

ФАКУЛЬТЕТ ЕКОНОМІКИ ТРАНСПОРТУ
Кафедра управління державними і корпоративними фінансами

М.О. Єрьоміна

ЕКОНОМЕТРИКА

Конспект лекцій

Харків – 2015

Єрємін М.О. Економетрика: Конспект лекцій. – Харків: УкрДУЗТ, 2015. – 40 с.

Цей конспект лекцій розкриває методологію економіко-математичного моделювання з використанням економетричних методів. Наведено основні етапи аналізу та розрахункові формули. Цей конспект лекцій може використовуватися як теоретичне джерело при написанні розділу дипломної роботи щодо фінансового аналізу та аналізу господарчої діяльності підприємства, використання моделей для прогнозування та аналізу економічних явищ.

Рекомендується для студентів економічних спеціальностей усіх форм навчання.

Іл. 1, бібліогр.: 18 назв.

Конспект лекцій розглянуто та рекомендовано до друку на засіданні кафедри управління державними і корпоративними фінансами 3 лютого 2015 р., протокол № 8.

Рецензент

проф. І.Ю. Зайцева

М.О. Єрємін

ЕКОНОМЕТРИКА

Конспект лекцій

Відповідальний за випуск Єрємін М.О.

Редактор Еткало О.О.

Підписано до друку 13.05.15 р.

Формат паперу 60x84 1/16. Папір писальний.

Умовн.-друк.арк. 1,5. Тираж 50. Замовлення №

Видавець та виготовлювач Українська державна академія залізничного транспорту,

61050, Харків-50, майдан Фейербаха, 7.

Свідоцтво суб'єкта видавничої справи ДК № 2874 від 12.06.2007 р.

ЗМІСТ

Вступ.....	4
...	
Тема 1. Предмет курсу «Економетрика». Роль та етапи економетричного дослідження. Метод експертних оцінок.....	5
..	
1.1 Економетрика як наука та історія її виникнення.....	5
1.2 Економетричні моделі та етапи проведення економетричного аналізу.....	6
....	
1.3 Метод експертних оцінок.....	8
Тема 2. Специфікація економетричної моделі.....	11
2.1 Вибір виду функції.....	11
2.2 Розгляд основних видів математичних функцій, що найчастіше зустрічаються в економетричному аналізі.....	13
Тема 3. Проста вибіркова лінійна регресія.....	16
3.1 Загальне поняття про вибірку лінійну регресію. Оцінка параметрів лінійної регресії.....	16
3.2 Визначення кореляційного зв'язку між факторами та результативним показником.....	18
3.3 Перевірка простої регресійної моделі на адекватність..	19
Тема 4. Загальна проста лінійна регресія.....	21
4.1 Загальна проста лінійна модель. Перевірка на значущість параметрів загальної простої лінійної регресії. Інтервали довіри.....	21

4.2 Прогнозування значення результативного показника загальної простої лінійної регресії.....	24
Тема 5. Загальна лінійна багатofакторна регресія.....	26
5.1 Класична лінійна багатofакторна модель. Етапи побудови багатofакторної регресійної моделі.....	26
5.2 Розрахунок невідомих параметрів багатofакторної регресії за методом найменших квадратів (МНК).....	28
5.3 Аналіз багатofакторної економетричної моделі.....	29
Тема 6. Мультиколінеарність.....	34
6.1 Поняття про мультиколінеарність та її вплив на оцінку параметрів моделі.....	34
6.2 Тестування наявності мультиколінеарності.....	35
Список літератури.....	39

ВСТУП

Основним завданням дисципліни «Економетрика» є вивчення методології економіко-математичного моделювання з використанням економетричних методів та ознайомлення з прикладами практичного використання прикладних економетричних моделей у фінансовій діяльності для проведення фінансового аналізу та аналізу господарчої діяльності підприємства, використання моделей для прогнозування та аналізу економічних явищ.

Знання цих питань необхідні майбутньому фахівцю, щоб об'єктивно оцінювати економічні процеси, які відбуваються в суспільстві, розуміти сутність і тенденції розвитку ринку, уміти розробляти й вирішувати актуальні питання теорії і практики розбудови фінансової політики галузі.

Курс починається із загальної характеристики науки «Економетрика», історії її виникнення, сутності та економічної складової у цій дисципліні, методів економетричного аналізу та моделювання.

Конспект лекцій складається з шести тем, у які включено глосарій основних понять тем та контрольні запитання.

Тема 1. Предмет курсу «Економетрика». Роль та етапи економетричного дослідження. Метод експертних оцінок

1.1 Економетрика як наука та історія її виникнення.

1.2 Економетричні моделі та етапи проведення економетричного аналізу.

1.3 Метод експертних оцінок.

Ключові слова: економетрика, кейнсіанська теорія споживання, економетрична модель, економетричний аналіз, метод експертних оцінок, коефіцієнт конкордації, критерій Персона.

1.1 Економетрика як наука та історія її виникнення

Економетрика порівняно молода галузь науки. Її виникнення можна датувати груднем 1930 року, коли в Клівленді (США) було засноване Міжнародне економетричне товариство. Воно визначило себе так: «Міжнародне товариство для розвитку економічної теорії і її зв'язку зі статистикою та математикою». В установчих документах товариства міститься і перше визначення сутності економетрики – «математична статистика для економістів».

Пізніше відомий економетрист Цві Грілліхес визначив сутність економетрики більш емоційно: «Економетрики є одночасно нашим телескопом і нашим мікроскопом для вивчення оточуючого нас економічного світу». Термін «економетрика» вперше запропонував львівський учений Петро Чомпа у книзі «Нариси економетрики і природної теорії бухгалтерії, яка ґрунтується на політичній економії», що вийшла друком у Львові в 1910 році. Цим терміном користувалися у своїх працях з економетрики вчені ще у 10 – 20-х роках минулого століття в США, Європі, СРСР. Уже тоді вони намагалися скласти так звані «барометри розвитку» економіки, поводження товарного і грошового ринків, циклічного розвитку виробництва товарів і послуг. Потужного поштовху надала економетричним дослідженням велика депресія в економіці США на початку 30-х років. Саме тоді була розроблена класична економетрична модель – виробнича функція Кобба-Дугласа.

Отже, економетрика як галузь економічної науки пройшла досить тривалий шлях формування найголовніших ознак самостійної дисципліни – предмета і методології. Великий внесок у розвиток економетричних методів моделювання належить лауреатам Нобелівської премії з економіки Р. Фрішу і Я. Тінбергену (1969 р.), Л. Клейну (1980 р.), Т. Хаавельмо (1989 р.), Дж. Хекману (2000 р.).

Економетрика поділяється на дві частини:

- 1) економетричні методи;
- 2) економетричні моделі економічних процесів і явищ.

Таким чином, **економетрика** – це наука, що вивчає кількісні закономірності та взаємозв'язки економічних об'єктів і процесів за допомогою математико-статистичних методів та моделей.

Економетричні методи можна умовно розбити на чотири групи. До першої групи входять методи оцінювання параметрів класичної економетричної моделі за методом найменших квадратів, їх верифікація. До другої групи належать методи оцінювання параметрів узагальненої моделі, коли порушуються деякі передумови використання методу найменших квадратів. До третьої групи входять методи оцінювання параметрів динамічних економетричних моделей, їх верифікація. Четверта група охоплює методи оцінювання параметрів економетричних моделей, які побудовані на основі системи одночасних структурних рівнянь.

1.2 Економетричні моделі та етапи проведення економетричного аналізу

Економетричні моделі кількісно описують зв'язок між вхідними показниками економічної системи, тобто факторами **X** та результативним показником **Y**. У загальному вигляді економетричну модель можна записати так:

$$Y=f(X,u), \quad (1.1)$$

де **X** — вхідні економічні показники;
u — випадкова складова.

Показники X найчастіше бувають детермінованими. Адитивна складова u є випадковою змінною, яка включає до себе помилки розрахунків, вимірювань та ті фактори, які невраховані у моделі, але вплив яких не можливо викреслити.

Існує багато відомих вам моделей. Наприклад, модель споживання, яка має вигляд:

$$C_j = f(r_j), \quad (1.2)$$

де C – споживання;

r – дохід.

До числа типових економіко-математичних моделей, які на сьогоднішній день розробляє і вивчає економетрика, належать: виробничі функції, функції попиту різних груп споживачів та цільові функції переваги споживачів, статистичні та динамічні міжгалузеві моделі виробництва, розподілу і споживання продукції, моделі загальної економічної рівноваги.

Крім того, сучасні економетричні методи широко використовуються для порівняння ефективності різноманітних економічних гіпотез та послідовного уточнення їх.

Етапи проведення економетричного аналізу

- 1 Формування теорії чи гіпотези.
- 2 Розроблення економетричної моделі для перевірки цієї теорії чи гіпотези.
- 3 Оцінка параметрів обраної моделі.
- 4 Перевірка моделі, статистичні висновки.
- 5 Прогнозування на основі отриманої моделі.
- 6 Застосування моделі (для контролю тощо).

Але доводиться аналізувати не тільки результативні показники, на які впливають фактори, що мають кількісне вимірювання. У таких випадках з метою вилучення факторів, які мають найбільший вплив на результативний показник, що аналізується, використовується метод експертних оцінок.

1.3 Метод експертних оцінок

Суть експертної процедури полягає в тому, що якомога більшій кількості експертів пропонується оцінити вплив кожного з установлених факторів на досліджуваній економічний показник шляхом розташування факторів у порядку убутання ступеня впливу. При цьому мінімальна кількість експертів становить кількість ранжованих факторів +2. Як правило, найбільш впливовому фактору присвоюється ранг 1 і далі, по мірі зменшення впливу, – ранги 2, 3 і т.ін.

Порядок проведення експертизи:

- 1) постановка задачі або висунення гіпотези;
- 2) відбір експертів із числа осіб, компетентних у досліджуваній галузі;
- 3) ранжирування факторів експертами;
- 4) обробка інформації, отриманої від експертів;
- 5) перевірка узгодженості і вірогідності експертних оцінок;
- 6) використання результатів дослідження.

При ранжируванні деякі фактори можуть одержати однаковий ранг, тому додатково для таких факторів експерту пропонують за результатами повторної експертизи уточнити місце, які одержать ці фактори при загальному упорядкуванні всіх факторів у порядку убутання рангів, тобто у порядку зменшення ступеня їх впливу на залежну змінну (При виконанні завдання місця факторів уточнюються студентом самостійно). У залежності від рангу і місця, які вони займають, фактори будуть мати ту або іншу вагу. У розрахунку приймається, що вага місця дорівнює його номеру, але для факторів, які мають однаковий ранг, вага рівняється середній арифметичній вазі відповідних місць.

Для перевірки узгодженості думок експертів визначається коефіцієнт конкордації:

$$W = \frac{S}{\frac{1}{12} m^2 (n^3 - n) - \frac{m}{12} \sum_{j=1}^m T_j}, \quad (1.3)$$

де S – сума квадратів відхилень;

m – кількість експертів;

n – кількість факторів;

T_j – показник, що враховує збіг рангів і розраховується за формулою

$$T_j = \sum_{k=1}^n (t_k^3 - t_k), \quad (1.4)$$

де t_k – число повторень рангу k при ранжируванні чинників j -м експертом.

Якщо збігу рангів не було, то $t_k=1$ та $T_j=0$.

Коефіцієнт конкордації змінюється в межах $0 \leq W \leq 1$.

Узгодженість вважається задовільною, якщо $W \geq 0,5$, якщо ж $W \geq 0,7$, то узгодженість вважається хорошою. При повному збігу думок експертів $W = 1$.

Вірогідність коефіцієнта конкордації перевіряється за **критерієм Пірсона**, розрахункове значення якого визначається за формулою

$$\chi_{\text{розр}}^2 = \frac{S}{\frac{1}{12} \cdot m \cdot n \cdot (n+1) - \frac{1}{(n-1)} \cdot \frac{1}{12} \cdot \sum_{j=1}^m T_j}. \quad (1.5)$$

Розрахункове значення $\chi_{\text{розр}}^2$ порівнюється з табличним $\chi_{\text{табл}}^2$, яке обирається із довідника критичних значень математичних показників.

При заданих рівні значущості і числі степенів вільності $\gamma = n - 1$, де n - кількість факторів, обираємо табличне значення критерію Пірсона.

Якщо розрахункове значення $\chi_{\text{розр}}^2$ перевищує табличне $\chi_{\text{табл}}^2$, то гіпотеза про наявність згоди думок експертів при ранжируванні факторів приймається, тобто можна стверджувати, що наявна не випадкова узгодженість думок експертів, тому за думкою обраних експертів можна здійснювати вилучення факторів, що найбільш впливають на результативний показник.

Для вилучення найбільш впливових факторів будується діаграма убування впливу факторів.

Глосарій

Економетрична модель — це функція чи система функцій, що описує кореляційно-регресійний зв'язок між економічними показниками, причому залежно від причинних зв'язків між ними один чи кілька із цих показників розглядаються як залежні змінні, а інші — як незалежні.

Екзогенні змінні – це незалежні змінні (фактори) від внутрішньої структури економічного явища та їхні величини задаються поза моделлю. Це змінні, які можуть бути змінені керівними органами (державним регулюванням чи керівництвом фірми). Ці керовані змінні, наприклад державні витрати та податки, є політичними інструментами. Якщо відомо структуру економічного процесу, то державні органи, змінюючи значення таких змінних, могли б робити заданими ендогенні змінні, тобто впливати на подальший розвиток процесу.

Ендогенні змінні – залежні змінні (результативні показники), які визначаються внутрішньою структурою досліджуваного економічного явища, тобто їхні величини обчислюють на основі економетричної моделі.

Рівняння регресії – рівняння, яке характеризує кореляційну залежність ендогенної змінної від екзогенних змінних.

Коефіцієнт конкордації – це показник, який використовується для перевірки узгодженості думок експертів.

Критерій Пірсона – це показник, за яким перевіряється вірогідність коефіцієнта конкордації.

Контрольні запитання

1 Розкрийте етапи проведення економетричного аналізу на прикладі кейнсіанської теорії споживання (функція залежності споживання від доходу).

2 Метод експертних оцінок як метод вилучення факторів, які мають найбільший вплив на показник, що досліджується.

3 Як перевірити узгодженість думок експертів при аналізі результатів досліджень за методом експертних оцінок?

4 Що визначається за значенням критерію Пірсона та у яких випадках необхідне його застосування?

5 Визначення ролі ПЕОМ у сучасній економетриці.

Тема 2. Специфікація економетричної моделі

2.1 Вибір виду функції.

2.2 Розгляд основних видів математичних функцій, що найчастіше зустрічаються в економетричному аналізі.

Ключові слова: специфікація моделі, види математичних функцій, проста регресія, багатofакторна регресія, параметри регресії, метод найменших квадратів.

2.1 Вибір виду функції

Використовуючи всі ті форми функцій, які можуть бути застосовані для вивчення взаємозв'язків, необхідно сформулювати теоретичні уявлення і прийняті гіпотези у вигляді математичних рівнянь. Ці рівняння встановлюють зв'язки між основними визначальними змінними за припущенням, що всі інші змінні є випадковими.

Економетрична модель базується на єдності двох аспектів — теоретичного, якісного аналізу взаємозв'язків та емпіричної інформації. Теоретична інформація відображається в специфікації моделі.

Специфікація моделі — це аналітична форма економетричної моделі. Вона складається з певного виду функції чи функцій, що використовуються для побудови моделей, має ймовірнісні характеристики, які притаманні стохастичним залишкам моделі, тобто обґрунтування лінійної або певного типу криволінійної форми регресії. Перевагу слід віддавати теоретичному аналізу форми залежностей, використовуючи як допоміжні графічні й аналітичні засоби обґрунтування.

З досвіду економетричних досліджень, а також на підставі якісного теоретичного аналізу взаємозв'язків між економічними показниками, можна навести клас функцій, які можуть описувати ці взаємозв'язки:

1) лінійна функція:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_m x_m; \quad (2.1)$$

2) степенева функція:

$$y = a_0 x_1^{a_1} x_2^{a_2} \dots x_m^{a_m} \rightarrow \ln y = \ln a_0 + a_1 \ln x_1 + a_2 \ln x_2 + \dots + a_m \ln x_m; \quad (2.2)$$

3) гіпербола:

$$y = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + \dots + \frac{a_m}{x_m} \rightarrow y = a_0 + a_1 z_1 + a_2 z_2 + \dots + a_m z_m, \quad (2.3)$$

де

$$z_j = \frac{1}{x_j}; \quad \dots \dots \dots (2.4)$$

4) квадратична функція:

$$y = a_0 + a_1 x_1^2 + a_2 x_2^2 + \dots + a_m x_m^2 \rightarrow y = a_0 + a_1 t_1 + a_2 t_2 + \dots + a_m t_m, \quad (2.5)$$

де

$$t_j = x_j^2. \quad (2.6)$$

У цих функціях:

- y** – залежна (пояснювана) змінна;
- x_j, j = 1, m** – незалежні змінні (фактори);
- a_j** – параметри функцій при j = 0, m.

Серед наведених щойно видів функцій три останні є нелінійними. Але за допомогою перетворення залежної і незалежних змінних ці функції можна звести до лінійного виду.

2.2 Розгляд основних видів математичних функцій, що найчастіше зустрічаються в економетричному аналізі

Лінійні функції найпоширеніші в економетричному моделюванні, тому обґрунтування економетричних методів розглянемо на базі лінійних моделей.

Маючи на увазі, що вибір аналітичної форми економетричної моделі не може розглядатись без конкретного переліку незалежних змінних, специфікація моделі передбачає відбір факторів для економетричного дослідження.

При цьому в процесі такого дослідження можна кілька разів повертатись до етапу специфікації моделі, уточнюючи перелік незалежних змінних та вид функції, що застосовується. Адже коли вид функції та її складові не відповідають реальним процесам, то йдеться про помилки специфікації.

Помилки специфікації моделі можуть бути трьох видів:

-ігнорування при побудові економетричної моделі істотної пояснюючої змінної;

-введення в модель незалежної змінної, яка не є істотною для вимірюваного зв'язку;

-використання невідповідних математичних форм залежності.

В економетричну модель включається обмежена кількість найбільш важливих факторів і вона набуває виду

$$\hat{y} = f(x_1, x_2, \dots, x_m), \quad (2.7)$$

де \hat{y} – математичне сподівання, або найбільш імовірне середнє з можливих значень y .

Таке спрощення дійсності зумовлює виникнення помилки апроксимації, або прогнозу, бо в такому разі

$$y = f(x_1, x_2, \dots, x_m) + e, \quad (2.8)$$

де e – випадкова складова y , що пояснюється впливом інших (не врахованих і невідомих) факторів.

Це означає, що економетричні моделі є стохастичними, ймовірнісними.

Стохастичні залежності в економетрії називаються також кореляційними.

Термін «**кореляція**» (лат. correlatio) ввів до наукової термінології англійський учений Ф. Гальтон у 1877 році. Цей термін буквально означає співвідношення, пропорцію, зв'язок, залежність.

Для визначення залежностей в економетрії використовуються такі означення:

y – залежна (ендогенна) змінна;

x_i – незалежні (екзогенні), пояснюючі змінні, або фактори.

Зазначене рівняння називається рівнянням регресії. Така назва введена англійським антропологом К. Пірсоном на початку ХХ сторіччя. Отже, предметом економетрії, по-перше, є методологія побудови – **специфікація** – рівнянь регресії в економіці.

Рівняння регресії можуть мати лінійну

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_m x_m + e, \quad (2.9)$$

і нелінійну форму, наприклад, степеневу

$$y = a_0 x_1^{a_1} x_2^{a_2} \dots x_m^{a_m} + \epsilon, \quad (2.10)$$

де a_i ($i = 0, m$) – параметри рівняння регресії, або коефіцієнти регресії. Вони визначають вплив змінення факторів x_i на рівень y .

Якщо модель лінійна, то a_i показує, на скільки одиниць змінюється y при збільшенні x_i на одиницю. У степеневій моделі коефіцієнт регресії a_i є відносною мірою змінення, він визначає, на скільки відсотків зміниться y , якщо x_i збільшиться на один відсоток.

Отже, по-друге, предметом економетрії є методи оцінювання параметрів рівнянь регресії.

Оскільки кількість незалежних змінних насправді невизначено велика й істинна аналітична форма залежностей y від x_i невідома, коефіцієнти регресії a_i і величину e визначити

абсолютно точно об'єктивно не можливо. Оцінювання цих параметрів моделі зводиться до визначення їх емпіричних, ймовірнісних оцінок, відповідно a_i і e . Тоді, наприклад, рівняння лінійної регресії набуває такого виду:

$$y = a_0 + a_1x_1 + \dots + a_mx_m + e. \quad (2.11)$$

Отже, підсумовуючи сказане, предметом економетрії є специфікація (визначення змінних та аналітичної форми) і оцінювання параметрів рівнянь регресії.

Глосарій

Специфікація моделі — це аналітична форма економетричної моделі. Вона складається з певного виду функції чи функцій, що використовуються для побудови моделей, має ймовірнісні характеристики, які притаманні стохастичним залишкам моделі, тобто обґрунтування лінійної або певного типу криволінійної форми регресії. Перевагу слід віддавати теоретичному аналізу форми залежностей, використовуючи як допоміжні графічні й аналітичні засоби обґрунтування

Параметри рівняння регресії або коефіцієнти регресії. Вони визначають вплив змінення фактора x_i на рівень y .

Проста регресія — це залежність між двома змінними, наприклад витратами на відпустку та складом родини; витратами на рекламу та обсягом продукції, що випускається; витратами на споживання та валовим національним продуктом.

Багатофакторна регресія — це залежність між двома та більше змінними, наприклад обсяг продажу та частина ринку, яку утримує фірма, якість продукції, середня заробітна плата населення у регіонах продажу та ін.

Випадкова складова — це складова економетричної моделі, яка пояснюється впливом (тобо включає) інших неврахованих факторів, помилкою у розрахунках та вимірюванні.

Контрольні питання

- 1 Параметри регресійної моделі як показники змін результативного показника при зміні факторів.
- 2 Дайте визначення специфікації моделі.
- 3 Визначте етапи специфікації моделі.
- 4 Поясніть методи вибору виду економетричної моделі.
- 5 Поясніть наявність випадкової величини у регресійних моделях.

Тема 3. Проста вибіркова лінійна регресія

3.1 Загальне поняття про вибірку лінійну регресію. Оцінка параметрів лінійної регресії.

3.2 Визначення кореляційного зв'язку між факторами та результативним показником.

3.3 Перевірка простої регресійної моделі на адекватність.

Ключові слова: проста вибіркова лінійна регресія, кореляція, коефіцієнт кореляції, адекватність моделі, коефіцієнт детермінації, критерій Фішера.

3.1 Загальне поняття про вибірку лінійну регресію. Оцінка параметрів лінійної регресії

Зазвичай перед початком економетричного аналізу, але після формування теорії щодо залежності, яка нас цікавить, висувається гіпотеза про існування лінійного зв'язку між фактором та результативним показником, тому що саме лінійна функція є базовою та шляхом математичних перетворень майже кожену функцію можливо привести до лінійного виду.

Прості парні лінійні регресійні моделі встановлюють лінійну залежність між двома змінними. При цьому одна із змінних вважається залежною y , тобто ендогенною змінною та розглядається як функція незалежної змінної x , тобто екзогенної.

У загальному вигляді проста вибіркова регресійна модель має вид

$$y = a_0 + a_1x + u, \quad (3.1)$$

де y – результативний показник (залежна змінна);
 x – фактор (незалежна змінна);
 a_0, a_1 - невідомі параметри регресійної моделі;
 u - випадкова величина (помилки або відхилення).

Регресійна модель називається лінійною, якщо вона лінійна за своїми параметрами. Її можна трактувати як пряму на площині, де a_0 – перетин з віссю ординат, тобто параметр a_0 показує, чому дорівнює результативний показник при значенні фактора 0, а a_1 – це параметр, за значенням якого можливо оцінити, на скільки зміниться результативний показник при збільшенні фактора на 1. При цьому за знаком параметра можливо визначити тип функції – пряма чи обернена. Так, при додатному значенні a_1 можливо дійти висновку, що при збільшенні фактора результативний показник також збільшується, а при від’ємному значенні цього параметра за умов збільшення незалежної змінної ендогенна змінна зменшується.

Таким чином, з метою подальшого аналізу обраної залежності необхідно оцінити (знайти) невідомі параметри a_0, a_1 цієї моделі. Для цього розглянемо пряму $\tilde{y} = a_0 + a_1x$.

Відхилення, або помилки, іноді називають залишками. Логічно, що пряму треба проводити таким чином, щоб сума квадратів помилок була мінімальною. У цьому й полягає критерій найменших квадратів.

Невідомі параметри a_0, a_1 можна знайти за формулами:

$$a_1 = \frac{\frac{1}{n} \left(\sum_{i=1}^n x_i y_i \right) - \bar{x} \cdot \bar{y}}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2}, \quad \text{де } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i. \quad (3.2)$$

При цьому параметр a_0 (перетин) можна знайти за формулою

$$a_0 = \bar{y} - a_1 \bar{x}. \quad (3.3)$$

3.2 Визначення кореляційного зв'язку між факторами та результативним показником

Критерієм, що характеризує щільність зв'язку між залежною змінною y і незалежною x , тобто на скільки значним є вплив змінної x на y , є **коефіцієнт кореляції**:

$$r_{yx} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 * \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (3.4)$$

Коефіцієнт кореляції є відносною мірою зв'язку між двома змінними. Значення коефіцієнта кореляції завжди перебуває в межах -1 та $+1$ ($-1 \leq r_{yx} \leq +1$). Додатне значення коефіцієнта кореляції свідчить про прямий, а від'ємне – про зворотний зв'язок між змінними. Коли коефіцієнт кореляції прямує за абсолютною величиною до 1 ($r_{yx} \rightarrow \pm 1$), то це свідчить про наявність міцного зв'язку, тобто щільність зв'язку велика. У протилежному випадку, коли коефіцієнт кореляції прямує до 0 ($r_{yx} \rightarrow 0$) - зв'язок слабкий або зовсім відсутній.

Знаки параметра a_1 та коефіцієнта кореляції повинні збігатися.

3.3 Перевірка простої регресійної моделі на адекватність

Наступним етапом проведення економетричного аналізу є підтвердження або спростування первинної гіпотези щодо лінійності зв'язку між обраними змінними. Для цього використовується коефіцієнт детермінації, який позначається R^2 . Коефіцієнт детермінації використовується як критерій адекватності моделі.

Коефіцієнт детермінації завжди додатний і перебуває в межах від нуля до одиниці ($0 \leq R^2 \leq 1$). Коефіцієнт детермінації дорівнює квадрату коефіцієнта кореляції:

$$R^2 = r_{yx}^2. \quad (3.5)$$

За допомогою коефіцієнта детермінації можна перевірити адекватність простої регресійної моделі. Якщо його значення близьке до одиниці, то можна вважати, що модель адекватна. Якщо його значення близьке до нуля, то модель неадекватна, тобто немає лінійного зв'язку між залежною та незалежною змінними.

$0 \leq R^2 < 0,45$ – модель неадекватна, тобто не існує лінійного зв'язку між змінними.

$0,55 < R^2 \leq 1$ – модель адекватна, тобто існує лінійний зв'язок між змінними.

$0,45 \leq R^2 \leq 0,55$ – невизначений коефіцієнт детермінації, тобто за його значенням неможливо визначити адекватність моделі.

У таких випадках необхідно використовувати інші показники. Отже, потрібен інший критерій, який би однозначно давав би відповідь на запитання про адекватність побудованої моделі. Найбільш поширеним з таких критеріїв є **критерій Фішера**.

Перевірка моделі за F-критерієм Фішера складається з певних етапів.

На першому етапі розраховуємо величину так званого F-відношення (розрахункове значення):

$$F_{(k-1, n-k)} = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)}, \quad (3.6)$$

де **n** – кількість спостережень;

k – кількість параметрів моделі (для простої регресії $k=2$).

На другому етапі задаємо рівень значущості **B**.

На третьому етапі за статистичними таблицями F-розподілу Фішера зі **(k-1, n-k)** степенями вільності та рівнем значущості **B** знаходимо критичне значення $F_{кр}$ (Критичне значення критерію Фішера визначається за довідником критичних значень математичних показників).

Якщо розраховане значення $F > F_{кр}$, то можливо зробити висновок, що побудована регресійна модель адекватна, тобто дійсно існує лінійний зв'язок між x та y . Якщо $F < F_{кр}$, то модель неадекватна, тобто зв'язок між досліджуваними змінними нелінійного виду.

Глосарій

Проста лінійна регресія — це модель, яка вважається лінійною, якщо вона лінійна за своїми параметрами. Її можна трактувати як пряму на площині.

Кореляція (лат. correlatio) – це співвідношення, зв'язок або залежність між змінними регресії.

Коефіцієнт кореляції — це критерій, що характеризує щільність зв'язку між залежною змінною y і незалежною x , тобто на скільки значним є вплив змінної x на y та є відносною мірою зв'язку між двома змінними.

Коефіцієнт детермінації — це критерій, який використовується для визначення адекватності моделі, тобто підтвердження або спростування первинної гіпотези щодо лінійності зв'язку між обраними змінними.

Критерій Фішера — це показник, за яким визначається адекватність моделей у випадках, коли коефіцієнт детермінації потрапляє у невизначений інтервал.

Контрольні запитання

1 Поясніть зв'язок між результативним показником y та параметром a_1 , тобто коефіцієнтом при змінній.

2 Поясніть зв'язок між результативним показником y та параметром a_0 , тобто вільним членом економетричної моделі.

3 Як оцінити щільність зв'язку між змінними та результативним показником y у регресійній моделі?

4 Поясніть значення адекватності моделі та як її перевірити?

5 У яких випадках необхідно використовувати критерій Фішера?

Тема 4. Загальна проста лінійна регресія

4.1 Загальна проста лінійна модель. Перевірка на значущість параметрів загальної простої лінійної регресії. Інтервали довіри.

4.2 Прогнозування значення результативного показника загальної простої лінійної регресії.

Ключові слова: загальна регресійна модель, t-тест Стьюдента, статистична значущість параметрів загальної регресійної моделі, прогнозування ендогенної змінної, точковий метод прогнозування значення результативного показника, інтервальний метод прогнозування значення результативного показника.

4.1 Загальна проста лінійна модель. Перевірка на значущість параметрів загальної простої лінійної регресії. Інтервали довіри

Вибіркові лінійні регресійні моделі правильні тільки для однієї вибірки. Якщо розглянути всю генеральну сукупність, з якої обирається вибірка, то правильною для неї буде модель

$$y = \alpha_0 + \alpha_1 x + \varepsilon, \quad (4.1)$$

де α_0, α_1 - дійсні параметри всієї генеральної сукупності;

ε - неспостережувана випадкова величина.

Цю модель називають **узагальненою регресійною моделлю**.

Якщо параметри вибіркової лінійної моделі розраховані за методом найменших квадратів, то при виконанні класичних припущень загальної лінійної моделі математичні сподівання параметрів a_0, a_1 дорівнюють значенням параметрів узагальненої моделі α_0, α_1 (яка є дійсною для всієї генеральної сукупності).

Для перевірки статистичної значущості параметрів узагальненої простої лінійної регресії a_0 і a_1 використовується **t-тест Стьюдента**, розрахункове значення t-критерію можна отримати за формулою

$$t^* = \frac{a_j}{\sqrt{\sigma_{a_j}^2}}, \quad (4.2)$$

$$\text{де } \sigma_{a_0}^2 = \sigma_u^2 \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}; \quad \sigma_{a_1}^2 = \sigma_u^2 \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}; \quad \sigma_u^2 = \frac{\sum_{i=1}^n u_i^2}{n - k}; \quad (4.3)$$

a_j – відповідний параметр вибіркової простої регресії;

σ_u^2 – оцінка дисперсії залишків;

$\sigma_{a_0}^2, \sigma_{a_1}^2$ - дисперсія оцінок параметрів моделі;

k – кількість оцінених параметрів (для простої регресії $k = 2$, бо оцінюються параметри a_0 і a_1).

Значення $t^*_{\text{розр}}$ порівнюється з табличним значенням, яке дає змогу знайти критичну область з $(n-k)$ степенями вільності. Для знаходження критичного значення потрібно задати рівень значущості β . Потім за таблицями t-розподілу Стьюдента (довідник критичних значень математичних показників) при заданому рівні значущості β та степенях вільності $(n-k)$ знайти відповідне критичне значення $t_{\beta/2}$ розрахункового значення t-статистики (рисунок 4.1).

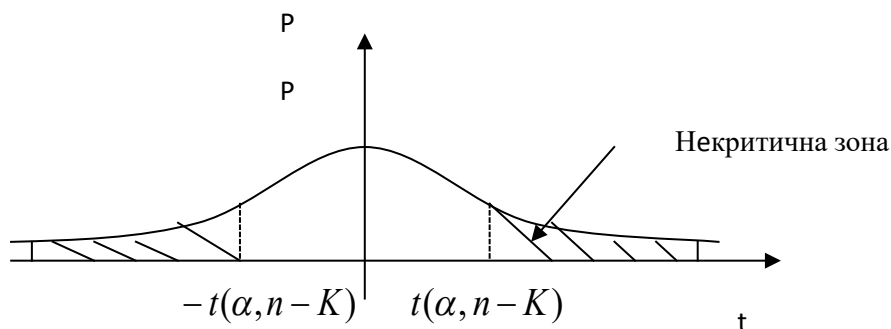


Рисунок 4.1 – Графічне зображення критичної зони для розрахункового значення t-статистики

Якщо значення t^* потрапляє в критичну зону, тобто $-t_{\beta/2} < t^* < t_{\beta/2}$, то можна зробити висновок, що з імовірністю $(1 - \beta)$ оцінка a_j є статистично незначущою (тобто приймається гіпотеза, що $\alpha_j = 0$). У протилежному випадку оцінка a_j є статистично значущою.

Для того, щоб визначити, як параметри a_0 і a_1 пов'язані з параметрами α_0 і α_1 , потрібно побудувати **інтервали довіри** для параметрів узагальненої регресійної моделі, тобто такі інтервали, у які із заданою ймовірністю потрапляють їхні значення. Для цього розраховуємо межі інтервалів:

$$\alpha_j = a_j \pm t_{\beta/2} \cdot \sigma_{a_j}, \quad (4.4)$$

де $\sigma_{a_j} = \sqrt{\sigma_{a_j}^2}$ - стандартна помилка оцінки параметра a_j .

4.2 Прогнозування значення результативного показника загальної простої лінійної регресії

Якщо модель адекватна, то можна знаходити прогнозні значення залежної змінної у виходячи з побудованої моделі. При цьому можна отримувати два типи прогнозів: точковий та інтервальний. **Точковий прогноз** дає теоретичне значення залежної змінної для відповідного значення x_{n+1} з побудованої вибіркової моделі без урахування всіх інших можливих факторів, які не враховані у моделі, помилок та відхилень:

$$\tilde{y}_{n+1} = a_0 + a_1 \cdot x_{n+1}. \quad (4.5)$$

При цьому, виходячи з узагальненої моделі, дійсне значення у для відповідного значення x_{n+1} буде дорівнювати

$$y_{n+1} = \alpha_0 + \alpha_1 \cdot x_{n+1} + \varepsilon_{n+1}, \quad (4.6)$$

де ε_{n+1} - значення випадкової величини, неспостережуваної для значення x_{n+1} . Дійсне значення y_{n+1} знайти неможливо, можна лише оцінити його за допомогою прогнозу.

Отже, застосування **інтервального методу прогнозу результативного показника** дає можливість визначити лише інтервал, у який можливо очікувати потрапляння значення y_{n+1} , але із урахуванням впливу всіх інших можливих факторів помилок та відхилень.

Такий інтервал довіри при заданому рівні значущості β для y_{n+1} можна буде знайти за формулою

$$(a_0 + a_1 \cdot x_{n+1}) \pm t_{\beta/2} \cdot \sigma_u \sqrt{\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}. \quad (4.7)$$

Глосарій

Загальна регресійна модель — це економетрична модель, яка дійсна для всієї сукупності екзогенних змінних та ендогенних змінних, що їм відповідають.

t-тест Стьюдента — це показник, за яким визначають статистичну значущість параметрів економетричної моделі.

Статистична значущість параметрів загальної регресійної моделі — параметри є статистично значущі, коли відповідне значення тесту Стьюдента не дорівнює 0, тобто вони не потрапляють у критичну зону.

Точковий метод прогнозування значення результативного показника — це метод, який дає теоретичне значення залежної змінної для відповідного значення x_{n+1} з побудованої вибіркової моделі без урахування всіх інших можливих факторів, які не враховані у моделі, помилок та відхилень.

Інтервальний метод прогнозування значення результативного показника — це економетрична модель, яка дійсна для всієї сукупності екзогенних змінних та ендогенних змінних, що їм відповідають. Дає можливість визначити лише

інтервал, у який можливо очікувати потрапляння значення Y_{n+1} , але із урахуванням впливу всіх інших можливих факторів помилок та відхилень.

Контрольні запитання

- 1 Дайте визначення загальної регресійної моделі.
- 2 Як визначити статистичну значущість параметрів загальної регресійної моделі?
- 3 Із застосуванням яких методів можливе прогнозування результативного показника загальної економетричної моделі?

Тема 5. Загальна лінійна багатофакторна регресія

5.1 Класична лінійна багатофакторна модель. Етапи побудови багатофакторної регресійної моделі.

5.2 Розрахунок невідомих параметрів багатофакторної регресії за методом найменших квадратів (МНК).

5.3 Аналіз багатофакторної економетричної моделі.

Ключові слова: лінійна багатофакторна модель, параметри багатофакторної регресії, множинна кореляція, множинна детермінація, коефіцієнт множинної кореляції, коефіцієнт множинної детермінації, оцінений коефіцієнт множинної детермінації.

5.1 Класична лінійна багатофакторна модель. Етапи побудови багатофакторної регресійної моделі

Узагальнена багатофакторна лінійна регресійна модель може бути записана у такому вигляді:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + e, \quad (5.1)$$

де y – залежна змінна;

x_1, x_2, \dots, x_n – незалежні змінні (або фактори);

$\beta_0, \beta_1, \dots, \beta_n$ – параметри моделі (константи), які потрібно оцінити;

e – неспостережувана випадкова величина.

Нагадаємо, що **узагальнена регресійна модель** – це модель, яка дійсна для всієї генеральної сукупності. Невідомі параметри узагальненої моделі є константами, а випадкова величина – неспостережувана, і ми можемо зробити тільки припущення відповідно до закону її розподілу. На відміну від узагальненої регресійної моделі, вибіркова модель будується для певної вибірки; невідомі параметри вибіркової моделі є випадковими величинами, математичне сподівання яких дорівнює параметрам узагальненої моделі (випадок класичної лінійної регресії), випадкові величини (помилки) можна оцінити, виходячи з вибіркових даних.

Вибіркова лінійна багатофакторна модель має такий вигляд:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + e, \quad (5.2)$$

де y – залежна змінна;

x_1, x_2, \dots, x_n – незалежні змінні (або фактори);

b_0, b_1, \dots, b_n – оцінки невідомих параметрів узагальненої моделі;

e – випадкова величина (помилка).

Лінійною регресійною моделлю називається модель, лінійна за своїми параметрами.

Багатофакторна лінійна регресійна модель має p незалежних змінних, або факторів, які впливають на залежну змінну y , та $(p+1)$ невідомих параметрів, які потрібно оцінити.

Етапи побудови багатофакторної регресійної моделі

Процес побудови багатофакторної регресійної моделі більш складний, ніж процес побудови простої лінійної регресії. Він складається з багатьох етапів. Серед них можна виділити такі:

- 1) вибір та аналіз усіх можливих факторів, які впливають на процес (або показник), що вивчається;
- 2) вибір та аналіз знайдених факторів;
- 3) математико-статистичний аналіз факторів;
- 4) вибір методу та побудова регресійної багатофакторної моделі;
- 5) оцінка невідомих параметрів регресійної моделі;
- 6) перевірка моделі на адекватність;
- 7) розрахунок основних характеристик та побудова інтервалів довіри;
- 8) аналіз отриманих результатів, висновки.

Метод побудови регресійної багатофакторної моделі неможливо відокремити від самої моделі, вони найтіснішим чином пов'язані між собою. Іншими словами, саме обраний метод впливає на остаточний вигляд регресійної моделі.

Оцінка невідомих параметрів $\beta_0, \beta_1, \dots, \beta_n$ здійснюється у лінійних регресійних моделях за **методом найменших квадратів**.

Після того, як параметри знайдено, проводиться перевірка моделей на адекватність за допомогою F – **критерію Фішера**, а також перевірка значущості знайдених параметрів за **t-критерієм Стюдента**. Якщо модель неадекватна, то необхідно повернутися до етапу побудови моделі і, можливо, від лінійної моделі перейти до нелінійної або ввести додаткові фактори.

Якщо модель адекватна, то можемо працювати далі: робити прогнозування, вивчати вплив окремих факторів на залежний показник, будувати інтервали довіри, аналізувати та інтерпретувати отримані результати. Для того, щоб розглянути, як можна проінтерпретувати параметри регресійної моделі, повернемося до загальної моделі багатофакторного регресійного аналізу та знайдемо математичне очікування обох частин. Виходячи з основних припущень, отримаємо

$$E(y_i \mid x_{1i}, x_{2i}, \dots, x_{ni}) = \beta_0 + \beta_1 x_{1i} + \dots + \beta_n x_{ni} . \quad (5.3)$$

Рівняння дає умовне математичне сподівання у при фіксованих значеннях x .

Параметри $\beta_0, \beta_1, \dots, \beta_n$ ще називають коефіцієнтами регресії. Кожний з них вимірює вплив відповідної змінної за умови, що всі інші залишаються, тобто є константами.

5.2 Розрахунок невідомих параметрів багатofакторної регресії за методом найменших квадратів (МНК)

Маємо ряд спостережень за залежною змінною $y = \{y_1, y_2, \dots, y_n\}$ та за незалежними змінними, або факторами:

$$\begin{aligned} X_1 &= \{X_{11}, X_{21}, \dots, X_{1n}\}; X_2 = \{X_{21}, X_{22}, \dots, X_{2n}\}; \\ & \\ X_n &= \{X_{n1}, X_{n2}, \dots, X_{pn}\}. \end{aligned} \quad (5.4)$$

На підставі цих спостережень побудуємо лінійну вибіркoву багатofакторну модель, а саме:

$$y = b_0 + b_1 x_{1i} + b_2 x_{2i} + \dots + \beta_n x_{ni}, \quad (5.5)$$

де y – залежна змінна;

x_1, x_2, \dots, x_p – незалежні змінні, або фактори;

b_0, b_1, \dots, b_p – невідомі параметри;

e – випадкова величина, або помилка.

Як і у випадку простої лінійної регресії, знайдемо невідомі параметри за методом найменших квадратів, тобто мінімізуючи суму квадратів відхилень фактичних даних від теоретичних (даних, які ми отримуємо з регресійної моделі):

$$F(b_0, b_1, \dots, b_n) = \min \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_{1i} - \dots - b_p x_{pi})^2. \quad (5.6)$$

Для того, щоб знайти мінімум виразу, необхідно прирівняти до нуля часткові похідні функції F за аргументами b_0, b_1, \dots, b_n . Отримаємо систему нормальних рівнянь. Зважаючи на досить громіздкий вигляд системи нормальних рівнянь у загальному випадку, ми не будемо її наводити. Загальний вираз для розрахунку невідомих параметрів моделі розглянемо пізніше,

коли повернемося до матричного підходу в багатофакторному аналізі. Зазначимо тільки, що перетин (параметр b_0) розраховується аналогічно до простої регресії за допомогою середніх значень:

$$b_0 = \bar{y} - b_1 \bar{x}_1 - \dots - b_p \bar{x}_p. \quad (5.7)$$

5.3 Аналіз багатофакторної економетричної моделі

Коефіцієнт множинної кореляції та детермінації

Корисною мірою ступеня відповідності даних $\{y_i, i=\overline{1, n}\}$, отриманих з регресійної моделі, фактичним даним $\{x_i, i=\overline{1, n}\}$ є коефіцієнт множинної кореляції, який визначається як коефіцієнт кореляції між y та x і має вигляд

$$R = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 (x_i - \bar{x})^2}}. \quad (5.8)$$

Квадрат коефіцієнта множинної кореляції, як і у випадку простої регресії, називають **коефіцієнтом детермінації** і позначають через R^2 . Можна показати, що вигляд коефіцієнта детермінації у випадку багатофакторної регресії ідентичний коефіцієнту детермінації простої лінійної регресії:

$$R^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (5.9)$$

Коефіцієнт детермінації R^2 та оцінений коефіцієнт детермінації \bar{R}^2

Важливою властивістю коефіцієнта детермінації R^2 є те, що він – неспадна функція від кількості факторів, які входять до моделі. Якщо кількість факторів зростає, то R^2 також зростає і ніколи не зменшується. Тобто, якщо ми додаємо новий фактор у

регресійну модель, це тільки збільшує значення коефіцієнта детермінації R^2 , що легко побачити з його визначення:

$$R^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}. \quad (5.10)$$

У цьому виразі знаменник не залежить від кількості факторів x , тоді як чисельник, навпаки, залежить. Інтуїтивно можна зрозуміти, що якщо кількість факторів x зростає, величина $\sum_{i=1}^n e_i^2$ спадає (або хоча б не зростає). Якщо ми порівнюватимемо дві регресійні моделі з однаковою залежною змінною, але різною кількістю факторів x , то, звичайно, віддамо перевагу тій, яка має більше значення R^2 .

Зразу ж постає питання, що робити, якщо ми хочемо порівняти значення коефіцієнтів детермінації в різних моделях. У таких випадках потрібно коригувати коефіцієнт кореляції з урахуванням кількості факторів x , які входять у різні моделі, тобто зменшити вплив залежності значення коефіцієнта детермінації від кількості факторів. Для цього вводиться спеціальний коефіцієнт детермінації, який має вигляд:

$$|R^2 = 1 - \frac{\frac{1}{(n-k)} \sum_{i=1}^n e_i^2}{\frac{1}{(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2}, \quad (5.11)$$

де k – кількість параметрів регресійної моделі, включаючи перетин.

На відміну від простого **коефіцієнта детермінації**, **оцінений коефіцієнт детермінації** коригується з урахуванням степенів вільності суми квадратів залишків та загальної суми квадратів.

Оцінений коефіцієнт детермінації \bar{R}^2 та коефіцієнт детермінації R^2 пов'язані між собою такою залежністю:

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n-1}{n-k}. \quad (5.12)$$

З виразу видно, якщо $k > 1$, то $\bar{R}^2 < R^2$. Крім того, якщо кількість факторів x зростає, оцінений коефіцієнт детермінації зростає повільніше, ніж просто коефіцієнт детермінації. Таким чином, зменшується вплив кількості факторів на величину коефіцієнта детермінації, тому на практиці більше використовують оцінений коефіцієнт детермінації, особливо при порівнянні різних регресійних моделей. Слід зазначити, що оцінений коефіцієнт детермінації може бути і від'ємним, на відміну від R^2 , який має додатне значення. Крім того, коли $R^2=1$, оцінений коефіцієнт кореляції також дорівнює одиниці. Коли \bar{R}^2 прямує до від'ємної величини, R^2 прямує до нуля.

Перевірка моделі на адекватність за f-критерієм Фішера

Для перевірки адекватності багатофакторної регресійної моделі, як і у випадку простої лінійної моделі, використовується **F-критерій Фішера**.

При цьому нуль-гіпотеза узагальнюється і має вигляд

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0. \quad (5.13)$$

Проти альтернативної гіпотези H_1 хоча б одне значення β_1 відмінне від нуля.

Якщо нуль-гіпотеза H_0 неправильна, то тоді правильна гіпотеза H_1 , тобто не всі параметри незначною мірою відрізняються від нуля, що дає підставу вважати, що відібрані фактори пояснюють зміну залежної величини y . Для перевірки H_0 - гіпотези розраховується F-статистика Фішера з p та $(n-p-1)$ степенями вільності:

$$F_{p,n-p-1} = \frac{\frac{\sum_{i=1}^n (x_i - \bar{y})^2}{p}}{\frac{\sum_{i=1}^n (y_i - x_i)^2}{n-p-1}}, \quad (5.14)$$

де p – кількість факторів, які увійшли в модель;
 n – загальна кількість спостережень.

За F-таблицями Фішера, як і у випадку простої регресії, знаходимо критичне значення $F_{кр}$ з p та $(n-p-1)$ степенями вільності, задавши попередньо рівень помилки $\alpha*100$ % (або рівень довіри $(1-\alpha)*100$ %).

Якщо $F > F_{кр}$, тоді нуль-гіпотеза відкидається, що свідчить про адекватність побудованої моделі. У протилежному випадку вона приймається і модель вважається неадекватною.

Глосарій

Лінійна багатофакторна модель — це лінійна регресійна модель, що має p незалежних змінних, або факторів, які впливають на залежну змінну y , та $(p+1)$ невідомих параметрів, при цьому $p=2$ або більше.

Параметри багатофакторної регресії — це незмінні коефіцієнти регресії. Вони визначають вплив змінення факторів x_i на рівень y .

Множинна кореляція — це відповідність побудованої регресії фактичним даним.

Коефіцієнт множинної детермінації — це показник, який оцінює якість побудованої моделі, індекс множинної детермінації.

Коефіцієнт множинної кореляції — це показник, який є корисною мірою ступеня відповідності побудованої регресії фактичним даним та який визначається як коефіцієнт кореляції між фактичним та теоретичним значеннями результативного показника.

Оцінений коефіцієнт множинної детермінації — це порівняний показник коефіцієнтів детермінації в різних моделях з

різними функціями залежності з урахуванням кількості факторів у моделях.

Контрольні запитання

1 Розкрийте поняття багатофакторної лінійної моделі та етапи її побудови.

2 Як визначити вплив змінення факторів x_i на рівень y ?

3 Як оцінити відповідність побудованої регресії фактичним даним?

4 Як оцінити якість побудованої моделі та ступеня відповідності побудованої регресії фактичним даним?

Тема 6. Мультиколінеарність

6.1 Поняття про мультиколінеарність та її вплив на оцінку параметрів моделі.

6.2 Тестування наявності мультиколінеарності.

Ключові слова: мультиколінеарність, суть мультиколінеарності, тестування наявності мультиколінеарності.

6.1 Поняття про мультиколінеарність та її вплив на оцінку параметрів моделі

Одна з передумов застосування методу найменших квадратів до оцінювання параметрів лінійних багатофакторних моделей — відсутність лінійних зв'язків між незалежними змінними моделі. Якщо такі зв'язки існують, то це явище називають **мультиколінеарністю**.

Суть мультиколінеарності полягає в тому, що в багатофакторній регресійній моделі дві або більше незалежних

змінних пов'язані між собою лінійною залежністю або, іншими словами, мають високий ступінь кореляції.

Наявність мультиколінеарності створює певні проблеми при розробленні моделей. Насамперед визначник матриці спостережень наближається до нуля, і оператор оцінювання за звичайним МНК стає надзвичайно чутливий до похибок вимірювань і похибок обчислень. При цьому МНК-оцінки можуть мати значне зміщення відносно дійсних оцінок узагальненої моделі, а в деяких випадках можуть стати взагалі беззмістовними.

Передусім потрібно зрозуміти природу мультиколінеарності.

Наприклад, коли вивчається залежність між ціною акції, дивідендами на акцію та отриманим прибутком на акцію, то дивіденди та отриманий прибуток на одну акцію мають високий ступінь кореляції. Іншими словами, виникає ситуація, коли два колінеарних фактори змінюються в одному напрямку. У такому разі майже неможливо оцінити вплив кожного з них на досліджуваний показник.

З'ясуємо, до яких наслідків може призвести мультиколінеарність. Це одне з найважливіших питань, яке потрібно зрозуміти при розробленні економетричних моделей.

Практичні наслідки мультиколінеарності:

- мультиколінеарність незалежних змінних (факторів) призводить до зміщення оцінок параметрів моделі, які розраховуються за методом найменших квадратів. На основі цих оцінок неможливо зробити конкретні висновки про результати взаємозв'язку між показником і факторами;

- збільшення дисперсії та коваріації оцінок параметрів, обчислених за методом найменших квадратів;

- збільшення довірчого інтервалу (оскільки збільшується середній квадрат відхилення параметрів);

- незначущість t-статистик.

Зауваження. Мультиколінеарність не є проблемою, якщо єдиною метою регресійного аналізу є прогноз (оскільки чим більше значення R^2 , тим точніший прогноз). Якщо метою аналізу є не прогноз, а дійсне значення параметрів, то мультиколінеарність перетворюється на проблему, оскільки її

наявність призводить до значних стандартних похибок оцінок параметрів.

6.2 Тестування наявності мультиколінеарності

Єдиного способу визначення мультиколінеарності, на жаль, немає. Зовнішні ознаки наявності мультиколінеарності такі:

- велике значення R^2 і незначущість t -статистики. Наявність цих двох факторів одночасно є «класичною» ознакою мультиколінеарності.

З одного боку, незначущість t -статистики Стюдента означає, що один або більше оцінених параметрів статистично незначуще відрізняються від нуля. З іншого боку, якщо значення R^2 велике, ми приймаємо з великою ймовірністю F -критерій Фішера, який відкидає нульову гіпотезу. Суперечність свідчить про наявність мультиколінеарності;

- велике значення парних коефіцієнтів кореляції.

Якщо значення хоча б одного коефіцієнта кореляції $r_{xy} > 0,8$, то мультиколінеарність є серйозною проблемою.

Зауважимо, що велике значення парних коефіцієнтів кореляції — достатня, але не необхідна умова наявності мультиколінеарності. Мультиколінеарність може бути навіть при відносно невеликих значеннях парних коефіцієнтів кореляції у більш ніж двофакторній регресійній моделі.

Для визначення мультиколінеарності здебільшого застосовують такі тести:

- **F-тест**, запропонований Глобером і Фарраром (він має й іншу назву – побудова допоміжної регресії);

- характеристичні значення та умовний індекс. Розглянемо їх більш детально.

Перший із них базується на тому, що за наявності мультиколінеарності один чи більше факторів пов'язані між собою лінійною залежністю. Одним із способів визначення щільності регресійного зв'язку є побудова регресійної залежності кожного фактора x_i з усіма іншими факторами. Тому F -тест має іншу назву – побудова допоміжної регресії. Обчислення відповідного коефіцієнта детермінації для цього допоміжного регресійного рівняння та його перевірка за допомогою

F-критерію дають змогу виявити лінійні зв'язки між незалежними змінними.

Нехай $R^2_{x_1x_2\dots x_m}$ — коефіцієнт детермінації в регресії, яка пов'язує фактор x_i з іншими факторами. Тоді F-тест виконується так, як зазначено у темі 3 цього конспекту лекцій, і за його значенням зробимо висновок щодо наявності лінійного зв'язку між факторами, що аналізуються і, як наслідок, корелюють між собою.

Але цей метод не є універсальним і має недолік: він не проводить чіткої межі між тим, що треба вважати «суттєвою» мультиколінеарністю, яку необхідно враховувати, і тим, коли нею можна знехтувати:

1) між незалежними змінними може існувати лінійна залежність, однак вона може й не бути явищем мультиколінеарності змінних, а тому не впливатиме на кількісні оцінки параметрів моделі, розрахованих за допомогою звичайного МНК;

2) якщо $F_{роз} > F_{табл}$, то X_k залежить від усіх інших незалежних змінних і треба вирішити питання про її виключення з переліку змінних;

3) якщо $t_{роз} > t_{кр}$, то X_k і X_j щільно пов'язані між собою;

4) аналізуючи F і t-критерії, робимо висновок, яку із змінних треба виключити з моделі (зрозуміло, якщо це можливо з економіко-логіко-теоретичних міркувань);

5) якщо виконавши пункти 2 – 4, ми не досягли мети, тобто не усунули мультиколінеарність, оцінку параметрів моделі слід обчислювати за допомогою іншого методу, наприклад методу головних компонентів (або однієї з його модифікацій).

Засоби усунення мультиколінеарності. Метод головних компонентів

Виявлення мультиколінеарності є лише частиною справи. Інша частина — як її усунути. Безпомилкових і абсолютно правильних порад немає, оскільки мультиколінеарність є прикладною проблемою.

Звичайно, усе залежить від ступеня мультиколінеарності, однак у будь-якому разі можна запропонувати кілька простих методів усунення мультиколінеарності:

- 1) використання додаткової або первинної інформації;
- 2) об'єднання інформації;
- 3) відкидання змінної з високою кореляцією;
- 4) перетворення даних (використання перших різниць);
- 5) збільшення кількості спостережень.

Які поради спрацюють на практиці, залежить від істотності проблеми та її характеру.

Якщо переліченими методами не вдається усунути мультиколінеарність, то для оцінювання параметрів багатовимірної моделі доцільно застосувати метод головних компонентів.

Зауважимо, що метод головних компонентів доцільно застосовувати, по-перше, для оцінювання параметрів моделей з великою кількістю факторів, по-друге, для моделей, у яких незалежні змінні (стовпці матриці спостережень X) мають однакові одиниці вимірювання.

Глосарій

Одна з передумов застосування методу найменших квадратів до оцінювання параметрів лінійних багатофакторних моделей — відсутність лінійних зв'язків між незалежними змінними моделі. Якщо такі зв'язки існують, то це явище називають **мультиколінеарністю**.

Суть мультиколінеарності полягає в тому, що в багатофакторній регресійній моделі дві або більше незалежних змінних пов'язані між собою лінійною залежністю або, іншими словами, мають високий ступінь кореляції.

Контрольні запитання

- 1 Що означає мультиколінеарність змінних?
- 2 До яких наслідків призводить мультиколінеарність?
- 3 Як впливає наявність мультиколінеарності на оцінку параметрів моделі?
- 4 За допомогою яких методів визначають наявність мультиколінеарності?
- 5 Що характеризують елементи кореляційної матриці R ?

6 Як усунути мультиколінеарність? Наведіть методи її усунення. 1

7 Яким методом можуть бути оцінені параметри моделі з мультиколінеарними змінними?

Список літератури

- 1 Краснікова І.Г, Лук'яненко Л.І. Економетрика: Підручник. – К.: Товариство “Знання”, КОО, 1998.
- 2 Краснікова І.Г, Лук'яненко Л.І. Економетрика: Практикум з використанням економетрики. – К.: Товариство “Знання”, КОО, 1998.
- 3 Наконечний С.І., Терещенко Т.О., Романюк Т.П. Економетрія [Текст]: Підручник. – 2-ге вид., допов. та перероб. – К.: КНЕУ, 2000.
- 4 Магнус Я.Р., Катышев П.К., Пересецкий А.А. Економетрика: Начальный курс. – М.: Дело, 1997.
- 5 Винн Р., Холден К. Введение в прикладной эконометрический анализ. – М., 1981.
- 6 Грубер Й. Эконометрия [Текст]: Учеб. пособие для студентов экономических специальностей. – К., 1996. – Т.1. Введение в эконометрию.
- 7 Дрейпер Смит. Прикладной регрессионный анализ. – М., 1998. – Т.1 – 2.
- 8 Фишер Ф. Проблема идентификации в эконометрии. – М., 1978.
- 9 Клас А., Герики К., Комен Ю., Шуян И. Введение в эконометрическое моделирование. – М., 1978.
- 10 Маленбо Э. Статистические методы в эконометрии. – М., 1975. – Вып.1. М., 1976. – Вып.2.
- 11 Тинтнер Г. Введение в эконометрию. – М.: 1965.
- 12 Кулинич Е.И. Эконометрия. – М., 1999.
- 13 Бородич С.А. Эконометрика: Учеб. пособие. — Минск: Новое знание, 2001.
- 14 Грубер Й. Економетрія: Вступ до множинної регресії та економетрії: У 2 т. — К.: Нічлава, 1998–1999.
- 15 Джонстон Дж. Эконометрические методы. — М.: Статистика, 1980.
- 16 Малков С.Ю. Математическое моделирование исторической динамики: подходы и модели // Моделирование социально-политической и экономической динамики / Ред. М. Г. Дмитриев. — М.: РГСУ, 2004. — С. 76 – 188.

17 Мышкис А.Д. Элементы теории математических моделей. — 3-е изд., испр. — М.: КомКнига, 2007. — 192 с [ISBN 978-5-484-00953-4](#)

18 Самарский А.А., Михайлов А.П. Математическое моделирование. Идеи. Методы. Примеры. — 2-е изд., испр. — М.:Физматлит, 2001.

