,96	-0,51	0,2601	-0,88	0,7744	0,4488
5	2,8	3,9	10,92	15,21	-0,21	0,0441	-0,28	0,0784	0,0588
6	3,1	4,2	13,02	17,64	0,09	0,0081	0,02	0,0004	0,0018
7	3,4	4,5	15,3	20,25	0,39	0,1521	0,32	0,1024	0,1248
8	4,5	5	22,5	25	0,89	0,7921	1,42	2,0164	1,2638
9	4,8	5,6	26,88	31,36	1,49	2,2201	1,72	2,9584	2,5628
10	5,4	6	32,4	36	1,89	3,5721	2,32	5,3824	4,3848
Всього	30,8	41,1	141,84	181,51		12,589		18,736	15,252

Рис. 1.5. Фрагмент документу MS Excel, в якому отримано результати обчислень

Матричний метод.

Представимо модель (1.2) у матричній формі

$$\boldsymbol{Y} = \boldsymbol{XA} + \boldsymbol{U}\,. \tag{1.9}$$

$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ – вектор результуючої змінної;

$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}$ – матриця пояснюючої змінної X, доповнена одиничним стовпцем;

$A = \begin{bmatrix} a_1 = a \\ a_2 = b \end{bmatrix}$ – вектор оціночних значень.

Тоді оцінки параметрів моделі знайдемо за формулою:

$$A = (X'X)^{-1} X'Y\,. \tag{1.10}$$

Нам необхідно знайти транспоновану матрицю пояснюючих змінних:

$$X' = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_m \end{bmatrix}.$$

Далі знайдемо добуток двох матриць X' та X:

$$X'X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_m \end{bmatrix} \cdot \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix}.$$

Після чого знаходимо обернену матрицю до знайденого добутку $[X'X]^{-1}$. Крім того нам необхідно мати добуток матриці X' та вектора Y.

Враховуючи числові дані прикладу, маємо:

$$Y = \begin{bmatrix} 1{,}2 \\ 1{,}5 \\ 1{,}9 \\ 2{,}2 \\ 2{,}8 \\ 3{,}1 \\ 3{,}4 \\ 4{,}5 \\ 4{,}8 \\ 5{,}4 \end{bmatrix} \qquad X = \begin{bmatrix} 1 & 2{,}5 \\ 1 & 2{,}8 \\ 1 & 3{,}0 \\ 1 & 3{,}6 \\ 1 & 3{,}9 \\ 1 & 4{,}2 \\ 1 & 4{,}5 \\ 1 & 5{,}0 \\ 1 & 5{,}6 \\ 1 & 6{,}0 \end{bmatrix}; \qquad A = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}.$$

Транспоновану матрицю X' можна знайти використовуючи табличний процесор MS Excel із допомогою функції ТРАНСП (TRANSPOSE) (рис. 1.6). Спочатку потрібно виділити порожні клітинки, кількість яких повинна збігатися з числом вихідних клітинок, але розташовані вони повинні бути в іншому напрямку. Не знімаючи виділення з порожніх клітинок, вибираємо функцію ТРАНСП та вводимо діапазон клітинок, який потрібно транспонувати, після чого натискаємо комбінацію клавіш CTRL+SHIFT+ENTER.

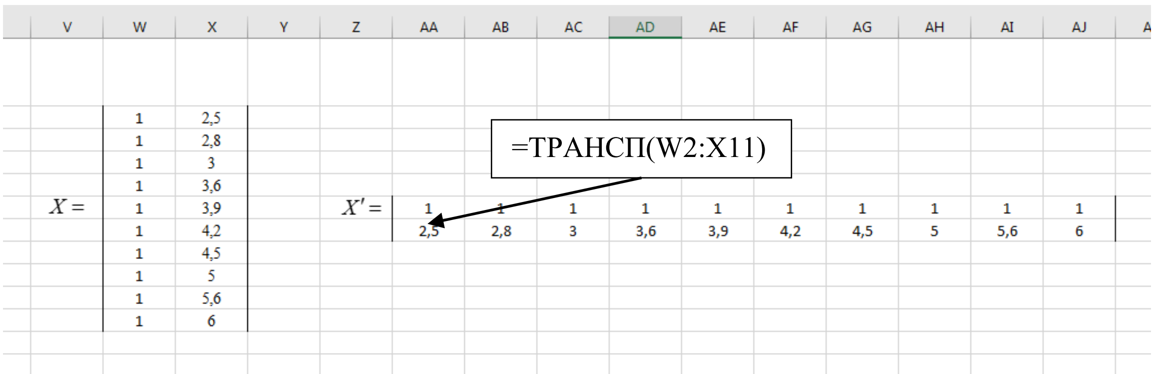


Рис. 1.6. Фрагмент документу MS Excel, в якому продемонстровано знаходження транспонованої матриці

Отже, $X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2{,}5 & 2{,}8 & 3{,}0 & 3{,}6 & 3{,}9 & 4{,}2 & 4{,}5 & 5{,}0 & 5{,}6 & 6{,}0 \end{bmatrix}$.

Нам необхідно виконати наступні числові процедури:

1. Знайдемо добуток двох матриць X' та X, використавши функцію МУМНОЖ (MMULT) в MS Excel (рис. 1.7). Спочатку виділяємо діапазон порожніх клітинок (розмір першої матриці (2×10), а другої – (10×2), як результат отримаємо матрицю розміром (2×2)), вибираємо функцію МУМНОЖ та вводимо діапазони клітинок першої та другої матриць через крапку з комою, після чого натискаємо комбінацію клавіш CTRL+SHIFT+ENTER.

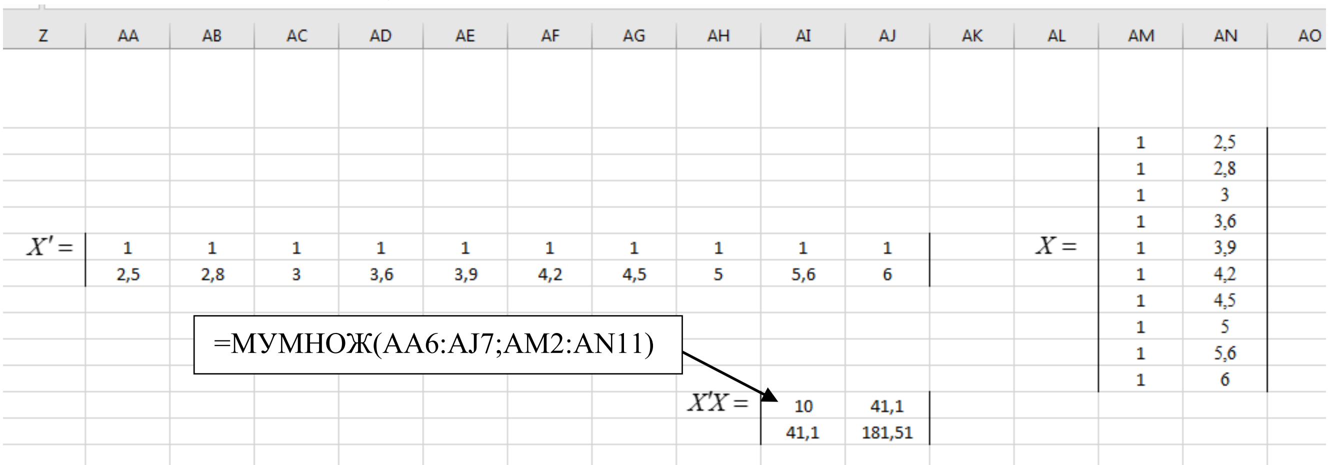


Рис. 1.7. Фрагмент документу MS Excel, в якому продемонстровано знаходження добутку двох матриць

$$X'X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2{,}5 & 2{,}8 & 3{,}0 & 3{,}6 & 3{,}9 & 4{,}2 & 4{,}5 & 5{,}0 & 5{,}6 & 6{,}0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2{,}5 \\ 1 & 2{,}8 \\ 1 & 3{,}0 \\ 1 & 3{,}6 \\ 1 & 3{,}9 \\ 1 & 4{,}2 \\ 1 & 4{,}5 \\ 1 & 5{,}0 \\ 1 & 5{,}6 \\ 1 & 6{,}0 \end{bmatrix} = \begin{bmatrix} 10 & 41{,}1 \\ 41{,}1 & 181{,}51 \end{bmatrix}.$$

2. Знайдемо обернену матрицю до отриманої у п. 1, використовуючи функцію МОБР (MINVERSE) в MS Excel (рис. 1.8). І знову виділяємо діапазон порожніх клітинок, вибираємо функцію МОБР та вводимо діапазони клітинок матриці до якої ми шукаємо

обернену, після чого натискаємо комбінацію клавіш CTRL+SHIFT+ENTER.

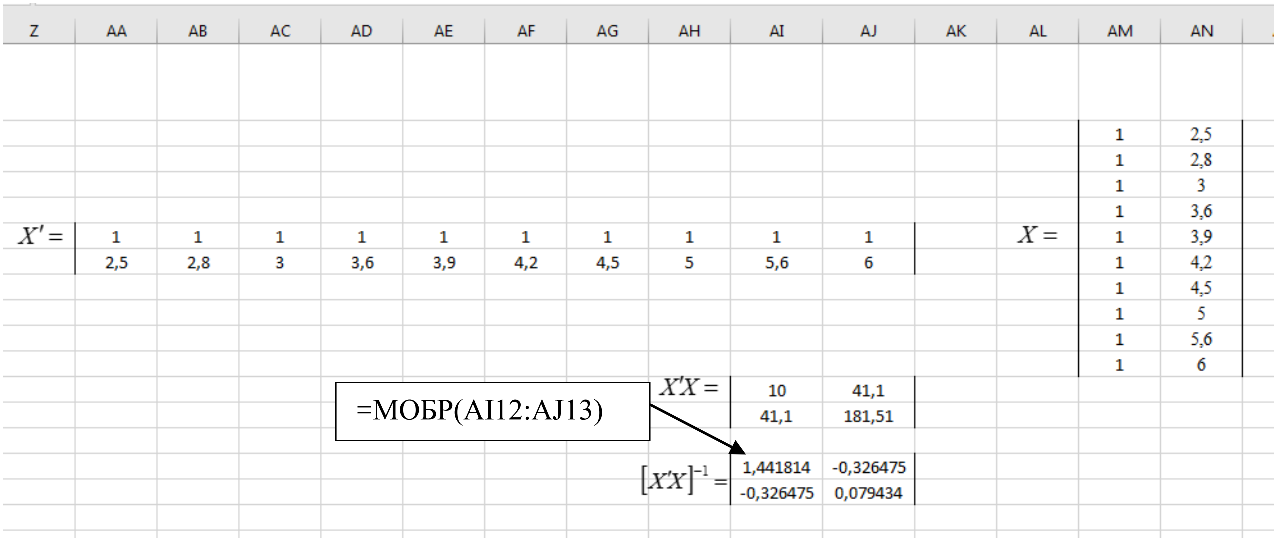


Рис. 1.8. Фрагмент документу MS Excel, в якому продемонстровано знаходження оберненої матриці

Отже,

$$[X'X]^{-1} = \begin{bmatrix} 1,441814 & -0,32648 \\ -0,32648 & 0,079434 \end{bmatrix}.$$

3. Знайдемо добуток матриці X' та вектора Y, використовуючи функцію МУМНОЖ (MMULT) в MS Excel:

$$X'Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2,5 & 2,8 & 3,0 & 3,6 & 3,9 & 4,2 & 4,5 & 5,0 & 5,6 & 6,0 \end{bmatrix} \cdot \begin{bmatrix} 1,2 \\ 1,5 \\ 1,9 \\ 2,2 \\ 2,8 \\ 3,1 \\ 3,4 \\ 4,5 \\ 4,8 \\ 5,4 \end{bmatrix} = \begin{bmatrix} 30,8 \\ 141,84 \end{bmatrix}.$$

Розмір матриці (2×10), а вектора – (10×1), як результат отримаємо вектор розміром (2×1).

4. Знайдемо добуток матриці $[X'X]^{-1}$ та вектора $[X'Y]$, знову ж таки за допомогою функції МУМНОЖ (MMULT) в MS Excel:

$$A = [X'X]^{-1} \cdot X'Y = \begin{bmatrix} 1,441814 & -0,32648 \\ -0,32648 & 0,079434 \end{bmatrix} \cdot \begin{bmatrix} 30,8 \\ 141,84 \end{bmatrix} = \begin{bmatrix} -1,8994 \\ 1,211534 \end{bmatrix}.$$

Таким чином, нами отримано значення оцінок наших параметрів $a_1 = a = -1{,}8994$, $a_2 = b = 1{,}211534$.

Отримані результати матричним способом співпадають із попередніми розрахунками. Таким чином $\hat{y} = -1{,}8994 + 1{,}211534x$.

IV. Геометрична інтерпретація оціночного рівняння.

Побудуємо на площині рівняння даної прямої $\hat{y} = -1{,}8994 + 1{,}211534x$ (рис. 1.1) і розглянемо його інтерпретацію.

Коефіцієнт при *x* (коефіцієнт нахилу) показує, що якщо вартість основних виробничих фондів збільшити на один млн. грн., то прибуток відповідно збільшиться на 1,211534 млн. грн. Тобто, можна зробити висновок, що збільшення факторного показника на одну одиницю (в одиницях виміру змінної *x*) приведе до збільшення (b>0) значення результативного показника на b одиниць (в одиницях виміру змінної *y*). Якщо b<0, то має місце протилежне.

Постійна величина *a* дає прогнозне значення *y*, при *x*=0. В залежності від конкретної ситуації це твердження може мати або не мати чіткого економічного змісту.

При проведенні інтерпретації рівняння регресії дуже важливо пам'ятати про три особливості. По-перше, *a* являється лише оцінкою α, а b – оцінкою β. Тому вся розглянута інтерпретація в дійсності представляє собою лише оцінку. По-друге, рівняння регресії відображає тільки загальну тенденцію для вибірки. Тому на кожне окреме спостереження впливає випадковість. По-третє, вірність інтерпретації у великій мірі залежить від правильності специфікації рівняння.

V. Обчислення загальної, пояснюючої та непояснюючої дисперсій.

Для знаходження даних дисперсій використаємо наступні формули:

$$\sigma^2_{заг.} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}; \quad \sigma^2_{поясн.} = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{n}; \quad \sigma^2_{непоясн.} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}; \tag{1.11}$$

Для проведення необхідних обчислень побудуємо таблицю 1.2 і перейдемо до знаходження числових значень дисперсій та відповідних їм середньоквадратичних відхилень.

Отже, маємо:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{30{,}8}{10} = 3{,}08; \quad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{41{,}1}{10} = 4{,}1,$$

$$\sigma^2_{\text{заг}} = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n} = \frac{18 \cdot 7632}{10} = 1{,}87632, \quad \sigma_{заг.} = \sqrt{1{,}87632} = 1{,}3698,$$

$$\sigma^2_{\text{поясн.}} = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{n} = \frac{18{,}4743}{10} = 1{,}84743, \quad \sigma_{поясн.} = \sqrt{1{,}84743} = 1{,}3592,$$

$$\sigma^2_{\text{непоясн.}} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n} = \frac{0{,}257663}{10} = 0{,}257663, \quad \sigma_{непоясн.} = 0{,}1605.$$

Таблиця 1.2

№ підпр.	y_i	x_i	$y_i - \bar{y}$	$(y_i - \bar{y})^2$	$\hat{y}_i$	$\hat{y}_i - \bar{y}$	$(\hat{y}_i - \bar{y})^2$	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$	$(x_i - \bar{x})^2$
1	1,2	2,5	-1,88	3,5344	1,1296	-1,9504	3,8041	0,0704	0,004956	2,5921
2	1,5	2,8	-1,58	2,4964	1,4930	-1,587	2,5186	0,007	0,000049	1,7161
3	1,9	3,0	-1,18	1,3924	1,7353	-1,3447	1,8082	0,1647	0,27126	1,2321
4	2,2	3,6	-0,88	0,7744	2,4621	-0,6179	0,3818	-0,2621	0,068696	0,2601
5	2,8	3,9	-0,28	0,0784	2,8256	-0,2544	0,0647	-0,0256	0,000655	0,0441
6	3,1	4,2	0,02	0,0004	3,1890	0,109	0,0119	-0,089	0,007921	0,0081
7	3,4	4,5	0,36	0,1296	3,5524	0,4724	0,2232	-0,1524	0,023226	0,1521
8	4,5	5,0	1,42	2,0164	4,1581	1,0781	1,1623	0,3419	0,116896	0,7921
9	4,8	5,6	1,72	2,9584	4,8849	1,8049	3,2577	-0,0849	0,007208	2,2201
10	5,4	6,0	2,32	5,3824	5,3695	2,2815	5,2418	-0,0305	0,000930	3,5721
Всього	30,8	41,1		18,7632			18,4743		0,257668	12,589

VI. Знаходження інтервалу довіри оціночного рівняння.

Позначимо параметр генеральної сукупності через δ, а його оцінку – d. Для знаходження довірчого інтервалу використаємо наступну формулу:

$$P(d - k\sigma_d \le \delta \le d + k\sigma_d) = 1 - \alpha, \tag{1.12}$$

де k – довірчий множник, який вказує частку стандартного відхилення, яка повинна бути врахована, щоб із заданою ймовірністю P довірчий інтервал $[d - k\sigma_d; d + k\sigma_d]$ покрив параметр генеральної сукупності.

Насамперед знайдемо довірчий інтервал для оціночного рівняння. Для цього нам необхідно мати похибку оцінки, яку знайдемо за формулою: $\Delta_{yx} = t_p \cdot \sigma_{пом}$, де t_p – ймовірнісний коефіцієнт, який при заданих рівнях ймовірності Р знаходиться за таблицями нормального розподілу. Значення t_p знаходимо з розв'язку рівняння

$2\Phi(t_p)=p$, де $\Phi(t_p)$ – інтегральна функція Лапласа.

Тоді довірчий інтервал для оціночного рівняння знайдемо з нерівності

$$\hat{y}_i - \Delta_{yx} \le y_i \le \hat{y}_i + \Delta_{yx} \tag{1.13}$$

Так, наприклад, для P=0,9 маємо: t_p=1,65 (P=0,95 маємо: t_p=1,96). Тоді $\Delta_{yx} = 1{,}65 \cdot 0{,}1605 = 0{,}2648$.

Довірчий інтервал знаходимо з нерівності:

$$\hat{y}_i - 0{,}2648 \le y_i \le \hat{y}_i + 0{,}2648,$$

$$-1{,}8989 + 1{,}2114x_i - 0{,}2648 \le y_i \le -1{,}8989 + 1{,}2114x_i + 0{,}2648$$

$$-2{,}1637 + 1{,}2114x_i \le y_i \le -1{,}6341 + 1{,}2114x_i.$$

Доповнимо рис. 1.1 геометричною інтерпретацією отриманих довірчих інтервалів. Нам необхідно побудувати дві додаткові граничні прямі, які відображають на площині верхню та нижню частину отриманої нерівності. Між даними прямими знаходиться наше оціночне рівняння. На графіку воно подано більш ширшою лінією.

VII. Знаходження довірчих інтервалів для оцінок.

Для встановлення зв’язку параметрів a та b із параметрами α та β, потрібно побудувати для них інтервали довіри. Процедура побудови інтервалів довіри є аналогічною попередній процедурі. Спочатку знаходимо граничні похибки оцінок відповідних параметрів за формулами:

$$\Delta_a = \sigma_a \cdot t_p, \Delta_b = \sigma_b \cdot t_p \tag{1.14}$$

де σ_a^2, σ_b^2 – відповідно, дисперсії оцінок a та b, які визначаються за формулами:

$$\sigma_a^2 = \frac{\sum_{i=1}^{n} x_i^2 \, \sigma^2_{\text{непоясн.}}}{n\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}, \quad \sigma_b^2 = \frac{\sigma^2_{\text{непоясн.}}}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2} \tag{1.15}$$

Отже, довірчий інтервал для оцінки α буде:

$$a - \Delta_a \le \alpha \le a + \Delta_a, \tag{1.16}$$

що означає: $P(a - \Delta_a \le \alpha \le a + \Delta_a) = 1 - \alpha$.

Відповідно, для параметра β маємо:

$$b - \Delta_b \le \beta \le b + \Delta_b \tag{1.17}$$

або

$$P(b - \Delta_b \le \beta \le b + \Delta_b) = 1 - \alpha.$$

Знайдемо значення σ_a^2, σ_b^2:

$$\sigma_a^2 = \frac{\sum_{i=1}^{n} x_i^2 \sigma^2_{непоясн.}}{n\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{181{,}51 \cdot 0{,}2577}{10 \cdot 12{,}589} = 0{,}3716,$$

$$\sigma_a = \sqrt{0{,}3716} = 0{,}6096.$$

$$\sigma_b^2 = \frac{\sigma^2_{непоясн.}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{0{,}2577}{12{,}589} = 0{,}0205,$$

$$\sigma_b = 0{,}1432.$$

Далі знайдемо

$$\Delta_a = \sigma_a \cdot t_p = 0{,}6096 \cdot 1{,}65 = 1{,}0058;$$

$$\Delta_b = \sigma_b \cdot t_p = 0{,}1432 \cdot 1{,}65 = 0{,}2363.$$

Отже, нами отримано наступні інтервали довіри для оцінки α:

$$\alpha \in [a - \Delta_a; a + \Delta_a];$$

$$\alpha \in [-1{,}8989 - 1{,}0058; -1{,}8989 + 1{,}0058],\ \alpha \in [-2{,}9047; 0{,}8931];$$

для оцінки β:

$$\beta \in [b - \Delta_b; b + \Delta_b],$$

$$\beta \in [1{,}2114 - 0{,}1432; 1{,}2114 + 0{,}1432],\ 1{,}0682 \le \beta \le 1{,}3546.$$

VIII. Знаходження значень коефіцієнтів детермінації та кореляції.

Для обчислення коефіцієнта детермінації використовуємо формулу:

$$R^2 = 1 - \frac{\sigma^2_{непоясн.}}{\sigma^2_{заг.}} = 1 - \frac{0{,}2577}{1{,}8763} = 0{,}8626, \qquad (1.18)$$

а коефіцієнт кореляції знайдемо за формулою:

$$r = R = \pm\sqrt{1 - \frac{\sigma^2_{непоясн.}}{\sigma^2_{заг.}}} = \sqrt{0{,}8626} = 0{,}9288. \qquad (1.19)$$

Отримане значення коефіцієнта кореляції вказує на ступінь тісноти зв'язку між вибраними факторами у вигляді прямолінійної залежності.

На практиці використовується наступна шкала ступеня тісноти зв'язку (табл. 1.3).

Таблиця 1.3

Шкала ступеня тісноти зв'язку коефіцієнта кореляції

Значення $\lvert r_{xy} \rvert$	**Характер зв'язку**
$\lvert r_{xy} \rvert < 0{,}3$	практично відсутній зв'язок
$0{,}3 \le \lvert r_{xy} \rvert < 0{,}5$	слабкий
$0{,}5 \le \lvert r_{xy} \rvert < 0{,}7$	помірний
$\lvert r_{xy} \rvert \ge 0{,}7$	сильний

Як бачимо, у нашому випадку має місце сильний зв'язок. Значення коефіцієнта детермінації показує, що 86,26 % варіація прибутку підприємств регіону пояснюється за рахунок варіації основних фондів. На частку неврахованих та випадкових факторів припадає лише 13,74 %.

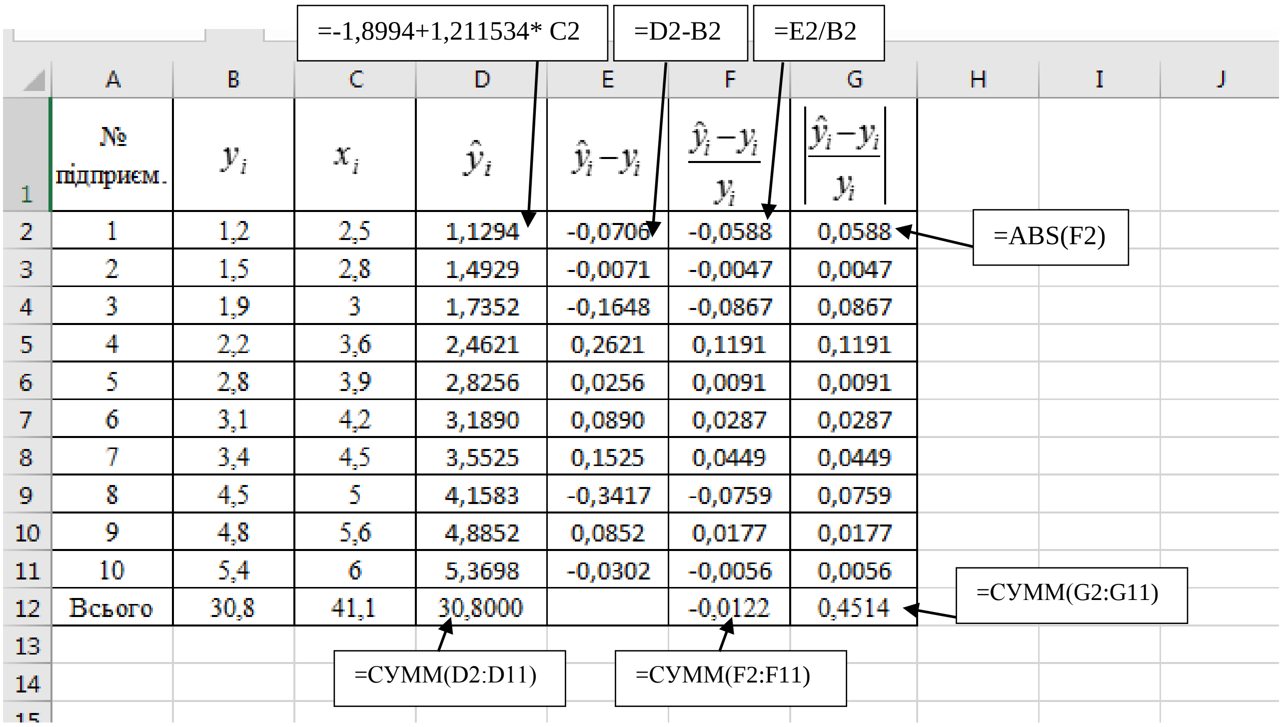
У нашому випадку коефіцієнт $b = 1{,}211534 > 0$. Якщо $b < 0$, то коефіцієнт кореляції приймає від'ємний знак, що свідчить про зворотній зв'язок між вибраними показниками.

IX. Знаходження показника середньої відсоткової помилки та її абсолютне значення.

Абсолютну середню відсоткову помилку обчислимо за формулою:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \cdot 100\% . \qquad (1.20)$$

Для того, щоб обчислити абсолютну відсоткову помилку для нашого випадку, знову скористаємось табличним процесором MS Excel (рис. 1.9).



№ підприєм.	y_i	x_i	$\hat{y}_i$	$\hat{y}_i - y_i$	$\frac{\hat{y}_i - y_i}{y_i}$	$\left\lvert\frac{\hat{y}_i - y_i}{y_i}\right\rvert$
1	1,2	2,5	1,1294	-0,0706	-0,0588	0,0588
2	1,5	2,8	1,4929	-0,0071	-0,0047	0,0047
3	1,9	3	1,7352	-0,1648	-0,0867	0,0867
4	2,2	3,6	2,4621	0,2621	0,1191	0,1191
5	2,8	3,9	2,8256	0,0256	0,0091	0,0091
6	3,1	4,2	3,1890	0,0890	0,0287	0,0287
7	3,4	4,5	3,5525	0,1525	0,0449	0,0449
8	4,5	5	4,1583	-0,3417	-0,0759	0,0759
9	4,8	5,6	4,8852	0,0852	0,0177	0,0177
10	5,4	6	5,3698	-0,0302	-0,0056	0,0056
Всього	30,8	41,1	30,8000		-0,0122	0,4514

Рис. 1.9. Фрагмент документу MS Excel, в якому позначені формули для здійснення розрахунків з метою обчислення показника середньої відсоткової помилки та її абсолютне значення

Отже, $MAPE = \frac{1}{10}\sum_{i=1}^{10}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \cdot 100\% = \frac{1}{10} \cdot 0{,}4514 \cdot 100\% = 4{,}514\%$.

Середню відсоткову помилку знайдемо за формулою:

$$MPE = \frac{1}{n}\sum_{i=1}^{n}\frac{\hat{y}_i - y_i}{y_i} \cdot 100\%. \qquad (1.21)$$

У нашому випадку маємо (рис. 1.9):

$$MPE = \frac{1}{10}\sum_{i=1}^{10}\frac{\hat{y}_i - y_i}{y_i} \cdot 100\% = \frac{1}{10} \cdot (-0{,}0122) \cdot 100\% = -0{,}122\%.$$

Для визначення ступеня якості отриманих моделей необхідно врахувати наступне:

- якщо значення *MAPE* менше 10%, то отримана модель дає високу точність прогнозу і в цілому якість моделі;
- якщо значення *MAPE* знаходиться в межах від 10% до 20% – модель дає добру точність;
- якщо значення *MAPE* знаходиться в межах від 20% до 50 – модель дає задовільну точність;
- якщо значення *MAPE* понад 50%, то отримана модель дає низьку, незадовільну точність прогнозу.

Отже, як бачимо, у нашому випадку $MAPE = 4{,}514\% < 10\%$, що свідчить про високу точність прогнозу і в цілому якість моделі.

X. Перевірка нульових гіпотез відносно коефіцієнту кореляції та кутового коефіцієнту.

Знайдений коефіцієнт кореляції є випадковою величиною. Виконаємо для нього перевірку нульової гіпотези, відповідно до якої коефіцієнт кореляції у генеральній сукупності дорівнює нулю. Тобто, відсутній кореляційний зв'язок між *y* та *x* у генеральній сукупності. Нам необхідно дослідити сумісність коефіцієнта кореляції *r* із нашої вибірки з нульовою гіпотезою.

Отже, маємо: нульова гіпотеза H_0: $r_{ген}=0$, альтернативна гіпотеза H_1: $r_{ген}\neq 0$.

Далі для заданої вибірки з $k=n-2$ ступенями вільності обчислимо значення статистики для критерія Ст'юдента:

$$t_{емп} = \frac{|r|\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0{,}9288\sqrt{10-2}}{\sqrt{1-0{,}8626}} = 7{,}1691. \qquad (1.22)$$

Для заданої ймовірності, наприклад, $P=0{,}95$ і $k=8$ ступенів вільності знаходимо табличне значення $t_{kp} = 2{,}306$.

Якщо $|t_{емп}| \geq t_{kp}$, то із надійністю P=0,95 гіпотезу H_0 необхідно відкинути і прийняти альтернативну гіпотезу H_1 про існування залежності між цими випадковими величинами. Оскільки, в нашому прикладі $|t_{емп}| \geq t_{kp}$,то нульова гіпотеза відхиляється і приймається альтернативна. Отже, в 95 % вибірок із генеральної сукупності коефіцієнт кореляції не дорівнює нулю.

Далі виконаємо перевірку нульової гіпотези відносно оцінки *b*, H_0: $b=0$, проти альтернативної H_1: $b\neq 0$. Для цього знаходимо емпіричне значення за формулою:

$$t_{емп} = \frac{|b|}{\sigma_в} = \frac{1{,}2114}{0{,}1432} = 8{,}4595. \qquad (1.23)$$

Оскільки емпіричне значення *t* більше критичного ($t_{емп} > t_{кр}$), то нульова гіпотеза відхиляється і робиться висновок, що кутовий коефіцієнт *b* розрахований за даною вибіркою вважається статистично значущим з ймовірністю $P=0{,}95$.

XI. Перевірка адекватності побудованої економетричної моделі.

Для оцінки рівня адекватності побудованої економетричної моделі експериментальним даним використовуємо критерій Фішера, тобто F-критерій. Він дозволяє оцінити, чи значно нахил b відрізняється від нуля, тобто перевірити гіпотезу H_0: $b=0$. Іншими словами, як краще апроксимувати дані за середнім значенням, чи регресійною прямою.

Альтернативна гіпотеза полягає в тому, що $b \neq 0$ і має вигляд: H_1: $b \neq 0$. Показник Фішера вираховується як відношення середнього квадрата, що пояснюється регресією, до середнього квадрата помилок:

$$F = \frac{КР}{КП} = \frac{\frac{СКР}{m}}{\frac{СКП}{n-m-1}} = \frac{СКР}{СКП} \cdot \frac{n-m-1}{m} = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}, \qquad (1.24)$$

де m – число незалежних змінних (для простої регресії $m=1$), тобто

$$F = \frac{R^2(n-2)}{1-R^2}. \qquad (1.25)$$

Отже, нам необхідно обчислити:

$$F = \frac{R^2(n-2)}{1-R^2} = \frac{0{,}8626 \cdot (10-2)}{1-0{,}8626} = 50{,}2242.$$

Нами отримано розрахункове значення $F_{роз.}$ = 50,2242.

Знайдемо табличне значення даного критерія ($F_{таб.}$) для рівня надійності P=0,95 і числа ступенів вільності $k_1 = m = 1$, $k_2 = n - m - 1 = 10 - 1 - 1 = 8$, $F_{таб}$ = 5,32.

Оскільки. $F_{роз} > F_{таб.}$, то отриману економетричну модель

$$\hat{y} = -1{,}8989 + 1{,}2114x$$

з надійністю P=0,95 можна вважати адекватною експериментальним даним і на її основі доцільно проводити всебічний економетричний аналіз.

Дану задачу можна розв'язати з допомогою програмного продукту STADIA, використавши при цьому процедуру "Простая регрессия" блоку "Регрессионный анализ" або MS Excel.

Завдання 2. Багатофакторна економетрична модель.

Припустимо, що деяка змінна y залежить від двох незалежних змінних x_1 та x_2. Тоді у випадку лінійної форми взаємозв’язку економетрична модель прийме вид:

$$y = b_0 + b_1 x_1 + b_2 x_2 + u, \tag{2.1}$$

де y – залежна змінна; x_1 та x_2 – незалежні змінні; b_0, b_1, b_2 – параметри моделі, для яких потрібно буде знайти оцінки; u – збурення або залишок.

Тоді оціночне рівняння для моделі (2.1) буде:

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2, \tag{2.2}$$

де a_0, a_1, a_2 відповідні оцінки параметрів b_0, b_1, b_2.

Приклад 2.1. Оцінити параметри економетричної моделі, яка описує залежність обсягу отриманого прибутку підприємствами регіону від розміру основних виробничих фондів та затрат праці. Вхідні дані приведено в табл. 2.1.

Таблиця 2.1.

Вплив вартості основних виробничих фондів та затрат праці на прибуток підприємства

Номер підприємства	Прибуток, млн. грн., y	Основні фонди, млн. грн., x_1	Затрати праці, млн. днів, x_2
1	1,2	2,5	4,0
2	1,5	2,8	4,2
3	1,9	3,0	3,6
4	2,2	3,6	4,6
5	2,8	3,9	4,3
6	3,1	4,2	5,1
7	3,4	4,5	5,3
8	4,5	5,0	4,8
9	4,8	5,6	5,4
10	5,4	6,0	5,8

Розв’язок даної задачі проведемо з допомогою наступних кроків.

І. Ідентифікація змінних.

Позначимо:

y – залежна змінна, прибуток підприємства, млн. грн.;

x_1 – незалежна змінна, вартість основних виробничих фондів підприємства, млн. грн.;

x_2 – незалежна змінна, затрати праці, млн. днів.

II. Специфікація моделі.

Попередній аналіз вхідної інформації (діаграми розсіювання) дає можливість зробити висновок про наявність лінійної форми зв'язку між вибраними економічними показниками:

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2, \tag{2.2}$$

III. Знаходження оціночних параметрів.

Розрахувати оцінки a_0, a_1, a_2 можна виконати з допомогою наступних методів:

- МНК (за системою нормальних рівнянь);
- матричним методом;
- стандартними процедурами прикладних програм (STATAN, MS Excel та ін.).

1. Для використання МНК (методу найменших квадратів) нам необхідно побудувати систему нормальних рівнянь, яка у нашому випадку прийме вигляд:

$$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} x_{i1} + a_2 \sum_{i=1}^{n} x_{i2} = \sum_{i=1}^{n} y_i \\ a_0 \sum_{i=1}^{n} x_{i1} + a_1 \sum_{i=1}^{n} x_{i1}^2 + a_2 \sum_{i=1}^{n} x_{i1} x_{i2} = \sum_{i=1}^{n} y_i x_{i1} \\ a_0 \sum_{i=1}^{n} x_{i2} + a_1 \sum_{i=1}^{n} x_{i1} x_{i2} + a_2 \sum_{i=1}^{n} x_{i2}^2 = \sum_{i=1}^{n} y_i x_{i2} \end{cases} \tag{2.3}$$

Таким чином, нами отримано систему рівнянь (2.3) з трьома невідомими величинами: a_0, a_1, a_2. З першого рівняння маємо:

$$a_0 = \bar{y} - a_1 \bar{x}_1 - a_2 \bar{x}_2. \tag{2.4}$$

Використовуючи даний вираз і два інші рівняння, після відповідних перетворень, знайдемо значення a_1:

$$a_1 = \frac{\operatorname{cov}(x_1, y)\operatorname{var}(x_2) - \operatorname{cov}(x_2, y)\operatorname{cov}(x_1, x_2)}{\operatorname{var}(x_1)\operatorname{var}(x_2) - \{\operatorname{cov}(x_1, x_2)\}^2}. \tag{2.5}$$

Аналогічно можна знайти значення для a_2, зробивши перестановку x_1 та x_2 у виразі (2.5).

Для того, щоб ми могли записати числову систему нормальних рівнянь для нашого прикладу, виконаємо деякі обрахунки в MS Excel (рис. 2.1).

	A	B	C	D	E	F	G	H	I	J
1	№ підприєм.	y_i	x_{i1}	x_{i2}	x_{i1}^2	x_{i2}^2	$y_i x_{i1}$	$y_i x_{i2}$	$x_{i1} x_{i2}$	y_i^2
2	1	1,2	2,5	4	6,25	16	3	4,8	10	1,44
3	2	1,5	2,8	4,2	7,84	17,64	4,2	6,3	11,76	2,25
4	3	1,9	3	3,6	9	12,96	5,7	6,84	10,8	3,61
5	4	2,2	3,6	4,6	12,96	21,16	7,92	10,12	16,56	4,84
6	5	2,8	3,9	4,3	15,21	18,49	10,92	12,04	16,77	7,84
7	6	3,1	4,2	5,1	17,64	26,01	13,02	15,81	21,42	9,61
8	7	3,4	4,5	5,3	20,25	28,09	15,3	18,02	23,85	11,56
9	8	4,5	5	4,8	25	23,04	22,5	21,6	24	20,25
10	9	4,8	5,6	5,4	31,36	29,16	26,88	25,92	30,24	23,04
11	10	5,4	6	5,8	36	33,64	32,4	31,32	34,8	29,16
12	Всього	30,8	41,1	47,1	181,51	226,19	141,84	152,77	200,2	113,6

=C2^2 =D2^2 =B2*C2 =B2*D2 =C2*D2 =B2^2

=СУММ(E2:E11) =СУММ(F2:F11) =СУММ(G2:G11) =СУММ(H2:H11) =СУММ(I2:I11) =СУММ(J2:J11)

Рис. 2.1. Фрагмент документу MS Excel, в якому здійснені розрахунки для подальшого запису числової системи нормальних рівнянь

Отже, числова система нормальних рівнянь матиме вигляд:

$$\begin{cases} 10a_0 + 41{,}1a_1 + 47{,}1a_2 = 30{,}8, \\ 41{,}1a_0 + 181{,}51a_1 + 200{,}2a_2 = 141{,}84, \\ 47{,}1a_0 + 200{,}2a_1 + 226{,}19a_2 = 152{,}77. \end{cases} \tag{2.6}$$

Для знаходження a_0, a_1, a_2 використаємо правило Крамера. Обчислимо визначники в MS Excel, застосувавши математичну функцію МОПРЕД (MDETERM) (рис. 2.2).

N	O	P	Q	R	S
					=МОПРЕД(O2:Q4)
	10	41,1	47,1		
$\Delta =$	41,1	181,51	200,2	=	109,384
	47,1	200,2	226,19		

Рис. 2.2. Фрагмент документу MS Excel, в якому здійснено обчислення визначників

Таким чином $\Delta = 109{,}384$. Аналогічно знаходимо

$$\Delta_{a_0} = \begin{vmatrix} 30{,}8 & 41{,}1 & 47{,}1 \\ 141{,}84 & 181{,}51 & 200{,}2 \\ 152{,}77 & 200{,}2 & 226{,}19 \end{vmatrix} = -105{,}993;$$

$$\Delta_{a_1} = \begin{vmatrix} 10 & 30{,}8 & 47{,}1 \\ 41{,}1 & 141{,}84 & 200{,}2 \\ 47{,}1 & 152{,}77 & 226{,}19 \end{vmatrix} = 153{,}5141; \quad \Delta_{a_2} = \begin{vmatrix} 10 & 41{,}1 & 30{,}8 \\ 41{,}1 & 181{,}51 & 141{,}84 \\ 47{,}1 & 200{,}2 & 152{,}77 \end{vmatrix} = -39{,}9251;$$

Отже,

$$a_0 = \frac{\Delta_{a_0}}{\Delta} = \frac{-105{,}993}{109{,}384} = -0{,}969 \approx -0{,}97, \quad a_1 = \frac{\Delta_{a_1}}{\Delta} = \frac{153{,}5141}{109{,}384} = 1{,}403442 \approx 1{,}4,$$

$$a_2 = \frac{\Delta_{a_2}}{\Delta} = \frac{-39{,}9251}{109{,}384} = -0{,}365 \approx -0{,}37.$$

Враховуючи отримані значення оціночних параметрів наша економетрична модель набуде вигляду:

$$y = -0{,}97 + 1{,}4x_1 - 0{,}37x_2 + u \tag{2.7}$$

2. Використаємо інструментарій матричної алгебри та табличний процесор MS Excel. Значення оцінок знайдемо за формулою:

$$A = \left(X'X\right)^{-1} X'Y \tag{2.8}$$

де X' – транспонована матриця до матриці X.

Введемо позначення:

$$A = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}; \quad Y = \begin{bmatrix} 1{,}2 \\ 1{,}5 \\ 1{,}9 \\ 2{,}2 \\ 2{,}8 \\ 3{,}1 \\ 3{,}4 \\ 4{,}5 \\ 4{,}8 \\ 5{,}4 \end{bmatrix}; \quad X = \begin{bmatrix} 1 & 2{,}5 & 4{,}0 \\ 1 & 2{,}8 & 4{,}2 \\ 1 & 3{,}0 & 3{,}6 \\ 1 & 3{,}6 & 4{,}6 \\ 1 & 3{,}9 & 4{,}3 \\ 1 & 4{,}2 & 5{,}1 \\ 1 & 4{,}5 & 5{,}3 \\ 1 & 5{,}0 & 4{,}8 \\ 1 & 5{,}6 & 5{,}4 \\ 1 & 6{,}0 & 5{,}8 \end{bmatrix}.$$

2.1) Знаходимо транспоновану матрицю X' використовуючи табличний процесор MS Excel із допомогою функції ТРАНСП (TRANSPOSE):

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2{,}5 & 2{,}8 & 3{,}0 & 3{,}6 & 3{,}9 & 4{,}2 & 4{,}5 & 5{,}0 & 5{,}6 & 6{,}0 \\ 4{,}0 & 4{,}2 & 3{,}6 & 4{,}6 & 4{,}3 & 5{,}1 & 5{,}3 & 4{,}8 & 5{,}4 & 5{,}8 \end{bmatrix}.$$

2.2) Знаходимо добуток двох матриць X' та X, використавши функцію МУМНОЖ (MMULT) в MS Excel:

$$\left[X' \cdot X\right] = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2,5 & 2,8 & 3,0 & 3,6 & 3,9 & 4,2 & 4,5 & 5,0 & 5,6 & 6,0 \\ 4,0 & 4,2 & 3,6 & 4,6 & 4,3 & 5,1 & 5,3 & 4,8 & 5,4 & 5,8 \end{bmatrix} \times \begin{bmatrix} 1 & 2,5 & 4,0 \\ 1 & 2,8 & 4,2 \\ 1 & 3,0 & 3,6 \\ 1 & 3,6 & 4,6 \\ 1 & 3,9 & 4,3 \\ 1 & 4,2 & 5,1 \\ 1 & 4,5 & 5,3 \\ 1 & 5,0 & 4,8 \\ 1 & 5,6 & 5,4 \\ 1 & 6,0 & 5,8 \end{bmatrix} =$$

$$= \begin{bmatrix} 10 & 41,1 & 47,1 \\ 41,1 & 181,5 & 200,2 \\ 47,1 & 200,2 & 226,2 \end{bmatrix}.$$

2.3) Знаходимо обернену матрицю до отриманої у п. 2.2, використовуючи функцію МОБР (MINVERSE) в MS Excel:

$$\left[X'X\right]^{-1} = \begin{bmatrix} 8,92 & 1,22 & -2,93 \\ 1,22 & 0,4 & -0,61 \\ -2,93 & -0,61 & 1,51 \end{bmatrix}.$$

2.4) Знаходимо добуток матриці X' та вектора Y, використавши функцію МУМНОЖ (MMULT) в MS Excel:

$$X' \cdot Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2,5 & 2,8 & 3,0 & 3,6 & 3,9 & 4,2 & 4,5 & 5,0 & 5,6 & 6,0 \\ 4,0 & 4,2 & 3,6 & 4,6 & 4,3 & 5,1 & 5,3 & 4,8 & 5,4 & 5,8 \end{bmatrix} \times \begin{bmatrix} 1,2 \\ 1,5 \\ 1,9 \\ 2,2 \\ 2,8 \\ 3,1 \\ 3,4 \\ 4,5 \\ 4,8 \\ 5,4 \end{bmatrix} = \begin{bmatrix} 30,8 \\ 141,84 \\ 152,77 \end{bmatrix}.$$

2.5) Знаходимо значення оператора оцінювання за допомогою функції МУМНОЖ (MMULT) в MS Excel:

$$A = \left[X'X\right]^{-1}\left[X'Y\right] = \begin{bmatrix} 8{,}92 & 1{,}22 & -2{,}93 \\ 1{,}22 & 0{,}4 & -0{,}61 \\ -2{,}93 & -0{,}61 & 1{,}51 \end{bmatrix} \cdot \begin{bmatrix} 30{,}8 \\ 141{,}84 \\ 152{,}77 \end{bmatrix} = \begin{bmatrix} -0{,}97 \\ 1{,}4 \\ -0{,}37 \end{bmatrix}.$$

Отже, ми отримали наступну економетричну модель:

$$y = -0{,}97 + 1{,}4x_1 - 0{,}37x_2 + u. \qquad (2.9)$$

Як бачимо отримані результати співпадають.

У випадку лінійної форми зв'язку коефіцієнт граничної продуктивності співпадає з відповідним оціночним значенням, тобто має місце

$$Г_j = \frac{\partial y}{\partial x_j} = a_j, \; j = \overline{1, m}. \qquad (2.10)$$

Так, наприклад, для нашої моделі $Г_1 = a_1 = 1{,}4$ означає, що при збільшенні вартості основних фондів на 1 млн. грн. прибуток підприємства зросте на 1,4 млн. грн.

Коефіцієнт еластичності знайдемо за допомогою формули:

$$E_j = \frac{\partial y}{\partial x_j} \cdot \frac{\bar{x}_j}{\bar{y}} = a_j \cdot \frac{\bar{x}_j}{\bar{y}}. \qquad (2.11)$$

Наприклад, $E_1 = a_1 \cdot \frac{\bar{x}_j}{\bar{y}} = 1{,}4 \cdot \frac{3{,}21}{3{,}08} = 1{,}46$ означає, що збільшення основних виробничих фондів на 1 % призведе до зростання прибутку на 1,46 %.

IV. Знаходження коефіцієнтів множинної кореляції та детермінації.

Для знаходження коефіцієнта множинної детермінації використаємо формулу:

$$R^2 = \frac{\sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2}, \qquad (2.12)$$

де $\bar{y}$ – середнє значення всіх y; $\hat{y}_i$ – розрахункові значення y.

Розрахунки для $\hat{y}_i$ проведемо в MS Excel (рис. 2.3).

№ підприєм.	x_{i1}	x_{i2}	$\hat{y}_i$
1	2,5	4	1,05
2	2,8	4,2	1,40
3	3	3,6	1,90
4	3,6	4,6	2,37
5	3,9	4,3	2,90
6	4,2	5,1	3,02
7	4,5	5,3	3,37
8	5	4,8	4,25
9	5,6	5,4	4,87
10	6	5,8	5,28
Всього	41,1	47,1	

=-0,97+1,4*B2-0,37*C2

Рис. 2.3. Фрагмент документу MS Excel, в якому обчислено розрахункові значення *y*

Для спрощення подальших розрахунків побудуємо таблицю і заповнимо її в MS Excel (рис. 2.4). Для потрібних обчислень знайдемо середнє значення всіх *y* за формулою:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} = \frac{\sum_{i=1}^{10} y_i}{10} = \frac{30,8}{10} = 3,08 .$$

=B2-3,08   =D2-3,08   =C2^2   =E2^2

№ підприєм.	y_i	$y_i - \bar{y}$	$\hat{y}_i$	$\hat{y}_i - \bar{y}$	$(y_i - \bar{y})^2$	$(\hat{y}_i - \bar{y})^2$
1	1,2	-1,88	1,05	-2,03	3,53	4,12
2	1,5	-1,58	1,40	-1,68	2,50	2,84
3	1,9	-1,18	1,90	-1,18	1,39	1,40
4	2,2	-0,88	2,37	-0,71	0,77	0,51
5	2,8	-0,28	2,90	-0,18	0,08	0,03
6	3,1	0,02	3,02	-0,06	0,0004	0,0032
7	3,4	0,32	3,37	0,29	0,102	0,08
8	4,5	1,42	4,25	1,17	2,02	1,38
9	4,8	1,72	4,87	1,79	2,96	3,21
10	5,4	2,32	5,28	2,20	5,38	4,86
Всього	30,8				18,736	18,428

=СУММ(F2:F11)   =СУММ(G2:G11)

Рис. 2.4. Фрагмент документу MS Excel, в якому здійнено розрахунки для знаходження коефіцієнтів детермінації та кореляції

Знайдені значення підставимо у нашу формулу, внаслідок чого коефіцієнт множинної детермінації прийме значення:

$$R^2 = \frac{18{,}428}{18{,}736} = 0{,}9836.$$

Отже, 98,36 % загальної дисперсії пояснюється загальною залежністю обсягу прибутку від вартості основних виробничих фондів та затрат праці. І тільки 1,64 % загальної дисперсії не може бути пояснено отриманою нами зележністю. Таким чином, можна зробити висновок, що побудована модель статистично значима.

Тоді коефіцієнт множинної кореляції буде:

$$R = \sqrt{R^2} = \sqrt{0{,}9836} = 0{,}9918.$$

Нами отримано досить високий коефіцієнт множинної кореляції. Це свідчить про те, що зв'язок між величиною прибутку, вартістю основних фондів та затратами праці досить тісний.

V. Знаходження коефіцієнтів парної кореляції.

Коефіцієнт парної кореляції використовується для вимірювання сили лінійних зв'язків різних пар змінних із заданої множини. Для їх знаходження використаємо формулу:

$$r_{yx} = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{\left[n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right]\left[n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}} \qquad (2.13)$$

Підставимо отримані значення із табл. 2.3 у вихідну формулу:

$$r_{01} = r_{y_1} = r_{yx_1} = \frac{10 \cdot 141{,}84 - 41{,}1 \cdot 30{,}8}{\sqrt{\left(10 \cdot 181{,}51 - 41{,}1^2\right)\left(10 \cdot 113{,}6 - 30{,}8^2\right)}} =$$

$$= \frac{1418{,}4 - 1265{,}88}{\sqrt{(1815{,}1 - 1689{,}21)(1136 - 948{,}64)}} = \frac{152{,}52}{\sqrt{125{,}89 \cdot 187{,}36}} =$$

$$= \frac{152{,}52}{\sqrt{23586{,}75}} = \frac{152{,}52}{153{,}58} = 0{,}993.$$

Нами отримано значення $r_{yx_1} = 0{,}993$. Це свідчить про те, що 99,3 % загальної дисперсії змінної y пояснюється економетричною моделлю $y = \alpha' + \beta' x_1$. Непояснена дисперсія викликана випадковою складовою, становить 0,7 %.

$$r_{02} = r_{y_2} = r_{yx_2} = \frac{10 \cdot 152{,}77 - 47{,}1 \cdot 30{,}8}{\sqrt{\left(10 \cdot 226{,}19 - 47{,}1^2\right)\left(10 \cdot 113{,}6 - 30{,}8^2\right)}} =$$

$$= \frac{1527{,}4 - 1450{,}68}{\sqrt{(2261{,}9 - 2218{,}41)(1136 - 948{,}64)}} = \frac{76{,}72}{\sqrt{43{,}49 \cdot 187{,}36}} =$$

$$= \frac{76{,}72}{\sqrt{8148{,}29}} = \frac{76{,}72}{90{,}27} = 0{,}85.$$

Маємо $r_{yx_2} = 0{,}85$, тобто 85 % загальної дисперсії змінної y пояснюється економетричною моделлю $y = \alpha'' + \beta'' x_2$, при цьому непояснена дисперсія складає 15 %.

$$r_{12} = r_{x_1 x_2} = \frac{10 \cdot 200{,}2 - 41{,}1 \cdot 47{,}1}{\sqrt{\left(10 \cdot 181{,}51 - 41{,}1^2\right)\left(10 \cdot 226{,}19 - 47{,}1^2\right)}} =$$

$$= \frac{2002 - 1935{,}81}{\sqrt{(1815{,}1 - 1689{,}21)(2261{,}9 - 2218{,}41)}} = \frac{66{,}19}{\sqrt{125{,}89 \cdot 43{,}49}} =$$

$$= \frac{66{,}19}{\sqrt{5474{,}956}} = \frac{66{,}19}{73{,}99} = 0{,}89.$$

Розраховане значення $r_{x_1 x_2} = 0{,}89$ вказує на те, що 89 % загальної дисперсії змінної x_1 пояснюється економетричною моделлю $x_1 = \alpha''' + \beta''' x_2$. Непояснена дисперсія складає 11 %. Тобто пояснююча змінна x_1 корелює з x_2, що свідчить про наявність мультиколінеарності.

VI. Знаходження коефіцієнтів частинної кореляції.

Коефіцієнти частинної кореляції так само представляють лінійні зв'язки ознак, але при цьому береться до уваги чистий зв'язок пари ознак за умови, що зв'язки всіх інших ознак з ознаками із даної пари не діють. Отже, частинною кореляцією між ознаками x_i та x_j називається кореляційна залежність між цими ознаками при фіксованих значеннях інших ознак.

Для знаходження коефіцієнтів частинної кореляції використаємо наступні альтернативні формули. Нехай число незалежних змінних $m = 2$. Тоді мають місце наступні формули:

$$r_{01.2} = r_{y1.2} = \frac{r_{y1} - r_{y2} r_{12}}{\sqrt{\left(1 - r_{y2}^2\right)\left(1 - r_{12}^2\right)}}, \quad r_{02.1} = r_{y2.1} = \frac{r_{y2} - r_{y1} r_{12}}{\sqrt{\left(1 - r_{y1}^2\right)\left(1 - r_{12}^2\right)}}, \qquad (2.14)$$

де $r_{y1.2}$ – коефіцієнт частинної кореляції між змінними y та x_1 при виключенні впливу x_2; $r_{y2.1}$ – коефіцієнт частинної кореляції між змінними y та x_2 при виключенні впливу x_1; r_{y1}, r_{y2}, r_{12} – відповідно парні коефіцієнти кореляції.

Аналогічно, для випадку $m = 3$ маємо:

$$r_{y1.23} = \frac{r_{y1.3} - r_{y2.3} r_{12.3}}{\sqrt{\left(1 - r_{y2.3}^2\right)\left(1 - r_{12.3}^2\right)}}; \quad r_{y2.13} = \frac{r_{y2.3} - r_{y1.3} r_{21.3}}{\sqrt{\left(1 - r_{y1.3}^2\right)\left(1 - r_{21.3}^2\right)}};$$

$$r_{y3.12} = \frac{r_{y3.2} - r_{y1.2} r_{31.2}}{\sqrt{\left(1 - r_{y1.2}^2\right)\left(1 - r_{31.2}^2\right)}}, \quad (2.15)$$

де $r_{y1.23}, r_{y2.13}, r_{y3.12}$ – коефіцієнти частинної кореляції третього порядку.

Враховуючи отримані значення коефіцієнтів парної кореляції перейдемо до обчислення відповідних значень коефіцієнтів частинної кореляції:

$$r_{01.2} = r_{y1.2} = \frac{r_{y1} - r_{y2} r_{12}}{\sqrt{\left(1 - r_{y2}^2\right)\left(1 - r_{12}^2\right)}} = \frac{0{,}993 - 0{,}85 \cdot 0{,}89}{\sqrt{\left(1 - 0{,}85^2\right)\left(1 - 0{,}89^2\right)}} = \frac{0{,}993 - 0{,}757}{\sqrt{(1 - 0{,}7225)(1 - 0{,}7921)}} =$$

$$= \frac{0{,}236}{\sqrt{0{,}2475 \cdot 0{,}2079}} = \frac{0{,}236}{\sqrt{0{,}0577}} = \frac{0{,}236}{0{,}24} = 0{,}983.$$

Нами отримано $r_{y1.2} = 0{,}983$, що свідчить про те, що 15 % дисперсії змінної y непоясненої змінної x_2 98,3 % пояснить введенням в модель змінної x_1.

$$r_{y2.1} = \frac{r_{y2} - r_{y1} r_{12}}{\sqrt{\left(1 - r_{y1}^2\right)\left(1 - r_{12}^2\right)}} = \frac{0{,}85 - 0{,}993 \cdot 0{,}89}{\sqrt{\left(1 - 0{,}993^2\right)\left(1 - 0{,}89^2\right)}} = \frac{0{,}85 - 0{,}884}{\sqrt{(1 - 0{,}986)(1 - 0{,}7921)}} =$$

$$= \frac{-0{,}034}{\sqrt{0{,}014 \cdot 0{,}2079}} = \frac{-0{,}034}{\sqrt{0{,}0029}} = \frac{-0{,}034}{0{,}054} = -0{,}63.$$

Як бачимо, що 0,7% дисперсії у непоясненої змінної x_1 63% пояснить введенням в модель змінної x_2.

Коефіцієнти частинної кореляції можна знаходити за допомогою матричної алгебри. Розглянемо випадок m=2. Тоді матриця парних коефіцієнтів кореляції матиме вигляд:

$$[R_2] = \begin{bmatrix} r_{00} & r_{01} & r_{02} \\ r_{10} & r_{11} & r_{12} \\ r_{20} & r_{21} & r_{22} \end{bmatrix}; \quad (2.16)$$

$$r_{01.2} = -\frac{R_{01}}{\sqrt{R_{00} R_{11}}}; \quad r_{02.1} = -\frac{R_{02}}{\sqrt{R_{00} R_{22}}}; \quad r_{12.0} = -\frac{R_{12}}{\sqrt{R_{11} R_{22}}} \quad (2.17)$$

де R_{kj} – алгебраїчні доповнення до елемента r_{kj} $\left(k = \overline{0, 2},\ j = \overline{0, 2}\right)$ матриці $[R_2]$.

VII. Знаходження незміщеної оцінки дисперсії залишків.

Незміщену оцінку дисперсії залишків розрахуємо за формулою:

$$\sigma_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-(m+1)} = \frac{1}{n-m-1}\cdot e'\cdot e = \frac{Y'Y - A'X'Y}{n-m-1}. \qquad (2.18)$$

У знаменнику даної формули міститься число ступенів вільності $n-(m+1)$, де n – обсяг вибірки, m – число пояснюючих змінних.

$$\sum_{i=1}^{n} e_i^2 = e'\cdot e = Y'\cdot Y - A'\cdot X'\cdot Y. \qquad (2.19)$$

Обчислимо добуток $Y'\cdot Y$:

$$Y'\cdot Y = \begin{bmatrix}1{,}2 & 1{,}5 & 1{,}9 & 2{,}2 & 2{,}8 & 3{,}1 & 3{,}4 & 4{,}3 & 4{,}8 & 5{,}4\end{bmatrix} \times \begin{bmatrix}1{,}2\\1{,}5\\1{,}9\\2{,}2\\2{,}8\\3{,}1\\3{,}4\\4{,}5\\4{,}8\\5{,}4\end{bmatrix} = 113{,}6.$$

$$A = \begin{bmatrix}-0{,}97\\1{,}4\\-0{,}36\end{bmatrix}.$$

$$A' = \begin{bmatrix}-0{,}97 & 1{,}4 & -0{,}36\end{bmatrix}.$$

Значення добутку $(X'\cdot Y)$ візьмемо раніше обчисленим:

$$X'\cdot Y = \begin{bmatrix}30{,}8\\141{,}84\\152{,}77\end{bmatrix}.$$

Тоді

$$A'\cdot X'\cdot Y = \begin{bmatrix}-0{,}97 & 1{,}4 & -0{,}36\end{bmatrix} \times \begin{bmatrix}30{,}8\\141{,}84\\152{,}77\end{bmatrix} = 112{,}175.$$

Звідси

$$\sum_{i=1}^{10} e_i^2 = 113{,}6 - 112{,}175 = 1{,}1425 \approx 1{,}4.$$

Отже, незміщена оцінка дисперсії залишків буде:

$$\sigma_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-m-1} = \frac{1,4}{10-2-1} = 0,2\,.$$

VIII. Провести оцінку якості економетричної моделі з допомогою процедури стандартних помилок оцінок параметрів моделі.

Стандартні помилки оцінок параметрів моделі обчислимо за формулою:

$$S_{a_j} = \sqrt{\hat{\sigma}^2_{a_j}},\ j = \overline{0,m}. \tag{2.20}$$

$$\hat{\sigma}^2_{a_j} = \sigma_e^2 \cdot C_{jj}\,, \tag{2.21}$$

де $\sigma_e^2 = \frac{Y' \cdot Y - A'X'Y}{n-m-1}$, а C_{jj} – елементи головної діагоналі матриці *C*.

Використаємо раніше знайдені значення $\left(X' \cdot X\right)^{-1}$.

$$C = \left(X' \cdot X\right)^{-1},\ \left[C\right] = \begin{bmatrix} 8,92 & 1,22 & -2,93 \\ 1,22 & 0,4 & -0,61 \\ -2,93 & -0,61 & 1,51 \end{bmatrix}.$$

Тоді

$$C_{00} = 8,92;\ C_{11} = 0,4;\ C_{22} = 1,51.$$

Значення виразу $Y' \cdot Y - A'X'Y$ ми вирахували:

$Y' \cdot Y - A'X'Y = 1,4$.

Нами знайдено $\sigma_e^2 = \frac{1,4}{10-2-1} = \frac{1,4}{7} = 0,2,$ тоді дисперсії оцінок a_j буде:

$$\hat{\sigma}^2_{a_0} = 0,2 \cdot 8,92 = 1,78; \qquad S_{a_0} = \sqrt{1,78} = 1,33,$$

$$\hat{\sigma}^2_{a_1} = 0,2 \cdot 0,4 = 0,08; \qquad S_{a_1} = \sqrt{0,08} = 0,28,$$

$$\hat{\sigma}^2_{a_2} = 0,2 \cdot 1,51 = 0,3; \qquad S_{a_2} = \sqrt{0,3} = 0,55.$$

Далі знайдемо відносний показник розсіювання δ_{aj}, тобто порівняємо кожну стандартну помилку S_{aj} з відповідним числовим значенням оцінки параметра:

$$\delta_{a_0} = \frac{S_{a_0}}{|a_0|} \cdot 100\% = \frac{1,33}{|-0,97|} = 137,11\%;$$

$$\delta_{a_1} = \frac{S_{a_1}}{|a_1|} \cdot 100\% = \frac{0,28}{|1,4|} = 20,0\%; \tag{2.22}$$

$$\delta_{a_2} = \frac{S_{a_2}}{|a_2|} \cdot 100\% = \frac{0,55}{|-3,7|} = 14,86\%.$$

Отже, стандартні помилки оцінок параметрів відносно рівня оцінок параметрів становлять відповідно 137,11 %, 20,0 % та 14,86 %, а це є підтвердженням зміщеності оцінок.

IX. Перевірка значущості економетричної моделі.

Для перевірки адекватності множинної регресійної моделі, як і у випадку парної регресії, використовується F-критерій Фішера. У даному випадку нульова гіпотеза узагальнюється:

$$H_0 : a_1 = a_2 = \ldots = a_m = 0.$$

Тоді альтернативною гіпотезою буде H_1: хоча б одне значення a_j відмінне від нуля. У випадку невиконання гіпотези H_0 приймається гіпотеза H_1. Отже, не всі параметри незначною мірою відрізняються від нуля. Це свідчить про те, що включені до моделі фактори пояснюють змінну результативного показника.

Для перевірки гіпотези H_0 використовують F-критерій Фішера, з $(m–1)$ та $(n–m–1)$ ступенями вільності:

$$F = F_{m-1,n-m-1} = \frac{\dfrac{\sum\limits_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2}{(m-1)}}{\dfrac{\sum\limits_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{(n-m-1)}} = \frac{\sum\limits_{i=1}^{n}\left(\hat{y}_i - \bar{y}\right)^2 \cdot (n-m-1)}{\sum\limits_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 \cdot (m-1)}, \qquad (2.23)$$

де m – кількість незалежних факторів, які включено до моделі разом із фіктивним, n – загальна кількість спостережень.

Має місце альтернативне представлення даного показника у матричній формі:

$$F_{m-1,n-m-1} = \frac{A'X'Y(n-m-1)}{(Y'Y - A'X'Y)(m-1)}. \qquad (2.24)$$

Далі для заданого рівня значущості α і ступенів вільності $k_1=m–1$ та $k_2=n–m–1$ знаходимо табличне значення критерія Фішера – $F_{табл.}(k_1,k_2,\alpha)$. Знайдене розрахункове значення критерія $F_{m-1,n-m-1} = F_{розр.}$ порівнюємо з табличним: якщо $F_{розр} > F_{табл.}$, тоді гіпотеза H_0 відхиляється і приймається альтернативна, що свідчить про адекватність побудованої моделі, іншими словами підтверджується наявність істотного зв’язку між залежною та незалежними змінними побудованої економетричної моделі. У протилежному випадку нульова гіпотеза приймається, і модель вважається неадекватною.

Для нашого випадку висуваємо нульову гіпотезу $H_0 : a_1 = a_2 = 0.$

Обчислимо розрахункове значення критерію за формулою:

$$F_{розр} = \frac{A'X'Y(n-m-1)}{(Y'Y - A'X'Y)(m-1)}. \qquad (2.25)$$

Значення виразів $A'X'Y$ та $Y'Y - A'X'Y$ візьмемо із раніше виконаних розрахунків:

$$A'X'Y = 112{,}175; \quad YY' - A'X'Y = 1{,}4.$$

Тоді $F_{розр} = \frac{112{,}175(10-2-1)}{1{,}4(2-1)} = \frac{112{,}175 \cdot 7}{1{,}4} = 80{,}125$.

Для рівня значущості α=0,05 і ступенів вільності $k_1 = m-1 = 2-1$ та $k_2 = m-n-1 = 10-2-1 = 7$ знаходимо табличне значення F-критерію Фішера

$$F_{табл} = (1;\ 7;\ 0{,}05) = 5{,}59.$$

Оскільки $F_{розр} > F_{табл}$,, то гіпотеза H_0 відхиляється і приймається H_1, а це свідчить про адекватність побудованої моделі, тобто існує істотний зв'язок між залежною та незалежними змінними побудованої економетричної моделі.

X. Перевірка значущості коефіцієнта детермінації.

Якість моделей множинної регресії можна оцінювати з допомогою коефіцієнтів детермінації.

При реалізації процедури перевірки значущості коефіцієнта детермінації висувається гіпотеза H_0 проти альтернативної H_1, зміст яких полягає в наступному.

H_0: суттєвої різниці між вибірковим коефіцієнтом детермінації та коефіцієнтом детермінації генеральної сукупності $R^2_{ген} = 0$ немає. Дана гіпотеза рівносильна гіпотезі $H_0 : a_1 = a_2 = \ldots = a_m = 0$, тобто жодна із пояснюючих змінних, включених у модель, не виявляють суттєвого впливу на залежну змінну.

H_1: вибірковий коефіцієнт детермінації суттєво більший від коефіцієнта детермінації генеральної сукупності $R^2_{ген} \neq 0$. Прийняття гіпотези H_1 означає, що хоча б одна із m пояснюючих змінних, включених до моделі, виявляє суттєвий вплив на результативний показник.

Для оцінки значущості множинного коефіцієнта детермінації, як і у випадку парної регресії, використаємо статистику F-критерія Фішера.

Між F-критерієм Фішера та множинним коефіцієнтом детермінації існує зв'язок:

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m-1}. \quad (2.26)$$

Вираз (2.26) показує, що коли $R^2 = 0$ то $F = 0$. Як бачимо, що між F та R^2 існує взаємозв'язок. Оскільки F-критерій є мірою адекватності регресійної моделі, то він є мірою значущості коефіцієнта множинної детермінації. Причому, при зростанні R^2, значення F-критерія буде зростати.

Знайдемо $F_{розр.} = \frac{0,9836}{1-0,9836} \cdot 7 = 419,8293$.

Оскільки $F_{розр} > F_{табл}$, то знайдений коефіцієнт детермінації $R^2 = 0,9836$ є значимий і приймається гіпотеза H_1.

XI. Перевірка значущості коефіцієнта кореляції.

Відомо, що коефіцієнт множинної кореляції є вибірковою характеристикою, тому при перевірці якості побудованої моделі доцільно провести оцінку його значущості. Дана оцінка ґрунтується на t-статистиці Ст'юдента:

$$t_{розр} = \frac{|R|\sqrt{n-m-1}}{\sqrt{1-R^2}}, \quad (2.27)$$

де R – коефіцієнт множинної кореляції; R^2 – коефіцієнт детермінації; $(n–m–1)$ – число ступенів вільності. Обчислене значення t-статистики за формулою (2.27) порівнюють з критичним значенням $t_{k,\alpha}$ знайденим за таблицею t-розподілу для рівня значущості α і числа ступенів вільності $k=n–m–1$.

Прийняття чи відхилення гіпотези про значущість коефіцієнта множинної кореляції проводиться за тими ж правилами, що і для випадку парної регресії.

Якщо $|t_{розр}| > t_{k,\alpha}$, то гіпотеза H_0 відхиляється і приймається гіпотеза H_1. У даному випадку можна зробити висновок про значущість коефіцієнта множинної кореляції між залежною та незалежними змінними. У протилежному випадку приймається нульова гіпотеза.

Значущість коефіцієнта множинної кореляції можна також оцінювати на основі проведення процедури перевірки значущості коефіцієнта детермінації, оскільки між ними є зв'язок: $R = \sqrt{R^2}$.

Знайдемо:

$$t_{розр} = \frac{0{,}9918}{0{,}1281} \cdot 2{,}645751 = 20{,}4844\,.$$

Табличне значення параметра t знаходимо з таблиці при α=0,05 та $k = m - n - 1 = 7$:

$$t_{табл.}(7;0{,}05) = 2{,}36\,.$$

Оскільки $t_{розр.} > t_{табл.}$, то гіпотеза H_0 відхиляється і приймається гіпотеза H_1, що свідчить про значущість коефіцієнта множинної кореляції.

XII. Перевірка значущості оцінок параметрів моделі.

Для розгляду значущості знайдених оцінок параметрів багатофакторної моделі побудуємо наступні гіпотези:

$H_0 : a_j = b_j$, що вказує на відсутність суттєвої різниці між оцінкою параметра регресії, отриманої за результатами вибірки, і дійсним значенням b_j (параметра регресії генеральної сукупності);

$H_1 : a_j \neq b_j$, що вказує на наявність суттєвої різниці між оцінкою параметра регресії та відповідним параметром генеральної сукупності.

Альтернативна гіпотеза може бути сформульованою таким чином:

$H_1 : a_j > b_j$ або $a_j < b_j$, тобто оцінка параметра суттєво більша або суттєво менша від параметра генеральної сукупності.

Для прийняття відповідних гіпотез використовується t-критерій Ст'юдента:

$$t_{розр} = t_j = \frac{|a_j - b_j|}{\sigma_{a_j}}, \; при \; k = n - m - 1, \qquad (2.28)$$

де a_j – оцінка параметра b_j, отримана за методом найменших квадратів; σ_{a_j} – середньоквадратичне відхилення оцінки j-го параметра; k – число ступенів вільності.

Обчислене значення t_j, порівнюється із критичним значенням $t_{k,\alpha}$, знайденим за таблицями при заданому рівні значущості α і числом степенів вільності k. Якщо $t_j > t_{k,\alpha}$, то a_j значно відрізняється від b_j, тобто не можна припускати, що вибірка взята з генеральної сукупності з параметром регресії b_j.

На практиці буває дуже складно вказати завчасно числове значення параметра регресії b_j генеральної сукупності, тому часом приходиться висувати інше припущення:

$H_0 : a_j = 0$, тобто пояснююча змінна x_j не виявляє суттєвого впливу на залежну змінну *y;*

$H_1 : a_j \neq 0$, тобто змінна x_j виявляє суттєвий вплив на *y*. У даному випадку для перевірки нульової гіпотези використовується *t*-статистика:

$$t_{розр} = t_j = \frac{|a_j|}{\sigma_{a_j}}, \tag{2.29}$$

яка має *k* ступенів вільності.

У матричній формі (2.29) прийме наступний вид:

$$t_j = \frac{|a_j|}{\sqrt{\sigma_e^2 C_{jj}}}, \ j = \overline{1,m}, \tag{2.30}$$

де C_{jj} – діагональний елемент матриці $(X'X)^{-1}$; σ_e^2 – дисперсія залишків.

Величина $\sigma_{a_j} = \sqrt{\sigma_e^2 C_{jj}}$ називається стандартною оцінкою *j*-го параметра моделі.

Знайдене за (2.30) значення t_j порівнюють із значенням $t_{k,\alpha}$. Якщо $t_j > t_{k,\alpha}$, то відповідна оцінка параметра економетричної моделі є достовірною.

Для нашого випадку нульова гіпотеза буде:

$$H_0 : a_j = 0; \ j = 0;1;2.$$

Перевірку цієї гіпотези здійснюємо за допомогою критерію Ст'юдента.

Використовуючи формулу (2.29) обчислимо розрахункове значення *t* для:

– параметра a_0:

$$t_0 = t_{розр} = \frac{|a_0|}{\sigma_{a_0}}.$$

Значення a_0, a_1, a_2 та $\sigma_{a_0}, \sigma_{a_1}, \sigma_{a_2}$ беремо із попередніх розрахунків $(a_0 = -0{,}97; \sigma_{a_0} = 1{,}33)$.

$$t_0 = t_{розр} = \frac{0{,}97}{1{,}33} = 0{,}73.$$

Табличне значення параметра *t* беремо з таблиці для рівня значимості $\alpha = 0{,}05$ та числа степенів вільності $k = n - m - 1 = 10 - 2 - 1 = 7$.

$$t_{табл} = t_{k,\alpha} = 2{,}36.$$

Оскільки $t_{розр} < t_{табл}$, то гіпотеза про рівність нулю параметра a_0 в генеральній сукупності не відхиляється.

– параметра a_1:

$$t_1 = t_{розр} = \frac{|a_1|}{\sigma_{a_1}},$$

$a_1 = 1{,}4$, $\sigma_{a_1} = 0{,}28$;

$$t_1 = t_{розр} = \frac{1{,}4}{0{,}28} = 5{,}00;$$

$$t_{табл} = 2{,}36.$$

Оскільки $t_{розр} > t_{табл}$, то нульова гіпотеза про рівність нулю параметра a_1 в генеральній сукупності відхиляється і оцінка параметра a_1 економетричної моделі є достовірною.

– параметра a_2:

$$t_2 = t_{розр} = \frac{|a_2|}{\sigma_{a_2}},$$

$a_2 = -0{,}37$, $\sigma_{a_2} = 0{,}55$;

$$t_2 = t_{розр} = \frac{|-0{,}37|}{0{,}55} = 0{,}67.$$

Оскільки $t_{розр} > t_{табл}$, то нульова гіпотеза про рівність нулю параметра a_2 в генеральній сукупності не відхиляється. Отже, тільки коефіцієнт a_1 в даній моделі є достовірним.

XIII. Знаходження довірчих інтервалів параметрів моделі.

Довірчі інтервали для параметрів b_j побудуємо на основі формули:

$$b_j = a_j \pm t_j \sqrt{\sigma_e^2 C_{jj}}. \tag{2.31}$$

Довірчі інтервали параметрів b_j економетричної моделі представлені у вигляді:

$$\left[a_j - t_j \sqrt{\sigma_e^2 c_{jj}}; a_j + t_j \sqrt{\sigma_e^2 c_{jj}}\right].$$

Значення a_0, a_1, a_2 розраховані раніше:

$$a_0 = -0{,}97; \quad a_1 = 1{,}4; \quad a_2 = -0{,}37.$$

Значення t_j теж нами розраховано:

$$t_0 = 0{,}73; \quad t_1 = 5{,}00; \quad t_2 = 0{,}67.$$

Значення виразів $\sqrt{\sigma_e^2 c_{jj}}$ нами знайдено:

$$\sqrt{\sigma_e^2 c_{00}} = 1{,}33; \quad \sqrt{\sigma_e^2 c_{11}} = 0{,}28; \quad \sqrt{\sigma_e^2 c_{22}} = 0{,}55.$$

Тоді довірчий інтервал:

1) параметра b_0:

$$[-0{,}97-0{,}73\cdot 1{,}33;-0{,}97+0{,}73\cdot 1{,}33],\ [-1{,}9409;0{,}0009];$$

2) параметра b_1:

$$[1{,}4-5\cdot 0{,}28;1{,}4+5\cdot 0{,}28],\ [0;2{,}8];$$

3) параметра b_2:

$$[-0{,}37-0{,}67\cdot 0{,}55;-0{,}37+0{,}67\cdot 0{,}55],\ [-0{,}7385;-0{,}0015].$$

XIV. Знаходження довірчого інтервалу коефіцієнта кореляції.

При побудові довірчого інтервалу для коефіцієнта кореляції генеральної сукупності ρ необхідно використати перетворення Фішера, завдяки якому розподіл параметра r може бути наближено приведений до нормального:

$$Z = 0{,}5\ln\frac{1+r}{1-r} = 1{,}1513\lg\frac{1+r}{1-r}. \qquad (2.32)$$

Значення σ_Z знаходимо з формули:

$$\sigma_Z = \frac{1}{\sqrt{n-3}}, \qquad (2.33)$$

де n – обсяг вибірки.

Довірчий множник в даному випадку є квантилем стандартного нормального розподілу $\lambda\alpha$. Тоді довірчі границі для величини Z при рівні значущості α будуть $Z \pm \sigma_Z$, а довірчий інтервал $[Z-\lambda_\alpha\sigma_Z;\, Z+\lambda_\alpha\sigma_Z]$.

Для рівня значущості α=0,05 квантиль стандартного нормального розподілу $\lambda_{0,05}$=1,96.

Обчислимо значення Z за формулою (2.32):

$$Z = 0{,}5\ln\frac{1+r}{1-r}.$$

Нами знайдено значення коефіцієнта кореляції r=0,9918. Тоді, використовуючи формулу (2.32), маємо:

$$Z = 0{,}5\ln\frac{1+0{,}9918}{1-0{,}99918} = 0{,}5\ln\frac{1{,}9897}{0{,}0103} = 0{,}5\cdot\ln 242{,}9 = 0{,}5\cdot 5{,}49 = 2{,}745.$$

Значення σ_Z знаходимо за формулою (2.33)

$$\sigma_z = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{10-3}} = \frac{1}{\sqrt{7}=} = \frac{1}{2{,}646} = 0{,}38.$$

Знайдемо довірчі границі для величини Z: $[Z-\lambda_\alpha\sigma_Z;\, Z+\lambda_\alpha\sigma_Z]$

Оскільки $\lambda\alpha$=1,96, тоді довірчий інтервал для Z:

$$[2{,}745-1{,}96\cdot 0{,}38;\ 2{,}745+1{,}96\cdot 0{,}38], \text{ або } [2.002;\ 3{,}4898].$$

Обчислимо значення коефіцієнта кореляції r для нижньої та верхньої меж Z. Враховуючи, що $\tanh x = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$, одержимо:

нижня межа: Z=2,002, тоді $r = \tanh 2{,}002 = \dfrac{e^{2{,}002} - e^{-2{,}002}}{e^{2{,}002} + e^{-2{,}002}} = 0{,}964$.

верхня межа: Z=3,4898, тоді $r = \tanh 3{,}4898 = \dfrac{e^{3{,}4898} - e^{-3{,}4898}}{e^{3{,}4898} + e^{-3{,}4898}} = 0{,}998$.

Отримаємо довірчий інтервал для коефіцієнта кореляції: [0,964; 0,998].

XV. Розрахунок прогнозних значень.

Нехай нам задано прогнозні показники:

$$X_П = \begin{bmatrix} 1 \\ 5{,}2 \\ 4{,}5 \end{bmatrix},$$

для нашої еконoметричної моделі $y = -0{,}97 + 1{,}4x_1 - 0{,}37x_2 + e$.

Знайдемо прогнозне значення прибутку:

$$\hat{y}_П = -0{,}97 + 1{,}4 \cdot 5{,}2 - 0{,}37 \cdot 4{,}5 = 4{,}645.$$

Вираз $\hat{Y}_П = X_П \cdot A$ розглядаємо в якості точкової оцінки математичного сподівання прогнозного значення $y_П$, а також як індивідуальне значення $Y_П$ для $X_П$.

Визначимо довірчий інтервал математичного сподівання прогнозних значень для рівня значущості $\alpha = 0{,}05$ за наступною формулою:

$$\hat{y}_П - t_\alpha \sigma_e \sqrt{X'_П [X'X]^{-1} X_П} \le M(y_П) \le \hat{y}_П + t_\alpha \sigma_e \sqrt{X'_П [X'X]^{-1} X_П} \qquad (2.34)$$

Враховуючи знайдені раніше значення, маємо:

$$\sigma_e = \sqrt{0{,}2} = 0{,}45,\ t_\alpha = 2{,}365,\ [X'X]^{-1} = \begin{bmatrix} 8{,}92 & 1{,}22 & -2{,}93 \\ 1{,}22 & 0{,}4 & -0{,}61 \\ -2{,}93 & -0{,}61 & 1{,}51 \end{bmatrix}.$$

Знайдемо

$$X'_П \cdot [X'X]^{-1} X_П = [1 \quad 5{,}2 \quad 4{,}5] \cdot \begin{bmatrix} 8{,}92 & 1{,}22 & -2{,}93 \\ 1{,}22 & 0{,}4 & -0{,}61 \\ -2{,}93 & -0{,}61 & 1{,}51 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 5{,}2 \\ 4{,}5 \end{bmatrix} = 8{,}0835,$$

$$\sqrt{X'_П \cdot [X'X]^{-1} X_П} = \sqrt{8{,}0835} = 2{,}8435.$$

Тоді

$$4{,}645 - 2{,}365 \cdot 0{,}45 \cdot 2{,}8435 \le M(y_П) \le 4{,}645 + 2{,}365 \cdot 0{,}45 \cdot 2{,}8435$$

$$1{,}618805 \le M(y_П) \le 7{,}671195.$$

Інтервальний прогноз індивідуального значення $\hat{y}_П$ знайдемо з формули:

$$\hat{y}_П - t_\alpha \sigma_e \sqrt{1 + X'_П [X'X]^{-1} X_П} \le y_П \le \hat{y}_П + t_\alpha \sigma_e \sqrt{1 + X'_П [X'X]^{-1} X_П}, \qquad (2.35)$$

$$4{,}645 - 2{,}365 \cdot 0{,}45 \cdot \sqrt{1 + 8{,}0835} \le y_П \le 4{,}645 + 2{,}365 \cdot 0{,}45 \cdot \sqrt{1 + 8{,}0835},$$

$$1{,}437473 \le y_П \le 7{,}852572.$$

Отже, з імовірністю $P = 0{,}95$ прогноз математичного сподівання потрапляє в інтервал $[1{,}618805; 7{,}671195]$, а прогноз індивідуального значення в інтервал $[1{,}437473; 7{,}852572]$.

Економічний зміст полягає в наступному: при встановленні у прогнозному періоді розмір основних фондів 5,2 млн. грн., а обсяг затрат праці на рівні 4,5 млн. днів, то середній прибуток потрапляє в інтервал $[1{,}62; 7{,}67]$, а окреме значення прибутку знаходиться в інтервалі $[1{,}44; 7{,}85]$.

XVI. Дослідження з допомогою нелінійних моделей.

Припустимо, що між дослідженими факторами існує залежність у вигляді функції Кобба-Дугласа:

$$\hat{y} = a_0 x_1^{a_1} x_2^{a_2} \qquad (2.36)$$

де $\hat{y}$ – прибуток, млн. грн.; x_1 –вартість основних виробничих фондів підприємства, млн. грн.; x_2 – затрати праці, млн. днів.

Для проведення даної залежності до лінійного виду треба виконати дію логарифмування за основою e:

$$\ln \hat{y} = \ln a_0 + a_1 \ln x_1 + a_2 \ln x_2. \qquad (2.37)$$

Зробивши заміну $y^* = \ln \hat{y}$, $a_0^* = \ln a_0$, $x_1^* = \ln x_1$, $x_2^* = \ln x_2$, перейдемо до лінійної форми зв'язку:

$$y^* = a_0^* + a_1 x_1^* + a_2 x_2^*. \qquad (2.38)$$

Нам необхідно знайти значення натурального логарифма $\hat{y}$, x_1 та x_2 (рис. 2.5). Для цього скористаємось функцією LN(число) в табличному процесорі MS Excel і занесемо знайдені значення до таблиці.

	A	B	C	D	E	F	G
					=LN(B2)	=LN(C2)	=LN(D2)
1	№ підприєм.	$\hat{y}_i$	x_{i1}	x_{i2}	$\ln \hat{y}_i$	$\ln x_{i1}$	$\ln x_{i2}$
2	1	1,2	2,5	4	0,1823	0,9163	1,3863
3	2	1,5	2,8	4,2	0,4055	1,0296	1,4351
4	3	1,9	3	3,6	0,6419	1,0986	1,2809
5	4	2,2	3,6	4,6	0,7885	1,2809	1,5261
6	5	2,8	3,9	4,3	1,0296	1,3610	1,4586
7	6	3,1	4,2	5,1	1,1314	1,4351	1,6292
8	7	3,4	4,5	5,3	1,2238	1,5041	1,6677
9	8	4,5	5	4,8	1,5041	1,6094	1,5686
10	9	4,8	5,6	5,4	1,5686	1,7228	1,6864
11	10	5,4	6	5,8	1,6864	1,7918	1,7579
12	Всього	30,8	41,1	47,1			

Рис. 2.5. Фрагмент документу MS Excel, в якому здійснено обчислення натуральних логарифмів $\hat{y}$, x_1 та x_2

Для знаходження оціночних параметрів використаємо інструмент “Аналіз даних” в MS Excel. Провівши регресійний аналіз отримаємо потрібні нам значення параметрів (рис. 2.6).

	A	B	C	D
1		ВЫВОД ИТОГОВ		
2				
3		*Регрессионная статистика*		
4		Множественный R	0,998	
5		R-квадрат	0,995	
6		Нормированный R-квад	0,994	
7		Стандартная ошибка	0,040	
8		Наблюдения	10	
9				
10			*Коэффициенты*	*Стандартная ошибка*
11		Y-пересечение	-0,7325	0,183
12		Переменная X 1	1,9914	0,094
13		Переменная X 2	-0,6425	0,186

Рис. 2.6. Фрагмент документу MS Excel, в якому здійснено регресійний аналіз для знаходження значень параметрів

Отже, $a_0^* = -0{,}7325$, $a_0 = e^{a_0} = e^{-0{,}7325} = 0{,}48$, $a_1 = 1{,}9914$, $a_2 = -0{,}6425$, при $R = 0{,}998$, $R^2 = 0{,}995$.

Отже, нами отримано наступну залежність у вигляді функції Кобба-Дугласа:

$$\hat{y} = 0{,}48 \cdot x_1^{1{,}9914} x_2^{-0{,}6425},$$

параметр $a_1 = 1{,}9914$ показує, що при збільшенні вартості основних виробничих фондів на 1% і закріпленні на середньому рівні впливу другого фактору забезпечується приріст прибутку на 1,9914 %. Враховуючи від'ємний знак другого параметра $a_2 = -0{,}6425$, маємо протилежне твердження. Так при збільшенні на 1% затрат праці на 0,6425 % зменшиться прибуток.

Завдання 3. Дослідження наявності мультиколінеарності.

Порушення четвертої передумови застосування МНК породжує явище мультиколінеарності, тобто має місце лінійного зв'язку між пояснюючими змінними.

Однією з ознак наявності мультиколінеарності між вибраними змінними є наближення до одиниці коефіцієнтів парної кореляції. Математично це означає, що $\left|r_{x_i x_j}\right| \to 1, i \neq j$, де $r_{x_i x_j}$ – коефіцієнт парної кореляції між x_i та x_j.

Для діагностування мультиколінеарності можна використати наступні характерні ознаки:

1. У першу чергу, аналізуючи матрицю r парних коефіцієнтів кореляції, вірніше ту її частину, яка відноситься до пояснюючих змінних, вважається, що наявність значень парних коефіцієнтів, за абсолютною величиною більших 0,75-0,80, або їх наближення до коефіцієнта множинної кореляції, свідчить про явище мультиколінеарності.

$$r = \begin{bmatrix} r_{x_1 x_1} & r_{x_1 x_2} & \dots & r_{x_1 x_m} \\ r_{x_2 x_1} & r_{x_2 x_2} & \dots & r_{x_2 x_m} \\ \vdots & \vdots & \vdots & \vdots \\ r_{x_m x_1} & r_{x_m x_2} & \dots & r_{x_m x_m} \end{bmatrix}. \tag{3.1}$$

2. Існування тісних лінійних статистичних зв'язків між пояснюючими змінними приводить до так званої слабої обумовленості матриці $X'X$, тобто близькості до нуля її визначника. Тому, якщо значення $\det(X'X)$ прямує до нуля, то це свідчить про наявність мультиколінеарності.

3. Аналіз кореляційної матриці r дає можливість лише у першому наближенні (відносно поверхнево) робити висновок про наявність або відсутність мультиколінеарності у вхідних даних. Більшої достовірності можна домогтися з допомогою процедури розрахунку значень коефіцієнтів часткової детермінації кожної пояснюючої змінної X_j з іншими змінними $X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_m$. Якщо значення R_j^2 наближається до одиниці, то це означає можливе існування мультиколінеарності.

4. Якщо в економетричній моделі більшість значень оцінюваних параметрів виявилися статистично незначимо відмінними від нуля

при високому рівні часткових коефіцієнтів детермінації R_j^2, і при цьому значення F-критерію значно відрізняється від нуля, то це вказує на наявність мультиколінеарності.

5. Якщо при побудові економетричної моделі методом покрокової регресії після деякого кроку істотно змінилося значення оціночних параметрів моделі при незначному підвищенні (чи зниженні) коефіцієнтів кореляції або детермінації, то включена на даному кроці змінна перебуває, очевидно, у лінійній залежності від інших, які введені до моделі раніше.

6. Про наявність мультиколінеарності свідчать деякі зовнішні ознаки побудованої моделі, що є її наслідком. До них у першу чергу, необхідно віднести:

- деякі оцінки параметрів a_j мають невірні з точки зору економічної теорії знаки чи невиправдано великі за абсолютною величиною значення;
- невелика зміна вхідних даних (додавання чи відкидання невеликої частки спостережень) приводить до суттєвої зміни оцінок коефіцієнтів моделі, в тому числі до можливої зміни знаків.

7. Для кількісної оцінки рівня мультиколінеарності можна використати розраховані характеристичні значення коваріаційної матриці *Cov(X,X)* та її умовний індекс.

Дані показники розраховуємо за наступними формулами:

$$k = \frac{\lambda \max}{\lambda \min}, \quad k_{ін} = \sqrt{k}, \tag{3.2}$$

де $\lambda \max, \lambda \min$ – відповідно найбільше та найменше власне значення коваріаційної матриці; k – характеристичне значення коваріаційної матриці; $k_{ін}$ – умовний індекс.

Якщо $0 \leq k \leq 1000$ ($0 \leq k_{ін} \leq 30$), то це свідчить про помірну мультиколінеарність, при $k \geq 1000$ ($k_{ін} \geq 30$) маємо високу мультиколінеарність.

На превеликий жаль, розглянуті вище ознаки мультиколінеарності не дають повної відповіді на запитання: коли мультиколінеарність – істина, і коли нею можна знехтувати.

Для більш повного вирішення проблеми мультиколінеарності використаємо метод Феррара-Глобера, основу якого складають три види статистичних критеріїв.

З допомогою цих статистичних критеріїв перевіряється мультиколінеарність не тільки всього масиву незалежних змінних, але й кожної пари зокрема. Мультиколінеарність всього рівняння встановлюється з допомогою критерія χ^2, знайденого на основі оцінок взаємної кореляційної матриці незалежних змінних. Для перевірки мультиколінеарності кожної незалежної змінної з усіма іншими обчислюється статистика F-критерію, а для кожної пари незалежних змінних - t-критерій.

Перейдемо до викладу алгоритму Феррара-Глобера з допомогою наступних кроків.

1. Нормалізація змінних.

Нехай $\{x_j,\ j=\overline{1,m}\}$ вектор j-ої незалежної змінної. Для обчислення нормалізованих значень векторів використаємо формулу:

$$x_{ij}^{*}=\frac{x_{ij}-\overline{x}_j}{\sigma_{x_j}\sqrt{n}},\quad i=\overline{1,n},\ j=\overline{1,m}, \tag{3.3}$$

де n – число спостережень кожного вектора; m – число незалежних змінних; $\overline{x}_j$ – середнє арифметичне вектора x_j,, $\overline{x}_j=\frac{\sum_{i=1}^{n}x_{ij}}{n}$; σ_{x_j} – стандартна помилка незалежної змінної x_j, $\sigma_{x_j}=\sqrt{\frac{\sum_{i=1}^{n}\left(x_{ij}-\overline{x}_j\right)^2}{n}}$.

У результаті виконання процедури нормалізації отримаємо матрицю, побудовану з нормалізованих векторів. Позначимо таку матрицю через X^{*}, а транспоновану до неї через – $X^{*'}$. Тобто, маємо

$$X^{*}=\begin{bmatrix} x_{11}^{*} & x_{12}^{*} & \dots & x_{1m}^{*}\\ x_{21}^{*} & x_{22}^{*} & \dots & x_{2m}^{*}\\ \vdots & \vdots & \vdots & \vdots\\ x_{n1}^{*} & x_{n2}^{*} & \dots & x_{nm}^{*}\end{bmatrix}. \tag{3.4}$$

2. Розрахунок кореляційної матриці.

Взаємна кореляційна матриця оцінюється із співвідношення

$$r_s=\frac{1}{n-1}\cdot\left[X^{*'}\cdot X^{*}\right]. \tag{3.5}$$

Дана матриця симетрична й має розмір $m\times m$.

3. Обчислення визначника кореляційної матриці та значення критерія χ^2.

Визначаємо визначник матриці r_s:

$$|r_s|=\det r_s\ , \tag{3.6}$$

де $|r_s|$ – визначник кореляційної матриці r_s..

Значення критерія χ^2 визначаємо за формулою:

$$\chi^2 = -\left[n - 1 - \frac{1}{6}(2m + 5)\right]\ln|r_s|. \tag{3.7}$$

Дана величина порівнюється із значенням χ^2 для $\frac{1}{2}m(m-1)$ ступенями вільності та рівня значущості α. Якщо $\chi^2_{факт.} > \chi^2_{табл.}$, то у множині статистичних даних існує мультиколінеарність.

4. Визначення матриці, оберненої до r_s.

Позначимо обернену матрицю через C. Тоді маємо:

$$C = r_s^{-1} = \begin{bmatrix} c_{11} & \dots & c_{1m} \\ \vdots & \vdots & \vdots \\ c_{m1} & \dots & c_{mm} \end{bmatrix}. \tag{3.8}$$

5. Обчислення значення F-критеріїв.

Для кожної незалежної величини знаходимо значення F-критерія за формулою:

$$F_j = (C_{jj} - 1)\frac{n - m}{m - 1},\ j = \overline{1, m}, \tag{3.9}$$

де C_{jj} - діагональні елементи матриці C. Знайдені значення F_j порівнюються з критичними величинами F-критерія при $n–m$ та $m–1$ ступенями вільності і при рівні значущості α. Якщо $F_j > F_{табл.}$, то відповідна j-та незалежна змінна мультиколінеарна з іншими.

Далі знаходимо коефіцієнт детермінації для кожної змінної x_j за формулою:

$$R^2_{x_j} = 1 - \frac{1}{c_{jj}},\ j = \overline{1, m}. \tag{3.10}$$

6. Знаходження часткових коефіцієнтів кореляції.

Ці коефіцієнти розраховуємо за формулою:

$$r_{jk} = \frac{-C_{jk}}{\sqrt{C_{kk} \cdot C_{jj}}},\ j = \overline{1, m},\ k = \overline{1, m},\ j \neq k, \tag{3.11}$$

де C_{jk} – позадіагональні елементи матриці C.

7. Обчислення t-критеріїв.

Для знаходження значень цих критеріїв використаємо формулу:

$$t_{jk} = r_{jk} \cdot \sqrt{\frac{n - m}{1 - r^2_{jk}}},\ j = \overline{1, m},\ k = \overline{1, m},\ j \neq k,$$

де t_{jk} – значення t-критеріїв між незалежними змінними x_j та x_k.

Далі фактичні значення t_{jk}-критеріїв порівнюються з табличними при $n-m$ ступенями вільності та рівні значущості α. Якщо знайдені значення $t_{jk} > t_{табл.}$, то між незалежними змінними x_j та x_k існує мультиколінеарність.

На завершення розгляду даного алгоритму приведемо висновки відносно його інтерпретації у прикладних дослідженнях.

1. Значення елементів матриці r_S дають нам первинну інформацію відносно мультиколінеарності незалежних змінних. Позадіагональні елементи рівні коефіцієнтам парної кореляції між незалежними змінними. Вони повинні бути меншими коефіцієнта множинної кореляції для загального рівняння.

2. Кількісною мірою загальної мультиколінеарності служить величина χ^2. Цей показник не повинен перевищувати табличного значення критерія χ^2 при рівні значущості α та $\frac{1}{2}m(m-1)$ ступенях вільності.

3. Величина F_j вказує на мультиколінеарність відповідних незалежних змінних. Якщо $F_j > F_{табл.}$, то відповідні їм незалежні фактори мультиколінеарні.

4. З допомогою значень t_{jk} можна уточнювати, які із пар незалежних факторів мультиколінеарні. Якщо деяке значення t_{jk} перевищує табличне значення критерія Стьюдента при $n-m$ ступенях вільності та рівні значущості α, то між парами x_j та x_k існує мультиколінеарність.

Приклад 3.1. Дослідити наявність мультиколінеарності між пояснюючими змінними, використовуючи алгоритм Феррара-Глобера. Для дослідження наведемо статистичну сукупність спостережень (табл. 3.1).

Таблиця 3.1.

Підприємство	Прибуток Y, тис. грн.	Вартість основних виробничих фондів X_1, тис. грн	Затрати трудових ресурсів X_2, тис. люд.-год.	Собівартість одиниці продукції X_3, грн.	Ритмічність виробництва X_4
1	200	110	70	1,23	0,6
2	550	140	90	0,95	0,8
3	800	215	115	0,55	0,7
4	920	310	240	0,4	0,72
5	980	300	300	0,38	0,5
6	300	80	65	1,55	0,6
7	155	92	50	1,38	0,4

8 9 10	164 172 1100	95 100 400	82 85 260	1,2 1,1 0,35	0,42 0,7 0,9

Розв'язування:

1.Обчислимо середні значення та стандартні відхилення пояснюючих змінних, користуючись табличним процесором MS Excel, зокрема функціями СРЗНАЧ та СТАНДОТКЛОН (рис. 3.1).

	A	B	C	D	E	F
1-3	Підприємство	Прибуток, Y, тис. грн.	Вартість основ-них виробничих фондів, X_1, тис. грн	Затрати тру-дових ресурсів, X_2, тис. люд.-год.	Собівартість одиниці продукції, X_3, грн.	Ритмічність виробництва, X4
4	1	200	110	70	1,23	0,6
5	2	550	140	90	0,95	0,8
6	3	800	215	115	0,55	0,7
7	4	920	310	240	0,4	0,72
8	5	980	300	300	0,38	0,5
9	6	300	80	65	1,55	0,6
10	7	155	92	50	1,38	0,4
11	8	164	95	82	1,2	0,42
12	9	172	100	85	1,1	0,7
13	10	1100	400	260	0,35	0,9
14	Середнє значення		184,2	135,7	0,909	0,634
15	Стандартне відхилення		114,80302	93,06754	0,45202	0,16167

=СРЗНАЧ(C4:C13) =СТАНДОТКЛОН(C4:C13)

Рис. 3.1. Фрагмент документу MS Excel, в якому обчислено середні значення та стандартні відхилення пояснюючих змінних

2. Нормалізуємо пояснюючі змінні. Обрахунки робимо в MS Excel, стандартна функція НОРМАЛИЗАЦИЯ (рис. 3.2).

	B	C	D	E	F
1-3	Прибуток, Y, тис. грн.	Вартість основ-них виробничих фондів, X_1, тис. грн	Затрати тру-дових ресурсів, X_2, тис. люд.-год.	Собівартість одиниці продукції, X_3, грн.	Ритмічність виробництва, X4
4	200	110	70	1,23	0,6
5	550	140	90	0,95	0,8
6	800	215	115	0,55	0,7
7	920	310	240	0,4	0,72
8	980	300	300	0,38	0,5
9	300	80	65	1,55	0,6
10	155	92	50	1,38	0,4
11	164	95	82	1,2	0,42
12	172	100	85	1,1	0,7
13	1100	400	260	0,35	0,9
14		184,2	135,7	0,909	0,634
15		114,80302	93,06754	0,45202	0,16167

=НОРМАЛИЗАЦИЯ(C4;C$14;C$15)

$X^* =$

	I	J	K	L
4	-0,6463	-0,7059	0,7101	-0,2103
5	-0,3850	-0,4910	0,0907	1,0268
6	0,2683	-0,2224	-0,7942	0,4082
7	1,0958	1,1207	-1,1261	0,5319
8	1,0087	1,7654	-1,1703	-0,8288
9	-0,9076	-0,7597	1,4181	-0,2103
10	-0,8031	-0,9208	1,0420	-1,4474
11	-0,7770	-0,5770	0,6438	-1,3237
12	-0,7334	-0,5448	0,4225	0,4082
13	1,8797	1,3356	-1,2367	1,6453

Рис. 3.2. Фрагмент документу MS Excel, в якому розраховано значення нормалізуючих пояснюючих змінних

$$X^* = \begin{bmatrix} -0{,}6463 & -0{,}7059 & 0{,}7101 & -0{,}2103 \\ -0{,}385 & -0{,}491 & 0{,}0907 & 1{,}0268 \\ 0{,}2683 & -0{,}2224 & -0{,}7942 & 0{,}4082 \\ 1{,}0958 & 1{,}1207 & -1{,}1261 & 0{,}5319 \\ 1{,}0087 & 1{,}7654 & -1{,}1703 & -0{,}8288 \\ -0{,}9076 & -0{,}7597 & 1{,}4181 & -0{,}2103 \\ -0{,}8031 & -0{,}9208 & 1{,}042 & -1{,}4474 \\ -0{,}777 & -0{,}577 & 0{,}6438 & -1{,}3237 \\ -0{,}7334 & -0{,}5448 & 0{,}4225 & 0{,}4082 \\ 1{,}8797 & 1{,}3356 & -1{,}2367 & 1{,}6453 \end{bmatrix}.$$

Транспонуємо нормалізовану матрицю X^* з допомогою функції ТРАНСП в MS Excel, отримуємо матрицю $X^{*'}$.

$$X^{*'} = \begin{bmatrix} -0{,}6463 & -0{,}385 & 0{,}2683 & 1{,}0958 & 1{,}0087 & -0{,}9076 & -0{,}8031 & -0{,}777 & -0{,}7334 & 1{,}8797 \\ -0{,}7059 & -0{,}491 & -0{,}2224 & 1{,}1207 & 1{,}7654 & -0{,}7597 & -0{,}9208 & -0{,}577 & -0{,}5448 & 1{,}3356 \\ 0{,}7101 & 0{,}0907 & -0{,}7942 & -1{,}1261 & -1{,}1703 & 1{,}4181 & 1{,}042 & 0{,}6438 & 0{,}4225 & -1{,}2367 \\ -0{,}2103 & 1{,}0268 & 0{,}4082 & 0{,}5319 & -0{,}8288 & -0{,}2103 & -1{,}4474 & -1{,}3237 & 0{,}4082 & 1{,}6453 \end{bmatrix}$$

Перемножимо матриці $X^{*'}$ та X^* з допомогою функції МУМНОЖ в MS Excel:

$$X^{*'} \cdot X^* = \begin{bmatrix} 9 & 8{,}3819 & -8{,}3801 & 4{,}7721 \\ 8{,}3819 & 9 & -7{,}9873 & 2{,}9178 \\ -8{,}3801 & -7{,}9873 & 9 & -4{,}5302 \\ 4{,}7721 & 2{,}9178 & -4{,}5302 & 9 \end{bmatrix}.$$

3. Знайдемо кореляційну матрицю r, помноживши кожний елемент отриманої матриці на $\frac{1}{n-1} = \frac{1}{10-1} = \frac{1}{9}$ (n – кількість статистичних спостережень).

$$r = \frac{1}{n-1}\left[X^{*'} X^*\right] = \begin{bmatrix} 1 & 0{,}9313 & -0{,}9311 & 0{,}5302 \\ 0{,}9313 & 1 & -0{,}8875 & 0{,}3242 \\ -0{,}9311 & -0{,}8875 & 1 & -0{,}5034 \\ 0{,}5302 & 0{,}3242 & -0{,}5034 & 1 \end{bmatrix}.$$

Кореляційну матрицю також можна одержати, використовуючи пакет “Аналіз даних”, команда “кореляція” в MS Excel (рис. 3.3).

	Столбец 1	Столбец 2	Столбец 3	Столбец 4
Столбец 1	1			
Столбец 2	0,9313	1		
Столбец 3	-0,9311	-0,8875	1	
Столбец 4	0,5302	0,3242	-0,5034	1

Рис. 3.3. Фрагмент документу MS Excel, в якому знайдено кореляційну матрицю до матриці X^*

4. Знайдемо визначник матриці r. Для цього використаємо функцію МОПРЕД в MS Excel. Отже, $\det r = 0{,}0085$. Оскільки даний визначник близький до нуля, то в масиві пояснюючих змінних може існувати мультиколінеарність.

5. Обчислимо критерій χ^2 за формулою (3.7); χ^2=32,5944.

Знайдене значення χ^2 порівнюємо з табличним $\chi^2_{табл.}$ = 12,592, так як маємо $\frac{1}{2}m(m-1) = \frac{1}{2}4\cdot 3 = 6$ ступенів вільності (m – кількість пояснюючих змінних) та рівень значущості α=0,05. $\chi^2 > \chi^2_{табл.}$, тому в масиві пояснюючих змінних існує мультиколінеарність.

6. Обчислимо F-критерії. Для їх визначення необхідно знайти матрицю C, яка є оберненою до матриці r. Для цього використаємо функцію МОБР в MS Excel.

$$C = \begin{bmatrix} 16{,}927 & -10{,}13 & 5{,}2314 & -3{,}0577 \\ -10{,}13 & 11{,}26 & 1{,}9145 & 2{,}6834 \\ 5{,}2314 & 1{,}9145 & 7{,}8505 & 0{,}5571 \\ -3{,}0577 & 2{,}6834 & 0{,}5571 & 2{,}0317 \end{bmatrix}$$

Для кожної незалежної пояснюючої змінної знаходимо значення F-критеріїв за формулою (3.9).

F_1	31,854
F_2	20,527
F_3	13,701
F_4	2,063

Знайдені значення порівнюються з критичними величинами F-критерія при n–m=10–4=6 і m–1=3 ступеня вільності та рівні значущості α=0,05($F_{табл.}$=4,76). Так як $F_1 > F_{табл.}$, $F_2 > F_{табл.}$, $F_3 > F_{табл.}$, то кожна із змінних X_1, X_2, X_3 мультиколінеарна з іншими.

7. Визначимо часткові коефіцієнти кореляції за формулою (3.11).

Часткові коефіцієнти кореляції показують тісноту зв'язку між двома пояснюючими змінними за умови, що інші змінні не впливають на цей зв'язок.

$r_{12.34}$	0,734
$r_{13.24}$	-0,454
$r_{23.14}$	-0,204
$r_{14.23}$	0,521
$r_{24.13}$	-0,561
$r_{34.12}$	-0,139

Враховуючи розраховані значення часткових коефіцієнтів кореляції, можна стверджувати, що зв'язок між змінною X_1 (вартість основних виробничих фондів) та змінною X_2 (витрати трудових ресурсів) є щільним, якщо не враховувати впливу собівартості виробництва одиниці продукції та ритмічності виробництва. Зв'язок між іншими парами пояснюючих змінних є слабким або середнім, якщо не враховувати впливу інших пояснюючих факторів.

8. Визначимо t-критерії за формулою $t_{jk} = r_{jk} \cdot \sqrt{\dfrac{n-m}{1-r_{jk}^2}}$.

Дані критерії використовуються для визначення мультиколінеарності двох пояснюючих змінних.

t_{12}	2,645
t_{13}	-1,247
t_{23}	-0,509
t_{14}	1,497
t_{24}	-1,66
t_{34}	-0,345

Обчислені t-критерії порівнюємо з табличним значенням при $n–m$=10–4=6 ступені вільності та рівні значущості α=0,05 ($t_{табл.}$=2,447). Оскільки $t_{12} > t_{табл.}$, то X_1 (вартість основних виробничих фондів) та X_2 (витрати трудових ресурсів) мультиколінеарні між собою.

Отже, для застосування однокрокового методу найменших квадратів необхідно звільнитися від мультиколінеарності. Для цього з економетричної моделі виключимо змінну X_2.

Аналогічно дослідимо наявність мультиколінеарності для змінних X_1, X_3, X_4.

Кореляційну матрицю одержимо, використовуючи пакет “Аналіз даних”, команда “кореляція” в MS Excel.

$$r = \begin{bmatrix} 1 & -0{,}9311 & 0{,}5302 \\ -0{,}9311 & 1 & -0{,}5034 \\ 0{,}5302 & -0{,}5034 & 1 \end{bmatrix}.$$

Розрахуємо χ^2=16,83.

Знайдене значення χ^2 порівнюємо з табличним $\chi^2_{табл.}$= 7,815, так як маємо $\frac{1}{2}m(m-1)=\frac{1}{2}3\cdot 2=3$ ступені вільності (m-кількість пояснюючих змінних) та рівень значущості α=0,05. $\chi^2 > \chi^2_{табл.}$, тому в масиві пояснюючих змінних існує мультиколінеарність.

Обчислимо F-критерії. Для їх визначення необхідно знайти матрицю C, яка є оберненою до матриці r:

$$C = \begin{bmatrix} 7{,}8159 & 6{,}9532 & -0{,}6443 \\ 6{,}9532 & 7{,}5251 & 0{,}101 \\ -0{,}6443 & 0{,}101 & 1{,}3925 \end{bmatrix}.$$

Для кожної незалежної пояснюючої змінної знаходимо значення F-критеріїв за формулою (3.9):

F_1	23,856
F_3	22,838
F_4	1,374

Знайдені значення порівнюються з критичними величинами F-критерія при n–m=10–3=7 та m–1=2 ступені вільності та рівні значущості α=0,05 ($F_{табл.}$=4,74). Так як $F_1 > F_{табл.}$, $F_3 > F_{табл.}$, то кожна із змінних X_1, X_3 мультиколінеарна з іншими.

Визначимо часткові коефіцієнти кореляції за формулою (3.11).

Часткові коефіцієнти кореляції показують тісноту зв’язку між двома пояснюючими змінними за умови, що інші змінні не впливають на цей зв’язок.

$r_{13.4}$	-0,907
$r_{14.3}$	0,195
$r_{34.1}$	-0,031

Враховуючи розраховані значення часткових коефіцієнтів кореляції, можна стверджувати, що зв’язок між змінною X_1 (вартість основних виробничих фондів) та змінною X_3 (собівартість одиниці продукції) є щільним, якщо не враховувати впливу ритмічності виробництва. Зменшення собівартості зумовлює збільшення прибутку.

Визначимо t-критерії, які використовуються для визначення мультиколінеарності двох пояснюючих змінних.

t_{13}	-5,686
t_{14}	0,527
t_{34}	-0,083

Обчислені t-критерії порівнюємо з табличним значенням при $n-m=10-3=7$ ступенів вільності та рівні значущості $\alpha=0{,}05$ ($t_{табл.}=2{,}365$). Оскільки $|t_{13}| > t_{табл.}$, то X_1 (вартість основних виробничих фондів) та X_3 (собівартість виробництва одиниці продукції) мультиколінеарні між собою. Для її уникнення звільнимося від змінної X_3. У результаті цих дій будемо розглядати економетричну модель для змінних Y, X_1, X_4 ($n=10$, $m=2$).

Аналогічно до попередніх розрахунків, маємо:

$$r = \begin{bmatrix} 1 & 0{,}5302 \\ 0{,}5302 & 1 \end{bmatrix}; \qquad \chi^2 = 2{,}751; \quad \chi^2_{табл.} = 3{,}841.$$

Так як 2,751<3,841, то мультиколінеарність між пояснюючими змінними відсутня. Проведені дослідження за алгоритмом Феррара-Глобера показали, що між початковою множиною пояснюючих факторів має місце мультиколінеарність. Для подальшого застосування МНК ми звільнимось від цього явища.

Завдання 4. Гетероскедастичність.

Друга передумова Гаусса-Маркова припускає існування сталої дисперсії для кожної випадкової величини u_i. Таке явище має назву гомоскедастичності. Значення u_i вектора збурення u незалежні між собою і мають постійну дисперсію. Математично таке твердження можна записати наступним чином:

$$M\left(u_i^2\right)=\sigma_u^2=const \tag{4.1}$$

У прикладних дослідженнях бувають випадки порушення умови гомоскедастичності, тобто умова (4.1) не виконується для конкретного випадку:

$$M\left(u_i^2\right)=\sigma_{u_i}^2 \neq const \text{ або } M\left(uu'\right)=\sigma_u^2 \cdot S, \tag{4.2}$$

де S – деяка матриця.

Така ситуація породжує проблему гетероскедастичності. У даному випадку отримані оцінки параметрів регресії за методом найменших квадратів будуть неефективними, хоча і незміщеними та обґрунтованими.

Отже, якщо дисперсія залишків змінюється для кожного спостереження або групи спостережень, тобто має місце (4.2), то таке явище називається гетероскедастичністю.

Перевірити модель на наявність гетероскедастичності можна з допомогою графічних та аналітичних методів. Серед аналітичних методів найбільш поширеними є наступні тести:

- рангової кореляції Спірмена;
- Гольдфельда-Кванта;
- Глейсера;
- μ-критерію.

<u>Тест рангової кореляції Спірмена</u>. Дана процедура є найбільш простою та доступною серед множини аналітичних методів. Її можна використовувати як для малих, так і для великих вибірок. Тут робиться припущення, що дисперсія випадкової величини буде або збільшуватися, або зменшуватися у залежності від збільшення значень пояснюючої змінної. Основу алгоритму даного тесту складає обчислення коефіцієнта рангової кореляції Спірмена:

$$r_S=1-\frac{6\sum_{i=1}^{n}d_i^2}{n^3-n}, \tag{4.3}$$

де d_i – різниця між рангами, які властиві двом характеристикам i-го об'єкта, тобто $d_i = rangx_i - rang|u_i|$; n – кількість об'єктів, що рангуються.

Алгоритм тесту рангової кореляції Спірмена складається із трьох кроків.

Припустимо, що $y_i = b_0 + b_1 x_i + u_i$.

Крок 1. Будуємо рівняння регресії $\hat{y} = a_0 + a_1 x$ і знаходимо відхилення u_i.

Крок 2. Нехтуючи знаком u_i, тобто розглядаємо $|u_i|$, ранжуємо їх та x_i у зростаючому чи спадному порядку і знаходимо за формулою (4.3) коефіцієнт r_S.

Крок 3. Перевіряємо значущість отриманого коефіцієнта рангової кореляції за t-критерієм Ст'юдента. Розрахункове значення t-критерія знаходимо за формулою:

$$t_{розр.} = \frac{r_S \sqrt{n-k-1}}{\sqrt{1-r_S^2}}, \tag{4.4}$$

де n – кількість спостережень; k – число пояснюючих змінних. Далі знаходимо табличне значення t-критерія, тобто $t_{\alpha,k'}$ (α означає рівень надійності, k' – ступінь вільності). Якщо $t_{розр.} > t_{\alpha,k'}$, то підтверджується гіпотеза про гетероскедастичність. Якщо дана нерівність не виконується, то має місце гомоскедастичність.

Коефіцієнт рангової кореляції Спірмена можна знайти з допомогою процедури рангової кореляції програмного продукту STADIA [11 розд. 6,5] або MS Excel.

<u>Приклад</u> 4.1. Дослідити наявність гетероскедастичності, використовуючи тест рангової кореляції Спірмена, для економетричної моделі, що досліджує залежність між прибутком підприємства та показником продуктивності праці (вхідні дані задані табл. 4.1).

***Розв’язування*.**

Таблиця 4.1.

Прибуток, тис.грн. Y	Продукти-вність праці, шт/год, X	$\hat{Y}$	$Y-\hat{Y}$ u	$\lvert Y-\hat{Y}\rvert$ $\lvert u\rvert$	Ранг X	Ранг $\lvert u\rvert$	d	d^2
210	23	236,6224	-26,6244	26,6244	4	4	0	0
560	51	597,2266	-37,2266	37,2266	6	6	0	0
820	67	803,285	16,7151	16,7151	7	1	-6	36
930	73	880,5569	49,4431	49,4431	8	7	-1	1
1210	92	1125,251	84,749	84,749	10	9	-1	1
970	89	1086,615	-116,615	116,615	9	10	1	1
285	25	262,3817	22,6184	22,6184	5	3	-2	4
153	18	172,2311	-19,2311	19,2311	2	2	0	0
162	13	107,8379	54,1622	54,1622	1	8	7	49
170	20	197,9884	-27,9884	27,9884	3	5	2	4

На основі статистичних даних, користуючись функцією “регресія” пакету MS Excel (рис. 4.1), отримано економетричну модель: $\hat{Y} = -59{,}5846 + 12{,}87865x$, r=0,9897.

	A	B
1	ВЫВОД ИТОГОВ	
2		
3	*Регрессионная статистика*	
4	Множественный R	0,9897
5	R-квадрат	0,9794
6	Нормированный R-квадрат	0,9769
7	Стандартная ошибка	61,4395
8	Наблюдения	10
9		
10		
11		*Коэффициенты*
12	Y-пересечение	-59,5846
13	Переменная X 1	12,87865
14		

Рис. 4.1. Фрагмент документу MS Excel, в якому проведено регресійний аналіз вибірки

Обчислюємо розрахункові значення ($\hat{Y}$, відхилення u, модуль u, ранги показників X і модуля u, їх різниці d та квадрат різниць d^2) і заносимо в табл. 4.1. Найменшим значенням X та $|u|$ надаємо значення рангу один, і далі присвоєння рівнів здійснюємо у порядку зростання показників. Ранжування показників можна виконати, з допомогою команди “ранжирование” пакету STADIA або MS Excel.

Обчислимо коефіцієнт рангової кореляції Спірмена по формулі (4.3).

$$r_S = \left|1 - \frac{6\cdot(1\cdot 3 + 4\cdot 2 + 36 + 49)}{10^3 - 10}\right| = \left|1 - \frac{6\cdot 96}{1000 - 10}\right| = \left|1 - \frac{576}{990}\right| \approx 0{,}582 .$$

Перевіримо значущість отриманого коефіцієнта рангової кореляції за t-критерієм Ст'юдента. Розрахункове значення t-критерія знаходимо за формулою (4.4):

$$t_{розр.} = \frac{0{,}582\cdot\sqrt{10-2}}{\sqrt{1-0{,}582^2}} \approx 2{,}024 .$$

Знаходимо табличне значення t-критерія для $k' = n - k - 1 = 10 - 2 = 8$ ступенів вільності й рівня надійності α=0,05 ($t_{табл.}$=2,306). Оскільки $t_{розр.} < t_{табл.}$ (2,024<2,306), то гіпотеза про гетероскедастичність не підтверджується.

Метод μ-критерія використовується, коли вхідна сукупність спостережень досить велика. Його алгоритм складається з наступних кроків:

1. Розбиваємо вхідні дані залежної змінної Y на K груп $\left(k = \overline{1, K}\right)$ відповідно до зміни рівня значень величин Y.

2. Розраховуємо для відповідної групи спостережень суму квадратів відхилень:

$$S_k = \sum_{i=1}^{N_k}\left(Y_{ik} - \overline{Y}_k\right)^2 ,\ k = \overline{1, K} , \qquad (4.5)$$

де N_k – кількість спостережень в кожній групі, $\sum_{k=1}^{K} N_k = n$; n – кількість спостережень у вхідній сукупності.

3. Знаходимо суму квадратів відхилень для всієї сукупності спостережень:

$$S = \sum_{k=1}^{K} S_k = \sum_{k=1}^{K}\sum_{i=1}^{N_k}\left(Y_{ik} - \overline{Y}_k\right)^2 \qquad (4.6)$$

4. Розраховуємо значення параметра a:

$$a = \prod_{k=1}^{K}\left(\frac{S_k}{N_k}\right)^{N_k/2} \Bigg/ \left(\frac{S}{n}\right)^{n/2} \qquad (4.7)$$

5. Знаходимо значення μ-критерію:

$$\mu = -2\ln a .$$

Якщо дисперсія всіх спостережень однорідна, то даний критерій наближено відповідатиме розподілу χ^2 при ступенях вільності $(k-1)$. Далі знаходимо табличне значення критерію χ^2 при рівні значущості α і $(k-1)$ ступенях вільності та позначимо його через $\chi^2_{табл}$. Якщо знайдене значення $\mu < \chi^2_{табл.}$ при заданому рівні значущості α і ступені вільності $(k-1)$, то явище гетероскедастичності відсутнє. Якщо дана нерівність не виконується, то спостерігається гетероскедастичність.

Приклад 4.2. Дослідити наявність гетероскедастичності, використовуючи метод μ-критерія при рівні значущості α=0.01 для економетричної моделі, що будується на таких статистичних даних (Y – прибуток, тис. грн.; X_1 – вартість основних засобів виробництва, тис. грн; X_2 – витрати трудових ресурсів, тис. люд. год, табл. 4.2).

Таблиця 4.2.

Y	2600	2210	2105	2304	1950	1728	1649	1514	1595	2298
X_1	815	655	519	748	485	503	556	608	684	852
X_2	143	162	185	104	138	142	154	71	105	238
Y	2508	2103	2510	1940	3000	4215	2950	3213	3500	3415
X_1	427	472	505	456	1100	925	900	1125	1300	1290
X_2	258	270	210	204	700	820	658	730	800	784

Розв'язування.

Використовуючи метод μ-критерія, розбиваємо вхідні дані залежної змінної Y на дві однакові групи, в кожній з яких маємо 10 показників. Розраховуємо для кожної групи спостережень суму квадратів відхилень по формулі (4.5). Суми квадратів відхилень знаходимо, використовуючи MS Excel, регресійний аналіз (рис. 4.2).

10	Дисперсионный анализ		
11		df	SS
12	Регрессия	2	595146,7
13	Остаток	7	600883,4
14	Итого	9	1196030,1

а)

10	Дисперсионный анализ		
11		df	SS
12	Регрессия	2	3698237,6
13	Остаток	7	616722,8
14	Итого	9	4314960,4

б)

Рис. 4.2. Фрагмент документу MS Excel, в якому проведено регресійний аналіз двох однакових груп вибірки: *а*) 10 перших спостережних даних вибірки; *б*) 10 останніх спостережних даних вибірки

S_1=1196030, S_2= 4314960, $S = S_1 + S_2$= 5510990.

Значення параметра a розраховуємо за формулою (4.7).

$$a=\frac{\left(\frac{1196030}{10}\right)^{5}\left(\frac{4314960}{10}\right)^{5}}{\left(\frac{5510990}{20}\right)^{10}},$$

$$a=0{,}145077.$$

Отже, $\mu=-2\ln a=-2\ln(0{,}145077)=-2\cdot(-1{,}93049)=3{,}86098$.

Знаходимо табличне значення критерію χ-квадрат для рівня значущості α=0,01 і k=2–1=1 ступенів вільності. Одержимо $\chi^2_{табл.}$=6,635. Оскільки $\mu<\chi^2_{табл.}$, то гетероскедастичність відсутня. Значить для подальшого дослідження можна використати економетричну модель, розраховану для загальної сукупності вибірки (MS Excel, регресійний аналіз) (рис. 4.3).

	A	B
1	ВЫВОД ИТОГОВ	
2		
3	*Регрессионная статистика*	
4	Множественный R	0,904018663
5	R-квадрат	0,817249742
6	Нормированный R-квадрат	0,795749712
7	Стандартная ошибка	326,7214469
8	Наблюдения	20
9		
10		
11		*Коэффициенты*
12	Y-пересечение	1635,480693
13	Переменная X 1	0,049315
14	Переменная X 2	2,306771
15		
16		

Рис. 4.3. Фрагмент документу MS Excel, в якому проведено регресійний аналіз вибірки

Вигляд економетричної моделі

$$\hat{y}=1635{,}481+0{,}049315x_1+2{,}306771x_2.$$

Параметричний тест Гольдфельда-Квандта. Даний метод використовується, коли вхідна сукупність спостережень невелика. В його основу покладено припущення відносно того, що дисперсія залишків зростає пропорційно до величини квадрата однієї з незалежних змінних, тобто має місце формула $M\left(u^2\right)=\sigma_u^2 X_{ij}^2$ для побудованої нами моделі виду

$$Y=XB+u. \tag{4.8}$$

Алгоритм даного тесту включає в себе наступні кроки:

1. Впорядковуємо вхідну сукупність спостережень відповідно до величини зростання елементів вектора X_j, який найбільш імовірно може викликати зміну дисперсії залишків.

2. Відкидаємо із вхідної сукупності C спостережень, які містяться в центрі вектора X_j. Вважається, що дисперсія залишків для них постійна. Величину C знаходимо за формулою:

$$C = \frac{4n}{15}, \tag{4.9}$$

де n – кількість елементів вектора X_j. У результаті такої процедури залишок $(n-c)$ спостережень поділиться на дві підвибірки однакового розміру $\frac{(n-c)}{2}$, одна з яких буде включати малі значення вектора X_j, а інша – великі.

3. Використовуючи МНК або стандартну процедуру продукту STADIA, MS Excel, побудуємо економетричні моделі для кожної з отриманих на попередньому кроці підвибірок, обсяг яких становить $\frac{n-c}{2} \geq m$, де m – кількість незалежних змінних. Як результат одержимо дві моделі виду:

$$Y_1 = XB_1 + u_1, \quad Y_2 = XB_2 + u_2. \tag{4.10}$$

4. Знаходимо суму квадратів залишків для першої та другої моделі:

$$S_1 = u_1' u_1, \quad S_2 = u_2' u_2, \tag{4.11}$$

де S_1, S_2 – сума квадратів залишків, відповідно для першої та другої моделі; u_1, u_2 – залишки, відповідно для першої та другої моделі.

5. Обчислюємо величину критерію Фішера:

$$F^* = \frac{S_2}{S_1}. \tag{4.12}$$

Далі знаходимо табличне значення ($F_{табл.}$) критерія Фішера для $v_1 = v_2 = \frac{(n-c-2m)}{2}$ ступенів вільності та рівня значущості α. Якщо $F^* > F_{табл.}$ ми допускаємо існування гетероскедастичності (відкидаємо гіпотезу H_0 щодо відсутності відмінності між дисперсіями u в двох підвибірках). Якщо $F^* \leq F_{табл.}$, то гетероскедастичність відсутня (приймається гіпотеза H_0). Чим більше значення F^*, тим більша гетероскедастичність залишків.

Приклад 4.3. Дослідити наявність гетероскедастичності, використовуючи параметричний тест Гольдфельда-Квандта для сукупності спостережень табл. 4.3.

Розв'язування.

Упорядкуємо вхідну сукупність спостережень відповідно до величини зростання елементів вектора X_1, який найбільш імовірно може викликати зміну дисперсії залишків, використовуючи процедуру сортування засобів ПК (табл. 4.3).

Таблиця 4.3.

Y	X_1	X_4
300	80	0,6
155	92	0,4
164	95	0,42
172	100	0,7
200	110	0,6
550	140	0,8
800	215	0,7
980	300	0,5
920	310	0,72
1100	400	0,9

Розрахуємо кількість спостережень, які слід відкинути із середини сукупності (формула 4.9) $C = \frac{4 \cdot 10}{15} \approx 2{,}67$. У нашому випадку викидаємо 5 та 6 спостереження. Тоді одержану вибірку ділимо на дві підвибірки однакового розміру $\frac{10-2}{2} = 4$. Рівняння економетричних моделей кожної підвибірки розраховуємо в MS Excel (рис. 4.4, рис. 4.5).

	A	B	C	D
1	ВЫВОД ИТОГОВ			
2				
3	*Регрессионная статистика*			
4	Множественный R	0,98713718		
5	R-квадрат	0,974439813		
6	Нормированный R-квадрат	0,923319438		
7	Стандартная ошибка	18,97389909		
8	Наблюдения	4		
9				
10	Дисперсионный анализ			
11		*df*	*SS*	*MS*
12	Регрессия	2	13724,7412	6862,3706
13	Остаток	1	360,0088	360,0088
14	Итого	3	14084,7500	
15				
16		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*
17	Y-пересечение	751,0495	123,4246	6,0851
18	Переменная X 1	-7,2659	1,2905	-5,6304
19	Переменная X 2	213,8618	75,8147	2,8208
20				

Рис. 4.4. Фрагмент документу MS Excel, в якому проведено регресійний аналіз для чотирьох перших спостережних значень вибірки

	A	B	C	D
1	ВЫВОД ИТОГОВ			
2				
3	*Регрессионная статистика*			
4	Множественный R	0,98296		
5	R-квадрат	0,96622		
6	Нормированный R-квадрат	0,89865		
7	Стандартная ошибка	39,76268		
8	Наблюдения	4		
9				
10	Дисперсионный анализ			
11		*df*	*SS*	*MS*
12	Регрессия	2	45218,92902	22609,46451
13	Остаток	1	1581,07098	1581,07098
14	Итого	3	46800,00000	
15				
16		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*
17	Y-пересечение	513,4468	111,2684	4,6145
18	Переменная X 1	1,7778	0,3607	4,9282
19	Переменная X 2	-153,0372	166,8012	-0,9175
20				

Рис. 4.5. Фрагмент документу MS Excel, в якому проведено регресійний аналіз для чотирьох останніх спостережних значень вибірки

Відповідні суми квадратів залишків першої та другої моделі S_1=360,0088, S_2=1581,071. Обраховуємо значення критерію по формулі (4.12).

$$F^* = \frac{1581{,}071}{360{,}0088} = 4{,}391756.$$

Знаходимо табличне значення критерію Фішера для $\nu_1 = \nu_2 = \frac{(n-c-2m)}{2} = \frac{10-2-2\cdot 3}{2} = 1$ ступенів вільності і рівня значущості α=0,05 ($F_{табл.}$=161,45). Оскільки $F^* < F_{табл.}$ (4,391756<161,45), то гетероскедастичність відсутня, і можна використовувати для знаходження оцінок параметрів моделі звичайний МНК.

Приклад 4.4. Дослідити наявність гетероскедастичності, використовуючи тест Глейзера, використовуючи дані задачі 4.1.

Розв'язування.

У завданні 4.1 нами була побудована економетрична модель $\hat{Y} = -59{,}5846 + 12{,}87865X$, r=0,9897 та знайдені відхилення $u = Y - \hat{Y}$.

Таблиця 4.4.

Прибуток, тис.грн. Y	Продуктивність праці, шт/год, X	$\hat{Y}$	$Y-\hat{Y}$ u	$\lvert Y-\hat{Y}\rvert$ $\lvert u\rvert$
210	23	236,6224	-26,6244	26,6244
560	51	597,2266	-37,2266	37,2266
820	67	803,285	16,7151	16,7151
930	73	880,5569	49,4431	49,4431
1210	92	1125,251	84,749	84,749
970	89	1086,615	-116,615	116,615
285	25	262,3817	22,6184	22,6184
153	18	172,2311	-19,2311	19,2311
162	13	107,8379	54,1622	54,1622
170	20	197,9884	-27,9884	27,9884

Виберемо $\delta = 1$, тобто розглянемо модель $\lvert u\rvert = \beta_0 + \beta_1 X$.

На основі статистичних даних, користуючись функцією “регресія” пакету MS Excel (рис. 4.6), отримано економетричну модель: $\lvert U\rvert = 12{,}1505 + 0{,}7089X$, r=0,6784.

	A	B	C	D	E	F	G	H	I
1	ВЫВОД ИТОГОВ								
2									
3	*Регрессионная статистика*								
4	Множественный R	0,678410525							
5	R-квадрат	0,460240841							
6	Нормированный R-кв	0,392770946							
7	Стандартная ошибка	25,26720212							
8	Наблюдения	10							
9									
10	Дисперсионный анализ								
11		*df*	*SS*	*MS*	*F*	*Значимость F*			
12	Регрессия	1	4355,012738	4355,012738	6,821425194	0,031044055			
13	Остаток	8	5107,452022	638,4315028					
14	Итого	9	9462,46476						
15									
16		*Коэффициенты*	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
17	Y-пересечение	12,15051322	15,07508899	0,805999436	0,443539783	-22,61270432	46,91373076	-22,61270432	46,91373076
18	Переменная X 1	0,708863838	0,271409635	2,611785825	0,031044055	0,082992097	1,33473558	0,082992097	1,33473558
19									

Рис. 4.6. Фрагмент документу MS Excel, в якому проведено регресійний аналіз вибірки

Отже, b_1=0,7089, $S_{b_1} = 0{,}27141$.

Із проведеного регресійного аналізу (рис. 4.6) бачимо, що t-статистика для параметра β_1 дорівнює 2, 6118.

Знаходимо табличне значення *t*-критерія для $k = n - k - 1 = 10 - 2 = 8$ ступенів вільності й рівня надійності α=0,05 ($t_{табл.}$=2,306). Оскільки $t_{розр.} < t_{табл.}$(2,6118>2,306), то гіпотеза про гетероскедастичність підтверджується.

Завдання 5. Економетричні моделі динаміки.

1. Часовий ряд та його складові.

При побудові економетричних моделей використовуються два типи вхідних даних:

- перший описує сукупність різних об'єктів у певний момент (період) часу;
- другий описує один об'єкт за ряд послідовних моментів (періодів) часу.

Така характеристика вхідних даних приводить до отримання двох видів моделей, відповідно просторових та часових рядів (рис. 5.1).

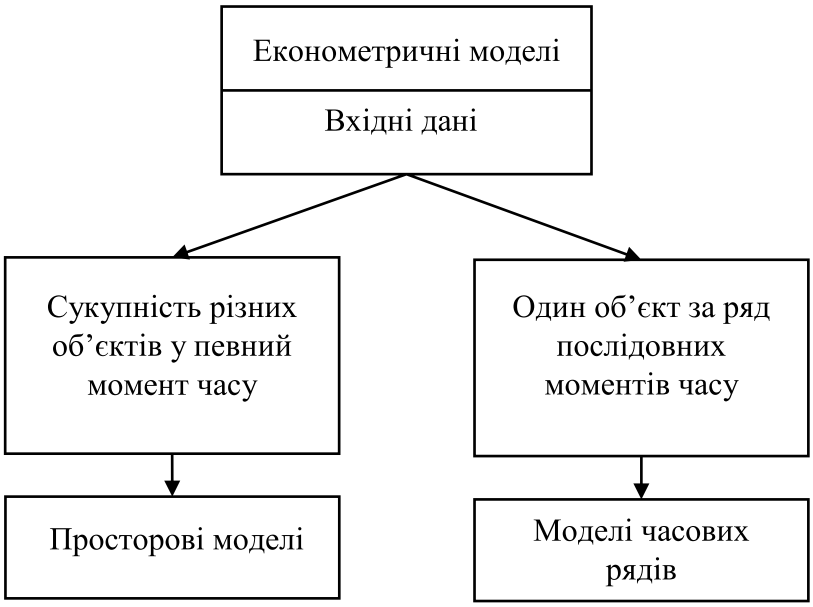


Рис. 5.1. Класифікація моделей за типом вхідних даних

Часовим рядом (рядом динаміки, динамічним рядом) називається впорядкована сукупність вимірювань будь-якого показника $\{(y_i, t_i), i = \overline{1, n}\}$, проведених за декілька послідовних моментів (періодів) часу. Окремі значення даного показника називається рівнями часового ряду.

Відносно часових характеристик розрізняють моментні (табл. 5.1) та інтервальні (табл. 5.2) ряди. Для першої групи економічні показники вказуються для певних моментів часу $t_1, t_2, \ldots t_n$, а для другої, вони відносяться до певних проміжків (інтервалів) часу, наприклад $[(t_0, t_1), (t_1, t_2), \ldots (t_{n-1}, t_n)]$.

Таблиця 5.1

Моментний часовий ряд

Дата	1.01	1.02	1.03	1.04	1.05	1.06	1.07	1.08	1.09
Обсяг ліквідних коштів, млн. грн.	120,3	140,5	110,0	160,3	180,4	150,6	190,2	200,1	160,5

Таблиця 5.2

Інтервальний часовий ряд

Місяць	Січень	Лютий	Березень	Квітень	Травень	Червень
Видано кредитів, млн. грн.	25,3	28,4	30,6	40,2	35,4	50,6

Якщо кожному моменту часу відповідають значення лише одного показника, ряд є одномірним. В іншому випадку – багатомірним.

Часові ряди можуть бути сформованими на основі абсолютних, середніх та відносних величин значень соціально-економічних показників.

При проведенні комплексного аналізу даних показників використовуються наступні основні характеристики динаміки часових рядів: абсолютний приріст, коефіцієнти зростання та приросту, темпи зростання та приросту, середні параметри різного виду.

Загальна класифікація часових рядів представлена на рис. 5.2.

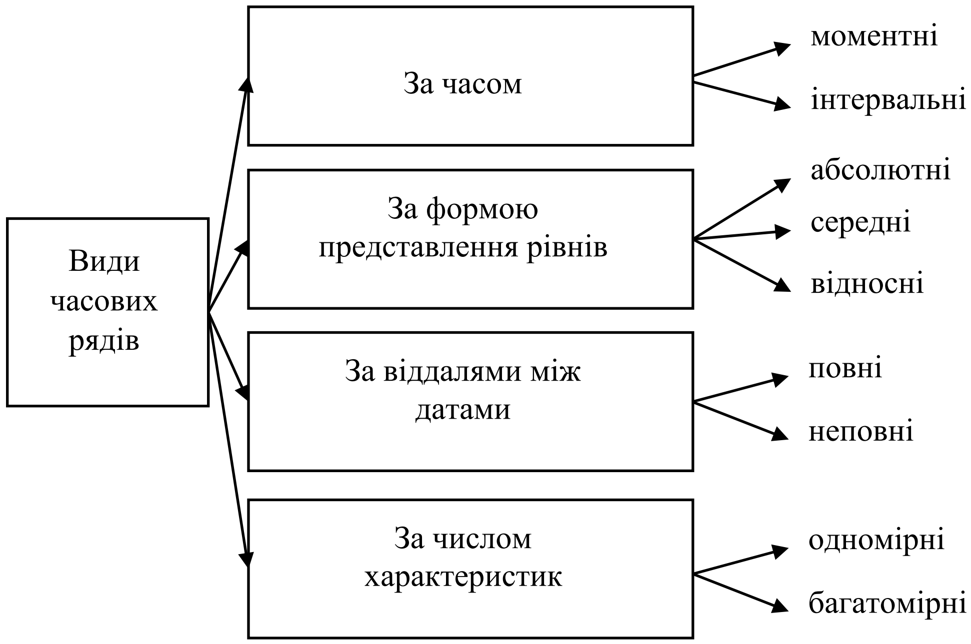
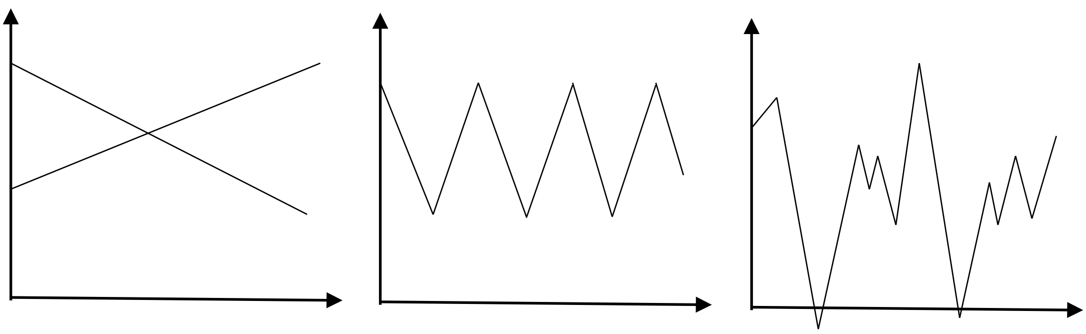


Рис. 5.2. Класифікація часових рядів

Формування рівнів часових рядів проходить під впливом множини факторів, які умовно можна поділити на наступні групи:

1) фактори, що описують тенденції (Т) ряду, тобто стійке збільшення або зменшення значень рівнів ряду (трендів);
2) фактори, що описують циклічні та сезонні (S) коливання ряду;
3) випадкові фактори (Е).



а) зростаюча (спадна) тенденція б) сезонна складова в) випадкова складова

Рис. 5.3. Основні складові часового ряду

Отже, складові часового ряду можна узагальнити і представити наступним чином:

1. Систематичні (регулярні) складові:

тренд (Т);

сезонні та циклічні коливання (S).

2. Випадкові складові (E).

У випадку, коли регулярна складова має постійний характер, то такий часовий ряд називається стаціонарним.

З позиції арифметичних дій фактичний рівень часового ряду можна представити як суму або добуток трендової, циклічної та випадкової складових.

Таки підхід дає можливість розглядати три види моделей часового ряду:

1) адитивна модель (часовий ряд представляється у вигляді суми перерахованих компонент);

$$Y_t = T_t + S_t + E_t; \qquad (5.1)$$

2) мультиплікативна модель (часовий ряд представляється у вигляді добутку перерахованих компонент);

$$Y_t = T_t \cdot S_t \cdot E_t; \qquad (5.2)$$

3) змішана

$$Y_t = T_t \cdot S_t + E_t. \qquad (5.3)$$

Адитивна модель застосовується для тих випадків, коли досліджуваний часовий ряд володіє приблизно однаковими змінами протягом всього періоду. Мультиплікативну модель доцільно використовувати у тих випадках, коли зміна часової послідовності збільшується зі зростанням рівня.

З допомогою операції логарифмування мультиплікативна модель може бути приведена до адитивної форми:

$$\ln Y_t = \ln T_t + \ln S_t + \ln E_t. \qquad (5.4)$$

Основна задача економетричного дослідження будь-якого часового ряду полягає у виявленні та проведенні кількісної оцінки кожної із зазначених вище компонент розвитку тих чи інших процесів. Дана задача тісно пов’язана із умовами правильного формування часових рядів:

- вибір діапазону інтервалу між сусідніми членами ряду;
- співставлення рівнів за відповідними ознаками та змістом.

Перелічимо основні етапи попереднього аналізу часових рядів:

1. Побудова графіка.
2. Розрахунок та аналіз основних статистичних показників: середнє, дисперсія, розмах варіації і т.д.
3. Розрахунок абсолютних і відносних показників динаміки.

4. Оцінювання автокореляційної і часткової автокореляційної функції часового ряду.

Основними завданнями аналізу часових рядів є:

1) визначення та опис характерних особливостей часового ряду;
2) підбір його статистичної моделі;
3) прогнозування майбутніх значень показників;
4) підготовка пропозицій з управління дослідженою системою.

Розглянемо основні методи дослідження часових рядів (рис. 5.4).

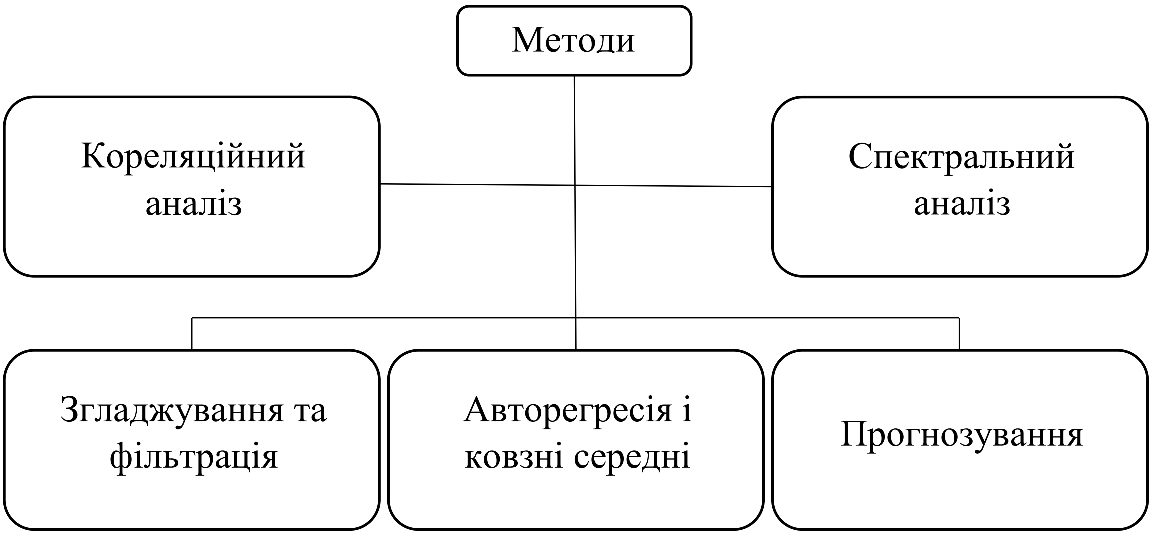


Рис. 5.4. Основні методи дослідження

Структурно часовий ряд складається із рівнів, які володіють наступними властивостями:

1. Рівні часового ряду зазвичай не є статистично незалежними.
2. Рівні часового ряду зазвичай не є однаково розпределіленими.
3. Рівні часового ряду можуть бути непорівняними.

Розглянемо основні типи моделей тренду (рис. 5.5):

- лінійна;
- поліноміальна;
- гіперболічна;
- степенева, логарифмічна;

* обернена логарифмічна, показникова;
* крива Гомперца;
* логістична.

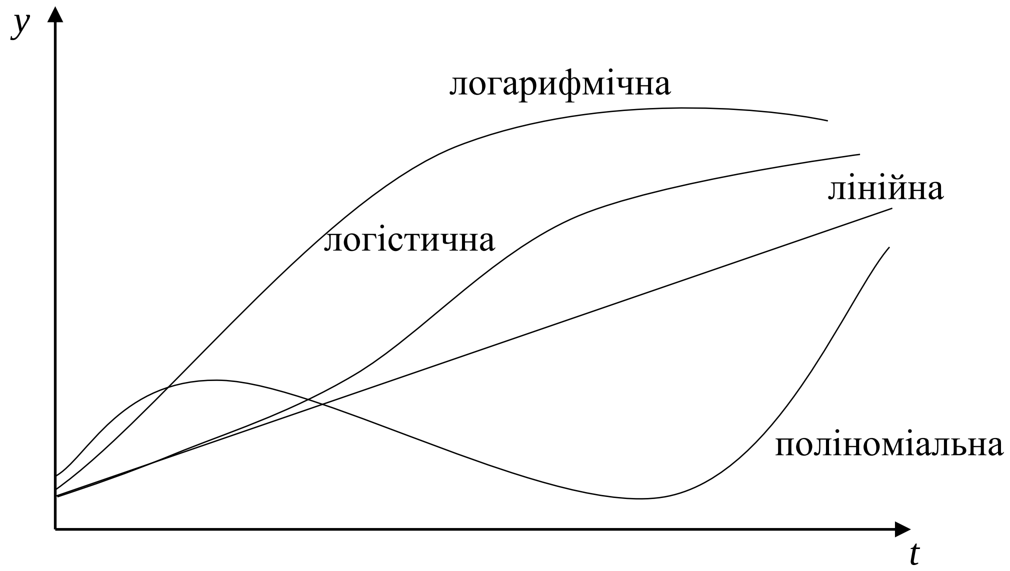


Рис. 5.5. Моделі тренду

Економетричний аналіз часових рядів з допомогою математичного інструментарію та відповідних програмних продуктів дає можливість розв’язати наступні задачі:

* дослідження структури часового ряду, яка як правило відображається закономірною зміною середнього рівня (тренду) та наявністю періодичних коливань;
* побудова математичної моделі економічного процесу, представленого часовим рядом;
* прогнозування майбутнього розвитку економічного процесу;
* дослідження причинно-наслідкових взаємозв’язків між економічними процесами, які проявляються у вигляді різноманітних взаємних кореляцій між часовими рядами.

Для розв’язання цих та інших задач економетричного аналізу часових рядів існує необхідна кількість різноманітних методів, серед яких можна відзначити наступні:

* методи кореляційного аналізу, які дають можливість виявити найбільш суттєві періодичні залежності та їх лаги (затримки) в одному процесі (автокореляція) або між декількома процесами (кроскореляція);

- методи спектрального аналізу дозволяють знаходити періодичні та квазіперіодичні залежності в даних;
- методи згладжування та фільтрації призначені для перетвореня часових рядів з метою усунення з них високоякісних або сезонних коливань;
- методи авторегресії та ковзного середнього є особливо корисними для опису та прогнозування процесів, які проявляють однорідні коливання навколо середнього значення.

Значна частина методів аналізу зорієнтована на стаціонарні процеси, статистичні властивості яких не змінюються протягом часу (середнє та дисперсія постійні у випадку нормального розподілу). Однак багато часових рядів мають нестаціонарний характер. У ряді випадків нестаціонарність можна усунути шляхом:

- вирахування тренду чи процесу зміни середнього значення, представленого деякою детермінованою функцією, підібраною шляхом процедур простої чи поліномної регресії;
- фільтрація нестаціонарним фільтром, нулі якого знаходяться тільки на одиничному колі (стохастичний тренд).

З метою стандартизації часових рядів буває доцільним провести її загальне чи сезонне центрування та нормування на стандартне відхилення шляхом відповідних операцій блоку перетворення даних. Центрування ряду спрямоване на усунення ненульового середнього значення, яке може завадити інтерпретації результатів (наприклад, при спектральному аналізі). Мета нормування – запобігти в обрахунках операцій з великими числами, що може призвести до переповнення результату.

Внаслідок приведених вище перетворень часового ряду можна побудувати його математичну модель, за якою здійснене прогнозування, тобто одержане певне продовження часового ряду. Проте, щоб результати прогнозу можна було би порівняти з вхідними даними, з ними треба провести перетворення, зворотні виконаним, що робиться засобами перетворень.

<u>2. Автокореляція рівнів часового ряду.</u>

Наявність у часовому ряді тенденції та циклічних коливань вказує на існування зв’язків між послідовними рівнями ряду, тобто значення кожного наступного рівня ряду залежить від попередніх. Можна стверджувати про дослідження кореляційної залежності між послідовними рівнями часового ряду, і яка має назву автокореляція рівнів часового ряду.

Ступінь тісноти автокореляційного зв’язку можна визначити з допомогою лінійного коефіцієнта кореляції між рівнями заданого часового ряду і рівнями даного часового ряду зсунутими на декілька кроків τ у часі. Величину зсув τ назвемо лагом. Він означає запізнення впливу факторів на досліджуваний показник і вказує на порядок коефіцієнта автокореляції.

При збільшенні величини лагу кількість пар значень, за якими розраховується коефіцієнт автокореляції, зменшується. Припускається, що максимальний лаг повинен не перевищувати величини $n/4$.

Для загального випадку формула розрахунку коефіцієнта автокореляції має вид:

$$r_\tau = \frac{\sum_{t=\tau+1}^{n}(y_t - \bar{y}_{1\tau})(y_{t-\tau} - \bar{y}_{2\tau})}{\sqrt{\sum_{t=\tau+1}^{n}(y_t - \bar{y}_{1\tau})^2 \sum_{t=\tau+1}^{n}(y_{t-\tau} - \bar{y}_{2\tau})^2}}, \tag{5.5}$$

де

$$\bar{y}_{1\tau} = \frac{\sum_{t=\tau+1}^{n} y_t}{n-\tau};\ \bar{y}_{2\tau} = \frac{\sum_{t=\tau+1}^{n} y_{t-\tau}}{n-\tau}. \tag{5.6}$$

Отже, коефіцієнт r_τ характеризує щільність зв’язку між первинним рядом динаміки і цим же рядом зсунутим на τ періодів.

Таблиця 5.3

Зсуви динамічного ряду

Зміна часу t	Рівень ряду y_t	$\tau = 1$	$\tau = 2$	$\tau = 3$
1	y_1	–	–	–
2	y_2	y_1	–	–
3	y_3	y_2	y_1	–
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n-2$	y_{n-2}	y_{n-3}	y_{n-4}	y_{n-5}
$n-1$	y_{n-1}	y_{n-2}	y_{n-3}	y_{n-4}
n	y_n	y_{n-1}	y_{n-2}	y_{n-2}

У табл. 5.3 подано зсунені часові ряди з лагами $\tau = 1,2,3$. Як бачимо, при збільшенні величини лагу τ, довжина корельованих рядів зменшується відповідно на τ, що приводить до зменшення кількості корельованих пар.

Запишемо формулу розрахунку коефіцієнта автокореляції першого порядку, поклавши у (5.5) $\tau = 1$:

$$r_1 = \frac{\sum_{t=2}^{n}\left(y_t - \bar{y}_{11}\right)\left(y_{t-1} - \bar{y}_{21}\right)}{\sqrt{\sum_{t=2}^{n}\left(y_t - \bar{y}_{11}\right)^2 \sum_{t=2}^{n}\left(y_{t-1} - \bar{y}_{21}\right)^2}}, \tag{5.7}$$

де

$$\bar{y}_{11} = \frac{\sum_{t=2}^{n} y_t}{n-1}; \; \bar{y}_{21} = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1}. \tag{5.8}$$

Коефіцієнт автокореляції першого порядку характеризує тісноту зв'язку між рівнями y_t та y_{t-1}.

Аналогічно можна записати формули визначення коефіцієнтів автокореляції більш високих порядків, наприклад другого порядку $\tau = 2$:

$$r_2 = \frac{\sum_{t=3}^{n}\left(y_t - \bar{y}_{12}\right)\left(y_{t-2} - \bar{y}_{22}\right)}{\sqrt{\sum_{t=3}^{n}\left(y_t - \bar{y}_{12}\right)^2 \sum_{t=3}^{n}\left(y_{t-2} - \bar{y}_{22}\right)^2}}, \tag{5.9}$$

де

$$\bar{y}_{12} = \frac{\sum_{t=3}^{n} y_t}{n-2}; \; \bar{y}_{22} = \frac{\sum_{t=3}^{n} y_{t-2}}{n-2}. \tag{5.10}$$

Приклад 5.1. Нам відомі умовні обсяги податкових надходжень (табл. 5.4) протягом 10 місяців (y_t, млн. грн.). Розрахувати коефіцієнт автокореляції рівнів для часового ряду податкових надходжень.

Розв'язування.

На першому кроці на основі ряду $\left\{y_t, t = \overline{1,10}\right\}$ побудуємо додатковий ряд $\left\{y_{t-1}, t = \overline{2,10}\right\}$ зсунувши початковий ряд на один період ($\tau = 1$).

Враховуючи, що $\tau = 1$ знайдемо коефіцієнт автокореляції рівнів ряду першого порядку, використавши (5.7) та (5.8).

Таблиця 5.4

№ п/п	y_t	y_{t-1}	$y_t - \bar{y}_{11}$	$y_{t-1} - \bar{y}_{21}$	$(y_t - \bar{y}_{11})(y_{t-1} - \bar{y}_{21})$	$(y_t - \bar{y}_{11})^2$	$(y_{t-1} - \bar{y}_{21})^2$
1	3						
2	4	3	-8,2222	-6,8889	56,64191	67,60457	47,45694
3	6	4	-6,2222	-5,8889	36,64191	38,71577	34,67914
4	8	6	-4,2222	-3,8889	16,41971	17,82697	15,12354
5	8	8	-4,2222	-1,8889	7,975314	17,82697	3,567943
6	10	8	-2,2222	-1,8889	4,197514	4,938173	3,567943
7	14	10	1,7778	0,1111	0,197514	3,160573	0,012343
8	16	14	3,7778	4,1111	15,53091	14,27177	16,90114
9	20	16	7,7778	6,1111	47,53091	60,49417	37,34554
10	24	20	11,7778	10,1111	119,0865	138,7166	102,2343
Разом	113	89	0,0002	0,0001	304,2222	363,5556	260,8889

$$\bar{y}_{11} = \frac{\sum_{t=2}^{10} y_t}{10-1} = 12,2222; \quad \bar{y}_{21} = \frac{\sum_{t=2}^{10} y_{t-1}}{10-1} = 9,8889;$$

$$r_1 = \frac{304,2222}{\sqrt{363,5556 \cdot 260,8889}} = 0,98782.$$

Розраховане значення $r_1 = 0,98782$ вказує на наявність дуже тісного зв’язку між податковими надходженнями поточного і попередніх місяців, і як наслідок, про доказ того, що у часовому ряді податкових надходжень присутня сильна лінійна тенденція.

Тому можна зробити висновок про те, що коли коефіцієнт автокореляції першого порядку прийме найбільше значення серед розрахованих, то даний ряд має тільки тенденцію.

Поклавши $\tau = 2$, $\tau = 3$ і т. д., ми можемо розрахувати коефіцієнти автокореляції другого, третього і більш високих порядків. Для цього формуємо нові таблиці аналогічні табл. 5.4.

Якщо при розрахунку коефіцієнтів автокореляції більш високих порядків виявиться, що найбільше значення прийме коефіцієнт порядку *m*, то можна зробити висновок про наявність у ряді циклічних коливань з періодом *m*.

Послідовність коефіцієнтів автокореляції першого, другого та вищих рівнів формує автокореляційну функцію часового ряду.

Графічне представлення залежності значень автокореляційної функції від величини лага називається корелограмою.

Кількісний аналіз автокореляційної функції та її корелограми дає можливість розрахувати величину лага, при якому автокореляція найбільш висока. Величина даного лага вказує на найбільшу тісноту зв’язку між поточним і попереднім рядами. Проведена процедура аналізу дає можливість виявити структуру часового ряду.

Отже, з допомогою коефіцієнта автокореляції можна оцінити наявність ступеня тісноти лінійного зв’язку між двома суміжними рядами. Проте, для часових рядів яким властива нелінійна тенденція, коефіцієнт автокореляції може наближатися до нуля. Крім того, за знаком коефіцієнта автокореляції ми не можемо робити висновок про існування зростаючої чи спадної тенденції у досліджуваному ряді.

3. Попередній аналіз часових рядів та їх згладжування.

Процедура попереднього аналізу часових рядів спрямована на виявлення та усунення аномальних значень рівнів часового ряду. Крім того, вона включає в себе перевірку гіпотези про наявність або відсутність тенденції у вхідному часовому ряді.

Аномальним рівнем часового ряду називається окреме значення рівня даного ряду, яке по-перше, не відповідає потенційним можливостям досліджуваного економічного процесу, і по-друге, має суттєвий вплив на кількісні характеристики часового ряду.

Розрізняють два види причин аномальних значень часового ряду: помилки першого та другого роду. Помилки першого роду носять технічний характер і вони утворюються в основному за рахунок процесів агрегування та дезагрегування економічних показників, а також при формуванні інформаційної бази та ін. Даний тип помилок можна виявляти та усувати з допомогою відповідних алгоритмів.

Помилки другого роду виникають завдяки впливу факторів, яким властивий об’єктивний характер та епізодичний прояв. Такий тип помилок неможливо усунути.

Одним із методів виявлення аномальних рівнів часових рядів є метод Ірвіна. В основу алгоритму даного методу покладено знаходження параметра:

$$I_t = \frac{|y_t - y_{t-1}|}{\sigma_y},\ t = \overline{2,n}, \qquad (5.11)$$

де I_t – критерій Ірвіна для періоду *t*;

σ_y – середньоквадратичне відхилення,

$$\sigma_y = \sqrt{\frac{\sum_{t=1}^{n}(y_t - \bar{y})^2}{n-1}}; \qquad (5.12)$$

$\bar{y}$ – середнє значення,

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n}. \qquad (5.13)$$

Розрахункове значення критерія Ірвіна для рівня значущості $\alpha = 0{,}05$ вибираємо із табл. 5.5.

Таблиця 5.5

Граничні значення критерія Ірвіна

n	2	3	10	20	30	50	100
I_t	2,8	2,3	1,5	1,3	1,2	1,1	1,0

Розраховані значення I_2, I_3 і т. д. порівнюються із табличним значенням критерію Ірвіна $I_{табл.}$. Якщо розрахункові значення $I_{розр.} > I_{табл.}$, то відповідне значення y_t рівня часового ряду вважаються аномальним.

Далі нам необхідно перейти до виявлення причин і встановлення їх роду. При виявленні наявності причин першого роду можна приступити до процедури їх усунення. Це можна виконати за допомогою заміни простою середньою арифметичною двох сусідніх рівнів $(y_{t-1}; y_{t+1})$ або з допомогою апроксимуючої кривої (трендової моделі).

<u>Приклад</u> 5.2. Задано часовий ряд поквартального надходження митних платежів за 2014-2017 рр. (табл. 5.6). Необхідно встановити наявність чи відсутність аномальних значень рівнів, використавши критерій Ірвіна.

Розв'язування.

Для виявлення аномальних рівнів заданого часового ряду скористаємось критерієм Ірвіна. Нам необхідно для кожного періоду розрахувати значення:

$$I_t = \frac{|y_t - y_{t-1}|}{\sigma_y}, \; t = \overline{2{,}16}.$$

Знайдемо $\sigma_y = \sqrt{\dfrac{\sum_{t=1}^{n}(y_t - \bar{y})^2}{n-1}} = \sqrt{\dfrac{1590{,}28}{15}} = 10{,}297$.

Тоді можемо знайти для кожного періоду t значення $I_t = \dfrac{|y_t - y_{t-1}|}{\sigma_y}$, які заносимо в табл. 5.6 в стовпчик 6.

Таблиця 5.6

	A	B	C	D	E	F	G	H
1	Рік	Квартал	Номер періоду, t	Обсяг надходжень митних платежів, млн. грн., y_t	$\lvert y_t - y_{t-1}\rvert$	Критерій Ірвіна, It	$y_t - \bar{y}$	$(y_t - \bar{y})^2$
2	2014	1	1	23,5			-11,48	131,82
3		2	2	20,2	3,3	0,32	-14,78	218,49
4		3	3	18,3	1,9	0,18	-16,68	278,26
5		4	4	29,8	11,5	1,12	-5,18	26,85
6	2015	1	5	32,5	2,7	0,26	-2,48	6,16
7		2	6	27,7	4,8	0,47	-7,28	53,02
8		3	7	25,6	2,1	0,20	-9,38	88,01
9		4	8	38,5	12,9	1,25	3,52	12,38
10	2016	1	9	40,3	1,8	0,17	5,32	28,29
11		2	10	36,1	4,2	0,41	1,12	1,25
12		3	11	33,6	2,5	0,24	-1,38	1,91
13		4	12	46,2	12,6	1,22	11,22	125,86
14	2017	1	13	48,3	2,1	0,20	13,32	177,39
15		2	14	44,6	3,7	0,36	9,62	92,52
16		3	15	42,4	2,2	0,21	7,42	55,04
17		4	16	52,1	9,7	0,94	17,12	293,05
18	Сума			559,7				1590,28

Із табл. 5.5 знайдемо значення $I_{табл.} \approx 1{,}42$.

Оскільки жодне розраховане значення I_t не перевищує граничного значення, то гіпотеза про відсутність аномальних значень рівнів ряду митних платежів може бути прийнята.

Виявити наявність тренда у вхідному часовому ряді можна з допомогою методів перевірки різниць середніх рівнів та Фостера-Ст'юарта.

Алгоритм методу середніх різниць дає можливість перевірити гіпотезу про наявність тренда середньої динамічного ряду і складається з наступних кроків:

Крок 1. Вхідний ряд $\{y_t, t = \overline{1,n}\}$ розбиваємо на дві приблизно рівні частини. До першої частини включаємо n_1 перших рівнів вхідного ряду, а до другої – n_2 інших рівнів. Причому має місце рівність $n_1 + n_2 = n$.

Крок 2. Знаходимо з цих частин середні значення та дисперсії:

$$\bar{y}_1 = \frac{\sum_{t=1}^{n_1} y_t}{n_1}; \qquad \bar{y}_2 = \frac{\sum_{t=n_1+1}^{n} y_t}{n_2} \tag{5.14}$$

$$\sigma_1^2 = \frac{\sum_{t=1}^{n_1}(y_t - \bar{y}_1)^2}{n_1 - 1}; \qquad \sigma_2^2 = \frac{\sum_{t=n_1+1}^{n}(y_t - \bar{y}_2)^2}{n_2 - 1} \tag{5.15}$$

Крок 3. Знаходимо розрахункове значення *F*-критерію Фішера:

$$F_{розр.} = \begin{cases} \sigma_1^2 / \sigma_2^2, & \sigma_1^2 > \sigma_2^2 \\ \sigma_2^2 / \sigma_1^2, & \sigma_2^2 > \sigma_1^2 \end{cases} \tag{5.16}$$

та табличне значення F_α з рівнем значущості α.

Якщо $F_{розр.} < F_\alpha$, то гіпотеза про рівність дисперсій приймається і переходимо до наступного етапу.

Для випадку $F_{розр.} \ge F_\alpha$ даний метод відповіді відносно наявності тренда не дає.

Табличне значення F_α залежить від рівня значимості α та довжини ряду *n* (табл. 5.7).

Таблиця 5.7

n	20	25	30
$F_{\alpha=0,05}$	2,12	1,96	1,84

Крок 4. Розраховуємо значення *t*-критерію Стьюдента:

$$t_{розр.} = \frac{|\bar{y}_1 - \bar{y}_2|}{\sigma\sqrt{\frac{1}{n_n} + \frac{1}{n_2}}}, \tag{5.17}$$

де σ – середньоквадратичне відхилення різниць середніх,

$$\sigma = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}}. \tag{5.18}$$

Далі знаходимо $t_{табл.}$ з заданим рівнем значущості α та відповідним числом ступенів вільності $(n_1 + n_2 - 2)$.

Значення $t_{табл.}$, при значеннях рівня значимості $\alpha = 0,05$ та довжини ряду *n* приведемо в табл. 5.8.

Таблиця 5.8

n	28	30	50	60
$t_{табл.}, \alpha = 0,05$	2,048	2,042	2,009	2,00

Якщо $t_{розр.} < t_{табл.}$, то тренда немає, у протилежному випадку тренд існує.

Характерною особливістю використання даного методу є те, що його можна використовувати тільки для часових рядів із наявністю монотонної тенденції.

Приклад 5.3. Відомий місячний обсяг надходжень акцизного збору за 2016-2017 рр. Виявити наявність тренда у вхідному часовому ряді з допомогою методу середніх рівнів.

Розв'язування.

Крок 1. Розбиваємо вхідний ряд $\{y_t, t=\overline{1,20}\}$ на дві рівні частини $n_1 = 10$ і $n_2 = 10$.

Крок 2. Знаходимо для кожної з цих частин середні значення $\bar{y}_1$, $\bar{y}_2$ та дисперсії σ_1^2, σ_2^2:

$$\bar{y}_1 = \frac{\sum_{t=1}^{10} y_t}{10} = 4{,}19; \qquad \bar{y}_2 = \frac{\sum_{t=11}^{20} y_t}{10} = 7{,}57;$$

$$\sigma_1^2 = \frac{\sum_{t=1}^{10}(y_t - \bar{y}_1)^2}{9} = \frac{16{,}289}{9} = 1{,}8099; \qquad \sigma_2^2 = \frac{\sum_{t=11}^{20}(y_t - \bar{y}_2)^2}{9} = \frac{8{,}06}{9} = 0{,}895667.$$

Крок 3. Знаходимо розрахункове значення F-критерію Фішера, маємо:

$$F_{розр.} = \frac{\sigma_1^2}{\sigma_2^2} = \frac{1{,}8099}{0{,}895667} = 2{,}020717.$$

Табличне значення $F_{табл.} = 2{,}12$ при $\alpha = 0{,}05$. Оскільки $F_{розр.} < F_{табл.}$, то гіпотеза про рівність дисперсій приймається і переходимо до наступного кроку.

Таблиця 5.9

	A	B	C	D	E	F	G	H
1	Рік	Місяць	Номер періоду, t	Акцизний збір, млн. грн., y_t	$\bar{y}_t^1$	$\bar{y}_t^2$	$(y_t - \bar{y}^1)^2$	$(y_t - \bar{y}^2)^2$
2	2016	1	1	2,1	2,1		4,3681	
3		2	2	2,5	2,5		2,8561	
4		3	3	3,2	3,2		0,9801	
5		4	4	3,6	3,6		0,3481	
6		5	5	3,9	3,9		0,0841	
7		6	6	4,5	4,5		0,0961	
8		7	7	5,1	5,1		0,8281	
9		8	8	5,4	5,4		1,4641	
10		9	9	5,6	5,6		1,9881	
11		10	10	6	6		3,2761	
12		11	11	6,1		6,1		2,1609
13		12	12	6,5		6,5		1,1449
14	2017	1	13	6,9		6,9		0,4489
15		2	14	7,2		7,2		0,1369
16		3	15	7,4		7,4		0,0289
17		4	16	7,5		7,5		0,0049
18		5	17	8,1		8,1		0,2809
19		6	18	8,3		8,3		0,5329
20		7	19	8,7		8,7		1,2769
21		8	20	9		9		2,0449
22	Сума				41,9	75,7		

Крок 4. Знайдемо значення *t*-критерію Стьюдента:

$$t_{розр.} = \frac{|\bar{y}_1 - \bar{y}_2|}{\sigma\sqrt{\frac{1}{n_n} + \frac{1}{n_2}}} = \frac{|4{,}19 - 7{,}57|}{1{,}16309 \cdot 0{,}447214} \approx 1{,}299626,$$

$$\sigma = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}} = \sqrt{\frac{9 \cdot 1{,}8099 + 9 \cdot 0{,}895667}{18}} \approx 1{,}16309.$$

Знайдемо $t_{табл.}(8;0{,}05) = 2{,}31$.

Оскільки $t_{розр.} < t_{табл.}$, то тренда не існує.

Метод Фостера-Ст'юарта одночасно дає можливість встановити наявність самого тренда та тренда дисперсії. У випадку існування тренда дисперсії дисперсія збільшується або зменшується.

Опишемо алгоритм даного методу:

Крок 1. Виконується процедура порівняння кожного рівня вихідного ряду, починаючи з другого, з усіма попередніми.

Дана процедура дає можливість сформувати дві числові послідовності:

$$k_t = \begin{cases} 1, & якщо\, y_t > y_{t-1} > \ldots > y_1 \; (y_t \text{ більше всіх попередніх рівнів}), \\ 0, & у\ протилежному\ випадку. \end{cases} \quad (5.19)$$

$$l_t = \begin{cases} 1, & якщо\, y_t < y_{t-1} < \ldots < y_1 \; (y_t \text{ менше всіх попередніх рівнів}), \\ 0, & у\ протилежному\ випадку. \end{cases} \quad (5.20)$$

для $t = 2, 3, \ldots, n.$

Крок 2. Обчислимо величини *S* та *d*, які характеризують зміну дисперсійного та середнього значення часового ряду:

$$S = \sum_{t=2}^{n}(k_t + l_t); \qquad d = \sum_{t=2}^{n}(k_t - l_t). \quad (5.21)$$

Таким чином розраховані величини є випадковими з математичним сподіванням рівним μ для значень *S* та рівним 0 для d.

Крок 3. Перевірка гіпотез про випадковість відхилення величин *S* та *d* від їх математичного сподівання на основі *t*-критерія Ст'юдента:

$$t_S = \frac{|S - \mu|}{\sigma_1}, \qquad \sigma_1 \approx \sqrt{2\ln n - 3{,}4253}, \quad (5.22)$$

$$t_d = \frac{|d - 0|}{\sigma_2}, \qquad \sigma_2 \approx \sqrt{2\ln n - 0{,}8456}\,. \qquad (5.23)$$

де σ_1 – середньоквадратичне відхилення величини S; σ_2 – середньоквадратичне відхилення величини d.

Значення μ, σ_1 та σ_2 знаходимо із табл. 5.10.

Таблиця 5.10

Значення μ, σ_1 та σ_2 для критерія Фостера-Ст’юарта

n	10	15	20	25	30	35	40	45	50
μ	3,853	4,636	5,195	5,632	5,990	6,294	6,557	6,79	6,998
σ_1	1,288	1,521	1,677	1,791	1,882	1,956	2,019	2,072	2,121
σ_2	1,964	2,153	2,279	2,373	2,447	2,509	2,561	2,606	2,645

Крок 4. Розраховані значення t_S та t_d порівнюються з табличними значеннями t-критерія Ст’юдента $t_{табл.}$ для рівня значущості α та з n ступенями вільності.

Якщо $t_{табл.}$ більше розрахункового значення, то відповідний тренд відсутній, тобто гіпотеза про відсутність тренда приймається. Наприклад, якщо $t_S < t_{табл.}$, а $t_d > t_{табл.}$, то тренд дисперсії відсутній, а тренд ряду буде.

<u>Приклад</u> 5.4. Заданий часовий ряд поквартального надходження до бюджету міста місцевих податків за період 2014-2017 рр. (табл. 5.11). Перевірити даний часовий ряд на наявність тренду методом Фостера-Ст’юарта.

Розв’язування.

Крок 1. Виконаємо процедуру порівняння кожного рівня вихідного ряду з усіма попередніми з допомогою формул (5.19) та (5.20) і знайдені значення k_t та l_t занесемо у табл. 5.11, стовпці 5 та 6.

Крок 2. Обчислимо величини S_t та d_t за формулами:

$$S_t = k_t + l_t; \;\; d_t = k_t - l_t, \; t = \overline{2{,}16};$$

і занесемо у табл. 5.11, стовпці 7 та 8, відповідно.

Далі знайдемо суми $S = \sum_{t=2}^{16} S_t = 14$, $d = \sum_{t=2}^{16} d_t = 14$.

Розраховані величини є випадковими величинами з математичним сподіванням $\mu = 4{,}636$ для значень S та рівним 0 для d.

Таблиця 5.11

Рік	Квартал	Номер періоду, t	Місцвий податок, млн. грн., y_t	k_t	l_t	S_t	d_t
2014	1	1	9,4	–	–	–	–
	2	2	17,5	1	0	1	1
	3	3	30,2	1	0	1	1
	4	4	51,4	1	0	1	1
2015	1	5	69,0	1	0	1	1
	2	6	73,2	1	0	1	1
	3	7	71,5	0	0	0	0
	4	8	79,1	1	0	1	1
2016	1	9	98,5	1	0	1	1
	2	10	128,9	1	0	1	1
	3	11	137,9	1	0	1	1
	4	12	155,6	1	0	1	1
2017	1	13	181,3	1	0	1	1
	2	14	198,5	1	0	1	1
	3	15	207,0	1	0	1	1
	4	16	222,4	1	0	1	1

Крок 3. Для перевірки гіпотез про випадковість відхилення величин S та d від їх математичного сподівання використаємо t-критерія Ст'юдента:

$$t_S = \frac{|S-\mu|}{\sigma_1} = \frac{|14-4{,}636|}{1{,}521} = 6{,}156, \; t_d = \frac{|d-0|}{\sigma_2} = \frac{14}{2{,}153} = 6{,}503,$$

де $\sigma_1 = 1{,}521 \, \sigma_2 = 2{,}153, \mu = 4{,}636$ (див. табл. 5.10).

Знайдені значення $t_S = 6{,}156$, $t_d = 6{,}503$ порівняємо із табличними значеннями критерія Ст'юдента $t_{табл.}(\alpha = 0{,}05;15) = 2{,}131$.

Оскільки $t_S > t_{табл.}$ та $t_d > t_{табл.}$, то гіпотеза про відсутність тенденції в середній та дисперсії відкидається, і робиться висновок про існування у вхідному ряді тенденції в середньому та дисперсії, що підтверджує існування тренду в обсягах надходжень місцевих податків.

Враховуючи наявність сильних коливань рівнів часових рядів та виявлення чітких тенденцій розвитку досліджуваного економічного явищу приходиться використовувати процедури згладжування або механічного вирівнювання часових рядів.

Процедура механічного вирівнювання окремих рівнів часового ряду базується на використанні фактичних значень сусідніх рівнів.

Припустимо, що нам задано часовий ряд

$$\{y_t, t = \overline{1,n}\} \tag{5.24}$$

для якого розглянемо алгоритм процедури згладжування даного ряду з допомогою методів ковзної середньої.

Метод ковзної середньої полягає в тому, що в основу розрахунку прогнозного показника на момент часу t будується з допомогою усереднювання значень рівнів за декілька періодів часу.

Визначаємо інтервал згладжування m ($m<n$). Бажано значення m вибирати непарним.

Для обчислення значень згладжених рівнів ряду використаємо формулу:

$$\bar{y}_t^K = \frac{1}{m}\sum_{i=t-p}^{t+p} y_i \,. \qquad (5.25)$$

де $\bar{y}_t^K$ – значення згладженого рівня ряду для періоду t, $p = \frac{m-1}{2}$ (для непарного m).

Після проведених обчислень ми отримуємо $n-m+1$ згладжених значень рівнів ряду. Причому, перші p і останні p значення рівнів ряду будуть відсутніми.

Для випадку, коли m парне для розрахунків використовуємо наступну формулу:

$$\bar{y}_t^K = \frac{1}{2p}\left[\frac{1}{2}y_{t-p} + \sum_{i=t-p+1}^{t+p-1} y_i + \frac{1}{2}y_{t+p}\right], \qquad (5.26)$$

де $m = 2p$.

Так, для випадку $m = 3$ маємо:

$$\bar{y}_t^K = \frac{y_{t-1} + y_t + y_{t+1}}{3};$$

а для $m = 5$:

$$\bar{y}_t^K = \frac{y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2}}{5}.$$

Метод зваженої ковзної середньої для процедури згладжування використовує рівні ваги w_i, які вибираються довільно (чим дальше від центра згладжування, тим вага менша), або з допомогою відповідних ваг. В основу алгоритму покладено формулу:

$$\bar{y}_t^З = \frac{1}{\sum_{i=t-p}^{t+p} w_i}\sum_{i=t-p}^{t+p} w_i y_i \,, \quad p+1 \le t \le n-p, \qquad (5.27)$$

де $\bar{y}_t^З$ – зважене ковзне середнє для періоду t, w_i – вага для i-ого інтервалу.

Тут апроксимація ряду в межах інтервалу згладжування здійснюється з використанням поліномів другого і вищих ступенів.

Наприклад, для поліномів другого і третього порядків числова послідовність ваг при $m=5$ буде $\{-3;12;17;12;-3\}$, а при $m=7$ – $\{-2;3;6;7;6;3;-2\}$.

Тоді для $m=5$ та $m=7$ формула (5.27) прийме вид:

$$\bar{y}_t^3=\frac{1}{35}\left(-3y_{t-2}+12y_{t-1}+17y_t+12y_{t+1}-3y_{t+2}\right) \text{ при } m=5;$$

$$\bar{y}_t^3=\frac{1}{21}\left(-2y_{t-3}+3y_{t-2}+6y_{t-1}+7y_t+6y_{t+1}+3y_{t+2}-2y_{t+3}\right) \text{ при } m=7.$$

Метод експоненціального згладжування для знаходження згладженого рівня використовує зважені значення попередніх рівнів ряду і в основу розрахунків покладена рекурентна формула:

$$\bar{y}_t^e=\alpha y_t+(1-\alpha)\bar{y}_{t-1}^e,\quad t=\overline{1,n}, \tag{5.28}$$

де $\bar{y}_t^e$ – експоненціальне ковзне середнє для періоду *t*, α – параметр згладжування $(0<\alpha<1)$.

Значення $\bar{y}_0^e$ приймаємо рівним значенню першого рівня ряду y_1, або рівним середньому арифметичному декількох перших рівнів ряду.

Приклад 5.5. Відома динаміка надходжень податку на прибуток підприємств до бюджету міста за 14 місяців (табл. 5.12).

Виконати згладжування заданого часового ряду податкових надходжень наступними методами:

а) проста ковзна середня з інтервалом згладжування $m=3$ та $m=7$;
б) зважена ковзна середня з інтервалом згладжування $m=5$;
в) експоненціальна ковзна середня з параметром $\alpha=0{,}2$.

Таблиця 5.12

Динаміка надходжень ППП до бюджету міста (млн. грн.)

Місяць, *t*	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Обсяг надходжень	10,2	14,5	8,2	15,4	14,2	17,1	16,5	15,6	21,2	18,1	8,3	14,5	16,3	22,2

***Розв'язування*.**

Результати розрахунків подамо в табл. 5.13.

а) Для тримісячної ковзної середньої маємо:

$$m=3,\ p=\frac{m-1}{2}=\frac{3-1}{2}=1.$$

Тоді формула (5.25) прийме вид:

$$\bar{y}_t^K=\frac{1}{3}\sum_{i=t-1}^{t+1}y_i,\quad t=\overline{2,13}.$$

Для даного випадку будуть відсутні $i = t - 1$ значення рівнів першого та останнього часового ряду.

$t = 2$, $\bar{y}_2^K = \frac{1}{3}(y_1 + y_2 + y_3) = \frac{1}{3}(10{,}2 + 14{,}5 + 8{,}2) = 10{,}97$;

$t = 3$, $\bar{y}_3^K = \frac{1}{3}(y_2 + y_3 + y_4) = \frac{1}{3}(14{,}5 + 8{,}2 + 15{,}4) = 12{,}7$;

і т. д.

$t = 13$, $\bar{y}_{13}^K = \frac{1}{3}(y_{12} + y_{13} + y_{14}) = \frac{1}{3}(14{,}5 + 16{,}3 + 20{,}2) = 17{,}0$.

У випадку семимісячної ковзної середньої маємо:

$$m = 7,\ p = \frac{m-1}{2} = \frac{7-1}{2} = 3.$$

Для даного випадку (5.25) прийме вид:

$$\bar{y}_t^K = \frac{1}{7}\sum_{i=t-3}^{t+3} y_i, \quad t = \overline{4{,}11}.$$

Як бачимо у табл. 5.12 при $m = 7$ будуть відсутні три перших і три останніх значень часового ряду.

Для $t = 4$, маємо:

$$\bar{y}_4^K = \frac{1}{7}\sum_{i=1}^{7} y_i = \frac{1}{7}(y_1 + y_2 + y_3 + y_4 + y_5 + y_6 + y_7) =$$

$$= \frac{1}{7}(10{,}2 + 14{,}5 + 8{,}2 + 15{,}4 + 14{,}2 + 17{,}1 + 16{,}5) = 13{,}73;$$

і т. д.

при $t = 11$

$$\bar{y}_{11}^K = \frac{1}{7}\sum_{i=8}^{14} y_i = \frac{1}{7}(y_8 + y_9 + y_{10} + y_{11} + y_{12} + y_{13} + y_{14}) =$$

$$= \frac{1}{7}(15{,}6 + 21{,}2 + 18{,}1 + 8{,}3 + 14{,}5 + 16{,}3 + 20{,}2) = 16{,}31.$$

б) обчислимо значення зваженого ковзного середнього з інтервалом згладжування $m = 5$. Для деякого випадку вагами будуть $(-3;12;17;12;-3)$, тоді формула (5.27) прийме вид:

$$\bar{y}_t^З = \frac{1}{35}(-3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+1} - 3y_{t+2}),\ \text{при}\ t = \overline{3{,}12},$$

причому в отриманому часовому ряді будуть відсутні перші два та два останні рівні ряду.

$\bar{y}_3^З = \frac{1}{35}(-3y_1 + 12y_2 + 17y_3 + 12y_4 - 3y_5) =$

$= \frac{1}{35}(-3 \cdot 10{,}2 + 12 \cdot 14{,}5 + 17 \cdot 8{,}2 + 12 \cdot 15{,}4 - 3 \cdot 14{,}2) = 12{,}14;$

і т. д.

$$\bar{y}_{12}^{3} = \frac{1}{35}\left(-3y_{10} + 12y_{11} + 17y_{12} + 12y_{13} - 3y_{14}\right) =$$

$$= \frac{1}{35}\left(-3\cdot 18{,}1 + 12\cdot 8{,}3 + 17\cdot 14{,}5 + 12\cdot 16{,}3 - 3\cdot 20{,}2\right) = 12{,}19.$$

в) побудуємо часовий ряд з допомогою експоненціального ковзного середнього, при $\alpha = 0{,}2$.

Формула для даного випадку буде:

$$\bar{y}_t^e = 0{,}2\cdot y_t + 0{,}8\cdot \bar{y}_{t-1}^e,\ t = \overline{1{,}14}.$$

Приймемо $\bar{y}_0^e = y_0 = 10{,}2$.

Тоді

$\bar{y}_1^e = 0{,}2\cdot y_1 + 0{,}8\cdot \bar{y}_0^e = 0{,}2\cdot 10{,}2 + 0{,}8\cdot 10{,}2 = 10{,}2$;

$\bar{y}_2^e = 0{,}2\cdot y_2 + 0{,}8\cdot \bar{y}_1^e = 0{,}2\cdot 14{,}5 + 0{,}8\cdot 10{,}2 = 11{,}06$;

$\bar{y}_3^e = 0{,}2\cdot y_3 + 0{,}8\cdot \bar{y}_2^e = 0{,}2\cdot 8{,}2 + 0{,}8\cdot 11{,}06 = 10{,}49$;

і т. д.

$\bar{y}_{14}^e = 0{,}2\cdot y_{14} + 0{,}8\cdot \bar{y}_{13}^e = 0{,}2\cdot 20{,}2 + 0{,}8\cdot 14{,}85 = 15{,}92$.

Отримані числові дані заносимо у табл. 5.13.

Таблиця 5.13

t	y_t	Прості ковзні середні		Зважені ковзні середні, $m=3$	Експоненціальні ковзні середні, $\alpha = 0{,}2$
		$m=3$	$m=7$		
1	10,2	–	–	–	10,20
2	14,5	10,97	–	–	11,06
3	8,2	12,70	–	12,14	10,49
4	15,4	12,60	13,73	12,45	11,47
5	14,2	15,57	14,50	15,92	12,02
6	17,1	15,93	15,46	16,17	13,03
7	16,5	16,40	16,87	16,19	13,73
8	15,6	17,77	15,86	17,49	14,10
9	21,2	18,30	15,90	19,73	15,52
10	18,1	15,87	15,79	16,33	16,04
11	8,3	13,63	16,31	11,99	14,49
12	14,5	13,03	–	12,19	14,49
13	16,3	17,00	–	–	14,85
14	20,2	–	–	–	15,92

Ефективним способом моделювання тенденції часового ряду є процедура побудови аналітичної функції, яка описує залежність рівнів динамічного ряду від часу або тренду. Таким чином, поряд з механічним вирівнюванням існує процедура аналітичного вирівнювання часового ряду. Зміст даної процедури полягає у знаходженні аналітичної функції $\hat{y} = f(x)$. Побудована функція $f(x)$

відображає основну тенденцію зміни часового ряду в динаміці і носить назву кривої росту.

В основу прогнозування з допомогою кривих росту покладено процес екстраполяції, який ґрунтується на збереженні у наступні часові періоди тенденції, яка мала місце у попередні періоди.

Розглянемо основні види кривих росту, які використовуються при побудові трендів:

- лінійна: $\hat{y}_t = a + bt$;
- гіперболічна: $\hat{y}_t = a + \frac{b}{t}$;
- експоненціальна: $\hat{y}_t = e^{a+bt}$;
- степенева: $\hat{y}_t = a \cdot t^b$;
- параболічна другого і вищих порядків: $\hat{y}_t = a + b_1 t + b_2 t^2 + \ldots + b_k t^k$;
- логістична: $\hat{y}_t = \frac{k}{1 + ae^{-bt}}$;
- крива Гомперца: $\hat{y}_t = k \cdot a^{b^t}$.

Процедура побудови оціночної функції $\hat{y}_t$ аналогічна до дослідження рівнянь парної регресії методом найменших квадратів чи відповідних програмних продуктів. Характерною особливістю є визначення залежних та незалежних змінних. В якості незалежної змінної приймаються моменти часу $t = 1, 2, \ldots, n$, а в якості залежної – фактичні рівні часового ряду.

Процедура прогнозування з використанням кривих росту проходить з допомогою наступних кроків:

1. Вибір однієї або декількох кривих, вид яких найбільш повно характеризують динаміку ряду.
2. Оцінка параметрів вибраних моделей.
3. Перевірка адекватності та оцінка точності вибраних моделей.
4. Розрахунок прогнозних показників (процес екстраполяції).

Для визначення типу тенденції можна використати наступні процедури. Якісний аналіз динамічного ряду, побудова діаграм процесів динаміки та їх графічний аналіз, розрахунок основних показників часового ряду та коефіцієнтів автокореляції рівнів.

4. Моделювання сезонних та циклічних коливань.

У більшості випадків економічним показникам властивий періодичний характер коливань із різними видами амплітуд та наявністю періодичності. Наявність періодичних коливань породжує клас моделей, які мають назву тренд-сезонними часовими рядами. При дослідженні даного типу моделей варто проводити розрахунок значень сезонної компоненти з допомогою методу ковзних середніх з подальшою побудовою адитивної або мультиплікативної моделі часового ряду:

- адитивна:

$$Y = T + S + E \tag{5.30}$$

- мультиплікативна:

$$Y = T \cdot S \cdot E \tag{5.31}$$

При виборі виду моделі необхідно провести графічний аналіз структури сезонних коливань. Якщо графічне представлення часового ряду дає можливість зробити висновок про наявність наближено постійної амплітуди коливань, тоді доцільно будувати адитивну модель. Для даної моделі значення сезонної компоненти приймається постійним для різних циклів.

У випадку зростання або спадання амплітуди сезонних коливань доречним буде здійснити побудову мультиплікативної моделі часового ряду, яка ставить рівні часового ряду в залежність від сезонної компоненти.

Для різних видів моделей нам необхідно розрахувати одні і ті ж значення T, S, Y для відповідних рівнів часового ряду.

Алгоритм процесу комплексного дослідження тренд-сезонних моделей включає у себе наступні кроки.

Крок 1. Визначення наявності у часовому ряді тренд та його ступені гладкості.

Крок 2. Побудова графіку динаміки часового ряду та оцінка наявності типу амплітуди коливань.

Крок 3. Враховуючи тип амплітуди коливань встановити вид моделі (адитивна або мультиплікативна).

Крок 4. При наявності амплітуди коливань виконати процедуру фільтрації сезонної компоненти.

Крок 5. Виконати аналіз динаміки сезонної хвилі.

Крок 6. Визначення факторів, які породжують сезонні коливання.

Крок 7. Розрахунок прогнозних показників на основі побудованої тренд-сезонної моделі.

Важливим моментом при дослідженні тренд-сезонних моделей будь-якого економічного процесу є формування достовірної інформаційної бази, яка, як правило, складається із місячних чи квартальних даних деякого показника за декілька років.

Табличну форму такої бази подамо наступним чином (табл. 5.14).

Таблиця 5.14

Квартал	Рік			
	1	2	…	k
I	y_{11}	y_{12}	y_{ij}	y_{1k}
II	y_{21}	y_{22}		y_{2k}
III	y_{31}	y_{32}		y_{3k}
IV	y_{41}	y_{42}		y_{4k}

де i – індекс кварталу, i=1,2,3,4; j – індекс року, j=1,2,…,k, y_{ij} – визначає значення рівня часового ряду для i-ого кварталу j-ого року.

Виявити сезонні коливання можна з допомогою наступних показників:

1) середній рівень показника за кожний квартал

$$\bar{y}_i = \frac{1}{k}\sum_{j=1}^{k} y_{ij}\ ,\ i = \overline{1,4}. \tag{5.32}$$

2) загальне середнє рівнів $\bar{y}$ за весь період

$$\bar{y} = \frac{1}{4k}\sum_{i=1}^{4}\sum_{j=1}^{k} y_{ij}\ ,\ i = \overline{1,4}. \tag{5.33}$$

або

$$\bar{y} = \frac{1}{4}\sum_{i=1}^{4}\bar{y}_i\ .$$

3) абсолютні відхилення середньомісячних показників від загальної середньої:

$$\Delta y_i = \bar{y}_i - \bar{y},\ i = \overline{1,4}. \tag{5.34}$$

Величину $S_i = \Delta y_i$ можна приймати як елемент адитивної сезонної складової.

Для випадку мультиплікативної залежності розраховуємо індекси сезонності

$$S_i = iH_i = \frac{\bar{y}_i}{\bar{y}}\cdot 100\%,\ i = \overline{1,4}, \tag{5.35}$$

де iH_i – мультиплікативна сезонна складова.

При розгляді місячних часових рядів значення $i = \overline{1,12}$.

Розглянемо алгоритм побудови адитивної тренд-сезонної моделі методом ковзних середніх з допомогою таких кроків.

Крок 1. Виконаємо процедуру згладжування часового ряду $\{y_t, t=\overline{1,n}\}$.

Запишемо вихідний ряд $\{y_{ij}, i=\overline{1,n}, j=\overline{1,k}\}$ у вигляді єдиного часового ряду

$y_1, y_2, \ldots, y_{4k}$ – для квартальних даних

або

$y_1, y_2, \ldots, y_{12k}$ – для місячних даних.

Проведемо згладжування ефекту квартального впливу на тенденцію ряду, знайшовши послідовні суми рівнів за кожні чотири квартали і виконаємо зсув на один момент часу з подальшим знаходженням середнього за чотири квартали – $\bar{y}_t^k$, $t=\overline{2,(4k-2)}$.

Далі нам необхідно знайти центрові ковзні середні

$$\bar{y}_t^{Ц} = \frac{1}{2}\left[\bar{y}_{t-1}^k + \bar{y}_t^k\right],\ t=\overline{2,(4k-2)}. \quad (5.36)$$

У випадку аналізу місячної інформації ($i=\overline{1,12}$) виконуємо процедуру згладжування простої ковзної середньої з інтервалом рівним 12 за формулою:

$$\bar{y}_t^K = \frac{1}{12}\left[0{,}5y_{t-6} + y_{t-5} + \ldots + y_t + \ldots + y_{t+5} + 0{,}5y_{t+6}\right],\ t=\overline{2,(12k-2)}. \quad (5.37)$$

Крок 2. Знаходимо оцінку сезонної компоненти для адитивної форми як різницю між фактичним рівнем вхідного ряду та центровим ковзним середнім:

$$\left(\bar{y}_t - \bar{y}_t^{Ц}\right),\ t=\overline{2,(4k-2)}. \quad (5.38)$$

Враховуючи значення $\left(\bar{y}_t - \bar{y}_t^{Ц}\right)$ для відповідних років *i*-ого кварталу, знайдемо середню оцінку сезонної компоненти $\bar{S}_i$, $i=\overline{1,4}$.

Далі знайдемо:

$$\sum_{i=1}^{4}\bar{S}_i = \bar{S}_1 + \bar{S}_2 + \bar{S}_3 + \bar{S}_4, \quad (5.39)$$

і перейдемо до знаходження величини коригуючого коефіцієнта за формулою:

$$k = \sum_{i=1}^{4}\bar{S}_i / 4. \quad (5.40)$$

Розрахуємо скориговані значення сезонної компоненти для кожного кварталу як різниці між середньою оцінкою $\bar{S}_i$ та коригуючим коефіцієнтом k:

$$\hat{S}_i = \bar{S}_i - k,\ i=\overline{1,4}. \quad (5.41)$$

Для адитивної моделі повинна виконуватися рівність: $\sum_{i=1}^{4} \hat{S}_i = 0$.

Враховуючи отримані значення $\{\hat{S}_i, i = \overline{1,4}\}$, побудуємо для кожного періоду t наступний часовий ряд $\{S_t, t = \overline{1,4k}\}$.

Крок 3. Знаходження оцінки сезонної компоненти для адитивної форми моделі:

$$T_t + E_t = Y_t - S_t, \; t = \overline{1,4k}. \tag{5.42}$$

Крок 4. Проводимо аналітичне вирівнювання ряду $\{T_t + E_t, t = \overline{1,4k}\}$ з допомогою лінійного тренду

$$T = a_0 + at, \; t - \text{час}. \tag{5.43}$$

Для знаходження числових значень T_t необхідно у знайдене рівняння покласти значення $t = \overline{1,4k}$.

Крок 5. Розрахуємо для адитивної форми значення рівнів ряду $\{T_t + S_t, t = \overline{1,4k}\}$.

Крок 6. Знайдемо абсолютну похибку з допомогою наступної формули:

$$E_t = Y_t - (T_t + S_t), \; t = \overline{1,4k}. \tag{5.44}$$

і далі обчислимо E_t^2.

Для оцінки загальної варіації рівнів часового ряду нам необхідно знайти

$$R^2 = 1 - \frac{\sum_{t=1}^{4k} E_t^2}{\sum_{t=1}^{4k} (y_t - \bar{y})^2}. \tag{5.45}$$

Крок 7. Знаходження прогнозних показників $y_{прогн.}$.

Для побудови мультиплікативної моделі всі операції на відповідних кроках потрібно у виразах з “-” замінити на ділення, а з “+” – на множення.

<u>Приклад</u> 5.6. Відомо надходження податку на прибуток підприємств до бюджету міста за квартальними даними за 2014-2017 рр. (табл. 5.15). Необхідно побудувати адитивну модель для заданого часового ряду.

Таблиця 5.15

t	Обсяг ППП, млрд. грн., y_t	Разом за 4 квартали	Ковзна середня за 4 квартали, $\bar{y}_t^k$	Центрова ковзна середня, $\bar{y}_t^Ц$	Оцінка сезонної компоненти, $\bar{y}_t - \bar{y}_t^Ц$
1	10,1	-	-	-	-
2	8,1	42,3	10,575	-	-
3	9,1	44	11	10,7875	-1,6875
4	15	45,1	11,275	11,1375	3,8625
5	11,8	46,4	11,6	11,4375	0,3625
6	9,2	45,6	11,4	11,5	-2,3
7	10,4	45,9	11,475	11,4375	-1,0375
8	14,2	46,3	11,575	11,525	2,675
9	12,1	46,4	11,6	11,5875	0,5125
10	9,6	47,3	11,825	11,7125	-2,1125
11	10,5	48,4	12,1	11,9625	-1,4625
12	15,1	49,6	12,4	12,25	2,85
13	13,2	50,6	12,65	12,525	0,675
14	10,8	50,3	12,575	12,6125	-1,8125
15	11,5	-	-	-	-
16	14,8	-	-	-	-

Побудований графік обсягу податкових надходжень в динаміці (рис. 5.6) показує на наявність сезонних коливань: різке зростання у четвертому та спад у другому кварталах.

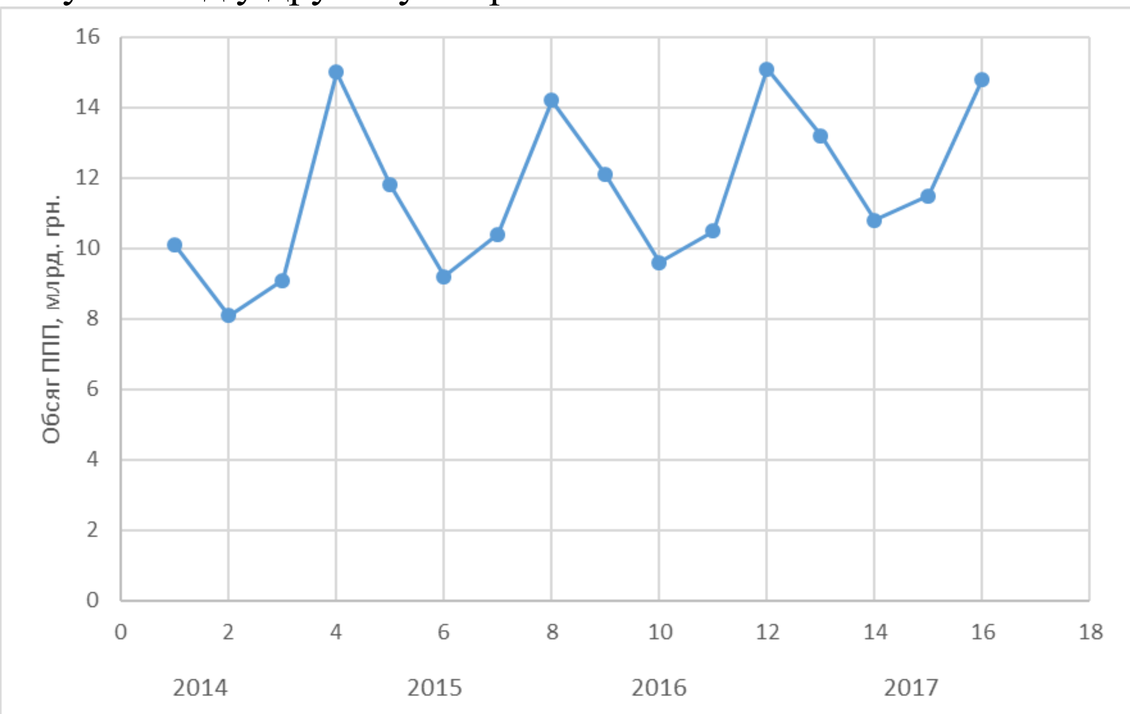


Рис. 5.6. Динаміка обсягу податкових надходжень

Даний ряд містить сезонні коливання з періодом 4 і наглядно можна побачити наявність приблизно рівної амплітуди коливань.

Тому можна стверджувати про відповідність даного часового ряду адитивній формі:

$$Y = T + S + E .$$

Побудуємо адитивну модель на основі отриманих даних з допомогою наступного алгоритму.

Крок 1. Виконаємо процедуру вирівнювання вхідного ряду $\{y_t, t = \overline{1,n}\}$ методом ковзного середнього.

а) Проведемо згладжування ефекту квартального впливу на тенденцію ряду. Знайдемо послідовні суми рівнів за кожні чотири квартали і здійснимо зсув на один момент часу (табл. 5.14, стовпець 3):

$$y_1 + y_2 + y_3 + y_4 = 10{,}1 + 8{,}1 + 9{,}1 + 15{,}0 = 42{,}3,$$

$$y_2 + y_3 + y_4 + y_5 = 8{,}1 + 9{,}1 + 15{,}0 + 11{,}8 = 44,$$

і т. д.

Тим самим ми знайшли укрупнені дані вихідного ряду.

б) Розділимо отримані суми на число 4 (кількість періодів, що укрупнювалися) і результат запишемо в стовпці 4 табл. 5.15:

$\bar{y}_2^k = 42{,}3 : 4 = 10{,}575$; $\bar{y}_3^k = 44 : 4 = 11$ і т. д.

в) Знайдемо центрові ковзні середні з допомогою середніх значень з двох послідовних ковзних середніх $\bar{y}_t^k$. Маємо:

$$\bar{y}_3^{Ц} = \frac{1}{2}\left(\bar{y}_2^k + \bar{y}_3^k\right) = \frac{1}{2}(10{,}575 + 11) = 10{,}7875,$$

$$\bar{y}_4^{Ц} = \frac{1}{2}\left(\bar{y}_3^k + \bar{y}_4^k\right) = \frac{1}{2}(11 + 11{,}275) = 11{,}1375,$$

і т. д.

Отримані значення $\bar{y}_t^{Ц}$ подамо в стовпці 5 табл. 5.15.

Крок 2. Знайдемо оцінки сезонної компоненти як різницю між фактичним рівнем вхідного ряду та центровими ковзними середніми $y_t - \bar{y}_t^{Ц}$ (табл. 5.14, стовпець 6):

$$y_3 - \bar{y}_3^{Ц} = 9{,}1 - 10{,}7875 = -1{,}6875,$$

$$y_4 - \bar{y}_4^{Ц} = 15{,}1 - 11{,}1375 = 3{,}8625,$$

і т. д.

Отримані значення в подальшому використаємо для оцінок сезонної компоненти (табл. 5.16).

Таблиця 5.16

Показники	Рік	Номер кварталу			
		I	II	III	IV
Оцінка сезонної компоненти	1	–	–	-1,6875	3,8625
	2	0,3625	-2,3	-1,0375	2,675
	3	0,5125	-2,1125	-1,4625	2,86
	4	0,675	-1,8125	–	–
Разом за *і*-ий квартал		1,55	-6,225	-4,1875	9,3875
Середня оцінка сезонної компоненти для *і*-ого кварталу, $\overline{S}_i$, $i=\overline{1,4}$		0,5167	-2,075	-1,3958	3,1291
Скорегована сезонна компонента, $\hat{S}_i, i=\overline{1,4}$		0,472917	-2,11875	-1,43958	3,085417

Знаходимо сумарні значення оцінок сезонної компоненти для кожного *і*-ого кварталу (табл. 5.16, рядок 5).

Далі знаходимо середню оцінку сезонної компоненти для *і*-ого кварталу $\overline{S}_i$, розділивши значення 5-ого рядка на 3 (табл. 5.16, рядок 6).

Знайдемо сумарне значення $\overline{S}_i$:

$$\sum_{i=1}^{4}\overline{S}_i=\overline{S}_1+\overline{S}_2+\overline{S}_3+\overline{S}_4=0{,}5167-2{,}075-1{,}3958+3{,}1291=0{,}175.$$

Перейдемо до знаходження величини коригуючого коефіцієнта за формулою:

$$k=\sum_{i=1}^{4}\overline{S}_i/4=0{,}175/4=0{,}04375.$$

Далі розрахуємо скореговані значення сезонної компоненти, як різницю між середньою оцінкою $\overline{S}_i$ та коригуючим коефіцієнтом k для *і*-ого кварталу:

$$\hat{S}_i=\overline{S}_i-k,\ i=\overline{1,4}.$$

Маємо (табл. 5.12, рядок 7):

$\hat{S}_1=\overline{S}_1-k=0{,}5167-0{,}04375=0{,}472917$,

$\hat{S}_2=\overline{S}_2-k=-2{,}075-0{,}04375=-2{,}11875$,

і т. д.

Враховуючи, що наша модель є адитивною, то повинна виконуватися умова про рівність нулю суми скорегованої сезонної компоненти: $\sum_{i=1}^{n}\hat{S}_i=0$.

Виконаємо перевірку даної умови:

$\hat{S}_1+\hat{S}_2+\hat{S}_3+\hat{S}_4=0{,}472917-2{,}11875-1{,}43958+3{,}685417=0$.

Розраховані скориговані значення сезонних компонент $\hat{S}_i(S_t)$ заносимо у табл. 5.14 стовпець 3 для відповідних кварталів.

Крок 3. Враховуючи форму адитивної моделі: $Y = T + S + E$, можемо отримати наступну рівність: $T + E = Y - S$.

Отже, маючи значення вхідного ряду Y_t та відповідні скориговані величини S_t можна знайти для кожного моменту часу сумарне значення тенденції та випадкової компоненти, як різницю між Y_t та S_t, тобто $T_t + E_t = Y_t - S_t$:

$T_1 + E_1 = Y_1 - S_1 = 10{,}1 - 0{,}4729 = 9{,}6271$,

$T_2 + E_2 = Y_2 - S_2 = 8{,}1 + 2{,}1188 = 10{,}2188$,

і т. д. (табл. 5.17, стовпець 4).

Крок 4. Для визначення компоненти T даної моделі нам необхідно провести аналітичне вирівнювання ряду $(T + E)$ з допомогою лінійного тренду:

$$T = a_0 + at,\ t - \text{час}.$$

Отже, маючи дані $T_t + E_t$ та $t = \overline{1{,}16}$, можемо скористатися процедурою «Регресія», де $(T_t + E_t)$ – залежна змінна, а t – незалежна.

Отримаємо наступне рівняння:

$$T = 10{,}28 + 0{,}1505t.$$

Для знаходження числових значень T_t необхідно у дане рівняння покласти значення $t = \overline{1{,}16}$.

Маємо:

$t = 1$, $T_1 = 10{,}28 + 0{,}1505 \cdot 1 = 10{,}4305$,

$t = 2$, $T_2 = 10{,}28 + 0{,}1505 \cdot 2 = 10{,}581$,

і т. д.

Результат заносимо в табл. 5.17, стовпець 5.

Крок 5. Враховуючи адитивний вигляд нашої моделі, визначимо значення рівнів ряду. Маючи окремо значення T_t та S_t, ми можемо знайти їх суму і записати у табл. 5.17, стовпець 6:

$T_1 + S_1 = 10{,}4305 + 0{,}4729 = 10{,}9034$,

$T_2 + S_2 = 10{,}581 - 2{,}1188 = 8{,}4622$,

і т. д.

Крок 6. Знайдемо абсолютну похибку з допомогою наступної формули $E_t = Y_t - (T_t + S_t)$:

$E_1 = Y_1 - (T_1 + S_1) = 10{,}1 - 10{,}9034 = -0{,}8034$,

$E_2 = Y_2 - (T_2 + S_2) = 8{,}1 - 8{,}4622 = -0{,}3622$, і т. д.

Результат заносимо у табл. 5.17, стовпець 7.

Нам необхідно знайти значення $\sum_{t=1}^{16}(y_t - \hat{y}_t)^2 = \sum_{t=1}^{16} E_t^2 = 4{,}020374$ (табл. 5.17, стовпець 8).

Для оцінки якості побудованої моделі нам потрібно мати значення $\sum_{t=1}^{16}(y_t - \bar{y})^2 = 77{,}46938$, де $\bar{y}$ – середнє значення вхідного ряду, $\bar{y} = \sum_{t=1}^{16} y_t / 16 = 11{,}59375$ (табл. 5.17, стовпець 9 та 10).

Таблиця 5.17

	A	B	C	D	E	F	G	H	I	J
1	t	Обсяг ППП, млрд. грн., y_t	S_t	$T_t + E_t = Y_t - S_t$	T_t	$T_t + S_t$	$E_t = Y_t - (T_t + S_t)$	E_t^2	$y_t - \bar{y}$	$(y_t - \bar{y})^2$
2	1	10,1	0,4729	9,6271	10,4305	10,9034	-0,8034	0,645	-1,494	2,231
3	2	8,1	-2,1188	10,2188	10,581	8,4622	-0,3622	0,131	-3,494	12,206
4	3	9,1	-1,44	10,54	10,7315	9,2915	-0,1915	0,037	-2,494	6,219
5	4	15	3,0854	11,9146	10,882	13,9674	1,0326	1,066	3,406	11,603
6	5	11,8	0,4729	11,3271	11,0325	11,5054	0,2946	0,087	0,206	0,043
7	6	9,2	-2,1188	11,3188	11,183	9,0642	0,1358	0,018	-2,394	5,730
8	7	10,4	-1,44	11,84	11,3335	9,8935	0,5065	0,257	-1,194	1,425
9	8	14,2	3,0854	11,1146	11,484	14,5694	-0,3694	0,136	2,606	6,793
10	9	12,1	0,4729	11,6271	11,6345	12,1074	-0,0074	0,000	0,506	0,256
11	10	9,6	-2,1188	11,7188	11,785	9,6662	-0,0662	0,004	-1,994	3,975
12	11	10,5	-1,44	11,94	11,9355	10,4955	0,0045	0,000	-1,094	1,196
13	12	15,1	3,0854	12,0146	12,086	15,1714	-0,0714	0,005	3,506	12,294
14	13	13,2	0,4729	12,7271	12,2365	12,7094	0,4906	0,241	1,606	2,580
15	14	10,8	-2,1188	12,9188	12,387	10,2682	0,5318	0,283	-0,794	0,630
16	15	11,5	-1,44	12,94	12,5375	11,0975	0,4025	0,162	-0,094	0,009
17	16	14,8	3,0854	11,7146	12,688	15,7734	-0,9734	0,948	3,206	10,280
18	Сума	185,5						4,020374		77,46938

Знайдемо величину $R^2 = 1 - \frac{\sum_{t=1}^{4k} E_t^2}{\sum_{t=1}^{4k}(y_t - \bar{y})^2} = 1 - \frac{4{,}020374}{77{,}46939} = 0{,}9481$.

Отже, можна зробити висновок, що адитивна модель пояснює 94,81% загальної варіації рівнів часового ряду надходжень ППП за 2014-2017 рр. і тільки на 5,19% іншими факторами включаючи стохастичну змінну.

Крок 7. Розрахуємо прогнозне значення обсягів поступлення ППП на перші квартали 2018 р. Спочатку знайдемо значення T_{17} при $t = 17$. Отримаємо $T_{17} = 10{,}28 + 0{,}1505 \cdot 17 = 12{,}8385$.

$Y_{прогн.}(t = 17) = T_{17} + S_{17} = 12{,}8385 + 0{,}4729 = 13{,}3114$.

5. Моделі з лаговими змінними.

Багатьом економічним процесам властивий характер запізненого ефекту від впливу деякого фактору на результативний показник.

Тобто корисність від фінансових дій проявляється не відразу у поточному моменті, а через деякий період або протягом певного часу. Приходиться проводити дослідження впливу взаємозв'язку факторів з врахуванням попередніх періодів. Все це є наслідком наявності ефекту запізнення в дії факторів або наявністю явища інерційності процесів дослідження чи відтермінування корисності.

Наприклад, прибуток банку в періоді t (y_t) залежить не тільки від обсягу наданих кредитів в періоді t (x_t), але і від кредитів, наданих у попередні періоди $t-1, t-2, \ldots, t-p$.

Вкладені кошти у цінні папери також не дають миттєвих доходів. Таке саме можна сказати і про інвестиції.

Припустимо, що залежна змінна y_t взаємопов'язана із значеннями пояснюючих змінних $\{x_{t-\tau}, \tau = \overline{0, p}\}$, тоді це твердження у формалізованій формі можна записати наступним чином:

$$y_t = f(x_t, x_{t-1}, \ldots, x_{t-p}) + u_t$$

або

$$y_t = a + a_0 x_t + a_1 x_{t-1} + \ldots + a_p x_{t-p} + u_t, \tag{5.46}$$

де $x_{t-1}, \ldots, x_{t-p}$ – лагові змінні.

Лагові змінні – пояснюючі змінні, які включені до економетричної моделі з деяким інтервалом запізнення у вигляді часового лагу. Отже, змінна $x_{t-\tau}$ у періоді $t-\tau$ буде лагова змінна, а величина τ – лаг (період зсуву).

Поряд із пояснюючими змінними в якості лагових можуть виступати й залежні змінні. Наприклад, рівень попиту на кредити або рівень ВВП в даному періоді визначається рівнями, які були у попередніх періодах.

Якщо економетрична модель включає в себе не лише поточні, а й попередні (лагові) значення незалежних змінних (x), то така модель називається дистрибутивною лаговою (або просто лаговою).

Приведемо наступну класифікацію економетричних моделей з врахуванням лагових змінних:

1) моделі з розподіленими лагами:

$$y_t = a + a_0 x_t + a_1 x_{t-1} + \ldots + a_p x_{t-p} + u_t \tag{5.47}$$

2) моделі з лаговими залежними змінними (моделі авторегресії):

$$y_t = a + b_1 y_{t-1} + b_2 y_{t-2} + \ldots + b_q y_{t-q} + u_t \tag{5.48}$$

3) моделі з лаговими залежними та незалежними змінними (авторегресійні моделі з розподіленими лагами):

$$y_t = a + a_0 x_t + a_1 x_{t-1} + \ldots + a_p x_{t-p} + b_1 y_{t-1} + b_2 y_{t-2} + \ldots + b_q y_{t-q} + u_t, \quad (5.49)$$

де p та q – величина лага.

У свою чергу перший клас моделей можна поділити в залежності від кількості попередніх періодів на такі два типи:

а) модель із скінченним числом лагів:

$$y_t = a + a_0 x_t + a_1 x_{t-1} + \ldots + a_p x_{t-p} + u_t \quad (5.50)$$

б) моделі із нескінченним числом лагів:

$$y_t = a + a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \ldots + u_t \quad (5.51)$$

Розглянемо інтерпретацію параметрів моделей з розподіленим лагом. Коефіцієнти $a_j, j = 0,1,2,\ldots$ називаються коефіцієнтами лага, а числова послідовність $\{a_j, j = 0,1,2,\ldots\}$ – структурою лага.

Значення коефіцієнта a_0 при x_t вказує на величину середньої абсолютної зміни y_t при зміні x_t на одну одиницю у періоді t і називається короткостроковим мультиплікатором $(\mu_0 = a_0)$.

Величини $\mu_1 = (a_0 + a_1)$, $\mu_2 = (a_0 + a_1 + a_2)$ і т. д. мають назву проміжних мультиплікаторів і вказують на зміну y_t протягом другого, третього і наступних періодів після зміни x_t на одну одиницю, а $\mu = \sum_{j=0}^{p} a_j$ утворює довгостроковий мультиплікатор, який характеризує максимальну сумарну зміну результативного показника через p періодів часу.

Якщо $a_j > 0, j = \overline{0, p}$, то можна розрахувати нормовані коефіцієнти лага:

$$\alpha_j = \frac{a_j}{\sum_{j=0}^{p} a_j}, j = \overline{0, p}, \quad (5.52)$$

де α_j – нормований коефіцієнт для j-ого лага.

Економічна інтерпретація нормованих коефіцієнтів полягає у встановленні пропорцій довгострокового впливу факторної ознаки у певні визначені періоди часу.

Отримана числова послідовність $\{\alpha_j, j = \overline{0, p}\}$ називається нормованою структурою лага і вона використовується в якості ваг для розрахунку середньої величини лага:

$$\bar{\tau} = \sum_{j=0}^{p} j \cdot \alpha_j . \quad (5.53)$$

Знайдене значення $\bar{\tau}$ вказує на середній інтервал часу, протягом якого буде відбуватися зміна y_t під впливом x в момент часу t. При малих значеннях τ y_t є більш чутливою до зміни x. У протилежному випадку фактор x меншими темпами впливає на результативний показник, і як наслідок його дія буде проходити протягом довгого періоду часу.

Наприклад, для моделі

$$y_t = 2{,}52 + 0{,}42x_t + 1{,}24x_{t-1} + 0{,}65x_{t-2}$$

короткостроковий мультиплікатор $\mu_0 = a_0 = 0{,}42$ і вказує на те, що при збільшенні x_t на одну одиницю приводить в середньому до збільшення y_t на 0,42 одиниці у періоді t_0. Протягом двох наступних періодів $(\mu_1 = a_0 + a_1 = 1{,}66)$ показник y_t зросте на 1,66 одиниць (проміжний мультиплікатор); через три роки $(\mu_2 = a_0 + a_1 + a_2 = 2{,}31)$ сумарна зміна y_t складе 2,31 одиниць (довгостроковий мультиплікатор).

Користуючись формулою (5.52) знайдемо нормовані коефіцієнти лага:

$$\alpha_0 = \frac{0{,}42}{2{,}31} = 0{,}18,\ \alpha_1 = \frac{1{,}24}{2{,}31} = 0{,}54,\ \alpha_2 = \frac{0{,}65}{2{,}31} = 0{,}28.$$

Можна зробити наступний висновок: на поточний період t припадає 18% впливу незалежної невідомої (x) на залежну змінну (y), через рік – 54%, через два роки – 28%. За формулою (5.53) середній лаг буде рівним:

$$\bar{\tau} = 0 \cdot 0{,}18 + 1 \cdot 0{,}54 + 2 \cdot 0{,}28 = 1{,}1 \text{ рік.}$$

Отже, протягом $\bar{\tau} = 1{,}1$ рік буде відбуватися зміна y під впливом фактора x.

Поряд із середньою величиною лагу існує медіанний лаг τ_m, який вказує на величину лага, для якого виконується умова: $\sum_{j=0}^{\tau_m} \alpha_j \approx 0{,}5$. Величина τ_m вказує на величину періоду протягом якого з моменту часу t буде реалізована половина загального впливу фактора на результат.

Розглянемо наступну виробничу ситуацію пов’язану із дослідженням впливу приросту доходу (x) та витрати на споживання (y).

Припустимо, що особа отримала підвищення заробітної плати на 600 грн. і вона поступово втрачає отриманий приріст доходу протягом певного періоду за наступним сценарієм:

1 місяць – 360 грн. (60%); 2 місяць – 120 грн. (20%) та 3 місяць – 60 грн. (10%).

У формалізованій формі даний сценарій можна описати з допомогою наступної моделі:

$$y_t = a + 0{,}6x_t + 0{,}2x_{t-1} + 0{,}1x_{t-2} + u_t \, .$$

Для нашого випадку довгостроковий мультиплікатор буде рівний 0,9 (0,6+0,2+0,1=0,9), що після підвищення приросту доходу на 1 грн. приріст на споживання становитиме 0,9 грн.

Враховуючи, що $a_j > 0\left(j = \overline{0{,}2}\right)$ можемо знайти нормовані коефіцієнти структури лага:

0,6:0,9=0,6667; 0,2:0,9=0,2222; 0,1:0,9=0,1111.

Опишемо економічний зміст отриманих значень. На поточний період припадає 66,67% загального впливу приросту доходу на приріст витрат на споживання, через місяць ця цифра становитиме 88,89%, і ще через місяць – 100%.

Важливе місце при побудові моделей розподіленого лага належить обґрунтуванню величини лага. З цією метою будується взаємна кореляційна функція у вигляді послідовності розрахованих коефіцієнтів кореляції за формулою:

$$r_\tau = \frac{(n-\tau)\sum\limits_{t=1}^{n-\tau} y_t x_{t+\tau} - \sum\limits_{t=1}^{n-\tau} y_t \sum\limits_{t=1}^{n-\tau} x_{t+\tau}}{\sqrt{\left[(n-\tau)\sum\limits_{t=1}^{n-\tau} y_t^2 - \left(\sum\limits_{t=1}^{n-\tau} y_t\right)^2\right]\cdot\left[(n-\tau)\sum\limits_{t=1}^{n-\tau} x_{t+\tau}^2 - \left(\sum\limits_{t=1}^{n-\tau} x_{t+\tau}\right)^2\right]}}, \qquad (5.54)$$

де τ – індекс лага.

Отримана числова послідовність $\left\{r_\tau, \tau = \overline{0, p}\right\}$ описує ступінь тісноти зв’язку y_t з незалежною змінною x_t, зсунутими один відносно одного на τ. За величину лага вибирається те значення τ, при якому r_τ приймає найбільше значення.

<u>Приклад</u> 5.7. Визначити величину лага для економетричної моделі взаємопов’язаних часових рядів, які характеризують обсяг випуску продукції (y, млн. грн.) та розмір інвестицій (x, млн. грн.) підприємства за останні десять років (табл. 5.18) та побудувати її модель.

Розв'язування. Для розрахунків використаємо формулу (5.54) і знайдемо значення r_τ при $\tau = 0{,}1{,}2{,}3{,}4{,}5$ і запишемо результати їх дій у табл. 5.18.

Таблиця 5.18

Рік	Обсяг виробництва продукції, млн. грн. y_t	Обсяг інвестиції, млн. грн. x_t
1	25,1	3,9
2	28,6	4,7
3	34,1	5,6
4	32,5	5,3
5	40,6	7,7
6	49,5	10,3
7	54,6	13,4
8	53,7	15,8
9	56,1	13,4
10	62,1	13,9

$$r_0 = \frac{10\sum_{t=1}^{10} y_t x_t - \sum_{t=1}^{10} y_t \sum_{t=1}^{10} x_t}{\sqrt{\left[10\sum_{t=1}^{10} y_t^2 - \left(\sum_{t=1}^{10} y_t\right)^2\right]\cdot\left[10\sum_{t=1}^{10} x_t^2 - \left(\sum_{t=1}^{10} x_t\right)^2\right]}} = 0{,}95819;$$

$$r_1 = \frac{9\sum_{t=1}^{9} y_t x_{t+1} - \sum_{t=1}^{9} y_t \sum_{t=1}^{9} x_{t+1}}{\sqrt{\left[9\sum_{t=1}^{9} y_t^2 - \left(\sum_{t=1}^{9} y_t\right)^2\right]\cdot\left[9\sum_{t=1}^{9} x_{t+1}^2 - \left(\sum_{t=1}^{9} x_{t+1}\right)^2\right]}} = -0{,}00563;$$

і т. д.

$$r_5 = \frac{5\sum_{t=1}^{5} y_t x_{t+5} - \sum_{t=1}^{5} y_t \sum_{t=1}^{5} x_{t+5}}{\sqrt{\left[5\sum_{t=1}^{5} y_t^2 - \left(\sum_{t=1}^{5} y_t\right)^2\right]\cdot\left[5\sum_{t=1}^{5} x_{t+5}^2 - \left(\sum_{t=1}^{5} x_{t+5}\right)^2\right]}} = 0{,}663725.$$

Таблиця 5.19

τ	0	1	2	3	4	5
r_τ	0,95819	-0,00563	0,888852	0,738157	0,608414	0,663725

За умови $\tau = 0$ ми отримуємо парний коефіцієнт кореляції $r_0 = 0{,}95819$. Із табл. 5.19 бачимо, що найбільше значення коефіцієнта взаємної кореляції $r_2 = 0{,}888852$ відповідає лагу $\tau = 2$. Це означає, що найбільший вплив росту обсягів виробництва буде спостерігатися через 2 роки. Модель розподіленого лага у даному випадку буде:

$$y_t = a + a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + u_t .$$

Оцінку параметрів економетричної моделі з розподіленими лагами можна виконати з допомогою методів Койка та Алмона.

Для методу Койка робиться припущення відносно коефіцієнтів a_j при лагових змінних, яке полягає у їх спаданні в геометричній прогресії:

$$a_j = a_0 \lambda^j ,\; j = 0,1,2,\ldots,\; 0 < \lambda < 1. \quad (5.55)$$

Тоді рівняння (5.47) прийме вид:

$$y_t = a + a_0 x_t + a_0 \lambda x_{t-1} + a_0 \lambda^2 x_{t-2} + \ldots + u_t \quad (5.56)$$

Як бачимо, для моделі (5.56) нам необхідно оцінити три параметри a, a_0 та λ. Звичайний МНК прямо використати немає можливості, оскільки ми маємо нелінійний випадок.

Далі нам необхідно виконати наступні процедури.

1) Визначаємо значення нової змінної

$$z_t = x_t + \lambda x_{t-1} + \lambda^2 x_{t-2} + \ldots + \lambda^p x_{t-p} \quad (5.57)$$

для кожного значення λ з малим кроком $\Delta\lambda = 0,01$ ($0 < \lambda < 1$). Величина p приймається такою, що дією наступних лагових значень x_{t-p+i} можна знехтувати.

2) Необхідно виконати оцінку рівняння регресії:

$$y_t = a + a_0 z_t + u_t . \quad (5.58)$$

3) За кінцеве значення λ приймається те, для якого коефіцієнт детермінації R^2 для рівняння (5.58). Тоді знайдені оцінки приймуть найбільше значення для a та a_0 для (5.58) будуть одночасно і для (5.56).

Крім того, для оцінки параметрів моделі (5.56) можна використати також наступні перетворення Койка.

Розглянемо модель (5.56) для періоду $(t-1)$

$$y_{t-1} = a + a_0 x_{t-1} + a_0 \lambda x_{t-2} + a_0 \lambda^2 x_{t-3} + \ldots + u_{t-1} \quad \Big| \times \lambda \quad (5.59)$$

і домножимо праву та ліву сторону на λ.

Далі від (5.56) віднімемо (5.59), одержимо:

$$\begin{aligned} y_t - \lambda y_{t-1} &= a + a_0 x_t + a_0 \lambda x_{t-1} + a_0 \lambda^2 x_{t-2} + \ldots + u_t - \\ &- \lambda a - a_0 \lambda x_{t-1} - a_0 \lambda^2 x_{t-2} - a_0 \lambda^3 x_{t-3} - \ldots - \lambda u_{t-1} = \\ &= a(1-\lambda) + a_0 x_t + \lambda y_{t-1} + (u_t - \lambda u_{t-1}) \end{aligned} \quad (5.60)$$

$$y_t = a(1-\lambda) + a_0 x_t + \lambda y_{t-1} + (u_t - \lambda u_{t-1})$$

Рівняння (5.60) представляє собою авторегресійну модель першого порядку. Знайшовши оцінки a, a_0 та λ, ми тим самим можемо записати економетричну модель виду (5.56).

В основі методу Алмона лежить гіпотеза про те, що коефіцієнти a_j апроксимуються поліномом відповідного ступеня від величини лага:

$$a_j = b_0 + b_1 j + b_2 j^2 + \ldots + b_p j^p \quad j = \overline{0, p}. \tag{5.61}$$

Розглянемо апроксимацію параметрів регресії з максимальним лагом $\tau = 3$:

$$y_t = a + a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + a_3 x_{t-3} + u_t. \tag{5.62}$$

Приймемо, що порядок апроксимації $p = 2$, тоді параметри моделі будуть:

$$a_j = b_0 + b_1 j + b_2 j^2. \tag{5.63}$$

Одержимо:

$$\begin{aligned} a_0 &= b_0 \\ a_1 &= b_0 + b_1 + b_2 \\ a_2 &= b_0 + 2b_1 + 4b_2 \\ a_3 &= b_0 + 3b_1 + 9b_2 \end{aligned} \tag{5.64}$$

Підставимо значення a_0, a_1, a_2 та a_3 у рівняння (5.62).

Одержимо:

$$\begin{aligned} y_t &= a + b_0 x_t + (b_0 + b_1 + b_2)x_{t-1} + (b_0 + 2b_1 + 4b_2)x_{t-2} + (b_0 + 3b_1 + 9b_2)x_{t-3} + \\ &+ u_t = a + b_0(x_t + x_{t-1} + x_{t-2} + x_{t-3}) + b_1(x_{t-1} + 2x_{t-2} + 3x_{t-3}) + \\ &+ b_2(x_{t-2} + 4x_{t-2} + 9x_{t-3}) + u_t \end{aligned} \tag{5.65}$$

Виконаємо заміну:

$$\begin{aligned} z_{0t} &= x_t + x_{t-1} + x_{t-2} + x_{t-3}, \\ z_{1t} &= x_{t-1} + 2x_{t-2} + 3x_{t-3}, \\ z_{2t} &= x_{t-1} + 4x_{t-2} + 9x_{t-3}. \end{aligned} \tag{5.66}$$

Одержимо наступне рівняння регресії:

$$y_t = a + b_0 z_{0t} + b_1 z_{1t} + b_2 z_{2t} + u_t. \tag{5.67}$$

Розрахувавши значення z_{0t}, z_{1t} та z_{2t} можна знайти оціночні значення параметрів a, b_0, b_1 та b_2 і після чого повернутися до початкових параметрів a_0, a_1, a_2 та a_3.

<u>Приклад</u> 5.8. Побудувати економетричну модель залежності величини прибутку банку (y_t, млн. грн.) та розміру наданих кредитів (x_t, млн. грн.) за останні 16 місяців (табл. 5.20).

Розв'язування. Використаємо алгоритм методу Алмона. Припустимо, що прибуток банку (y_t) залежить від величини виділених кредитів (x_t) у поточному і трьох попередніх місяцях:

$$y_t = a + a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + a_3 x_{t-3} + u_t . \qquad (5.68)$$

Використаємо формулу (5.63), (5.64) та виконаємо заміну (5.66).

Нам необхідно знайти оцінки для рівняння (5.67). Для цього необхідно розрахувати за формулою (5.66) значення нових змінних z_{0t}, z_{1t} та z_{2t} і записати їх у табл. 5.20.

Як бачимо, їх значення у перших трьох періодах будуть відсутніми. Нам необхідно буде знайти оцінки із врахуванням залежностей за 13 останніх місяці (з 4 по 16 місяці).

Таблиця 5.20

	A	B	C	D	E	F
1	Місяць	Прибуток, млн. грн. y_t	Кредити, млн. грн. x_t	z_{0t}	z_{1t}	z_{2t}
2	1	65,8	38,1			
3	2	75,2	40,6			
4	3	78,1	43,4			
5	4	74,9	40,1	162,2	238,9	548,7
6	5	80,1	46,1	170,2	248,7	579,1
7	6	81,2	47,2	176,8	256,5	597,1
8	7	82,4	48,5	181,9	259,7	592,5
9	8	88,1	49,2	191	281,2	652,2
10	9	91,2	51,9	196,8	287,8	668
11	10	95,6	56,1	205,7	295,8	685,2
12	11	103,4	62,1	219,3	307,5	706,5
13	12	115,6	67,2	237,3	330	753,6
14	13	126,1	68,9	254,3	359,7	820,5
15	14	135,8	76,2	274,4	389,6	896,6
16	15	146,1	80,4	292,7	415,6	956,6
17	16	151,2	79,4	304,9	439,5	1005,3

Використовуючи стандартні програмні продукти, отримуємо: $a = -16{,}524$, $b_0 = 0{,}582$, $b_1 = 0{,}0025$, $b_2 = -0{,}0088$.

Отже, рівняння (5.67) буде:

$y_t = -16{,}524 + 0{,}582 z_{0t} + 0{,}0025 z_{1t} - 0{,}0088 z_{2t} + u_t$, $R = 0{,}9985$.

Маючи значення b_0, b_1 та b_2 ми можемо знайти значення a_0, a_1, a_2 та a_3:

$$a_0 = b_0 = 0{,}582,$$

$$a_1 = b_0 + b_1 + b_2 = 0{,}582 + 0{,}0025 - 0{,}0088 = 0{,}5757,$$

$$a_2 = b_0 + 2b_1 + 4b_2 = 0{,}582 + 2 \cdot 0{,}0025 - 4 \cdot 0{,}0088 = 0{,}5518,$$

$$a_3 = b_0 + 3b_1 + 9b_2 = 0{,}582 + 3 \cdot 0{,}0025 - 9 \cdot 0{,}0088 = 0{,}5103.$$

Отже, наша економетрична модель розподіленого лагу прийме вид:

$$\hat{y}_t = -16{,}524 + 0{,}582x_t + 0{,}5757x_{t-1} + 0{,}5518x_{t-2} + 0{,}5103x_{t-3}, \ R = 0{,}9985.$$

Завдання 6. Природа автокореляції та методи її усунення.

Однією з передумов класичного регресійного аналізу є припущення про незалежність залишків економетричної моделі. Якщо дана умова порушується, тобто має місце: $M(u_i u_j) \neq 0$, при $i \neq j$, то як наслідок виникає явище автокореляції.

Під автокореляцією будемо розуміти взаємозв'язок послідовних елементів часового чи просторового ряду.

Основними причинами автокореляції є:

- помилки специфікації;
- інерція;
- ефект павутини;
- згладжування даних.

Рис. 6.1. Види автокореляції

Нехтування автокореляцією залишків при оцінюванні параметрів моделі може привести до наступних наслідків:

1. Оцінки параметрів моделі можуть бути незміщеними, але неефективними, тобто вибіркові дисперсії вектора оцінок A можуть бути невиправдано великими.
2. Статистичні критерії t та F-статистики, які розраховані для лінійної моделі, практично не можуть бути використаними в

дисперсійному аналізі, оскільки вибіркові дисперсії знаходяться не за уточненими формулами.

3. Неефективність оцінок параметрів економетричної моделі породжує, як правило, прогнози з дуже великою вибірковою дисперсією (неефективні прогнози).

Розглянемо загальний випадок часового ряду у вигляді лінійної регресії:

$$y_t = a_0 + a_1 x_{t1} + \ldots + a_m x_{tm} + u_t,\ t = \overline{1,n}. \tag{6.1}$$

Якщо випадкові залишки в різні моменти часу корелюють між собою, і не виконується третя умова Гаусса-Маркова, то між значеннями випадкових залишків у різні моменти часу існує залежність, яка для загального випадку буде:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots + \rho_s u_{t-s} + e_t, \tag{6.2}$$

де ρ_i називається i-м коефіцієнтом автокореляції, $i = \overline{1,s}$; e_t – випадкова величина; t – індекс періоду, $t = \overline{1,n}$.

Як бачимо, випадкова величина u_t визначається значенням цієї ж самої величини з запізненням, максимальний порядок якого становить S.

Отже, при виконанні рекурентного співвідношення (6.2) кажуть, що послідовність u_t утворює авторегресійний процес S-го порядку.

Зауважимо, що значення коефіцієнта ρ_i невідоме і в процесі дослідження підлягає оцінюванню. Крім того, вважають, що інші умови Гаусса-Маркова для випадкової величини u_t із рівності (6.1) є справедливими:

$$M(u_t) = 0,\ D(u_t) = \sigma_u^2 = const,\ t = \overline{1,n}.$$

Відносно випадкової величини e_t так само припускаємо виконання всіх умов Гаусса-Маркова.

Розглянемо більш детальніше проблему автокореляції залишків для випадку двох змінних. Тобто, наша економетрична модель має вигляд:

$$y_t = a_0 + a_1 x_t + u_t,\ t = \overline{1,n}, \tag{6.3}$$

і для неї виконується умова:

$$u_t = \rho u_{t-1} + e_t,\ t = \overline{1,n}. \tag{6.4}$$

Отже, залишок u_t утворює автокореляцію першого порядку. Якщо $\rho>0$, то автокореляція рахується додатною, а якщо $\rho<0$ – від’ємною. Додатна автокореляція виникає, якщо залишки характеризуються трендом або циклічними коливаннями, а від’ємна – якщо залишки приймають то додатні, то від’ємні значення. При $\rho=0$

автокореляція відсутня. Крім того $|\rho|<1$. Коефіцієнт ρ характеризує силу зв'язку залишків у період t від їх рівня в періоді $t-1$.

Автокореляція 2-ого порядку має вид:

$$u_t=\rho_1 u_{t-1}+\rho_2 u_{t-2}+e_t, \qquad (6.5)$$

Для загального випадку маємо:

$$u_t=\rho_1 u_{t-1}+\rho_2 u_{t-2}+\ldots+\rho_s u_{t-s}+e_t, \qquad (6.6)$$

де ρ_i – i-й коефіцієнт автокореляції, $i=\overline{1,s}$.

Фальшиву автокореляцію можна представити наступним чином:

$$y_t=a_0+a_1 x_{t1}+a_2 x_{t2}+u_t, \qquad (6.7)$$

$$y_t=a_0+a_1 x_{t1}+u_t^*, \qquad (6.8)$$

де

$$u_t^*=f\left(a_2 x_{t2}+u_t\right) \qquad (6.9)$$

Як бачимо сама величина x_2 є автокорельованою змінною.

Перевірку наявності автокореляції можна здійснити за допомогою наступних критеріїв:

- критерій Дарбіна-Уотсона;
- критерій фон Неймана;
- нециклічний коефіцієнт автокореляції;
- циклічний коефіцієнт автокореляції.

Критерій Дарбіна-Уотсона. При використанні даного критерія формулюється нульова гіпотеза про відсутність автокореляції – $H_0:\rho=0$. Альтернативна гіпотеза може бути побудованою на основі використання односторонньої критичної області – $H_1:\rho>0$, тобто існує додатна, або $H_1:\rho<0$, тобто існує від'ємна автокореляція, або на основі двосторонньої критичної області – $H_1:\rho\neq 0$.

Статистика Дарбіна-Уотсона визначається за формулою:

$$DW=d=\frac{\sum_{t=2}^{n}\left(u_t-u_{t-1}\right)^2}{\sum_{t=1}^{n}u_t^2} \qquad (6.10)$$

при виконанні наступних умов:

1) випадкові помилки u_t розподілені нормально;

2) регресія містить постійний член (неоднорідна);

3) матриця X (матриця незалежних змінних) нестатична.

Між статистикою d і коефіцієнтом авторегресії ρ існує наближене співвідношення:

$$d\approx 2-2\rho. \qquad (6.11)$$

Оскільки $-1<\rho<1$, то для загального випадку має місце $0<d<4$.

Розподіл статистики d (6.10), як випадкової величини, залежить від кількості спостережень у вибірці n, числа пояснюючих змінних у рівнянні регресії m, конкретних значень пояснюючих змінних і рівня значущості α. Ці обставини свідчать про неможливість побудови таблиці критичних значень d-статистики. Тому Дарбіном і Уотсоном було доведено, що існує нижня та верхня границі для критичного значення d, які не залежать від значень вибірки, а визначаються тільки об’ємом вибірки n, числом пояснюючих змінних m і рівнем значущості α.

Позначимо нижню границю величини d через $d_н$,, а верхню – через $d_в$. Тому в таблиці d-статистики приведені значення $d_н$ та $d_в$ для рівнів значущості α=0,01 (α=0,025, α=0,05).

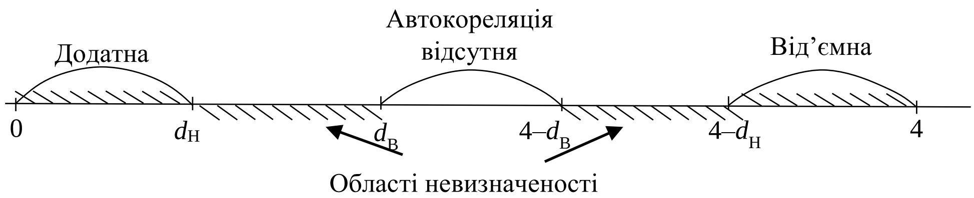


Рис. 6.2. Області автокореляційного зв’язку для критерію Дарбіна-Уотсона

Критерій фон Неймана. Алгоритм даного критерію полягає у знаходженні величини Q за формулою:

$$Q=\frac{\sum_{t=2}^{n}\left(u_t-u_{t-1}\right)^2 \Big/ (n-1)}{\sum_{t=1}^{n}u_t^2 \Big/ n}=\frac{\sum_{t-2}^{n}\left(u_t-u_{t-1}\right)^2}{\sum_{t=1}^{n}u_t^2}\cdot\frac{n}{n-1}=d\,\frac{n}{n-1}. \tag{6.12}$$

При великому числі спостережень маємо:

$$\lim_{n\to\infty}Q=\lim_{n\to\infty}d\,\frac{n}{n-1}=d\,, \tag{6.13}$$

тобто критерій Дарбіна-Уотсона та фон Неймана співпадають.

Зауважимо, що критерій Неймана дає лише точкову оцінку. Для цього необхідно знайти табличне значення даного критерію $Q_{табл}$ при заданому рівні значущості α та числі спостережень n. Якщо $Q_{факт}<Q_{табл}$, то має місце додатна автокореляція.

Нециклічний коефіцієнт автокореляції. При перевірці автокореляції можна використати також нециклічний коефіцієнт автокореляції для одиничного лага, який виражає ступінь взаємозв'язку залишків кожного наступного значення з попереднім, тобто ряду $u_1, u_2, ..., u_{n-1}$ з рядом $u_2, u_3, ..., u_n$.

Для знаходження кількісного значення даного коефіцієнта використаємо формулу:

$$r^* = \frac{\sum_{t=2}^{n}(u_t u_{t-1}) - \frac{1}{n-1}\left(\sum_{t=2}^{n}u_t\right)\cdot\left(\sum_{t=2}^{n}u_{t-1}\right)}{\sqrt{\left[\sum_{t=2}^{n}u_t^2 - \frac{1}{n-1}\left(\sum_{t=2}^{n}u_t\right)^2\right]\left[\sum_{t=2}^{n}u_{t-1}^2 - \frac{1}{n-1}\left(\sum_{t=2}^{n}u_{t-1}\right)^2\right]}}, \quad (6.14)$$

причому $\left|r^*\right| < 1$.

Якщо $r^* \in (-1;0)$, то має місце від'ємна автокореляція залишків, а при $r^* \in (0;1)$ – додатна. Якщо r^* лежить в як завгодно малому околі точки нуль, то автокореляція відсутня.

Використання нециклічного коефіцієнта автокореляції пов'язане з певними труднощами встановлення імовірнісного розподілу r^*. Через це в прикладних дослідженнях часто використовують циклічний коефіцієнт автокореляції r^0.

Циклічний коефіцієнт автокореляції. Величина циклічного коефіцієнта автокореляції r^0 виражає ступінь взаємозв'язку двох рядів $u_1, u_2, ..., u_{n-1}, u_n$ та $u_2, u_3, ..., u_n, u_1$ і обчислюється за формулою:

$$r^0 = \frac{\sum_{t=2}^{n}(u_t u_{t-1}) + u_n u_1 - \frac{1}{n}\left(\sum_{t=1}^{n}u_t\right)^2}{\sum_{t=1}^{n}u_t^2 - \frac{1}{n}\left(\sum_{t=1}^{n}u_t\right)^2}. \quad (6.15)$$

У табл. 6.1 приведені критичні значення циклічних коефіцієнтів автокореляції для одного лагу. Дана таблиця дає можливість зробити перевірку гіпотези про рівність нулю коефіцієнта автокореляції для будь-якого лагу.

Таблиця 6.1.

Критичні значення циклічного коефіцієнта автокореляції r^0

Число спостережень	*Додатна автокореляція*		*Від'ємна автокореляція*	
	α=0,05	α=0,01	α=0,05	α=0,01
10	0,360	0,525	-0,564	-0,705
15	0,328	0,475	-0,462	-0,597
20	0,299	0,432	-0,399	-0,524
25	0,276	0,398	-0,356	-0,473
30	0,257	0,370	-0,325	-0,433

При великих n $(n\to\infty)$ ми маємо справу з досить довгими рядами. У такому випадку вплив циклічних рядів на кількісну величину r^0 незначний і можна припустити, що імовірнісний розподіл r^* наближається до розподілу r^0.

Розглянемо ряд окремих випадків. Припустимо, що перший і останній члени ряду співпадають, тобто має місце $u_1=u_n$. Тоді, порівнюючи формули (6.14) та (6.15), приходимо до висновку, що $r^*= r^0$.

Враховуючи вибраний рівень значущості α та довжину ряду n, знаходимо $r^0_{табл}$, порівнюємо його з $r^0_{факт}=r^0$. Якщо $r^0_{факт}\ge r^0_{табл}$, то можна стверджувати про існування автокореляційних залишків.

Для подальших досліджень покладемо, що $\overline{u}=0$, тобто середнє залишків дорівнює нулю, і як наслідок будемо мати $\sum_{t=2}^{n}u_t\approx\sum_{t=2}^{n}u_{t-1}\approx 0$.

Враховуючи дане припущення, отримаємо наближену формулу для обрахунку циклічного коефіцієнта автокореляції:

$$r^0\approx\frac{\sum_{t=2}^{n}u_t\cdot u_{t-1}}{\sum_{t=1}^{n}u_t^2}. \qquad (6.16)$$

При проведенні прикладних досліджень часом доцільно замість циклічного коефіцієнта автокореляції (6.15) використовувати коефіцієнт автокореляції виду:

$$r'=\frac{\frac{1}{n-1}\sum_{t=2}^{n}u_t\cdot u_{t-1}}{\frac{1}{n}\sum_{t=1}^{n}u_t^2}\ . \qquad (6.17)$$

Якщо з допомогою перелічених критеріїв встановлена автокореляція залишків, то перед нами стоїть завдання виявити можливі причини її появи і далі побудувати таку модель, для якої буде менш ймовірною поява автокореляції залишків.

Приклад 6.1. Дослідити наявність автокореляції, використовуючи статистичні дані (табл. 6.2).

Таблиця 6.2.

/п								
г, шт, Y	5012	4968	5412	4768	6234	5987	4653	7945
од. товару, грн.,	560	620	595	535	360	565	600	275
ı								
г, шт, Y	5536	4812	5237	6978	4556	5768	6034	
од. товару, грн.,	540	665	485	350	680	450	380	

Розв'язування. Використовуючи стандартний пакет EXCEL (регресія), знайдемо економетричну модель:

$$\hat{Y} = 9028{,}8971 - 6{,}7276x.$$

Обчислимо статистику Дарбіна-Уотсона за формулою (6.10).

Для цього розширимо початкову таблицю, додавши додаткові стовпці (табл. 6.3).

Таблиця 6.3.

№ п/п	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Попит, шт, Y	5012	4968	5412	4768	6234	5987	4653	7945
Ціна од. товару, грн., X	560	620	595	535	360	565	600	275
$\hat{Y}$	5261,438	4857,782	5025,972	5429,628	6606,959	5227,8	4992,334	7178,806
u_i	-249,438	110,2182	386,028	-661,628	-372,959	759,1999	-339,334	766,1943
u_i^2	62219,38	12148,05	149017,7	437752	139098,6	576384,5	115147,5	587053,8
№ п/п	**9**	**10**	**11**	**12**	**13**	**14**	**15**	
Попит, шт, Y	5536	4812	5237	6978	4556	5768	6034	
Ціна од. товару, грн., X	540	665	485	350	680	450	380	
$\hat{Y}$	5395,99	4555,04	5766,009	6674,235	4454,126	6001,475	6472,407	
u_i	140,0098	256,9604	-529,009	303,7647	101,8745	-233,475	-438,407	
u_i^2	19602,73	66028,66	279850	92273,02	10378,41	54510,45	192200,8	

Із табл. 6.3 $\sum_{t=1}^{15} u_t^2 = 2793665{,}464$. Обчислимо:

$$\sum_{t=2}^{15}(u_t - u_{t-1})^2 = (110{,}2182 - (-249{,}438))^2 + (386{,}028 - 110{,}2182)^2 + (-661{,}628 - 386{,}028)^2 +$$
$$+(-372{,}959 - (-661{,}628))^2 + (759{,}1999 - (-372{,}959))^2 + (-339{,}334 - 759{,}1999)^2 +$$
$$+(766{,}1943 - (-339{,}334))^2 + (140{,}0098 - 766{,}1943)^2 + (256{,}9604 - 140{,}0098)^2 +$$
$$+(-529{,}009 - 256{,}9604)^2 + (303{,}7647 - (-529{,}009))^2 + (101{,}8745 - 303{,}7647)^2 +$$
$$+(-233{,}475 - 101{,}8745)^2 + (-438{,}407 - (-233{,}475))^2 = 129352{,}7 + 76071{,}08 + 1097584 +$$
$$+83329{,}83 + 1281784 + 1206777 + 1222193 + 392107{,}1 + 13677{,}46 + 617747{,}2 + 693511{,}3 +$$
$$+40759{,}67 + 112459{,}1 + 41997{,}28 = 7009350{,}046.$$

Тоді d= 7009350,046 / 2793665,464 ≈ 2,509016.

Використовуючи табличні значення статистики Дарбіна-Уотсона для n=15 та α=0,05 при двох факторах-аргументах, одержуємо $d_н$=0,95, $d_в$=1,54. Здійснюючи порівняльний аналіз, робимо висновок, що залишки мають від'ємну автокореляцію.

Отже, в побудованій нами економетричній моделі виду (6.3) та (6.4) виявлено наявність автокореляції у залишках. Нам необхідно здійснити перерахунок оціночних параметрів або зробити спробу покращити специфікацію моделі.

З формули (6.4) знайдемо:

$$e_t = u_t - \rho u_{t-1} \tag{6.18}$$

Запишемо (6.3) для періоду $(t-1)$:

$$y_{t-1} = a_0 + a_1 x_{t-1} + u_{t-1}, \quad \left| \times \rho \right. \tag{6.19}$$

Далі помножимо ліву і праву частину (6.19) на величину ρ, і одержану рівність віднімемо від (6.3).

У результаті таких дій отримаємо наступну економетричну модель:

$$y_t - \rho y_{t-1} = a_0(1-\rho) + a_1(x_t - \rho x_{t-1}) + (u_t - \rho u_{t-1}). \tag{6.20}$$

Нам необхідно знайти оцінки для ρ, a_0 та a_1. Як бачимо, нами отримано задачу нелінійного програмування.

Для оцінки параметрів моделі з автокорельованими залишками можна використати наступні методи:

1) узагальнений метод Ейткина;
2) метод перетворення вхідної інформації;
3) метод перших різниць;
4) метод Кохрана-Орката;
5) метод Дарбіна;
6) метод Халдрета-Лу.

Алгоритм узагальненого методу Ейткина.

1. Оцінка параметрів МНК.
2. Дослідження з допомогою відповідних критеріїв на наявність автокореляції.
3. Формування матриці коваріації залишків S.

$$S = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & 1 \end{bmatrix}. \tag{6.21}$$

Значення ρ знайдемо із формули:

$$\rho \approx r'_{скор.} = \frac{n}{n-1} \cdot \frac{\sum_{t=2}^{n} u_t u_{t-1}}{\sum_{t=1}^{n} u_t^2} + \frac{m+1}{n}, \tag{6.22}$$

де $r'_{скор.}$ – скорегований циклічний коефіцієнт кореляції r' на величину зміщення $\frac{m+1}{n}$.

4. Знаходження оберненої матриці S^{-1}:

$$S^{-1} = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & -\rho & 0 & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & \cdots & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{6.23}$$

5. Оцінка параметрів методом Ейткина. Оператор оцінювання має вид:

$$A = \left(X'S^{-1}X\right)^{-1} X'S^{-1}Y \tag{6.24}$$

6. Отримавши нову економетричну модель переходимо до пункту 2.

<u>Приклад</u> 6.2. Оцінити параметри моделі з автокорельованими залишками прикладу 6.1 методом Ейткена.

Розв'язування. Використаємо метод Ейткена для оцінювання параметрів економетричної моделі з автокорельованими залишками.

Використовуючи формулу (6.22) маємо:

$$\rho \approx r'_{скор.} = \frac{15}{15-1} \cdot \frac{\sum_{t=2}^{15} u_t u_{t-1}}{\sum_{t=1}^{15} u_t^2} + \frac{1+1}{15}.$$

Знайдемо суму добутків $\sum_{t=2}^{15} u_t u_{t-1}$, для цього доповнимо табл. 6.3 прикладу 6.1:

Таблиця 6.4.

ı								
	-249,438	110,2182	386,028	-661,628	-372,959	759,1999	-339,334	766,1943
	62219,38	12148,05	149017,7	437752	139098,6	576384,5	115147,5	587053,8
$u_t u_{t-1}$	–	-27492,62	42547,31	-255407,07	246760,35	-283150,58	-257622,28	-259995,74
ı								
	140,0098	256,9604	-529,009	303,7647	101,8745	-233,475	-438,407	
	19602,73	66028,66	279850	92273,02	10378,41	54510,45	192200,8	
$u_t u_{t-1}$	107274,68	35976,97	-135934,26	-160694,14	30945,88	-23785,12	102356,97	

Із табл. 6.4 $\sum_{t=2}^{15} u_t u_{t-1} = -838219,64$.

Таким чином $\rho \approx \frac{15}{14} \cdot \frac{-838219{,}64}{2793665{,}464} + \frac{2}{15} = 1{,}0714 \cdot (-0{,}30) + 0{,}133 \approx -0{,}188141$.

Отже, матриця S матиме вигляд:

$$S = \begin{bmatrix}
1 & -0{,}1881 & 0{,}0354 & -0{,}0067 & 0{,}0013 & -0{,}0002 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0{,}1881 & 1 & -0{,}1881 & 0{,}0354 & -0{,}0067 & 0{,}0013 & -0{,}0002 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0{,}0354 & -0{,}1881 & 1 & -0{,}1881 & 0{,}0354 & -0{,}0067 & 0{,}0013 & -0{,}0002 & 0 & 0 & 0 & 0 & 0 & 0 \\
-0{,}0067 & 0{,}0354 & -0{,}1881 & 1 & -0{,}1881 & 0{,}0354 & -0{,}0067 & 0{,}0013 & -0{,}0002 & 0 & 0 & 0 & 0 & 0 \\
0{,}0013 & -0{,}0067 & 0{,}0354 & -0{,}1881 & 1 & -0{,}1881 & 0{,}0354 & -0{,}0067 & 0{,}0013 & -0{,}0002 & 0 & 0 & 0 & 0 \\
-0{,}0002 & 0{,}0013 & -0{,}0067 & 0{,}0354 & -0{,}1881 & 1 & -0{,}1881 & 0{,}0354 & -0{,}0067 & 0{,}0013 & -0{,}0002 & 0 & 0 & 0 \\
0 & -0{,}0002 & 0{,}0013 & -0{,}0067 & 0{,}0354 & -0{,}1881 & 1 & -0{,}1881 & 0{,}0354 & -0{,}0067 & 0{,}0013 & -0{,}0002 & 0 & 0 \\
0 & 0 & -0{,}0002 & 0{,}0013 & -0{,}0067 & 0{,}0354 & -0{,}1881 & 1 & -0{,}1881 & 0{,}0354 & -0{,}0067 & 0{,}0013 & -0{,}0002 & 0 \\
0 & 0 & 0 & -0{,}0002 & 0{,}0013 & -0{,}0067 & 0{,}0354 & -0{,}1881 & 1 & -0{,}1881 & 0{,}0354 & -0{,}0067 & 0{,}0013 & -0{,}0002 \\
0 & 0 & 0 & 0 & -0{,}0002 & 0{,}0013 & -0{,}0067 & 0{,}0354 & -0{,}1881 & 1 & -0{,}1881 & 0{,}0354 & -0{,}0067 & 0{,}0013 \\
0 & 0 & 0 & 0 & 0 & -0{,}0002 & 0{,}0013 & -0{,}0067 & 0{,}0354 & -0{,}1881 & 1 & -0{,}1881 & 0{,}0354 & -0{,}0067 \\
0 & 0 & 0 & 0 & 0 & 0 & -0{,}0002 & 0{,}0013 & -0{,}0067 & 0{,}0354 & -0{,}1881 & 1 & -0{,}1881 & 0{,}0354 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -0{,}0002 & 0{,}0013 & -0{,}0067 & 0{,}0354 & -0{,}1881 & 1 & -0{,}1881 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -0{,}0002 & 0{,}0013 & -0{,}0067 & 0{,}0354 & -0{,}1881 & 1
\end{bmatrix}$$

Знайдемо обернену матрицю до S, використовуючи функцію МОБР (MINVERSE) в MS Excel:

$$S^{-1} = \begin{bmatrix}
1{,}0367 & 0{,}195 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0{,}195 & 1{,}0734 & 0{,}195 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0{,}195 & 1{,}0734 & 0{,}195 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0{,}195 & 1{,}0734 & 0{,}195 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0{,}195 & 1{,}0734 & 0{,}195 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0{,}195 & 1{,}0734 & 0{,}195 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0{,}195 & 1{,}0734 & 0{,}195 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0{,}195 & 1{,}0734 & 0{,}195 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}195 & 1{,}0734 & 0{,}195 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}195 & 1{,}0734 & 0{,}195 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}195 & 1{,}0734 & 0{,}195 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}195 & 1{,}0734 & 0{,}195 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}195 & 1{,}0734 & 0{,}195 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}195 & 1{,}0367
\end{bmatrix}$$

За умовою задачі 6.1 матриці Y та X матимуть вигляд:

$$Y = \begin{bmatrix} 5012 \\ 4968 \\ 5412 \\ 4768 \\ 6234 \\ 5987 \\ 4653 \\ 7945 \\ 5536 \\ 4812 \\ 5237 \\ 6978 \\ 4556 \\ 5768 \\ 6034 \end{bmatrix}; \qquad X = \begin{bmatrix} 1 & 560 \\ 1 & 620 \\ 1 & 595 \\ 1 & 535 \\ 1 & 360 \\ 1 & 565 \\ 1 & 600 \\ 1 & 275 \\ 1 & 540 \\ 1 & 665 \\ 1 & 485 \\ 1 & 350 \\ 1 & 680 \\ 1 & 450 \\ 1 & 380 \end{bmatrix}.$$

Знаходимо транспоновану матрицю X' використовуючи табличний процесор MS Excel із допомогою функції ТРАНСП (TRANSPOSE):

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 560 & 620 & 595 & 535 & 360 & 565 & 600 & 275 & 540 & 665 & 485 & 350 & 680 & 450 & 380 \end{bmatrix}.$$

Знаходимо добуток двох матриць X' та S^{-1}, використавши функцію МУМНОЖ (MMULT) в MS Excel:

$$X' S^{-1} = \begin{bmatrix} 1{,}23 & 1{,}46 & 1{,}46 & 1{,}46 & 1{,}46 & 1{,}46 & 1{,}46 & 1{,}46 & 1{,}46 & 1{,}46 & 1{,}46 & 1{,}46 & 1{,}46 & 1{,}46 & 1{,}23 \\ 701{,}48 & 890{,}78 & 863{,}95 & 760{,}53 & 600{,}97 & 793{,}71 & 807{,}87 & 517{,}53 & 762{,}97 & 913{,}73 & 718{,}57 & 602{,}92 & 885{,}94 & 689{,}77 & 481{,}71 \end{bmatrix}$$

Аналогічним чином знаходимо добуток матриць $X' S^{-1} X$ та $X' S^{-1} Y$:

$$X' S^{-1} X = \begin{bmatrix} 21{,}49 & 10992{,}44 \\ 10992{,}44 & 5832937{,}03 \end{bmatrix}; \qquad X' S^{-1} Y = \begin{bmatrix} 120226{,}3877 \\ 60124937{,}7089 \end{bmatrix}.$$

За допомогою функції МОБР (MINVERSE) в MS Excel знаходимо $\left(X' S^{-1} X\right)^{-1}$:

$$\left(X' S^{-1} X\right)^{-1} = \begin{bmatrix} 1{,}2937 & \text{-}0{,}0024 \\ \text{-}0{,}0024 & 0 \end{bmatrix}.$$

Отже, згідно формули (6.24) та, використавши функцію МУМНОЖ (MMULT) в MS Excel, знайдемо оператор оцінювання А:

$$A = \left(X' S^{-1} X\right)^{-1} X' S^{-1} Y = \begin{bmatrix} 1{,}2937 & \text{-}0{,}0024 \\ \text{-}0{,}0024 & 0 \end{bmatrix} \cdot \begin{bmatrix} 120226{,}3877 \\ 60124937{,}7089 \end{bmatrix} = \begin{bmatrix} 8949{,}8251 \\ \text{-}6{,}5585 \end{bmatrix}.$$

Отже, економетрична модель матиме вигляд:

$$\hat{Y}^* = 8949{,}8251 - 6{,}5585x.$$

Знайдемо розрахункові значення $\hat{Y}^*$ на основі побудованої економетричної моделі та визначимо залишки (табл. 6.5):

Таблиця 6.5.

№ п/п	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Попит, шт, Y	5012	4968	5412	4768	6234	5987	4653	7945
$\hat{Y}^*$	5277,05	4883,54	5047,50	5441,01	6588,75	5244,26	5014,71	7146,23
u_i^*	-265,05	84,46	364,50	-673,01	-354,75	742,74	-361,71	798,77
$\left(u_i^*\right)^2$	70250,90	7133,98	132859,99	452945,28	125850,88	551668,32	130832,47	638034,10
$u_t^* u_{t-1}^*$		-22386,79	30786,69	-245312,67	238754,18	-263491,82	-268656,15	-288921,40
№ п/п	**9**	**10**	**11**	**12**	**13**	**14**	**15**	
Попит, шт, Y	5536	4812	5237	6978	4556	5768	6034	
$\hat{Y}^*$	5408,22	4588,40	5768,94	6654,34	4490,03	5998,49	6457,58	
u_i^*	127,78	223,60	-531,94	323,66	65,97	-230,49	-423,58	
$\left(u_i^*\right)^2$	16327,87	49995,47	282958,62	104755,82	4352,65	53124,28	179423,48	
$u_t^* u_{t-1}^*$	102067,32	28571,31	-118939,69	-172167,25	21353,35	-15206,29	97630,65	

Обчислимо статистику Дарбіна-Уотсона, врахувавши, що $\sum_{t=1}^{15} u_t^2 = 2800514{,}1$ та

$$\sum_{t=2}^{15}\left(u_t - u_{t-1}\right)^2 = 122158{,}46 + 78420{,}5954 + 1076430{,}611 + 101287{,}789 + 1204502{,}837 +$$
$$+ 1219813{,}083 + 1346709{,}363 + 450227{,}3319 + 9180{,}7298 + 570833{,}469 + 732048{,}939 + :$$
$$+ 66401{,}781 + 87889{,}522 + 37286{,}46299 = 7103190{,}968.$$

Тому d^*= 7103190,968 / 2800514,1 ≈ 2,5364.

Врахувавши, що $d_н$=0,95, $d_в$=1,54, та, здійснивши порівняльний аналіз, робимо висновок, що ми не звільнились від автокореляції залишків.

Алгоритм методу перетворення вхідної інформації.

1. Оцінка параметрів МНК.
2. Дослідження на наявність автокореляції.
3. Перетворення вхідної інформації та знаходження параметра ρ.

Робимо заміну у формулі (6.20):

$$y_t^* = y_t - \rho y_{t-1}, \qquad x_t^* = x_t - \rho x_{t-1},$$
$$a_0^* = a_0(1-\rho), \qquad e_t = u_t - \rho u_{t-1}.$$

Отримуємо:

$$y_t^* = a_0^* + a_1 x_t^* - e_t . \tag{6.25}$$

4. Знаходимо значення ρ та виконуємо поправку Прайса-Винстена:

$$x_1^* = x_1\sqrt{1-\rho^2}\,, \qquad y_1^* = y_1\sqrt{1-\rho^2}\,. \tag{6.26}$$

5. Знаходимо матриці перетворень:

$$T_1 = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \cdots & 0 \\ -\rho & 1 & 0 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{(n\times m)} \qquad T_2 = \begin{bmatrix} -\rho & 1 & 0 & 0 & \cdots & 0 \\ 0 & -\rho & 1 & 0 & \cdots & 0 \\ 0 & 0 & -\rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}_{(n-1)\times n} \tag{6.27}$$

6. Виконуємо процедури обчислень:

$$X^* = T_1 X = \begin{bmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}\cdot x_1' & \cdots & \sqrt{1-\rho^2}\cdot x_1^m \\ 1-\rho & x_2' - \rho x_1' & \cdots & x_2^m - \rho x_1^m \\ \vdots & \vdots & \vdots & \vdots \\ 1-\rho & x_n' - \rho x_{n-1}' & \cdots & x_n^m - \rho x_{n-1}^m \end{bmatrix}_{(n\times n)} \tag{6.28}$$

$$Y^* = T_1 Y = \begin{bmatrix} \sqrt{1-\rho^2}\cdot y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ \vdots \\ y_n - \rho y_{n-1} \end{bmatrix}_{(n\times 1)} \tag{6.29}$$

7. Для перетворених даних використаємо оператор оцінювання МНК:

$$\hat{A} = \left(X^{*'} X^* \right)^{-1} X^{*'} Y^* \tag{6.30}$$

Отримаємо $\hat{a}_0^*$ та $\hat{a}_1$.

8. Знаходимо значення оцінок

$$a_0 = \frac{\hat{a}_0^*}{1-\rho}\,, \qquad a_1 = \hat{a}_1 \tag{6.31}$$

9. Отримуємо економетричну модель.

<u>Приклад</u> 6.3. Оцінити параметри економетричної моделі з автокорельованими залишками прикладу 6.1 методом перетворення вихідної інформації.

Розв'язування. Згідно методу перетворення вихідної інформації сформуємо матрицю T_1, попередньо скориставшись відомим нам із прикладу 6.2 значенням $\rho \approx -0{,}188141$, а також обчислимо

$$\sqrt{1-\rho^2} = \sqrt{1-(-0{,}188141)^2} = \sqrt{1-0{,}035397} = \sqrt{0{,}9646029} \approx 0{,}9821.$$

Отже,

$$T_1 = \begin{bmatrix}
0{,}9821 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1
\end{bmatrix}$$

Використавши функцію МУМНОЖ (MMULT) в MS Excel, обчислимо

$$X^* = T_1 X = \begin{bmatrix}
0{,}9821 & 549{,}9995 \\
1{,}1881 & 725{,}3591 \\
1{,}1881 & 711{,}6476 \\
1{,}1881 & 646{,}9441 \\
1{,}1881 & 460{,}6556 \\
1{,}1881 & 632{,}7309 \\
1{,}1881 & 706{,}2998 \\
1{,}1881 & 387{,}8848 \\
1{,}1881 & 591{,}7388 \\
1{,}1881 & 766{,}5963 \\
1{,}1881 & 610{,}1139 \\
1{,}1881 & 441{,}2485 \\
1{,}1881 & 745{,}8494 \\
1{,}1881 & 577{,}9361 \\
1{,}1881 & 464{,}6636
\end{bmatrix}; \qquad
Y^* = T_1 Y = \begin{bmatrix}
4922{,}4956 \\
5910{,}9640 \\
6346{,}6858 \\
5786{,}2205 \\
7131{,}0575 \\
7159{,}8726 \\
5779{,}4017 \\
8820{,}4213 \\
7030{,}7823 \\
5853{,}5500 \\
6142{,}3357 \\
7963{,}2958 \\
5868{,}8497 \\
6625{,}1716 \\
7119{,}1988
\end{bmatrix}$$

Далі згідно алгоритму метода перетворення вхідної інформації знайдемо добуток $\left(X^{*'}X^{*}\right)$. Для цього використаємо функцію МУМНОЖ (MMULT) в MS Excel, одержимо:

$$X^{*'}X^{*}=\begin{bmatrix} 20{,}7281 & 10603{,}3401 \\ 10603{,}3401 & 5626467{,}7752 \end{bmatrix};$$

$$\left(X^{*'}X^{*}\right)^{-1}=\begin{bmatrix} 1{,}341 & \text{-}0{,}003 \\ \text{-}0{,}003 & 0 \end{bmatrix};$$

$$X^{*'}Y^{*}=\begin{bmatrix} 115970{,}7181 \\ 57996687{,}2193 \end{bmatrix}.$$

Таким чином, оператор оцінювання:

$$\hat{A}=\left(X^{*'}X^{*}\right)^{-1}X^{*'}Y^{*}=\begin{bmatrix} 1{,}341 & \text{-}0{,}003 \\ \text{-}0{,}003 & 0 \end{bmatrix}\cdot\begin{bmatrix} 115970{,}7181 \\ 57996687{,}2193 \end{bmatrix}=\begin{bmatrix} 8949{,}8251 \\ \text{-}6{,}5585 \end{bmatrix}.$$

Отже, економетрична модель матиме вигляд:

$$\hat{Y}^{*}=8949{,}8251-6{,}5585x.$$

Окрім того, знайдений оператор оцінювання методом перетворення вхідної інформації ідентичний оператору оцінювання, що був знайдений методом Ейткена у прикладі 6.2. І попередньо проведений порівняльний аналіз із використанням критерію Дарбіна-Уотсона, дає можливість зробити висновок, що ми не звільнились від автокореляції залишків.

Тому наступним кроком буде знаходження матриці T_2:

$$T_2=\begin{bmatrix}
0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}1881 & 1
\end{bmatrix}$$

Знаходимо:

$$X^* = T_2 X = \begin{bmatrix} 1{,}1881 & 725{,}3591 \\ 1{,}1881 & 711{,}6476 \\ 1{,}1881 & 646{,}9441 \\ 1{,}1881 & 460{,}6556 \\ 1{,}1881 & 632{,}7309 \\ 1{,}1881 & 706{,}2998 \\ 1{,}1881 & 387{,}8848 \\ 1{,}1881 & 591{,}7388 \\ 1{,}1881 & 766{,}5963 \\ 1{,}1881 & 610{,}1139 \\ 1{,}1881 & 441{,}2485 \\ 1{,}1881 & 745{,}8494 \\ 1{,}1881 & 577{,}9361 \\ 1{,}1881 & 464{,}6636 \end{bmatrix}; \qquad Y^* = T_2 Y = \begin{bmatrix} 5910{,}964 \\ 6346{,}6858 \\ 5786{,}2205 \\ 7131{,}0575 \\ 7159{,}8726 \\ 5779{,}4017 \\ 8820{,}4213 \\ 7030{,}7823 \\ 5853{,}5500 \\ 6142{,}3357 \\ 7963{,}2958 \\ 5868{,}8497 \\ 6625{,}1716 \\ 7119{,}1988 \end{bmatrix};$$

$$X^{*\prime} X^* = \begin{bmatrix} 19{,}7635 & 10063{,}1625 \\ 10063{,}1625 & 5323968{,}3165 \end{bmatrix};$$

$$\left(X^{*\prime} X^* \right)^{-1} = \begin{bmatrix} 1{,}347 & -0{,}003 \\ -0{,}003 & 0 \end{bmatrix};$$

$$X^{*\prime} Y^* = \begin{bmatrix} 111136{,}1285 \\ 55289317{,}0639 \end{bmatrix}.$$

Отже,

$$\hat{A} = \left(X^{*\prime} X^* \right)^{-1} X^{*\prime} Y^* = \begin{bmatrix} 1{,}347 & -0{,}003 \\ -0{,}003 & 0 \end{bmatrix} \cdot \begin{bmatrix} 111136{,}1285 \\ 55289317{,}0639 \end{bmatrix} = \begin{bmatrix} 8929{,}6851 \\ -6{,}4936 \end{bmatrix}.$$

Економетрична модель матиме вигляд:

$$\hat{Y}^* = 8929{,}6851 - 6{,}4936x.$$

Як бачимо, оцінки параметрів моделі знайдені із використанням матриць T_1 і T_2 дещо відрізняються між собою, проте і в першому, і в другому випадках ми не звільнились від автокореляції залишків.

Алгоритм методу перших різниць.

Даний метод застосовується для випадку, коли автокореляція залишків досить висока і припускається, що $\rho = 1$. Звідси (6.20) прийме вид:

$$y_t - y_{t-1} = a_1 \left(x_t - x_{t-1} \right) + u_t \tag{6.32}$$

1. Виконуємо заміну:

$\Delta y_t = y_t - y_{t-1}$, $\Delta x_t = x_t - x_{t-1}$, $e_t = u_t - u_{t-1}$.

Отримаємо:

$$\Delta y_t = a_1 \Delta x_t + e_t. \tag{6.33}$$

2. З допомогою МНК знаходимо оцінки для (6.33):

$$\hat{a}_1,\ \hat{a}_0 = \overline{y} - \hat{a}_1 \overline{x}. \qquad (6.34)$$

Алгоритм методу Кохрана-Оркатта.

1. Оцінюється початкова регресія МНК:

$$y_t = a_0 + a_1 x_t + u_t. \qquad (6.35)$$

Знаходимо оцінки $\hat{a}_0$, $\hat{a}_1$:

$$\hat{y}_t = \hat{a}_0 + \hat{a}_1 x_t. \qquad (6.36)$$

2. Знаходимо залишки:

$$u_t = y_t - \hat{y}_t. \qquad (6.37)$$

3. Оцінюємо авторегресію залишків МНК:

$$u_t = \rho u_{t-1} + e_t. \qquad (6.38)$$

Знаходимо оцінку $\rho = \hat{\rho}_1$.

4. Вводимо нові змінні:

$$y'_t = y_t - \hat{\rho}_1 y_{t-1},\quad x'_t = x_t - \hat{\rho}_1 x_{t-1}, \qquad (6.39)$$

$$y'_1 = \sqrt{1-\hat{\rho}_1^{\ 2}} \cdot y_1,\quad x'_1 = \sqrt{1-\hat{\rho}_1^{\ 2}} \cdot x_1. \qquad (6.40)$$

5. Будуємо нову модель:

$$\hat{y}'_t = \hat{a}'_0 + \hat{a}'_1 x'_t. \qquad (6.41)$$

6. Знаходимо залишки:

$$u'_t = y'_t - \hat{y}'_t. \qquad (6.42)$$

7. Оцінюємо авторегресію залишків МНК:

$$u'_t = \rho_1 u'_{t-1} + e'_t. \qquad (6.43)$$

Знаходимо оцінку для $\rho_1 = \hat{\rho}_2$.

8. Вводимо нові змінні:

$$y''_t = y'_t - \hat{\rho}_2 y'_{t-1},\quad x''_t = x'_t - \hat{\rho}_2 x'_{t-1}, \qquad (6.44)$$

$$y''_1 = \sqrt{1-\hat{\rho}_2^{\ 2}} \cdot y'_1,\quad x''_1 = \sqrt{1-\hat{\rho}_2^{\ 2}} \cdot x'_1. \qquad (6.45)$$

9. Побудуємо нову модель:

$$\hat{y}''_t = \hat{a}''_0 + \hat{a}''_1 x''_t. \qquad (6.46)$$

10. Знаходимо залишки:

$$u''_t = y''_t - \hat{y}''_t. \qquad (6.47)$$

11. Оцінюємо авторегресію залишків МНК:

$$u''_t = \rho_2 u''_{t-1} + e''_t. \qquad (6.48)$$

Знаходимо оцінку для $\rho_2 = \hat{\rho}_3$.

12. Продовжуємо даний ітераційний процес:

$$\left|\hat{\rho}_k - \hat{\rho}_{k-1}\right| \le \alpha. \qquad (6.49)$$

13. Знайдене значення $\rho = \hat{\rho}_k$ підставимо у (6.20), отримаємо:

$$y_t - \hat{\rho}_k y_{t-1} = a_0(1-\hat{\rho}_k) + a_1(x_t - \hat{\rho}_k x_{t-1}) + u_t - \hat{\rho}_k u_{t-1}. \quad (6.50)$$

Виконаємо заміну:

$$y_t^* = y_t - \hat{\rho}_k y_{t-1},\ a_0^* = a_0(1-\hat{\rho}_k),\ x_t^* = x_t - \hat{\rho}_k x_{t-1},\ e_t = u_t - \hat{\rho}_k u_{t-1}.$$

Одержимо:

$$y_t^* = a_0^* + a_1 x_t^* + e_t. \quad (6.51)$$

14. Знаходимо МНК оцінки для a_0^* та a_1:

$$a_0 = \frac{\hat{a}_0^*}{1-\hat{\rho}_k},\ a_1 = \hat{a}_1. \quad (6.52)$$

Алгоритм методу Дарбіна.

1. Представимо формулу (6.20) наступним чином:

$$y_t = a_0(1-\rho) + a_1 x_t + \rho y_{t-1} - \rho a_1 x_{t-1} + e_t. \quad (6.53)$$

Виконаємо заміну:

$$a_0^* = a_0(1-\rho), \quad a_1^* = -\rho a_1. \quad (6.54)$$

Одержимо:

$$y_t = a_0^* + a_1 x_t + \rho y_{t-1} - a_1^* x_{t-1} + e_t. \quad (6.55)$$

Тоді для (6.55) оціночним рівнянням буде:

$$\hat{y}_t = \hat{a}_0^* + \hat{a}_1 x_t + \hat{\rho}_1 y_{t-1} - \hat{a}_1^* x_{t-1}. \quad (6.56)$$

З допомогою МНК знайдемо оцінки: $\hat{a}_0^*$, $\hat{a}_1$, $\hat{\rho}_1$ та $\hat{a}_0^*$.

Звідси $a_0 = \dfrac{\hat{a}_0^*}{1-\hat{\rho}_1}$, $a_1 = \hat{a}_1$, $\rho = \hat{\rho}_1$.

2. Знайдене значення $\rho = \hat{\rho}_1$ використаємо для перетворень:

$$y_t^* = y_t - \hat{\rho}_1 y_{t-1},\ x_t^* = x_t - \hat{\rho}_1 x_{t-1}. \quad (6.57)$$

Для отриманих перетворених даних використаємо МНК.

Алгоритм методу Хилдрета-Лу.

1. Перетворене рівняння (6.20) оцінюється для кожного значення $\rho \in [-1;1]$ із заданим кроком $\Delta\rho$, $\rho_i = \rho_{i-1} + \Delta\rho$.

$$y_t - \rho_i y_{t-1} = a_0(1-\rho_i) + a_1(x_t - \rho_i x_{t-1}) + e_t. \quad (6.58)$$

$$y_t^* = y_t - \rho_i y_{t-1},\ a_0^* = a_0(1-\rho_i),\ x_t^* = x_t - \rho_i x_{t-1}.$$

Оціночне рівняння буде:

$$y_t^* = a_0^* + a_1 x_t^* + e_t. \quad (6.59)$$

Знаходимо $\hat{a}_0^*$, $\hat{a}_1$, $a_0 = \dfrac{\hat{a}_0^*}{1-\rho_i}$.

2. Значення ρ_i вибираємо із умови, що сума квадратів залишків в перетвореному рівнянні буде мінімальною.

Коефіцієнти регресії обчислюються при знайденому значенні ρ_i.

Завдання 7. Економетричні моделі на основі системи одночасних рівнянь.

У залежності від характеру постановки економічної проблеми та мети дослідження економетрична модель може бути представлена в різних видах. Для визначення видів моделей необхідно провести пояснення відносно форми її класифікації з врахуванням існуючих взаємозв’язків і співвідношень між економічними показниками, а також показати переваги чи недоліки між запропонованими видами економетричних моделей.

У попередніх завданнях нами вивчалися односторонні стохастичні причинні відносини між економічними явищами та процесами, і нас цікавили методи оцінювання тільки одного регресійного рівняння:

$$\hat{y} = f\left(x_1, x_2, ..., x_m\right). \qquad (7.1)$$

У нашому випадку виходили з того, що залежна змінна y пояснюється незалежними змінними $x_1, x_2, ..., x_m$, а зворотній зв’язок відсутній.

Однак в економічних дослідженнях досить рідко зустрічаються подібні односторонні стохастичні причинні відносини. В більшості випадків доводиться мати справу з розглядом системи відносин, оскільки вона більш адекватно описує існуючі взаємозалежності та взаємозв’язки між економічними явищами.

Для подальших досліджень розглянемо наступних три види економетричних систем рівнянь:

- система незалежних рівнянь, коли кожна залежна змінна y_i розглядається як функціональна залежність однієї множини пояснюючих чинників:

$$y_i = f_i\left(x_1, x_2, ..., x_n, u_i\right), i = \overline{1, m}. \qquad (7.2)$$

- система рекурсивних рівнянь, коли в кожному наступному i-ому рівнянні системи залежна змінна y_i представляється функцією усієї множини пояснюючих змінних та залежними змінними попередніх рівнянь:

$$\begin{aligned} y_1 &= f_1\left(x_1, x_2, ..., x_n, u_1\right) \\ y_2 &= f_2\left(x_1, x_2, ..., x_n, y_1, u_2\right) \\ &\dots\dots\dots\dots\dots\dots \\ y_m &= f_m\left(x_1, x_2, ..., x_n, y_{m-1}, u_m\right). \end{aligned} \qquad (7.3)$$

– система одночасних рівнянь, коли поряд із множиною незалежних змінних залежна змінна y_i в одному рівнянні входить у ліву, а в інших рівняннях у праву частину:

$$y_i = f_i\left(x_1, x_2, ..., x_n, y_1, ...y_{m-1}, y_{m+1}, ...y_m, u_i\right), i = \overline{1, m}. \tag{7.4}$$

Система регресійних рівнянь, яка описує одночасні багатосторонні стохастичні причинні зв’язки між економічними процесами та явищами, називається системою одночасних рівнянь.

Розглянемо оцінку параметрів рівняння функції споживання в простій кейнсіанській моделі формування доходів:

$$C_t = \alpha + \beta V_t + u_t, \tag{7.5}$$

$$V_t = C_t + I_t, \tag{7.6}$$

де C_t – обсяг споживання у періоді t; V_t – обсяг сукупного доходу у періоді t; I_t – обсяг інвестицій у періоді t; α та β – параметри регресії, які необхідно оцінити; u_t – випадкова величина у періоді t.

Підставимо (7.5) у (7.6) і після відповідних перетворень отримуємо:

$$V_t = \alpha + \beta V_t + u_t + I_t,\ V_t = \frac{\alpha}{1-\beta} + \frac{I_t}{1-\beta} + \frac{u_t}{1-\beta} \tag{7.7}$$

Перших дві складових правої частини показують, що сукупний рівень доходу залежить від постійної складової обсягу споживання і від обсягу інвестицій. У той же час рівень сукупного доходу залежить від випадкової величини u_t (збурення). Якщо обсяг інвестицій збільшиться на одиницю, то сукупний дохід зросте на 1/(1-β) одиниць. Величина 1/(1-β) називається мультиплікатором.

Модель формування доходу вказує на те, що обсяг споживання залежить від сукупного доходу та випадкової величини. У той же час на основі (7.6) можна стверджувати, що сукупний дохід залежить від обсягу споживання. Тому в подальшому необхідно виразити залежність від I_t та u_t. Для цього підставимо (7.6) у (7.5). Після відповідних перетворень отримаємо:

$$C_t = \alpha + \beta\left(C_t + I_t\right) + u_t, C_t = \alpha + \beta C_t + \beta I_t + u_t,$$

$$C_t = \frac{\alpha}{1-\beta} + \frac{\beta I_t}{1-\beta} + \frac{u_t}{1-\beta}. \tag{7.8}$$

Рівняння (7.7) і (7.8), які є складовими вихідної моделі (7.5) і (7.6), називаються структурними.

1. *Структурна форма економетричних моделей*.

У більшості випадків, системи одночасних структурних рівнянь містять у собі рівняння лінійного виду. Усунути наявність нелінійних зв'язків можна з допомогою лінійної апроксимації.

Оскільки економетричні моделі в загальному випадку будуються на основі часових рядів, то динаміка економічних зв'язків враховується з допомогою часових лагів або лагових змінних.

Запишемо загальну економетричну модель на основі системи одночасних рівнянь для періоду t:

$$
\begin{aligned}
&y_{1t} = b_{11}y_{1t} + \ldots + b_{1k}y_{kt} + a_{10}x_{0t} + \ldots + a_{1m}x_{mt} + u_{1t} \\
&y_{2t} = b_{21}y_{1t} + \ldots + b_{2k}y_{kt} + a_{20}x_{0t} + \ldots + a_{2m}x_{mt} + u_{2t} \\
&\vdots \\
&y_{kt} = b_{k1}y_{1t} + \ldots + b_{kk}y_{kt} + a_{k0}x_{0t} + \ldots + a_{1k}x_{mt} + u_{kt},
\end{aligned}
\tag{7.9}
$$

де y_{it} – i-та змінна для періоду t, яка повинна бути пояснена з допомогою моделі $\left(i=\overline{1,k};\, t=\overline{1,T}\right)$; x_{jt} – змінні j-го виду для періоду t, які характеризуються одностороннім причинним зв'язком, тобто вони пояснюють залежні змінні; x_{0t}=1 – фіктивна змінна моделі; b_{it} та a_{jt} - параметри моделі, які в окремих випадках можуть бути рівними нулю, якщо відповідна змінна не входить у рівняння; u_{it} – випадкові величини (збурення), які також можуть дорівнювати нулю, якщо відповідне рівняння є тотожністю.

Представимо систему рівнянь (7.9) у матричній формі:

$$
Y = BY + AX + u, \tag{7.10}
$$

де

$$
Y = \begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{kt} \end{bmatrix};\quad B = \begin{bmatrix} b_{11} & \cdots & b_{1k} \\ b_{21} & \cdots & b_{2k} \\ \cdots & \cdots & \cdots \\ b_{k1} & \cdots & b_{kk} \end{bmatrix};\quad A = \begin{bmatrix} a_{10} & \cdots & a_{1m} \\ a_{20} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots \\ a_{k0} & \cdots & a_{km} \end{bmatrix};
$$

$$
X = \begin{bmatrix} x_{0t} \\ x_{1t} \\ \vdots \\ x_{mt} \end{bmatrix};\quad u = \begin{bmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{kt} \end{bmatrix};\quad t = \overline{1,T}.
$$

Вираз (7.10) представляє собою систему одночасних рівнянь у матричній формі. Одночасний характер моделі очевидний: залежна змінна одного рівняння виступає як пояснююча змінна в інших або пояснюючі змінні в одному чи декількох рівняннях включені в інше рівняння системи, як загальні змінні.

Економетрична модель виду (7.9), яка безпосередньо відображає структуру взаємозв'язків між змінними величинами, називається структурною формою економетричної моделі.

Отже, структурна форма моделі відображає одно- та багатосторонні стохастичні причинні відношення між економічними показниками в їх безпосередньому вигляді і містить усю існуючу інформацію про залежності між ними.

Зрозуміло, що поряд із структурними рівняннями економетрична модель може містити так звані визначаючі рівняння – тотожності. Такі тотожності потрібні для більш адекватного відображення реальної дійсності та повного охоплення змінних одночасними співвідношеннями. Тотожності не містять у собі збурюючих змінних і невідомих параметрів.

Параметри структурної форми моделі називаються структурними параметрами і їх неможливо визначити звичайним МНК, оскільки права частина системи одночасних рівнянь містить ендогенні змінні. Використання МНК привело би до порушення передумов про причинно-наслідкові залежності між факторними та результативними ознаками, а також ряду інших передумов звичайного МНК.

2. *Повна економетрична модель*.

Економетрична модель називається повною, якщо вона має наступні властивості:

а) охоплює ті змінні, які виявляють суттєвий вплив на сумісно залежні змінні, а збурюючі змінні мають випадковий характер;

б) містить стільки рівнянь, скільки в неї є сумісних залежних змінних. Тому кожна сумісно залежна змінна може бути пояснена з допомогою відповідного рівняння;

в) система рівнянь має однозначний розв'язок відносно сумісних незалежних змінних.

Модель повинна бути повною в тих випадках, якщо необхідно кількісно описати певний економічний процес або якщо вона використовується для розрахунку прогнозних стратегій.

3. *Приведена форма економетричної моделі.*

Якщо економетрична модель повна, то існує зворотна матриця $(E\text{-}B)^{-1}$. Завдяки цьому можна знайти розв'язок системи (7.9) або (7.10).

Знайдемо розв'язок системи рівнянь (7.9) відносно сумісно залежних величин. Як результат отримаємо наступну систему:

$$\begin{aligned}
&y_{1t} = d_{10}x_{0t} + ... + d_{1m}x_{mt} + e_{1t} \\
&y_{2t} = d_{20}x_{0t} + ... + d_{2m}x_{mt} + e_{2t} \\
&\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\
&y_{kt} = d_{k0}x_{0t} + ... + d_{km}x_{mt} + e_{kt}
\end{aligned} . \qquad (7.11)$$

Далі знайдемо розв’язок системи (7.11) відносно y. Маємо:

$$Y - BY = AX + u; (E - B)Y = AX + u; Y = (E - B)^{-1} AX + (E - B)^{-1} u .$$

Нехай $(E\text{-}B)^{-1}A= D$ та $(E\text{-}B)^{-1}u=e$, тоді отримаємо розв’язок системи у матричній формі:

$$Y=DX+ e . \qquad (7.12)$$

Економетрична модель, яка представляється системою рівнянь виду (7.11), називається приведеною формою моделі або прогнозною. З (7.11) виходить, що сумісно залежні змінні є лінійними функціями від напередвизначених та збурюючих змінних. Крім того, коефіцієнтами правих частин є комбінація всіх структурних коефіцієнтів спільно залежних та відповідно напередвизначених змінних в усіх структурних рівняннях (7.9).

Враховуючи складність вираження коефіцієнтів рівнянь у прогнозній формі, можна зробити висновок, що вони втрачають по відношенню до напередвизначених змінних свою автономність, яка властива структурним рівнянням. Проте з іншого боку, кожне рівняння у приведеній формі характеризується певною автономністю відносно сумісно залежних змінних, оскільки кожне з цих рівнянь містить поточне значення тільки однієї ендогенної змінної, яка визначається як функція напередвизначених змінних. Очевидно, що взаємозв’язок сумісно залежних змінних при переході від структурної до прогнозної форми поширюються на напередвизначені змінні та збурення.

Коефіцієнти рівнянь у прогнозній формі відображають безпосередній і побічний вплив напередвизначених на сумісно залежні змінні. Тут іде мова відносно загального ефекту. У той же час, структурні коефіцієнти виражають тільки безпосередній вплив напередвизначених змінних, тобто мають частковий ефект. У такому розумінні економічна інтерпретація коефіцієнтів рівняння в прогнозній формі більш реальна, в порівнянні зі структурними коефіцієнтами.

Кожне рівняння в прогнозній формі представляє собою множинну регресію, до якої для оцінки параметрів моделі можна застосувати МНК.

Якщо оцінки коефіцієнтів приведеної форми і значення наперед визначених змінних припадають на період прогнозу, то на основі моделі знаходять прогнозні значення сумісно залежних змінних. І навпаки, структурна форма для прогнозу непридатна, оскільки в кожному структурному рівнянні міститься декілька сумісно незалежних змінних, для яких не можуть бути вказані значення на прогнозний період, тому що вони тільки підлягають оцінці.

Недоліком прогнозної форми є те, що від кількісної оцінки в даній формі не в усіх випадках можна перейти до структурної моделі. Разом з тим за заданою числовою моделлю в структурній формі може бути завжди знайдена прогнозна форма.

Якщо параметри структурної форми однозначно виражаються через параметри прогнозної форми системи реєстрації, то така економетрична модель називається ідентифікованою. Якщо число оцінюваних параметрів сукупної форми реєстрації більше числа оцінюваних параметрів прогнозної форми, тобто число оцінюваних параметрів прогнозної форми регресії більше числа рівнянь, то така система називається неідентифікованою.

Припустимо, що нам задана модель у структурній формі:

$$\begin{aligned} y_1 &= b_{12} y_2 + a_{11} x_1 \\ y_2 &= b_{21} y_1 + a_{22} x_2 \end{aligned}.$$

Тоді приведену форму моделі подамо у наступному виді:

$$\begin{aligned} y_1 &= \delta_{11} x_1 + \delta_{12} x_2 \\ y_2 &= \delta_{21} x_1 + \delta_{22} x_2 \end{aligned}.$$

Виконаємо наступні алгебраїчні перетворення. Знайдемо y_2 з першого рівняння.

$$y_2 = \frac{(y_1 - a_{11} x_1)}{b_{12}},$$

$$\begin{cases} y_2 = \dfrac{(y_1 - a_{11} x_1)}{b_{12}} \\ y_2 = b_{21} y_1 + a_{22} x_2 \end{cases},$$

$$\frac{(y_1 - a_{11} x_1)}{b_{12}} = b_{21} y_1 + a_{22} x_2,$$

$$y_1 - a_{11} x_1 = b_{12} b_{21} y_1 + b_{12} a_{22} x_2,$$

$$y_1 (1 - b_{12} b_{21}) = a_{11} x_1 + b_{12} a_{22} x_2,$$

$$y_1 = \frac{a_{11}}{1 - b_{12} b_{21}} x_1 + \frac{b_{12} a_{22}}{1 - b_{12} b_{21}} x_2.$$

Тоді перше рівняння у приведеній формі:

$$y_1 = \delta_{11} x_1 + \delta_{12} x_2,$$

де $\delta_{11} = \frac{a_{11}}{1 - b_{12}b_{21}}$, $\delta_{12} = \frac{b_{12}a_{22}}{1 - b_{12}b_{21}}$.

Аналогічні перетворення виконаємо y_1, знайшовши його з другого рівняння.

$$y_2 = \frac{b_{21}a_{11}}{1 - b_{12}b_{21}} x_1 + \frac{a_{22}}{1 - b_{12}b_{21}} x_2 .$$

$$y_2 = \delta_{21}x_1 + \delta_{22}x_2 ,$$

де $\delta_{21} = \frac{b_{21}a_{11}}{1 - b_{12}b_{21}}$, $\delta_{22} = \frac{a_{12}}{1 - b_{12}b_{21}}$.

4. *Економетрична модель із взаємозалежних змінних (інтердепедентна модель).*

Економетрична модель називається інтердепедентною, якщо вона може бути представлена у вигляді системи структурних рівнянь, в яких змінні одночасно задовольняють декільком рівностям. Як наслідок, в інтердепедентній моделі змінні є багатосторонньо залежними.

5. *Рекурсивна економетрична модель.*

Якщо в економетричній моделі матриця параметрів B при внутрішніх змінних Y_i має трикутний вид, то система рівнянь називається рекурсивною. Рекурсивна модель може бути представлена наступним чином:

$$\begin{aligned}
&Y_{1t} = a_{10}x_{0t} + a_{11}x_{1t} + \ldots + a_{1m}x_{1m} + u_{1t} \\
&Y_{2t} = b_{21}Y_{1t} + a_{20}x_{0t} + a_{21}x_{1t} + \ldots + a_{2m}x_{mt} + u_{2t} \\
&Y_{3t} = b_{31}Y_{1t} + b_{32}Y_{2t} + a_{30}x_{0t} + a_{31}x_{1t} + \ldots + a_{3m}x_{mt} + u_{3t} \\
&\ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\
&Y_{kt} = b_{k1}Y_{1t} + \ldots + b_{k,k-1}Y_{k-1,t} + a_{k0}x_{0t} + \ldots + a_{km}x_{mt} + u_{kt}
\end{aligned} \tag{7.13}$$

Для даної системи матриця параметрів B має вигляд:

$$B = \begin{bmatrix} -1 & 0 & 0 & \cdots & 0 & 0 \\ b_{21} & -1 & 0 & \cdots & 0 & 0 \\ b_{31} & b_{32} & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{k1} & b_{k2} & b_{k3} & \cdots & b_{k,k-1} & -1 \end{bmatrix}$$

Рекурсивна модель володіє такими властивостями:

а) відповідним розміщенням ендогенних змінних і структурних рівнянь можна досягти того, що в першому структурному рівнянні буде тільки одна ендогенна змінна, а в наступних будуть кожен раз додаватися інші. Тобто, в рекурсивній моделі спостерігається одностороння залежність між внутрішніми змінними. Наприклад, Y_{2t} залежить від Y_{1t}, але Y_{1t} не залежить від Y_{2t};

б) матриця дисперсій і коваріацій збурюючих змінних є діагональною:

$$M\left(u_t u_t^{'}\right)=\begin{bmatrix}\sigma_{11}^t & 0 & \cdots & 0\\ 0 & \sigma_{22}^t & \cdots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \cdots & \sigma_{kk}^t\end{bmatrix} \qquad (7.14)$$

Як наслідок маємо, що збурюючі змінні різних рівнянь у момент часу t стохастично незалежні одна від одної, тобто вони некорельовані;

в) збурюючі змінні рівнянь неавтокорельовані, тобто

$$M\left(u_{it}u_{it-\tau}\right)=0,\quad i=\overline{1,k},\ t=\overline{1,T},\quad \tau\neq 0\,. \qquad (7.15)$$

6. *Блочно-рекурсивна економетрична модель.*

За певних умов, які накладаються на матрицю структурних коефіцієнтів B і коваріаційно-дисперсійну матрицю збурюючих змінних, має місце блочно-рекурсивна модель. Така модель виникає при наявності великого числа пояснюючих змінних, що в свою чергу призводить до розбиття на підмоделі. Даний процес полегшує виконання процедури статистичного оцінювання параметрів і забезпечується структурою матриці B. Отже, економетричну модель назвемо блочно-рекурсивною, якщо при великому числі пояснювальних змінних є можливість розбити структурну матрицю B на підматриці у вигляді блоків. Як результат наша модель буде розкладена на декілька підмоделей. Блочно-рекурсивна форма особливо властива економетричним моделям регіональної економіки, де відповідні галузі виробництва та складові економічної системи утворюють окремі взаємозв’язані блоки.

7. *Економетрична модель із системи незалежних рівнянь.*

Система незалежних рівнянь – частковий випадок рекурсивної моделі, коли матриця B є одиничною, $B=E$. У даному випадку в кожному рівнянні міститься тільки одна ендогенна змінна, яка підлягає поясненню та не залежить від ендогенних змінних інших рівнянь. Таким чином, кожне рівняння незалежне від інших. Структурна та приведена форма такої моделі співпадають.

Отже, економетрична модель називається моделлю із системи незалежних рівнянь, якщо матриця $B=E$. У загальному випадку систему незалежних регресій можна записати наступним чином:

$$
\begin{aligned}
y_{1t} &= a_{10}x_{0t} + a_{11}x_{1t} + \cdots a_{1m}x_{mt} + u_{1t} \\
y_{2t} &= a_{20}x_{0t} + a_{21}x_{1t} + \cdots a_{2m}x_{mt} + u_{2t} \\
\ldots & \;\; \cdots \;\; \cdots \;\; \cdots \;\; \cdots \;\; \cdots \;\; \cdots \\
y_{kt} &= a_{kt}x_{0t} + a_{k1}x_{1t} + \cdots a_{km}x_{mt} + u_{kt},
\end{aligned}
\tag{7.16}
$$

або у матричній формі

$$Y = AX + u. \tag{7.17}$$

Параметри системи незалежних регресій оцінюються з допомогою МНК, або стандартними процедурами програмного продукту STADIA.

Приклад 7.1. Побудувати економетричну модель попиту та пропозиції системи незалежних регресій і знайти точку рівноваги на основі статистичних даних, приведених у таблиці 7.1.

Таблиця 7.1

Ціна, грн., x_i	5	10	15	20	25	30	35	40	45	50
Попит на даний вид товару, y_{1i}	270	240	190	170	145	120	115	110	100	90
Пропозиція да-ного товару, y_{2i}	95	105	112	123	141	153	176	208	244	288

Розв'язування. Припустимо, що залежність між ціною, попитом та пропозицією можна описати з допомогою експоненти виду:

$$y_1 = e^{a_0 + a_1 x} + u_1 \text{ та } y_2 = e^{b_0 + b_1 x} + u_2.$$

Для знаходження оцінок параметрів використаємо процедуру "Проста регресія" системи STADIA. У результаті отримаємо:

ПРОСТАЯ РЕГРЕССИЯ. Файл: wwq1.std

Переменные: x1, x2

Модель: экспонента Y = EXP(a0+a1*x)

Коэфф. a0 a1

Значение 5,648 -0,02438

Ст.ошиб. 0,05101 0,001644

Значим. 0 0

Источник Сум.квадр. Степ.св Средн.квадр.

Регресс. 1,226 1 1,226

Остаточн 0,04461 8 0,005576

Вся 1,27 9

Множеств R R^2 R^2прив Ст.ошиб. F Значим

0,98228 0,96488 0,96049 0,074674 219,8 0

Гипотеза 1: <Регрессионная модель адекватна экспериментальным данным>

ПРОСТАЯ РЕГРЕССИЯ. Файл wwq1.std:

Переменные: x1, x3

Модель: экспонента Y = EXP(a0+a1*x)

Коэфф. a0 a1

Значение 4,367 0,02441

Ст.ошиб. 0,03661 0,00118

Значим. 0 0

Источник Сум.квадр. Степ.св Средн.квадр.

Регресс. 1,229 1 1,229

Остаточн 0,02298 8 0,002872

Вся 1,252 9

Множеств R R^2 R^2прив Ст.ошиб. F Значим

0,99078 0,98164 0,97935 0,053595 427,7 0

Гипотеза 1: <Регрессионная модель адекватна экспериментальным данным>

Отже, $y_1 = e^{5.648-0.02438x} + u_1, \quad y_2 = e^{4.367+0.02441x} + u_2$.

Знайдемо точку рівноваги попиту та пропозиції

$$e^{5.648-0.02438x} + u_1 = e^{4.367+0.02441x},$$
$$5.648 - 0.02438x = 4.367 + 0.02441x,$$
$$x = 26{,}25\,грн. \quad y_1 = y_2 = 149{,}6.$$

Умова рівноваги між попитом і пропозицією наступає при ціні x=26,25 грн. (рис.7.1), точка A(26,25;149,6).

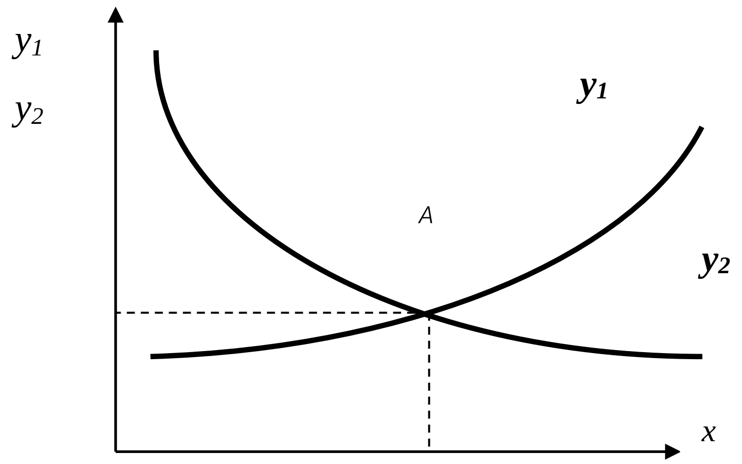


Рис. 7.1. Графік попиту та пропозиції

Припустимо, що нам задана економетрична модель виду (7.10). Тоді перед нами стоїть задача визначення її структурних параметрів на основі даних спостережень над сумісно залежними та наперед визначеними змінними. Дана ситуація пов'язана із проблемою ідентифікації, причина якої породжується взаємозв'язками економічних явищ і, як наслідок, взаємозалежностями змінних. Економетрична модель ідентифікується, якщо ідентифікуються структурні рівняння. При цьому необхідно враховувати, що ідентифікація окремого рівня залежить не стільки від самого рівня, скільки від виду всіх структурних рівнянь моделі. Ідентифікованість структурних рівнянь означає, що шляхом лінійної комбінації деяких або всіх рівнянь моделі неможливо отримати ні одного рівняння, яке б суперечило моделі і параметри якого відрізнялись би від параметрів структурних рівнянь, які підлягають оцінці.

Отже, ідентифікація – це однозначність відповідності між приведеною та структурною формами моделі, що в свою чергу дає можливість провести процедуру оцінки структурних коефіцієнтів на основі приведених.

З позиції ідентифікованості структурні моделі можна поділити на три види:

- ідентифіковані;
- неідентифіковані;
- надідентифіковані.

Рівняння структурної форми економетричної моделі є точно ідентифікованим, якщо всі невідомі його коефіцієнти однозначно відновлюються за коефіцієнтами приведеної форми без будь-яких обмежень на значення останніх.

Рівняння структурної форми називається надідентифікованим, якщо всі невідомі в ньому коефіцієнти відновлюються за коефіцієнтами приведеної форми, причому деякі з вхідних параметрів можуть приймати одночасно декілька числових значень, що відповідають одній і тій самій приведеній формі.

Рівняння структурної форми називається неідентифікованим, якщо хоча б один коефіцієнт не може бути відновлений за коефіцієнтами приведеної форми.

Дамо визначення видів економетричних моделей.

Модель вважається ідентифікованою, якщо кожне рівняння системи ідентифіковане. У даному випадку число параметрів структурної моделі рівне числу параметрів приведеної форми. Оцінювання проводиться із допомогою НМНК.

Якщо хоча б одне із рівнянь системи неідентифіковане, тоді і вся модель вважається неідентифікованою. Число параметрів структурної моделі більше числа параметрів приведеної моделі. Неможливо провести процедуру оцінювання.

Ідентифікована модель, яка містить хоча б одне надідентифіковане рівняння являється надідентифікованою. Число параметрів структурної моделі менше числа параметрів приведеної моделі. Для оцінювання використовується ДМНК.

Для встановлення ідентифікованості рівняння необхідне виконання необхідної та достатньої умов.

Необхідною умовою є критерій правила рахунку:

1) $D+1=H$ – рівняння ідентифіковане;
2) $D+1<H$ – рівняння неідентифіковане;
3) $D+1>H$ – рівняння надідентифіковане;

де D – число екзогенних змінних, які містяться в системі, але не входять в дане рівняння; H – число ендогенних змінних у рівнянні.

Достатня умова ідентифікації закладена у ранговій умові. Якщо визначник матриці, побудованої із коефіцієнтів при змінних, відсутніх у цьому рівнянні, не рівний нулю, а ранг цієї матриці не менший від кількості ендогенних змінних системи без одиниці, тоді дане рівняння є точно ідентифікованим.

<u>Приклад</u> 7.2. Виконати процедуру ідентифікації для наступних моделей.

Модель №1:

1) $R_t = a_1 + b_{11}M_t + b_{12}Y_t + u_1$,
2) $Y_t = a_2 + b_{21}R_t + b_{22}I_t + u_2$,
3) $I_t = a_3 + b_{33}R_t + u_3$,

де R_t – облікова ставка у періоді *t*; Y_t – ВВП країни у періоді *t*; I_t – інвестиції у періоді *t*; M_t – грошова маса у періоді *t*.

Розв'язування.

У нашому випадку ендогенними величинами будуть R_t, Y_t та I_t, а екзогенними – M_t.

		$D+1$		H
1) $H=2$	$D=0$	1	<	2 неідентифікована

2) H=3 D=1	2	<	3 неідентифікована
3) H=2 D=1	2	=	2 ідентифікована

Модель №2:

1) $C_t = a_1 + b_{11}Y_t + b_{12}I_t + u_1$,

2) $I_t = a_2 + b_{11}Y_{t-1} + u_2$,

3) $T_t = a_3 + b_{31}Y_t + u_3$,

4) $Y_t = C_t + I_t + G_t$,

де C_t – витрати на кінцеве споживання у періоді *t*; Y_t – сукупний дохід у періоді *t*; I_t – інвестиції у періоді *t*; T_t – витрати на заробітну плату у періоді *t*; G_t – державні витрати у періоді *t*.

Маємо: ендогенні – C_t, I_t, T_t та Y_t; екзогенні – Y_{t-1}, G_t.

	D+1		H
1) H=3 D=2	3	=	3 ідентифікована
2) H=1 D=1	2	>	1 надідентифікована
3) H=2 D=2	3	>	2 надідентифікована
4) H=3 D=1	2	<	3 неідентифікована

Приклад 7.3. Виконати повну ідентифікацію моделі грошового ринку України, тобто перевіримо необхідну та достатню умови. Маємо наступну модель:

$$R_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 M_t + u_{1t},$$

$$Y_t = \beta_0 + \beta_1 I_t + \beta_2 G_t + u_{2t},$$

$$I_t = \gamma_0 + \gamma_1 R_t + u_{3t},$$

де R_t – облікова ставка у періоді *t*; Y_t – ВВП країни у періоді *t*; I_t – інвестиції у періоді *t*; M_t – грошова маса у періоді *t*; G_t – державні витрати у періоді *t*.

Розв'язування.

Наша модель має три ендогенні змінні: Y_t, R_t, I_t та дві екзогенні: M_t, G_t.

Перевіримо виконання необхідної умови ідентифікації кожного рівняння:

	D+1		H
1) H=2 D=1	2	=	2 ідентифікована
2) H=2 D=1	2	=	2 ідентифікована
3) H=2 D=2	3	>	2 надідентифікована

Далі нам необхідно перевірити достатню умову ідентифікації. Для цього побудуємо матрицю коефіцієнтів при змінних моделі, перенісши ліву частину у праву. Маємо:

$$\alpha_0 + \alpha_1 Y_t + \alpha_2 M_t - R_t + u_{1t} = 0,$$

$$\beta_0 + \beta_1 I_t + \beta_2 G_t - Y_t + u_{2t} = 0,$$

$$\gamma_0 + \gamma_1 R_t - I_t + u_{3t} = 0.$$

Результати дій представимо у вигляді наступної таблиці:

	Y_t	R_t	I_t	M_t	G_t
1-рівняння	α_1	-1	0	α_2	0
2-рівняння	-1	0	β_1	0	β_2
3-рівняння	0	γ_1	-1	0	0

Відповідно до достатньої умови ідентифікації визначник матриці при змінних, що не входять до даного рівняння, не повинен бути рівним нулю, а ранг матриці повинен дорівнювати числу ендогенних змінних моделі мінус одиниця. Для нашого випадку $rang = 3 - 1 = 2$.

Для першого рівняння, матриця коефіцієнтів при змінних, що не входять до нього, буде:

$$A_1 = \begin{pmatrix} \beta_1 & \beta_2 \\ -1 & 0 \end{pmatrix},$$

$$rangA_1 = 2, \ \det A_1 = \begin{vmatrix} \beta_1 & \beta_2 \\ -1 & 0 \end{vmatrix} = \beta_2 \neq 0.$$

Достатня умова для першого рівняння виконується.

Для другого рівняння маємо:

$$A_2 = \begin{pmatrix} -1 & \alpha_2 \\ \gamma_1 & 0 \end{pmatrix},$$

$$rangA_2 = 2, \ \det A_2 = \begin{vmatrix} -1 & \alpha_2 \\ \gamma_1 & 0 \end{vmatrix} = \gamma_1 \alpha_2 \neq 0.$$

Достатня умова виконується для другого рівняння.

Для третього рівняння маємо:

$$A_3 = \begin{pmatrix} \alpha_1 & \alpha_2 & 0 \\ -1 & 0 & \beta_2 \end{pmatrix},$$

$rangA_3 = 2$. Визначник квадратної матриці не дорівнює нулю, а $\det A_3^* = \begin{vmatrix} \alpha_1 & \alpha_2 \\ -1 & 0 \end{vmatrix} = \alpha_2 \neq 0$.

Достатня умова виконується.

Отже, нами доведено, що перші два рівняння є ідентифікованими, тому для їх оцінки використаємо непрямий МНК

(НМНК). Оскільки третє рівняння є надідентифікованим, оцінити його можна за допомогою двокрокового МНК (ДМНК).

Алгоритм НМНК складається із наступних кроків:

1) на основі структурної форми системи одночасних рівнянь будується її приведена форма, всі параметри якої виражаються через структурні коефіцієнти;

2) приведені коефіцієнти кожного рівняння оцінюються звичайним НМК;

3) на основі оцінок приведених коефіцієнтів системи одночасних рівнянь визначаються оцінки структурних коефіцієнтів через приведене рівняння.

Припустимо, що нас цікавить взаємозв'язок між ендогенними величинами: Y_{1t} – експорт, Y_{2t} – імпорт та екзогенними величинами: x_{1t} – національний дохід України, x_{2t} – зовнішній товарообіг країн ЄС.

Нехай між заданими ендогенними та екзогенними величинами існує лінійна форма зв'язку у вигляді взаємопов'язаних регресій:

$$\begin{aligned} Y_{1t} &= b_{12}Y_{2t} + a_{10} + a_{11}x_{1t} + u_{1t} \\ Y_{2t} &= b_{21}Y_{1t} + a_{20} + a_{22}x_{2t} + u_{2t} \end{aligned} \qquad (7.18)$$

Представимо систему регресій у прогнозній (приведеній) формі:

$$\begin{aligned} Y_{1t} &= c_{10} + c_{11}x_{1t} + c_{12}x_{2t} + e_{1t} \\ Y_{2t} &= c_{20} + c_{21}x_{1t} + c_{22}x_{2t} + e_{2t} \end{aligned} \qquad (7.19)$$

Для оцінки параметрів прогнозної форми системи регресій використаємо МНК для кожної з них.

Позначимо через:

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \cdots & \cdots \\ 1 & x_{1n} & x_{2n} \end{bmatrix}; \quad Y = \begin{bmatrix} y_{11} & y_{21} \\ y_{12} & y_{22} \\ \cdots & \cdots \\ y_{1n} & y_{2n} \end{bmatrix}$$

Тоді отримаємо наступну систему нормальних рівнянь:

$$X'X = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \end{bmatrix} \times \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{2i}x_{1i} \\ \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{1i}x_{2i} & \sum_{i=1}^{n} x_{2i}^2 \end{bmatrix} \qquad (7.20)$$

$$X'Y=\begin{bmatrix}1 & 1 & \cdots & 1\\ x_{11} & x_{12} & \cdots & x_{1n}\\ x_{21} & x_{22} & \cdots & x_{2n}\end{bmatrix}\times\begin{bmatrix}y_{11} & y_{21}\\ y_{12} & y_{22}\\ \vdots & \vdots\\ y_{1n} & y_{2n}\end{bmatrix}=\begin{bmatrix}\sum\limits_{i=1}^{n} y_{1i} & \sum\limits_{i=1}^{n} y_{2i}\\ \sum\limits_{i=1}^{n} y_{1i}x_{1i} & \sum\limits_{i=1}^{n} y_{2i}x_{1i}\\ \sum\limits_{i=1}^{n} y_{1i}x_{2i} & \sum\limits_{i=1}^{n} y_{2i}x_{2i}\end{bmatrix} \qquad (7.21)$$

Отже, матриця параметрів прогнозної системи регресій буде:

$$C=(X'X)^{-1}\cdot(X'Y)=\begin{bmatrix}c_{10} & c_{20}\\ c_{11} & c_{21}\\ c_{12} & c_{22}\end{bmatrix} \qquad (7.22)$$

Нам відомо, що якщо параметри структурної форми (7.18) одночасно визначаються через параметри прогнозної форми (7.19) системи регресій, то така економетрична модель є ідентифікованою. Якщо число оцінюваних параметрів структурної форми регресій більше числа оцінюваних параметрів прогнозної форми, тобто число оцінюваних параметрів прогнозної форми регресії більше числа рівнянь, то система є неіндефікованою.

Параметри прогнозної форми є комбінацією всіх параметрів структурної форми регресій. За параметрами прогнозної форми не можна робити висновок про взаємозалежність ендогенних величин, оскільки при переході до прогнозної форми регресії вони розподіляються на екзогенні величини та відхилення. З іншого боку, структурна форма не придатна для визначення прогнозних значень ендогенних величин, тому що в правій частині регресії знаходяться значення ендогенних величин.

Розглянемо методику розрахунку параметрів структурної форми регресії через параметри прогнозної форми. Економетричну модель (7.18) представимо у такому вигляді:

$$\begin{aligned}-Y_{1t}+b_{12}Y_{2t}+a_{10}+a_{11}x_{1t}+u_{1t}&=0\\ -Y_{2t}+b_{21}Y_{1t}+a_{20}+a_{22}x_{2t}+u_{2t}&=0,\end{aligned} \qquad (7.23)$$

або у матричній формі:

$$B\cdot Y+A\cdot X+U=0. \qquad (7.24)$$

$$B=\begin{bmatrix}-1 & b_{12}\\ b_{21} & -1\end{bmatrix},\quad A=\begin{bmatrix}a_{10} & a_{11} & 0\\ a_{20} & 0 & a_{22}\end{bmatrix};$$

$$X=\begin{bmatrix}1 & x_{11} & x_{21}\\ 1 & x_{12} & x_{22}\\ \dots & \dots & \dots\\ 1 & x_{1n} & x_{2n}\end{bmatrix},\quad Y=\begin{bmatrix}y_{11} & y_{21}\\ y_{12} & y_{22}\\ \dots & \dots\\ y_{1n} & y_{2n}\end{bmatrix},\quad U=\begin{bmatrix}u_{11} & u_{21}\\ u_{12} & u_{22}\\ \dots & \dots\\ u_{1n} & u_{2n}\end{bmatrix},$$

де B – матриця параметрів при ендогенних величинах структурної форми системи регресій, A – матриця параметрів при екзогенних величинах структурної форми системи регресій, X – матриця екзогенних величин, Y – матриця ендогенних величин. Тоді прогнозна форма системи регресій матиме вигляд:

$$Y = C \cdot X + e, \tag{7.25}$$

або у розгорнутій формі:

$$\begin{aligned} Y_1 &= c_{10} + c_{11}x_1 + c_{12}x_2 + e_1 \\ Y_2 &= c_{20} + c_{21}x_1 + c_{22}x_2 + e_2 \end{aligned} \tag{7.26}$$

Запишемо структурну форму регресій у вигляді матричного рівняння:

$$\begin{bmatrix} -1 & b_{12} \\ b_{21} & -1 \end{bmatrix} \times \begin{bmatrix} c_{10} & c_{11} & c_{12} \\ c_{20} & c_{21} & c_{22} \end{bmatrix} = -\begin{bmatrix} a_{10} & a_{11} & 0 \\ a_{20} & 0 & a_{22} \end{bmatrix}. \tag{7.27}$$

Знайдемо добуток матриць B і C та прирівняємо його до матриці A:

$$\begin{aligned} -c_{10} + b_{12}c_{20} &= -a_{10}, \qquad & b_{21}c_{10} - c_{20} &= -a_{20}, \\ -c_{11} + b_{12}c_{21} &= -a_{11}, \qquad & b_{21}c_{11} - c_{21} &= 0, \\ -c_{12} + b_{12}c_{22} &= 0, \qquad & b_{21}c_{12} - c_{12} &= -a_{22}. \end{aligned} \tag{7.28}$$

З цієї системи знайдемо параметри оцінки елементів матриць B і A:

$$\begin{aligned} & b_{12} = \frac{c_{12}}{c_{22}}; \quad b_{21} = \frac{c_{21}}{c_{11}}; \\ & a_{10} = c_{10} - \frac{c_{12} \cdot c_{20}}{c_{22}}; \; a_{11} = c_{11} - \frac{c_{12} \cdot c_{21}}{c_{22}}; \\ & a_{20} = c_{20} - \frac{c_{21} \cdot c_{10}}{c_{11}}; \; a_{22} = c_{22} - \frac{c_{21} \cdot c_{12}}{c_{11}}. \end{aligned} \tag{7.29}$$

Приклад 7.4. Виконати оцінку наступної моделі:

$$\begin{aligned} Y_{1t} &= b_{12}Y_{2t} + a_{10} + a_{11}x_{1t} + u_{1t} \\ Y_{2t} &= b_{21}Y_{1t} + a_{20} + a_{22}x_{2t} + u_{2t} \end{aligned},$$

де Y_{1t} – вартість основних виробничих фондів; Y_{2t} – кількість працюючих у періоді t; x_{1t} – обсяг інвестицій; x_{2t} – обсяг продукції.

Дані для розрахунків приведено у табл. 7.2.

Таблиця 7.2

Період, t	Вартість ОВФ, Y_{1t}	Кількість працюючих, Y_{2t}	Обсяг інвестицій, x_{1t}	Обсяг продукції, x_{2t}
1	70	3,9	2,0	30
2	75	4,1	2,4	32
3	76	4,2	2,5	34

4	83	4,6	2,6	36
5	82	4,5	2,6	37
6	74	4,1	1,9	34
7	72	4,2	1,7	32
8	76	4,3	1,8	36
9	78	4,4	2,0	39
10	80	4,6	2,1	40

Розв'язування. Виконаємо ідентифікацію даної моделі. Маємо:
ендогенні змінні – Y_{1t}, Y_{2t};
екзогенні змінні – x_{1t}, x_{2t}.

	D+1		H
1) H=2 D=1	2	=	2 ідентифікована
2) H=2 D=1	2	=	2 ідентифікована

Перевіримо достатню умову ідентифікованості.
Маємо:

$$\begin{array}{l} b_{12}Y_{2t} + a_{10} + a_{11}x_{1t} - Y_{1t} + u_{1t} = 0 \\ b_{21}Y_{1t} + a_{20} + a_{22}x_{2t} - Y_{2t} + u_{2t} = 0 \end{array}.$$

Побудуємо наступну матрицю:

	Y_{1t}	Y_{2t}	x_{1t}	x_{2t}
1-рівняння	-1	b_{12}	a_{11}	0
2-рівняння	b_{21}	-1	0	a_{22}

Як бачимо визначники матриць відповідно будуть: $\det A_1 = a_{22} \neq 0$, $\det A_2 = a_{11} \neq 0$. При цьому $rangA_1 = rangA_2 = 1$, що відповідає умові: число екзогенних мінус одиниця. Отже, нами доведено, що обидва рівняння є ідентифікованими. Тому для подальшої оцінки можна використати НМНК.

Визначимо параметри приведеної форми. Відповідні матриці приймуть наступний вид:

$$Y = \begin{pmatrix} 70 & 3,9 \\ 75 & 4,1 \\ 76 & 4,2 \\ 83 & 4,6 \\ 82 & 4,5 \\ 74 & 4,1 \\ 72 & 4,2 \\ 76 & 4,3 \\ 78 & 4,4 \\ 80 & 4,6 \end{pmatrix}; \qquad X = \begin{bmatrix} 1 & 2,0 & 30 \\ 1 & 2,4 & 32 \\ 1 & 2,5 & 34 \\ 1 & 2,6 & 36 \\ 1 & 2,6 & 37 \\ 1 & 1,9 & 34 \\ 1 & 1,7 & 32 \\ 1 & 1,8 & 36 \\ 1 & 2,0 & 39 \\ 1 & 2,1 & 40 \end{bmatrix}.$$

Знайдемо $X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2,0 & 2,8 & 2,5 & 2,6 & 2,6 & 1,9 & 1,7 & 1,8 & 2,0 & 2,1 \\ 30 & 32 & 34 & 36 & 37 & 34 & 32 & 36 & 39 & 40 \end{bmatrix}.$

Знайдемо:

$$\left(X' \cdot X\right) = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 2,0 & 2,8 & 2,5 & 2,6 & 2,6 & 1,9 & 1,7 & 1,8 & 2,0 & 2,1 \\ 30 & 32 & 34 & 36 & 37 & 34 & 32 & 36 & 39 & 40 \end{bmatrix} \times \begin{bmatrix} 1 & 2,0 & 30 \\ 1 & 2,4 & 32 \\ 1 & 2,5 & 34 \\ 1 & 2,6 & 36 \\ 1 & 2,6 & 37 \\ 1 & 1,9 & 34 \\ 1 & 1,7 & 32 \\ 1 & 1,8 & 36 \\ 1 & 2,0 & 39 \\ 1 & 2,1 & 40 \end{bmatrix} =$$

$$= \begin{pmatrix} 10 & 21,6 & 350 \\ 21,6 & 47,68 & 757,4 \\ 350 & 757,4 & 12342 \end{pmatrix}.$$

Знайдемо:

$$\left(X'X\right)^{-1} = \begin{bmatrix} 16,0565 & -1,62302 & -0,35574 \\ -1,62302 & 0,997312 & -0,01518 \\ -0,35574 & -0,01518 & 0,011101 \end{bmatrix}.$$

Знайдемо: $X' \cdot Y = \begin{bmatrix} 766 & 42,9 \\ 1662,8 & 92,93 \\ 26904 & 1507,3 \end{bmatrix}.$

Знайдемо: $C' = \left(X'X\right)^{-1} \cdot \left(X'Y\right) = \begin{bmatrix} 29,7871 & 1,795005 \\ 2,791258 & 0,177261 \\ 0,918394 & 0,060346 \end{bmatrix}.$

Знайдемо: $C = \begin{bmatrix} 29,7871 & 2,791258 & 0,918394 \\ 1,795005 & 0,177261 & 0,060346 \end{bmatrix}.$

Нами отримано зведену модель:

$$\begin{aligned} Y_{1t} &= 29{,}7871 + 2{,}791258x_{1t} + 0{,}918394x_{2tt} \\ Y_{2t} &= 1{,}795005 + 0{,}177261x_{1t} + 0{,}060346x_{2t} \end{aligned}.$$

Матриця B та A у структурній формі мають вид:

$$B = \begin{bmatrix} -1 & b_{12} \\ b_{21} & -1 \end{bmatrix}, \quad A = \begin{bmatrix} a_{10} & a_{11} & 0 \\ a_{20} & 0 & a_{22} \end{bmatrix};$$

Запишемо структурну форму регресій у вигляді матричного рівняння:

$$B \cdot C = -A$$

або

$$\begin{bmatrix} -1 & b_{12} \\ b_{21} & -1 \end{bmatrix} \times \begin{bmatrix} c_{10} & c_{11} & c_{12} \\ c_{20} & c_{21} & c_{22} \end{bmatrix} = -\begin{bmatrix} a_{10} & a_{11} & 0 \\ a_{20} & 0 & a_{22} \end{bmatrix}.$$

Маємо: $\begin{bmatrix} -1 & b_{12} \\ b_{21} & -1 \end{bmatrix} \times \begin{bmatrix} 29{,}7871 & 2{,}791258 & 0{,}918394 \\ 1{,}795005 & 0{,}177261 & 0{,}060346 \end{bmatrix} = -\begin{bmatrix} a_{10} & a_{11} & 0 \\ a_{20} & 0 & a_{22} \end{bmatrix}.$

Перемножимо у лівій частині дві матриці і прирівняємо до правої частини і далі необхідно розв'язати систему рівнянь відносно b_{12}, b_{21}, a_{10}, a_{11}, a_{20} та a_{22}.

Використаємо формули (7.29):

$$b_{12} = \frac{c_{12}}{c_{22}} = \frac{0{,}918394}{0{,}060346} = 15{,}2188; \quad b_{21} = \frac{c_{21}}{c_{11}} = \frac{0{,}177261}{2{,}791258} = 0{,}063506;$$

$$a_{10} = c_{10} - \frac{c_{12} \cdot c_{20}}{c_{22}} = 2{,}469269; \; a_{11} = c_{11} - \frac{c_{12} \cdot c_{21}}{c_{22}} = 0{,}093557;$$

$$a_{20} = c_{20} - \frac{c_{21} \cdot c_{10}}{c_{11}} = -0{,}09665; \; a_{22} = c_{22} - \frac{c_{21} \cdot c_{12}}{c_{11}} = 0{,}002023.$$

Отже, наша модель у структурній формі прийме вид:

$$\begin{aligned} \hat{Y}_{1t} &= 15{,}2188Y_{2t} + 2{,}469269 + 0{,}093557x_{1t} \\ \hat{Y}_{2t} &= 0{,}063506Y_{1t} + 0{,}09665 + 0{,}002023x_{2t} \end{aligned}.$$

У тих випадках, коли рівняння структурної форми моделі надідентифіковане, то для оцінки параметрів регресій використовують двокроковий метод найменших квадратів (ДМНК). Він являє собою звичайний МНК для оцінки параметрів структурного рівняння, яку проводять у два етапи. ДМНК дозволяє оцінити параметри одного регресійного рівняння з врахуванням його взаємозв'язків з іншими рівняннями.

Припустимо, що нам задано структурне рівняння виду:

$$y_i = Y_i b_i + X_i a_i + u_i, i = \overline{1, m}, \tag{7.30}$$

де y_{i} – вектор спостережень над сумісно залежною змінною, яка визначається за допомогою i-го структурного рівняння; Y_i – матриця, яка складається із векторів ендогенних змінних, що входять у праву частину i-го рівняння системи; X_{i} – матриця, яка складається з векторів ендогенних змінних системи, що входять у i-те рівняння (включаючи вільний член); X – матриця, яка складається із векторів всіх ендогенних змінних системи; b_i – вектор оцінок параметрів залежних змінних, які містяться у матриці Y_i; a_{i} – вектор оцінок параметрів наперед визначених змінних матриці X_i; u_i – вектор залишків i-го структурного рівняння для всіх періодів спостережень.

Нехай на основі правила рахунку рівняння (7.30) надідентифіковане. Сумісно залежні змінні, які містяться в матриці Y_i не є стохастично незалежними відносно залишків i-го структурного рівняння u_i. Тому безпосереднє використання МНК приведе до необґрунтованості оцінок. Основна ідея ДМНК полягає в заміні матриці Y_i матрицею оцінок $\hat{Y}_i$ (матриця значень регресії). Завдяки даній процедурі, змінні, які містяться в матриці, набувають характеру наперед визначених змінних, і тоді використання МНК дає задовільний результат.

Отже, на першому кроці використання ДМНК полягає у визначенні матриці значень регресій $\hat{Y}_i$. Для цього будується зведена форма сумісно залежних змінних матриці Y_i:

$$Y_i = XC_i + V_i. \tag{7.31}$$

Для побудови зведеної форми (7.31) повинні бути задані всі наперед визначені змінні моделі.

Матрицю значень регресій $\hat{Y}_i$ одержуємо із (7.31) шляхом наступних перетворень:

$$\hat{Y}_i = Y_i - V_i = XC_i + V_i - V_i = XC_i. \tag{7.32}$$

Значення регресій матриці $\hat{Y}_i$ незалежні від збурюючих змінних приведеної та структурної форм, оскільки вони є лінійними функціями тільки від наперед визначених змінних. Таким чином, окремі рівняння (7.32) являють собою множинну регресію, для якої виконуються передумови регресійного аналізу.

Далі використаємо МНК для оцінювання параметрів матриці C_i:

$$C_i = (X'X)^{-1} X'Y_i. \tag{7.33}$$

Тепер підставимо (7.33) в (7.32), як результат одержимо матрицю значень регресії:

$$\hat{Y}_i = (X'X)^{-1} X'Y_i. \tag{7.34}$$

Отже, поставлена задача на першому кроці виконана.

На другому етапі матрицю Y_i в (7.30) замінюють матрицею $\hat{Y}_i$ з врахуванням (7.32) та (7.31):

$$y_i = \left(\hat{Y}_i + V_i\right)b_i + X_i a_i + u_i = Y_i b_i + X_i a_i + u_i + V_i b_i$$

або
$$y_i = \hat{Y}_i b_i + X_i a_i + e_i, \tag{7.35}$$

де $e_i = u_i + V_i b_i$.

У результаті таких перетворень рівняння (7.35) в правій частині містить тільки наперед визначені змінні, оскільки матриці X_i містять тільки наперед визначені змінні, а елементи матриці $\hat{Y}_i$ “наперед визначені” через (7.34). При цьому значення регресій $\mathbf{y}_i$ більше не корелюють із залишками $\mathbf{e}_i$. Таким чином, вираз (7.35) являє собою рівняння множинної регресії, для якого виконуються передумови регресійного аналізу. Невідомі параметри регресій a_i та b_i можуть бути оцінені з допомогою МНК.

Процедура дворазового використання МНК може бути представлена у вигляді однієї формули. Для цього необхідно побудувати систему нормальних рівнянь для рівняння регресій (7.35).

Покладемо:

$$Z_i = \left(\hat{Y}_i X_i\right) \quad \text{та} \quad d_i = \begin{bmatrix} b_i \\ a_i \end{bmatrix}. \tag{7.36}$$

Тоді (7.35) прийме вигляд:

$$y_i = Z_i d_i + e_i. \tag{7.37}$$

Звідси одержимо наступну систему нормальних рівнянь:

$$Z_i' Z_i d_i = Z_i' y_i. \tag{7.38}$$

Далі підставимо (7.36) в (7.38), як результат маємо:

$$\begin{bmatrix} \hat{Y}_i'\hat{Y}_i & \hat{Y}_i X_i \\ X_i'\hat{Y}_i & X_i' X_i \end{bmatrix} \cdot \begin{bmatrix} b_i \\ a_i \end{bmatrix} = \begin{bmatrix} \hat{Y}_i' y_i \\ X_i' y_i \end{bmatrix}. \tag{7.39}$$

Враховуючи (7.34) отримаємо наступне рівняння для знаходження оцінок ДМНК:

$$\begin{bmatrix} b_i \\ a_i \end{bmatrix} = \begin{bmatrix} Y_i' X\left(X'X\right)^{-1} X'Y_i & \hat{Y}_i X_i \\ X_i'\hat{Y}_i & X_i' X_i \end{bmatrix}^{-1} \cdot \begin{bmatrix} Y_i' X\left(X'X\right)^{-1} X'y_i \\ X_i' y_i \end{bmatrix}. \tag{7.40}$$

Формула (7.40) є результативною формулою використання ДМНК до *i*-го структурного рівняння. Як бачимо, матриця значень регресій $\hat{Y}_i$, знайдена на першому етапі використання даного методу, не міститься в (7.40) в явному вигляді. У дану формулу входять матриці та вектори спостережень. Перевага двокрокового методу полягає, по-перше, в тому, що його можна використати до надідентифікованих рівнянь, і, по-друге, в тому, що нерозглянуті нами структурні рівняння моделі не повинні бути точно специфіковані. Зрозуміло, що повинні бути відомими всі наперед визначені змінні моделі і вказані результати спостережень над ними. Недолік методу полягає в тому, що в оцінках містяться не залишки i-го структурного рівняння – u_i, а залишки рівняння, отриманого на другому етапі – e_i.

Представимо процедуру оцінювання параметрів економетричної моделі у вигляді наступного алгоритму.

1. Проводиться перевірка кожного рівняння моделі на ідентифікованість. Якщо рівняння надідентифіковане, то для оцінювання їх параметрів використовуємо оператор оцінювання:

$$\begin{bmatrix} b_i \\ a_i \end{bmatrix} = \begin{bmatrix} Y_i'X\left(X'X\right)^{-1}X'Y_i & Y_i'X_i \\ X_i'Y_i & X_i'X_i \end{bmatrix}^{-1} \cdot \begin{bmatrix} Y_i'X\left(X'X\right)^{-1}X'y_i \\ X_i'y_i \end{bmatrix}.$$

Далі необхідно розрахувати наступні показники (матриці чи вектори).

2. $A_1 = (X' \cdot X)^{-1}$; 3. $A_2 = (Y_i' \cdot X)$; 4. $A_3 = (X' \cdot Y_i)$; 5. $A_4 = A_2 \cdot A_1 \cdot A_3$;

6. $A_5 = (X_i' \cdot Y_i)$; 7. $A_6 = (Y_i' \cdot X_i)$; 8. $A_7 = (X_i' \cdot X_i)$; 9. $Q_i = \begin{bmatrix} A_4 & A_6 \\ A_5 & A_7 \end{bmatrix}$;

10. Q_i^{-1}; 11 $A_8 = (X' \cdot y_i)$; 12. $A_9 = A_2 \cdot A_1 \cdot A_8$; 13. $A_{10} = (X_i' \cdot y_1)$:

14. $A_{11} = \begin{bmatrix} A_9 \\ A_{10} \end{bmatrix}$; 15. $A_{12} = Q_i^{-1} \cdot A_{11}$.

У результаті нами отримано вектор A_{12} оцінки параметрів економетричної моделі.

Приклад 7.5. Побудувати економетричну модель, яка містить регресійні рівняння валової продукції та прибутку підприємств регіону. Вхідні дані представлені в табл. 7.3.

Розв’язування.

Виконаємо ідентифікацію змінних величин моделі: Y_1 , Y_2 – ендогенні змінні; X_1, X_2, X_3 – екзогенні змінні.

Економетричну модель представимо наступним чином:

$$\begin{aligned} Y_1 &= f_1(Y_2, X_1, X_2, u_1) \\ Y_2 &= f_2(Y_1, X_1, X_3, u_2) \end{aligned}.$$

Для оцінки параметрів моделі використаємо оператор ДМНК:

$$\begin{bmatrix} b_i \\ a_i \end{bmatrix} = \begin{bmatrix} Y_i'X\left(X'X\right)^{-1}X'Y_i & Y_i'X_i \\ X_i'Y_i & X_i'X_i \end{bmatrix}^{-1} \cdot \begin{bmatrix} Y_i'X\left(X'X\right)^{-1}X'y_i \\ X_i'y_i \end{bmatrix}.$$

Таблиця 7.3.

№ п-ва	Валова продукція, млн.грн.	Прибуток, млн. грн	Вартість основних виробничих фондів, млн. грн.	Ціна прод. тис. грн.	Затрати праці, млн. л. год.
	Y_1	Y_2	X_1	X_2	X_3
1	12,3	4,3	22,3	3,2	14,2
2	14,5	8,2	28,4	4,3	12,3
3	16,7	10,1	30,6	5,6	13,4
4	18,3	12,1	35,4	5,9	11,5
5	20,1	18,1	40,8	5,4	9,6
6	25,2	22,1	52,4	5,0	10,6
7	28,3	20,1	60,3	6,8	8,4
8	34,0	26,1	68,5	6,3	6,5
9	38,2	35,4	76,3	7,6	6,4
10	40,0	36,1	80,5	8,2	5,8

Відповідно до даного оператора оцінювання запишемо матриці вхідних змінних і розрахуємо необхідні матриці на основі побудованого алгоритму.

$$X = \begin{bmatrix} 1 & 22,3 & 3,2 & 14,2 \\ 1 & 28,4 & 4,3 & 12,3 \\ 1 & 30,6 & 5,6 & 13,4 \\ 1 & 35,4 & 5,9 & 11,5 \\ 1 & 40,8 & 6,4 & 9,6 \\ 1 & 52,4 & 5 & 10,2 \\ 1 & 60,3 & 6,8 & 8,4 \\ 1 & 68,5 & 6,3 & 6,5 \\ 1 & 76,3 & 7,6 & 6,4 \\ 1 & 80,5 & 8,2 & 5,8 \end{bmatrix}$$

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 22,3 & 28,4 & 30,6 & 35,4 & 40,8 & 52,4 & 60,3 & 68,5 & 76,3 & 80,5 \\ 3,2 & 4,3 & 5,6 & 5,9 & 6,4 & 5 & 6,8 & 6,3 & 7,6 & 8,2 \\ 14,2 & 12,3 & 13,4 & 11,5 & 9,6 & 10,2 & 8,4 & 6,5 & 6,4 & 5,8 \end{bmatrix}$$

$$Y_1 = \begin{bmatrix} 12,3 \\ 14,5 \\ 16,7 \\ 18,3 \\ 20,1 \\ 25,2 \\ 28,3 \\ 34 \\ 38,2 \\ 40 \end{bmatrix}; \quad y_1 = \begin{bmatrix} 4,3 \\ 8,2 \\ 10,1 \\ 12,1 \\ 18,1 \\ 22,1 \\ 20,1 \\ 26,2 \\ 35,4 \\ 36,1 \end{bmatrix}; \quad X_1 = \begin{bmatrix} 1 & 22,3 & 3,2 \\ 1 & 28,4 & 4,3 \\ 1 & 30,6 & 5,6 \\ 1 & 35,4 & 5,9 \\ 1 & 40,8 & 6,4 \\ 1 & 52,4 & 5 \\ 1 & 60,3 & 6,8 \\ 1 & 68,5 & 6,3 \\ 1 & 76,3 & 7,6 \\ 1 & 80,5 & 8,2 \end{bmatrix}$$

$$X_i' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 22,3 & 28,4 & 30,6 & 35,4 & 40,8 & 52,4 & 60,3 & 68,5 & 76,3 & 80,5 \\ 3,2 & 4,3 & 5,6 & 5,9 & 6,4 & 5 & 6,8 & 6,3 & 7,6 & 8,2 \end{bmatrix}$$

$$X' \cdot X = \begin{bmatrix} 10 & 495,5 & 59,3 & 98,3 \\ 495,5 & 28534,05 & 3178,39 & 4316,27 \\ 59,3 & 3178,39 & 371,79 & 547,93 \\ 98,3 & 4316,27 & 547,93 & 1048,35 \end{bmatrix}$$

$$Y_1' = [12,3 \quad 14,5 \quad 16,7 \quad 18,3 \quad 20,1 \quad 25,2 \quad 28,3 \quad 34 \quad 38,2 \quad 40]$$

$$y_1' = [4,3 \quad 8,2 \quad 10,1 \quad 12,1 \quad 18,1 \quad 22,1 \quad 20,1 \quad 26,2 \quad 35,4 \quad 36,1]$$

$$A_2 = Y_1' \cdot X = [247,6 \quad 14155,64 \quad 1582,8 \quad 2172,44]$$

$$A_1 = \begin{bmatrix} 71,856 & -0,4680 & -1,6109 & -3,9689 \\ -0,468 & 0,0043 & -0,0030 & 0,0278 \\ -1,6109 & -0,003 & 0,1936 & 0,0622 \\ -3,9689 & 0,0278 & 0,0622 & 0,2263 \end{bmatrix}, \quad A_3 = \begin{bmatrix} 247,6 \\ 14155,64 \\ 1582,8 \\ 2172,44 \end{bmatrix}.$$

$$A_2 \cdot A_1 = [-5,1247 \quad 0,5078 \quad 0,2286 \quad 0,3425] \text{ . } A_4 = A_2 \cdot A_1 \cdot A_3 = 7025,464$$

$$A_5 = (X_1' \cdot Y_1) = \begin{bmatrix} 247,6 \\ 14155,64 \\ 1582,8 \end{bmatrix}, \quad A_6 = (Y_1' \cdot X_1) = [247,6 \quad 14155,64 \quad 1582,8].$$

$$A_7 = (X_1' \cdot X_1) = \begin{bmatrix} 10 & 495,5 & 59,3 \\ 495,5 & 28534,05 & 3178,39 \\ 59,3 & 3178,39 & 371,79 \end{bmatrix}.$$

$$Q = \begin{bmatrix} 7025,464 & 247,6 & 14155,64 & 1582,8 \\ 247,6 & 10 & 495,5 & 59,3 \\ 14155,64 & 495,5 & 28534,05 & 3178,39 \\ 1582,8 & 59,3 & 3178,39 & 371,79 \end{bmatrix}.$$

$$Q^{-1} = \begin{bmatrix} 1,9287 & -1,7029 & -0,8984 & -0,2593 \\ -1,7029 & 3,7482 & 0,8120 & -0,2903 \\ -0,8984 & 0,8121 & 0,4193 & 0,1101 \\ -0,2593 & -0,2903 & 0,1101 & 0,2113 \end{bmatrix}.$$

$$A_2 \cdot A_1 = [-5,1247 \quad 0,5078 \quad 0,2286 \quad 0,3425].$$

$$A_8 = (X' \cdot Y_1) = \begin{bmatrix} 192,8 \\ 11584,12 \\ 1270,87 \\ 1611,31 \end{bmatrix}, A_9 = 5736,769, A_{10} = (X_1' \cdot Y_1) = \begin{bmatrix} 192,8 \\ 11584,12 \\ 1270,87 \end{bmatrix},$$

$$A_{11} = \begin{bmatrix} A_9 \\ A_{10} \end{bmatrix} = \begin{bmatrix} 5733,186 \\ 192,8 \\ 11584,12 \\ 1270,87 \end{bmatrix}, A_{12} = Q^{-1} \cdot A_{11} = \begin{bmatrix} 0,1741 \\ -8,7797 \\ 0,3738 \\ 0,8799 \end{bmatrix}.$$

Запишемо регресійне рівняння прибутку для підприємств регіону, яке оцінюється з допомогою ДМНК:

$$Y_2 = -8{,}7797 + 0{,}1741Y_1 + 0{,}3738X_1 + 0{,}8799X_3.$$

Аналогічно можна розрахувати регресійне рівняння випуску валової продукції.

Література:

1. Доля В.Т. Економетрія: навч. посібник / В.Т. Доля; Харк. нац. акад. міськ. госп-ва. – Х.: ХНАМГ, 2010. – 171 с.
2. Доугерти К. Введение в эконометрику: Пер. с англ. – М.: ИНФРА-М, 1997. – 402 с.
3. Економіко-математичне моделювання: Навч. посібник / За ред. О.Т. Іващука. – Тернопіль: ТНЕУ, Економічна думка, 2008. – 704 с.
4. Економетрика : Підручник / [О. І. Черняк, О. В. Комашко, А. В. Ставицький, О. В. Баженова] За ред.. О. І. Черняка. – К. : ВПЦ «Київський університет», 2010. – 359 с.
5. Єлейко В. Основи економетрії. – Львів: "Марка ЛТД", 1995. – 191 с.
6. Іващук О.Т. Математичні методи та моделі в управлінні виробництвом.: Навч.посіб. – К.: ІСДО, 1993. – 180 с.
7. Іващук О.Т., Кулаічев О.П. Методи економетричного аналізу даних у системі STADIA.: Навч.посіб. – Т.: ТАНГ, 2001. – 151 с.
8. Кейн Э. Экономическая статистика и эконометрия. Введение в количественный экономический анализ. – М.: Статистика, 1997. – 254 с.
9. Корольов О.А. Економетрія: Навч.посіб. – К.: КНТЕУ, 2000. – 660 с.
10. .Лугінін В. М. Економетрія: навч. посіб. / В. М. Лугінін. – К. : ЦНЛ, 2008. – 312 с.
11. Лук'яненко І.Г., Краснікова Л.І. Економетрика: Підручник. – К.: Знання, 1998. – 494 с.
12. Магнус Я.Р. Катышев П.К., Пересецкий А.А. Эконометрика.: Начальный курс. – М.: Дело, 1997. – 248 с.
13. Наконечний С.І., Терещенко Т.О., Романюк Т.П. Економетрія: Підручник. – Вид. 2-ге доповн. Та перероб. – К.: КНЕУ, 2000. – 296 с.
14. Толбатов Ю.А. Економетрика: Підручник. – К.: Четверта хвиля, 1997. – 320 с.
15. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере. – М.: ИНФРА-М, 1995. – 384 с.

Навчальне видання

Іващук Олег Тимофійович

Дзюбановська Наталія Володимирівна

Економетрика

Методичні рекомендації для підготовки до практичних занять