ДЕРЖАВНА СЛУЖБА СТАТИСТИКИ УКРАЇНИ
НАЦІОНАЛЬНА АКАДЕМІЯ СТАТИСТИКИ, ОБЛІКУ ТА АУДИТУ

КАФЕДРА СТАТИСТИКИ, ІТ ТА МАТЕМАТИЧНИХ МЕТОДІВ В ЕКОНОМІЦІ

ОПОРНИЙ КОНСПЕКТ ЛЕКЦІЙ

з начальної дисципліни

ЕКОНОМЕТРИКА

Київ
2024 рік

Опорний конспект з дисципліни «Економетрика» для здобувачів вищої освіти першого (бакалаврського) рівня галузі знань / Укладач О.В. Ставицький. Київ: НАСОА, 2024. 88 с.

Укладач: Ставицький О.В., доцент кафедри статистики, ІТ та математичних методів в економіці, к.е.н., доцент.

Рецензенти: д.е.н., с.н.с. Пішеніна Т. І., завідувач кафедри економіки та підприємництва Київського інституту бізнесу та технологій;
к.ф.-м.н., доц. Дереглазов Л.В., доцент кафедри статистики, ІТ та математичних методів в економіці НАСОА.

Затверджено на засіданні кафедри статистики, ІТ та математичних методів в економіці Протокол від «28» серпня 2024 року № 1

Схвалено Вченою радою обліково-статистичного факультету НАСОА
Протокол від «30» серпня 2024 року № 1

©Ставицький О.В.
©НАСОА, 2024 рік

ЗМІСТ

ТЕМА 1. Вступ до економетрики 5

ТЕМА 2. Парна лінійна регресія 16

ТЕМА 3. Множинна лінійна регресія 24

ТЕМА 4. Мультиколінеарність в економетричних моделях 30

ТЕМА 5. Гетероскедастичність в економетричних моделях..... 41

ТЕМА 6. Автокореляція в економетричних моделях..... 52

ТЕМА 7. Модель Коба-Дугласа..... 64

ТЕМА 8. Вибіркові розділи прикладної економетрики…….....73

ТЕМА 9. Економетричний аналіз виробничих функцій, інтерпретація результатів…….79

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ 87

ВСТУП

Попередні чотири десятки років характерні, зокрема, бурхливим розвитком наукової дисципліни, яка має назву економетрія або економетрика. Всесвітнє визнання робіт, виконаних з допомогою економетричних методів, відображалося великим числом лауреатів Нобелівських премій по економіці. Останні досягнення економічної науки висувають нові вимоги до вищої професійної освіти. Економетрія досліджує методи оцінювання кількісних взаємозв'язків між економічними показниками. У зв'язку з тим, що основою математичного інструментарію економетрії є теорія імовірностей і математична статистика, в першому параграфі наведені основні закони розподілу випадкових величин, з якими доводиться постійно працювати.

Метою навчальної дисципліни «Економетрика» є формування системи фундаментальних знань щодо методів оцінювання параметрів залежностей, які характеризують кількісні взаємозв'язки між економічними величинами.

У процесі вивчення цієї дисципліни студенти набувають навичок вивчення економетричних методів та моделей, набуття вмінь використання їх у практиці управління економічними процесами на різних ієрархічних рівнях національної економіки.

ТЕМА 1. Вступ до економетрики

План

1. Поняття економетрики. Історичні відомості.
2. Предмет, цілі, завдання та структура курсу.
3. Загальне поняття економетричної моделі.

1.1. Поняття економетрики. Історичні відомості

Сьогодні в будь-якій області економіки (управлінні, фінансово-кредитній сфері, маркетингу, обліку, аудиті) спеціалістам доводиться працювати в умовах недостатності і неповноти початкових даних. Аналіз такої інформації потребує спеціальних методів, що базуються на економетричних моделях, концепціях і прийомах.

Перші спроби кількісних досліджень в економіці припадають на XVII ст. Вони були пов'язані з представниками нового напряму в економічній теорії — політичної арифметики. Вільям Петті, Чарльз Давенант, Г. Кінг використовували конкретні економічні дані у своїх дослідженнях, в першу чергу, при обчисленні національного доходу. Цей напрямок спонукав до пошуку економічних законів, за аналогією з фізичними, астрономічними та іншими природно-науковими законами. При цьому існування невизначеності в економіці ще не усвідомлювалася.

Важливим етапом виникнення економетрики став розвиток статистичної теорії в працях Ф. Гальтона, К. Пірсона, Ф. Еджворта. Ці вчені випередили перші застосування парної кореляції. Так, Дж. Е. Юл визначав зв'язок між рівнем бідності і формами допомоги бідним. Г. Хукер, в свою чергу, вимірював зв'язок між рівнем одруженості і добробутом, в якому використовувалося кілька індикаторів добробуту, також він досліджував часові ряди економічних змінних.

З 1830-х років найбільш розвинені країни почали відчувати незрозумілі з точки зору тогочасної економічної науки потрясіння — занепад ділової активності, виникнення масового безробіття. Швидкий промисловий розвиток і урбанізація виявили величезний пласт невирішених соціальних проблем. Вже в кінці XIX століття неокласична теорія стала сприйматись як занадто віддалена від дійсності. Теорія могла стати переконливою у тому випадку, якщо вона б змогла пояснити зміни, що відбуваються в економіці. Для її практичного застосування були потрібні кількісні вираження базових економічних термінів.

1911 р. виходить книга американського економіста Г. Мура «Закони заробітної плати: есе по статистичній економіці». Цю роботу історик статистики Єлісєєва І. І. називає першою працею з економетрики. У своєму дослідженні Г. Мур провів аналіз ринку праці, статистично перевірив теорію продуктивності Дж. Кларка та виклав основи стратегії об'єднанняпролетаріату. Г. Мур показав, що з допомогою складних математичних конструкцій, заснованих на фактичних даних, можна розробити основу для соціальної політики. В цей же час італійський економіст Р. Беніні вперше використовував множинну регресію при оцінці функції попиту.

Значний внесок у становлення економетрики внесли дослідження циклічності економіки. Першим циклічність економіки встановив К. Жугляр. Він виявив 7-11-річні цикли інвестицій. Відразу після нього С. Кітчин виявив 3-5-річну періодичність оновлення оборотних коштів, С. Кузнець, лауреат Нобелівської премії з економіки за 1971 рік, виявив 15-20-річні цикли в будівництві, а М. Кондратьєв виявив відомі «довгі хвилі» тривалістю 45-60 років.

Важливим етапом формування економетрики стала розробка економічних барометрів. Розробка економічних барометрів засноване на ідеї, що існують показники, які змінюються раніше за інші й тому можуть служити сигналами змін останніх. Першим і найвідомішим став Гарвардський барометр, який був створений в 1903 році під керівництвом У. Персонса та У. Мітчелла. Він складався з кривих, що характеризують фондовий, товарний і грошовий ринки. Кожна з цих кривих являла собою середню арифметичну з вхідних в неї декількох показників. Ці ряди попередньо оброблялися шляхом виключення трендів, сезонних коливань та приведення коливань окремих кривих до порівняльного масштабу. Успіх використання гарвардського барометра викликав появу багатьох аналогічних барометрів в інших країнах. Проте, приблизно з 1925 р. він втратив свою чутливість. Його крах пояснюється появою потужного регулюючого чинника в економіці США. В цих умовах основним методом макроекономічного аналізу стає метод побудови міжгалузевого балансу В. В. Леонтьєва. В цей же час почали будуватися економічні моделі, що використовують методи гармонійного аналізу. Ці методи були перенесені в економіку з астрономії, метеорології та фізики .

Більш детально про становлення економетрики як науки за посиланням

https://uk.wikipedia.org/wiki/%D0%95%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D1%96%D1%8F

Термін “економетрія” вперше запровадив польський учений Павло Чомпа, опублікувавши у Львові в 1910 р. книгу “Нариси економетрії і природної теорії бухгалтерії, яка ґрунтується на політичній економії”. Однак це поняття не набуло поширення, оскільки на той час не було фундаментальних праць у цій галузі науки.

У 30-х роках XX століття склалися всі передумови для виокремлення економетрії в окрему науку. Спеціалісти, що займалися розвитком економетричної науки, повинні були використовувати в тій чи іншій мірі математику і статистику. Тому виникла необхідність в появі нового терміну, що поєднував всі дослідження в даному напрямку.

У сучасній світовій науці загальноприйнятим став термін «економетрія». Слово «економетрія» є комбінацією двох слів - «економіка» та «метрика» (від грецького «метрон»). Таким чином, сам термін підкреслює специфіку, зміст економетрики як науки: кількісне вираження тих зв’язків та залежностей, які розкриті і обґрунтовані економічною теорією.

У світовій науці економетрія займає гідне місце. Свідченням цього є присудження за найбільш видатні розробки в цій області Нобелівських премій з економіки Рагнару Фрішу і Яну Тільбергену (1969), Лоуренсу Клейну (1980), Трюгве Хаавельмо (1989), Роберту Лукасу (1995), Джеймсу Хекманом і Даніелю Мак- Фаддену (2000).

Зародження економетрії стало наслідком міждисциплінарного підходу до вивчення економіки. Ця наука виникла в результаті взаємодії і об’єднання в особливий «сплав» трьох компонент: економічної теорії, статистичних і математичних методів. Згодом до них приєднався розвиток обчислювальної техніки як умова розвитку економетрії.

Основні результати економічної теорії носять якісний характер, а економетрика вносить в них емпіричний зміст. Математична економіка виражає економічні закони у вигляді математичних співвідношень, а економетрика здійснює дослідну перевірку цих законів. Економічна статистика дає інформаційне забезпечення досліджуваного процесу у вигляді вихідних статистичних даних і економічних показників, а економетрика, використовуючи традиційні математико-статистичні і спеціально розроблені методи, проводить аналіз кількісних взаємозв’язків між цими показниками.

Найбільш часто використовують визначення економетрії, яке запропонував російський вчений С. А. Айвазян.

Економетрія – це самостійна наукова дисципліна, яка об'єднує сукупність теоретичних результатів, прийомів, методів і моделей, призначена для того, щоб

на базі економічної теорії, економічної статистики, математико-статистичного інструментарію надавати конкретне кількісне вираження загальним закономірностям, що обумовлені економічною теорією взаємозв'язків економічних явищ і процесів.

Більшість базових понять економетрії мають два визначення: «економічне» і «математичне». Подібна двоїстість має місце і в формулюванні результатів. Характер наукових робіт по економетрії варіюється від «класичних» економічних робіт, в яких майже не використовується математичний апарат, до солідних математичних праць, що використовують достатньо тонкий апарат сучасної математики.

Застосування математичних методів у економіці дає змогу виокремити та формально описати найважливіші, найсуттєвіші зв'язки економічних змінних і об'єктів, а також індуктивним шляхом отримати нові знання про об'єкт. Крім того, мовою математики можна точно та компактно відображати твердження економічної теорії, формулювати її поняття та висновки.

1.2 Предмет, цілі, завдання та структура курсу

Об'єктом економетрії є сукупність різних соціально-економічних процесів, що протікають в економічній системі.

Предмет економетрії – це економетричні методи і моделі, що дозволяють визначити і дослідити кількісні взаємозв'язки між соціально-економічними процесами і явищами.

Метою економетричного дослідження є аналіз реальних економічних систем і процесів, що в них відбуваються, за допомогою економетричних методів і моделей, їх застосування при прийнятті науково обґрунтованих управлінських рішень.

Основне *завдання економетрії* – побудова і аналіз основних економічних моделей, оцінка параметрів, перевірка значимості моделі, прогнозування на основі побудованої моделі.

Мета вивчення курсу "Економетрика" – навчитися аналізувати інформаційні потоки в соціально-економічних системах, прогнозувати їх поведінку оцінювати та будувати економетричні моделі різного рівня.

Задачі, що *вирішуються за допомогою економетрії*, можна класифікувати за трьома параметрами: за кінцевими прикладними цілями, за рівнем ієрархії і за профілем досліджуваної економічної системи.

За кінцевими прикладними цілями виділяють:

– прогноз економічних і соціально-економічних показників, які характеризують стан і поведінку економічної системи;

– імітація можливих сценаріїв соціально-економічного розвитку системи, що аналізується.

За рівнем ієрархії економічної системи, що аналізується, виділяють:

– макрорівень, тобто країни в цілому, міжкраїнний аналіз;

– мезорівень – регіони всередині країни, галузі, корпорації і т.д.;

– мікрорівень – підприємства, фірми, сім'ї.

За профілем економетричного моделювання: дослідження може бути направлене на вирішення проблем попиту і пропозиції, ціноутворення, інвестиційної, фінансової чи соціальної політики чи на рішення певного комплексу проблем.

Економетрика поділяється на дві частини: *економетричні методи* та *економетричні моделі* економічних процесів і явищ.

У цьому курсі вивчається здебільшого матеріал, що входить до першої частини, тобто *економетричні методи*. їх можна умовно розбити на чотири *групи*. *Перша група* – це методи оцінювання параметрів класичної економетричної моделі, насамперед метод найменших квадратів (МНК). *Друга група* – це методи оцінювання параметрів узагальненої моделі, коли порушуються деякі передумови використання методу найменших квадратів. До *третьої групи* входять методи оцінювання параметрів динамічних економетричних моделей. *Четверта група* охоплює методи оцінювання параметрів економетричних моделей, які побудовані на основі системи одночасних структурних рівнянь.

1.3 Загальне поняття економетричної моделі

Нагадаємо, що математичне моделювання — це метод дослідження процесів або явищ шляхом створення їхніх математичних моделей і дослідження цих моделей.

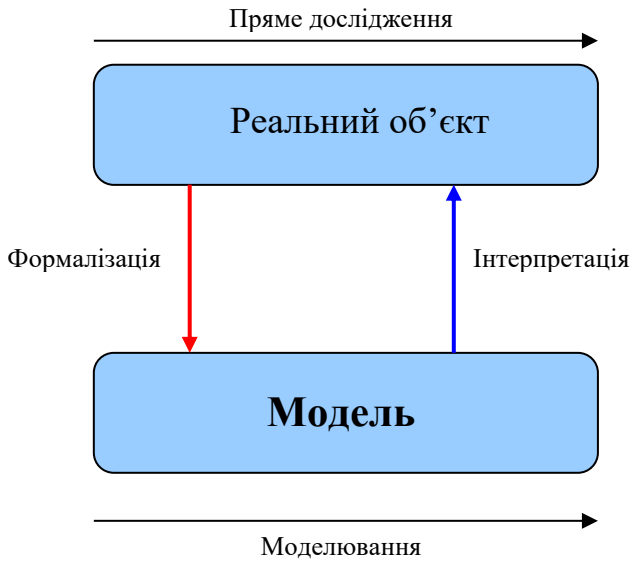


Рисунок 1.1 – Процес моделювання

Формалізація – процес подання інформації про об'єкт, процес, явище в формалізованому вигляді. Формалізація – це метод відображення певної області у вигляді формальної системи, коли форма виділяється у якості особливого предмета дослідження незалежно від змісту. Такий метод полегшує вивчення предмета.

Інтерпретація – це відношення, що переносить формули формальної системи на змінні та зв’язки реальної системи.

Як правило, модель має форму рівняння чи системи рівнянь, що характеризують виокремлені дослідником взаємозалежності між економічними показниками. Економетрична модель, що пояснює поведінку одного показника, складається з одного рівняння, а модель, що характеризує зміну кількох показників, — із такої самої кількості рівнянь. У моделі можуть бути також тотожності, що відбивають функціональні зв'язки в певній економічній системі. Оскільки така модель поєднує не лише теоретичний, якісний аналіз взаємозв'язків, а й емпіричну інформацію, то в ній, на відміну від просто економічної моделі, завжди присутні стохастичні залишки. Саме ймовірнісні характеристики залишків моделі зумовлюють якість тієї чи іншої аналітичної форми моделі.

Отже, сформулюємо таке означення економетричної моделі.

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.

У загальному випадку рівняння в економетричній моделі має вигляд:

$$Y = f(x_1, x_2, .., x_n, u) \quad (1.1)$$

де Y – результат, або залежна змінна, змінювання якої описує дане рівняння;

$x_1, x_2, ..., x_n$ – фактори, або незалежні змінні, що визначають поведінку Y;

u – випадкова змінна, що містить ту частину руху Y, що не пояснюється змінними $x_1, x_2, ..., x_n$, і має випадковий характер.

Незалежні змінні $x_1, x_2, ..., x_n$, що задані заздалегідь чи за межами моделі, називаються *екзогенними* змінними (регресорами). Це змінні, які входять в модель, але розглядаються незалежно від модельованого явища.

Залежна змінна Y, що визначається як розв'язок рівняння, називається *ендогенною* змінною (регресандом). Це змінна, яка визначається тільки явищем для якого будується модель. Значення цієї змінної будується в процесі функціонування аналізованої соціально-економічної системи, причому в істотній мірі, під впливом екзогенних змінних і, звичайно, у взаємодії один з одним.

Зауважимо, що в багатьох економетричних моделях є такі екзогенні 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.

Існують різні класифікації економетричних моделей (залежно від ознаки).

1. За способом врахування часу: *статичні* і *динамічні моделі.* У статичних моделях зв'язки розглядаються у фіксований момент часу і часові зміни в них ролі не відіграють. У динамічній моделі, навпаки, взаємозв'язки вивчаються в розвитку й час є необхідним фактором змін.

2. За рівнем агрегування змінних: моделі з мікро- чи макроекономічними показниками.

3. За способом відображення змінних: у постійних цінах, у поточних цінах, в абсолютних значеннях показників, у приростах показників.

4. За кількістю змінних: одно- чи багатофакторні моделі.

5. За кількістю рівнянь: моделі з одним рівнянням чи моделі з системою рівнянь.

6. За агрегованим часом спостережень: річні, квартальні, щомісячні дані, щотижневі тощо.

Економетричні моделі в порівнянні з аналітичними більш точні і докладні, не вимагають грубих допущень і спрощень, дозволяють врахувати велику кількість чинників. Основні їх недоліки – громіздкість, великі витрати машинного часу при їх побудові та аналізі і значні труднощі пошуку оптимальних рішень, які доводиться шукати «навпомацки», шляхом здогадів та спроб (на відміну від більш пристосованих до оптимізаційних завдань аналітичних моделей). Найбільш ефективна методика економіко-математичних досліджень – це спільне застосування аналітичних та економетричних моделей. Аналітична модель дає можливість в загальних рисах розібратися в явищі, намітити якби контури основних закономірностей. Уточнення ж цих закономірностей – прерогатива економетричних моделей.

Побудова і дослідження економетричних моделей мають ряд особливостей. Ці особливості пов'язані з тим, що економетричні моделі є стохастичними. Вони кількісно описують кореляційний зв'язок між економічними величинами.

Умови побудови якісної економетричної моделі:

1) достатня сукупність спостережень (на кожну незалежну змінну мінімум 7 спостережень) ;

2) однорідність сукупності спостережень;

3) точність вхідних даних.

До основних етапів економетричного моделювання належать:

1) ідентифікація змінних (визначення залежних і незалежних змінних - факторів);

2) специфікація економетричної моделі (вибір виду аналітичної залежності);

Розглядаючи проблему вибору конкретного виду функціонала зауважимо, що в практиці економетричних досліджень використовується досить широке коло функціональних залежностей між змінними. Наведемо деякі, найбільш часто використовувані, їх види:

Найпростішою є лінійна форма між двома змінними, але можливі й інші форми (рис. 1.2 – 1.7).

Лінійна форма залежності

$$Y = a + b \cdot x$$

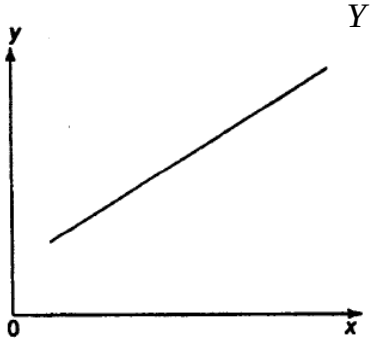


Рис. 1.2

Показникова форма залежності

$$Y = a + b^x$$

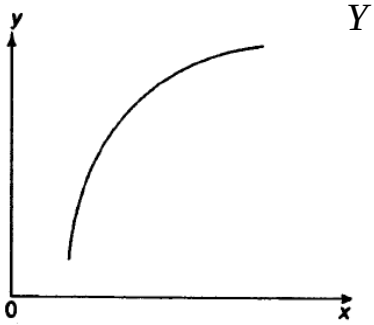


Рис. 1.3

Степенева форма залежності

$$Y = a + x^b$$

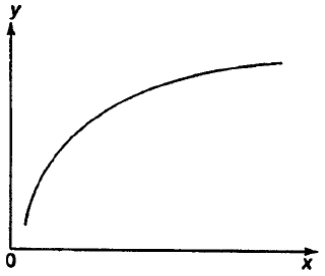


Рис. 1.4

Гіперболічна форма залежності

$$Y = a + \frac{b}{x}$$

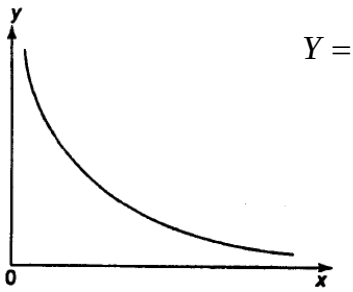


Рис. 1.5

Параболічна форма залежності

$$Y = a + b \cdot x + c \cdot x^2$$

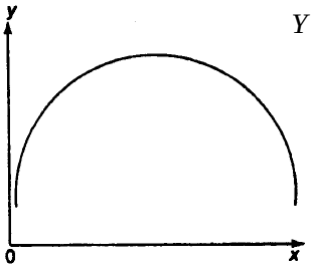


Рис. 1.6

??? форма залежності

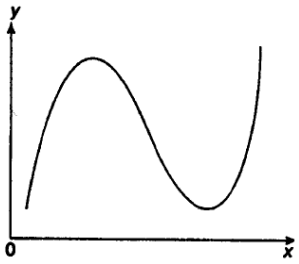


Рис. 1.7

Якщо економетрична модель вимірює зв’язок між двома змінними, то кожну пару спостережень над цими змінними можна зобразити у двовимірній системі координат. У результаті дістаємо кореляційне поле точок.

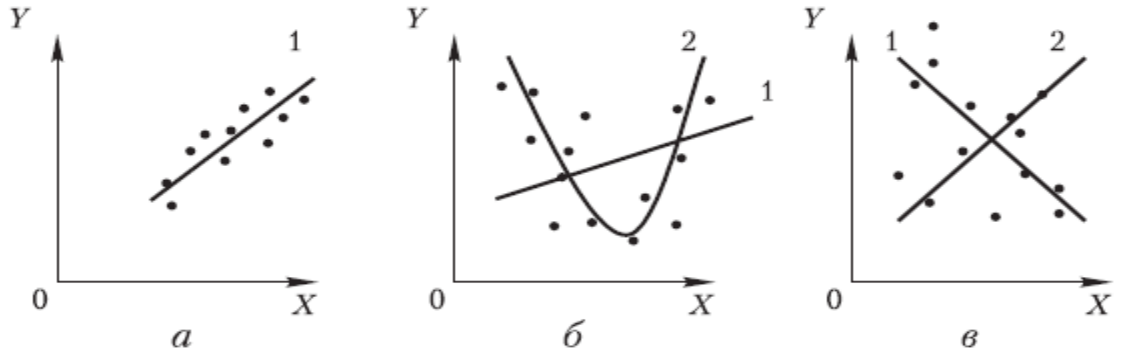


Рисунок 1.8 – Кореляційне поле

На рис.1.8 проілюстровано три ситуації. На графіку 1.8, *а* взаємозв'язок між *X* і *Y* близький до лінійного, і пряма 1 досить добре узгоджується з емпіричними точками. Тому щоб описати залежність між *X* і *Y*, доцільно вибрати лінійну функцію $Y = a + b \cdot x$.

На графіку 1.8, *б* реальний взаємозв'язок між *X* і *Y*, найімовірніше, описується квадратичною функцією $Y = a + b \cdot x + c \cdot x^2$ (лінія 2).

На графіку 1.8, *в* явний взаємозв'язок між *X* і *Y* відсутній. Тому, щоб краще вибрати форму зв'язку, необхідно, можливо, збільшити кількість спостережень (точок кореляційного поля) або скористатися іншими способами вимірювання показників.

3) вибір методу оцінювання параметрів моделі;

4) верифікація моделі – перевірка її якості;

5) застосування моделі (прогноз пояснюваних змінних на основі моделі).

Наведені на етапі 2 форми залежності кількісно описують взаємозв'язок між економічними показниками лише в середньому, а кожне індивідуальне значення відрізнятиметься від обчисленого за модельною функцією, оскільки на цей зв'язок впливають і інші фактори, серед яких є випадкові, не враховані при вимірюванні.

Щоб урахувати наявність впливу факторів, які не входять до економетричної моделі, вводиться стохастична складова *u*. Математичний аналіз цієї складової дає змогу зробити висновок про те, чи можна вважати її випадковою величиною, чи вона містить систематичну частину відхилень, яка може бути зумовлена наявністю тих чи інших помилок у моделюванні. Економетрична модель у такому разі подається у вигляді

$$Y = a + b \cdot x + u \qquad (1.5)$$

У класичній лінійній економетричній моделі змінна *u* інтерпретується як випадкова змінна, що має розподіл з математичним сподіванням, яке дорівнює нулю, і постійною дисперсією. Це дає змогу розглядати змінну *u* як стохастичне збурення (помилку, відхилення). Згідно з центральною граничною теоремою стохастична складова економетричної моделі розподілена за нормальним законом.

Причини, які спонукають появу *випадкових збурень u* можуть бути такими:

1. Слід пам'ятати, що будь-яка економетрична модель є в певній мірі спрощенням реальної ситуації, яка в дійсності являє собою складне переплетіння різних факторів, багато з яких практично неможливо врахувати в моделі.

Так, наприклад, попит на товар буде визначатися як його ціною, так і ціною на ті товари, які можуть його заміняти, ціною на супроводжуючі товари, доходами споживачів та їх вподобаннями і т.ін. Однак в цьому переліку не враховуються традиції як релігійні, так і національні, особливості кліматичних умов і багато інших факторів. При цьому ще виникає проблема визначення факторів, які за певних умов будуть домінуючими, а якими можна знехтувати. В ряді випадків існують фактори, які не можна використати в моделі тому, що для них проблематично одержати необхідні статистичні дані. Наприклад, величина заощаджень родини визначається не лише доходами її членів, а й їхнім здоров'ям, інформацію про що в цивілізованих країнах тримають в таємниці. Окрім цього, багато факторів мають випадковий характер (погода, стихійні лиха), які посилюють неоднозначність.

2. Неправильно вибрана функціональна залежність. Це може трапитися внаслідок недостатнього дослідження процесу, який підлягає моделюванню. Так, виробнича функція, яка описує залежність Y від одного фактора X, може бути виражена лінійним співвідношенням:

$$Y = \beta_0 + \beta_1 X\,, \tag{1.6}$$

хоча насправді, при більш ретельному дослідженні, стане відомо, що співвідношення між Y та X матиме нелінійний характер, наприклад:

$$Y = \beta_0 X^{\beta_1}\,. \tag{1.7}$$

Вибір форм функціональної залежності між змінними називають *специфікацією* моделі.

3. Можуть бути невірно вибрані пояснюючі змінні.

4. В багатьох моделях залежність між факторами має складну форму зв'язку між цілими комплексами подібних величин. Так, при дослідженні залежності попиту Y в якості пояснюючої вибирають змінну, яка уособлює складну комбінацію індивідуальних запитів, що мають на неї певний вплив поряд

із факторами, які враховані в моделі. Здійснюється так зване агрегування пояснюючих змінних, що може бути однією із причин появи в моделі випадкового збудника *u*.

5. Можуть бути допущені помилки при аналізі та обробці статистичних даних, які також сприятимуть появі *u*.

6. Як правило, будь-яка статистична інформація є обмеженою і, крім цього, більшість моделей описуються неперервними функціями, але при цьому використовуються вибіркові дані, які мають дискретну структуру.

7. Слід також зважити на наявність людського фактора, який в тій чи іншій мірі обов'язково є присутнім в будь-якому економічному процесі, але врахувати який в моделі поки що практично неможливо. В певних ситуаціях цей фактор може навіть якісну модель деформувати до примітивного рівня.

ТЕМА 2. Парна лінійна регресія

План

2.1. Алгоритм побудови парної лінійної регресії та дослідження її якості.
2.2. Побудова рівняння парної лінійної регресії.
2.3. Оцінка тісноти зв'язку між статистичними величинами.
2.4. Оцінка якості побудованої моделі.
2.5. Оцінка статистичної значущості рівняння парної лінійної регресії та його параметрів.
2.6. Довірчі інтервали параметрів парної лінійної регресії.

2.1 Алгоритм побудови парної лінійної регресії та дослідження її якості

Алгоритм побудови парної лінійної регресії та дослідження її якості. Оцінка параметрів парної лінійної регресії. Коефіцієнт кореляції. Коефіцієнт детермінації. Середня похибка апроксимації. Статистична значущість рівняння регресії. Критерій Фішера. Статистична значущість параметрів рівняння регресії. Критерій Стьюдента. Довірчі інтервали. Точковий та інтервальний прогноз.

Для побудови парної лінійної регресії та дослідження її якості слід виконати такі дії:

1. Побудувати емпіричне рівняння регресії.
2. Оцінити тісноту зв'язку між досліджуваними статистичними величинами.

3. Оцінити якість побудованої моделі на основі коефіцієнта детермінації.
4. Оцінити статистичну значущість рівняння в цілому (критерій Фішера).
5. Оцінити статистичну значущість параметрів регресії та коефіцієнта кореляції.
6. Проаналізувати довірчі інтервали для кожного з параметрів.

Якщо нас задовольняють результати п.3-6, то продовжуємо досліджувати рівняння парної лінійної регресії далі.

7. Дати інтерпретацію отриманому рівнянню парної лінійної регресії (проінтерпретувати коефіцієнт b_1 та середній коефіцієнт еластичності).
8. Скористатися отриманим рівнянням парної лінійної регресії для точкового та інтервального прогнозу.

Розглянемо поступово всі зазначені етапи.

2.2. Побудова рівняння парної лінійної регресії

За вибіркою обмеженого обсягу можна побудувати так зване емпіричне рівняння регресії:

$$\tilde{Y}_i = b_0 + b_1 x_i, \tag{2.1}$$

$$\begin{cases} b_0 = \bar{y} - b_1 \bar{x}; \\ b_1 = \dfrac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}. \end{cases} \tag{2.2}$$

де $\overline{xy} = \dfrac{1}{n}\sum_{i=1}^{n} x_i y_i$, $\bar{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$.

2.3 Оцінка тісноти зв’язку між статистичними величинами

Тісноту лінійного зв’язку між парою статистичних величин оцінює *лінійний коефіцієнт парної кореляції*, який розраховується за формулами:

$$r_{xy} = b_1 \frac{\sigma_x}{\sigma_y} \tag{2.3}$$

або

$$r_{xy} = \frac{\mathrm{cov}(x, y)}{\sigma_x \sigma_y}. \tag{2.4}$$

Значення $|r_{xy}|$ змінюється в діапазоні від -1 до +1. Додатне значення свідчить про прямий лінійний зв'язок, від'ємне – про обернений. Чим ближче абсолютне значення коефіцієнта до одиниці, тим тісніший зв'язок. Вважається, що зв'язок досить сильний, якщо $|r_{xy}| > 0{,}7$, і слабкий, якщо $|r_{xy}| < 0{,}3$.

Довідка: $\mathrm{cov}(x, y) = \overline{x \cdot y} - \bar{x} \cdot \bar{y}$ - коваріація ознак ***x*** та ***y***, $\sigma_x^2 = \overline{x^2} - \bar{x}^2$ - дисперсія ознаки ***x***.

Для оцінки *статистичної значущості коефіцієнта кореляції* між двома змінними використовують критерій Стьюдента.

Статистична значущість деякої величини означає, що вона отримана не випадково. Якщо б ми повторили дослідження у схожих умовах, то отримали б аналогічний результат.

Алгоритм перевірки статистичної значущості коефіцієнта кореляції містить такі кроки:

1) Оцінюють дисперсію коефіцієнта кореляції за формулою:

$$m_{r_{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n - 2}}. \tag{2.5}$$

2) Обчислюють розрахункові значення t-статистики Стьюдента для кожної з величин за формулами:

$$t_r = \frac{r_{xy}}{m_r}. \qquad (2.6)$$

3) Знаходять табличне значення *t*-критерія Стьюдента у табличному процесорі MsExcel за допомогою функції *Стьюдраспробр(α;n-m-1)*. Параметри функції: α – рівень значущості (для економічних досліджень $\alpha = 0{,}05$); n – кількість спостережень, за якими будується модель; m – кількість незалежних змінних моделі.

4) Порівнюють $t_{розр}$ і $t_{табл}$. Якщо $t_{розр} > t_{табл}$, то коефіцієнт кореляції, а отже і лінійний зв'язок між змінними ***x*** та ***y***, є статистично значущим.

2.4. Оцінка якості побудованої моделі

Для оцінки якості побудованої моделі використовують коефіцієнт детермінації та середню похибку апроксимації

Коефіцієнт детермінації для парної лінійної регресії зв'язаний з коефіцієнтом кореляції формулою:

$$R^2 = r_{xy}^2. \qquad \mathbf{(2.7)}$$

Коефіцієнт детермінації R^2 приймає значення з відрізку [0, 1]. Якщо $R^2 > 0{,}75$, вважається, що модель досить гарно (адекватно) описує реальні дані.

Коефіцієнт детермінації, помножений на 100%, показує, скільки відсотків змін результативного фактору Y пояснюється зміною незалежного фактору X.

Середня похибка апроксимації характеризує відхилення розрахованих (модельних) значень від фактичних. Розраховується за формулою:

$$\overline{A} = \frac{1}{n}\sum\left|\frac{y_i - Y_i}{y_i}\right| * 100\%, \qquad (2.8)$$

де n – кількість спостережень, за якими будується модель.

Для $\overline{A}$ максимально допустиме значення 10%. Іноді в економічних дослідженнях за умови $R^2 > 0{,}75$ припускається $\overline{A} \le 20\%$.

Якщо модель отримала задовільні оцінки щодо її якості, можна перейти до економічного тлумачення отриманого рівняння.

У загальному випадку коефіцієнт регресії b_1 (див. формулу 2.1) показує, як у середньому зміниться результативна ознака Y, якщо факторна ознака (X) збільшиться на одиницю.

Розраховують також середній коефіцієнт еластичності за формулою:

$$\overline{E} = b_1 \frac{\overline{x}}{\overline{y}}. \qquad (2.9)$$

Середній коефіцієнт еластичності показує на скільки середньому по сукупності відхилиться результат ***Y*** від своєї середньої величини при відхиленні фактора ***x*** на 1% від свого середнього значення.

2.5. Оцінка статистичної значущості рівняння парної лінійної регресії та його параметрів

Для перевірки *статистичної значущості рівняння* в цілому застосовують критерій Фішера. Послідовність дій така:

1) Обчислюють величину:

$$F_{розр} = \frac{S^2_{факт}}{S^2_{зал}}, \qquad (2.10)$$

де $S^2_{факт} = \frac{\sum_{i=1}^{n}(\tilde{Y}_i - \bar{y})^2}{\text{m}}$,

$$S^2_{зал} = \frac{\sum_{i=1}^{n}(y_i - \tilde{Y}_i)^2}{\text{n - m - 1}},$$

n – кількість спостережень, за якими будується модель;

m – кількість незалежних змінних моделі.

2) Знаходять табличне значення F-критерія Фішера. Його можна розрахувати у табличному процесорі MsExcel за допомогою функції *Fраспробр(α;k₁;k₂)*. Параметри функції: α – рівень значущості (для економічних досліджень $\alpha = 0{,}05$); $k_1 = m$; $k_2 = n - m - 1$ (n – кількість спостережень, за якими будується модель; m – кількість незалежних змінних моделі).

3) Порівнюють $F_{розр}$ з табличним значенням F-статистики Фішера. Якщо $F_{розр} > F_{табл}$, то рівняння парної лінійної регресії (2.1) є статистично значущим. У протилежному випадку рівняння (2.1) буде статистично незначущим.

Слід зазначити, що величина $F_{розр}$ звязана з коефіцієнтом детермінації формулою:

$$F_{розр} = \frac{\text{r}^2_{\text{xy}}}{1 - \text{r}^2_{\text{xy}}}(n - m - 1), \qquad (2.11)$$

де n – кількість спостережень, за якими будується модель;

m – кількість незалежних змінних моделі.

Для оцінки *статистичної значущості параметрів моделі* парної лінійної регресії використовують критерій Стьюдента і виконують такі дії:

1) Оцінюють дисперсію параметрів моделі за формулами:

$$m_{b_0} = \sqrt{S^2_{зал} \frac{\sum_{i=1}^{n} x_i}{n^2 \sigma_x^2}}, \quad (2.12)$$

$$m_{b_1} = \sqrt{\frac{S^2_{зал}}{n \cdot \sigma_x^2}}. \quad (2.13)$$

2) Обчислюють розрахункові значення t-статистики Стьюдента для кожної з величин за формулами:

$$t_{b_0} = \frac{b_0}{m_{b_0}}, \quad (2.14)$$

$$t_{b_1} = \frac{b_0}{m_{b_1}}, \quad (2.15)$$

3) Знаходять табличне значення *t*-критерія Стьюдента.

4) Порівнюють $t_{розр}$ і $t_{табл}$. Якщо розраховане значення t-статистики Стьюдента більше за табличне, то відповідна величина визнається статистично значущою. У протилежному випадку приймається гіпотеза про те, що дана величина є статистично незначущою.

2.6 Довірчі інтервали параметрів парної лінійної регресії

Для розрахунку довірчих інтервалів для кожного з параметрів парної лінійної регресії необхідно розрахувати відповідні граничні похибки за формулами:

$$\Delta_{b_0} = t_{табл} \mathrm{m}_{\mathrm{b}_0}, \tag{2.16}$$

$$\Delta_{b_1} = t_{табл} \mathrm{m}_{\mathrm{b1}}. \tag{2.17}$$

Тоді довірчі інтервали можна записати так:

$$b_0 - \Delta_{b_0} \le b_0 \le b_0 + \Delta_{b_0}; \tag{2.18}$$

$$b_1 - \Delta_{b_1} \le b_1 \le b_1 + \Delta_{b_1}. \tag{2.19}$$

Якщо у межі довірчого інтервалу потрапляє нуль, тобто нижня межа інтервалу від'ємна, а верхня – додатна, то відповідний параметр вважається нульовим, так як він не може одночасно приймати від'ємне і додатне значення.

2.7. Точковий та інтервальний прогноз

Для здійснення точкового прогнозу значення $x_{пр}$ підставляють у формулу (2.1).

Для інтервального прогнозу слід знайти оцінку дисперсії та відповідну граничну похибку за формулами:

$$m_{Y_p} = \sqrt{S^2_{зал} \cdot (1 + \frac{1}{n} + \frac{\left(x_{пр} - \bar{x}\right)^2}{n \cdot \sigma_x^2})}, \tag{2.20}$$

$$\Delta_Y = m_Y \cdot t_{табл}. \tag{2.21}$$

Тоді інтервальний прогноз запишеться так:

$$Y_{np} - \Delta_{Y_p} \le Y_{np} \le Y_p + \Delta_{Y_p}. \qquad (2.22)$$

ТЕМА 3. Метод найменших квадратів

План

3.1. Сутність методу найменших квадратів.
3.2. Алгоритм методу найменших квадратів.
3.3. Рівняння регресії в стандартизованому масштабі.
3.4. Передумови застосування методу найменших квадратів 1МНК.
3.5. Підходи до побудови багатофакторної регресійної моделі.

3.1 Сутність методу найменших квадратів

У загальному випадку парна лінійна регресія (з двома змінними)є лінійною функцією між залежною змінною Y і однією пояснюючою змінною X має вигляд:

$$Y = a_0 + a_1 \cdot X + u \qquad (1)$$

При цьому співвідношення (1) називається теоретичною лінійною регресійною моделлю; a_0 і a_1 – теоретичні параметри (коефіцієнти) регресії.

Параметр a_1 (коефіцієнт регресії) — величина іменована, має розмірність результативної ознаки і розглядається як ефект впливу X на Y. Параметр a_0 — вільний член рівняння регресії, це значення Y при $X = 0$. Якщо межі варіації X не містять нуля, то цей параметр має лише розрахункове значення.

Нехай за вибіркою (x_i, y_i), $i = 1,2,\ldots, n$, потрібно визначити оцінки $\hat{a}_0$ і $\hat{a}_1$ емпіричного рівняння регресії (1), тобто підібрати такі значення коефіцієнтів рівняння, щоб сума квадратів відхилень була мінімальною (рис. 1).

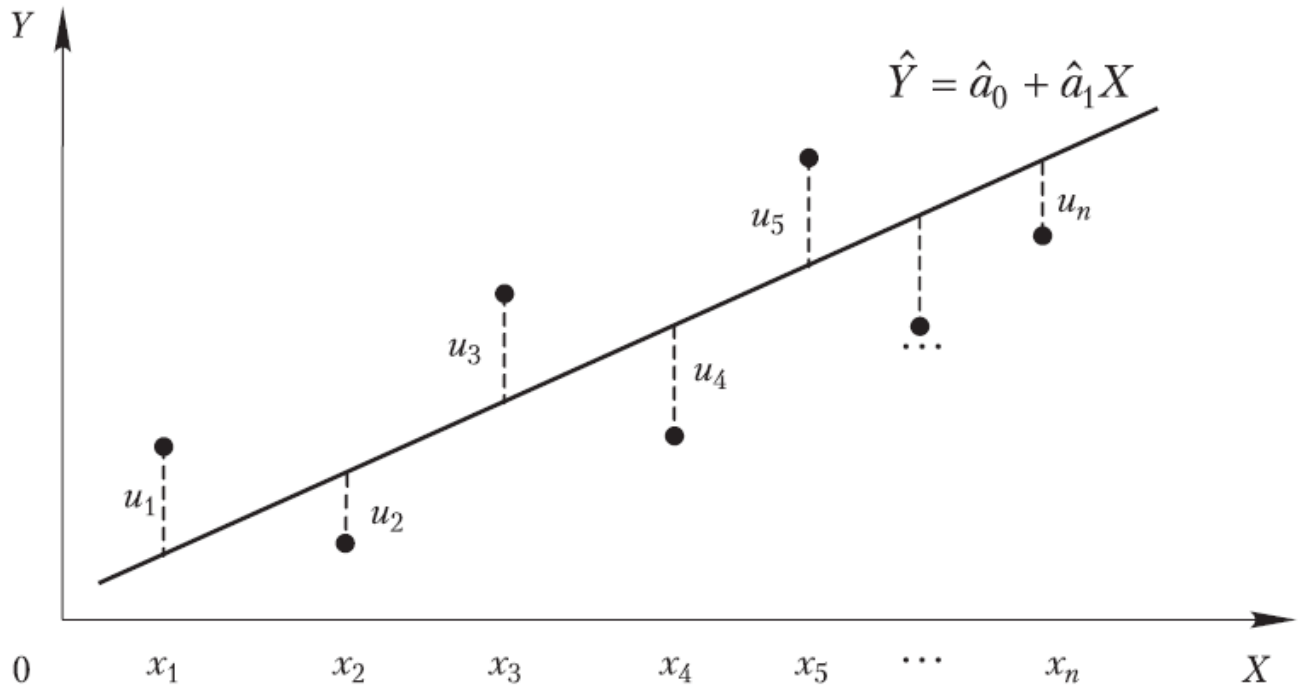


Рисунок 1

У цьому разі $\sum_{i=1}^{n} u_i^2$ є квадратичною функцією двох параметрів $\hat{a}_0$ і $\hat{a}_1$, оскільки x_i, y_i ($i = 1,2,\ldots,n$) – відомі дані спостережень:

$$\sum_{i=1}^{n} u_i^2 = Q(\hat{a}_0, \hat{a}_1) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - \hat{a}_0 - \hat{a}_1 x_i)^2 \to \min \qquad (2)$$

Лінійна множинна регресія – рівняння зв’язку з декількома незалежними змінними:

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_m x_m + u \qquad (3)$$

Рівняння лінійної множинної регресії має відповідати таким самим умовам:

$$\sum_{i=1}^{n} u_i^2 = Q(a_0, a_1) = \sum_{i=1}^{n} (y_i - Y_i)^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_{1i} - a_2 x_{2i} - \ldots - a_m x_{mi})^2 \to \min \qquad (4)$$

Неважко помітити, що квадратична функція Q неперервна, опукла та обмежена знизу ($Q \geq 0$), тобто має мінімум.

Необхідною умовою існування мінімуму неперервно диференційованої функції кількох змінних є рівність нулю її частинних похідних:

3.2 Алгоритм методу найменших квадратів

Для оцінки параметрів рівняння множинної регресії застосовують метод найменших квадратів (1 МНК). Для лінійних рівнянь будується наступна система нормальних рівнянь, рішення якої дозволяє отримати оцінки параметрів регресії:

$$\begin{cases} \sum y = na_0 + a_1\sum x_1 + a_2\sum x_2 + ... + a_m\sum x_m, \\ \sum yx_1 = a_0\sum x_1 + a_1\sum x_1^2 + a_2\sum x_1x_2 + ... + a_m\sum x_mx_1, \\ ..... \\ \sum yx_m = a_0\sum x_m + a_1\sum x_1x_m + a_2\sum x_2x_m + ... + a_m\sum x_m^2. \end{cases} \quad (5)$$

Для її рішення може бути використаний метод визначників:

$$a_0 = \frac{\Delta a_0}{\Delta}, \quad a_1 = \frac{\Delta a_1}{\Delta}, \quad ..., \quad a_m = \frac{\Delta a_m}{\Delta}, \quad (6)$$

де $\Delta = \begin{vmatrix} n & \sum x_1 & \sum x_2 & ... & \sum x_m \\ \sum x_1 & \sum x_1^2 & \sum x_2x_1 & ... & \sum x_mx_1 \\ \sum x_2 & \sum x_1x_2 & \sum x_2^2 & ... & \sum x_mx_2 \\ ... & ... & ... & ... & ... \\ \sum x_m & \sum x_1x_m & \sum x_2x_m & ... & \sum x_m^2 \end{vmatrix}$ – визначник системи;

$\Delta a_0, \Delta a_1, ..., \Delta a_m$ – частинні визначники, що отримані шляхом заміни відповідного стовпця матриці визначника системи даними лівої частини системи.

Коефіцієнт a_i рівняння множинної лінійної регресії показує, на скільки зміниться результат Y, якщо відповідний фактор x_i зміниться на одиницю (в тих одиницях вимірювання, у яких він введений у модель)

Приклад 1. Нехай маємо специфікацію моделі у вигляді лінійної функції:

$$Y = a_0 + a_1X_1 + a_2X_2 + u \quad (7)$$

Звідси слідує, що система нормальних рівнянь для моделі (7) буде мати вигляд:

$$\begin{cases} \sum y = na_0 + a_1 \sum x_1 + a_2 \sum x_2, \\ \sum yx_1 = a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2, \\ \sum yx_2 = a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2. \end{cases} \quad (8)$$

3.3 Рівняння регресії в стандартизованому масштабі

Інший вид рівняння множинної регресії – рівняння регресії в стандартизованому масштабі:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + ... + \beta_m t_{x_m}, \quad (9)$$

де $t_y = \frac{y - \bar{y}}{\sigma_y}, \quad t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$ – стандартизовані змінні;

β_i – стандартизовані коефіцієнти регресії.

зв'язок між коефіцієнтами «чистої» регресії $\hat{a}_i$ та стандартизованими коефіцієнтами β_i існує зв'язок:

$$\beta_i = a_i \frac{\sigma_{x_i}}{\sigma_y}. \quad (10)$$

Стандартизовані коефіцієнти регресії показують на скільки «сігм» зміниться в середньому результат (*Y*), якщо відповідний фактор x_i зміниться на одну «сігму» при незмінному середньому рівні інших факторів.

3.4 Передумови застосування методу найменших квадратів 1МНК

Оцінювати параметри економетричної моделі за допомогою однокрокового (звичайного) методу найменших квадратів (1 МНК) можна за умов:

1) математичне сподівання залишків, тобто середня величина випадкових значень, дорівнює нулю

$$M(u)=0 \tag{11}$$

Якщо ця передумова не виконується, то йдеться про помилки специфікації.

Специфікацією моделі називають її аналітичну форму, яка складається з певного виду вибраної функції чи системи функцій для змінних. До помилок специфікації приводять:

– відсутність у моделі основної пояснювальної (незалежної) змінної, що призводить до зміщення оцінок параметрів і може привести до хибних висновків щодо значень параметрів;

– введення в модель пояснювальної змінної, яка не є істотною для вимірювального зв'язку, що може привести до неправильного встановленого кількісного зв'язку між змінними;

– використання невідповідних аналітичних форм вибраних функцій для моделі, що як і при першій помилці специфікації може привести до зміщення оцінок параметрів моделі.

2) *значення u_i в матриці залишків u незалежні між собою і мають постійну дисперсію σ^2:*

$$M(uu')=\sigma^2 E \tag{12}$$

де E – одинична матриця;

u' – матриця, транспонована до матриці u.

Дисперсія відображає «розсіювання» випадкових значень u_i навколо їх математичного сподівання.

Наявність сталої (постійної) дисперсії залишків називається *гомоскедастичністю*. Ця властивість може виконуватись лише тоді, коли залишки u є помилками вимірювання.

3) *незалежні змінні моделі не пов'язані із залишками*:

$$M(X'u)=0 \tag{13}$$

При порушенні цієї умови для оцінювання параметрів моделі використовують не 1 МНК, а інші методи.

4) *незалежні змінні моделі утворюють незалежну систему векторів, тобто ці змінні незалежні між собою*:

$$|X'X| \neq 0 \tag{14}$$

Якщо незалежні змінні пов’язані між собою, то це явище називають мультиколінеарністю і воно є небажаним, так як робить оцінку параметрів за допомогою 1 МНК ненадійною чутливою до вибраної специфікації моделі.

3.5 Підходи до побудови багатофакторної регресійної моделі

На кожний економічний показник впливає безліч факторів. При побудові регресійного рівняння виникає питання, які саме з них слід уводити в модель. Причому при використанні моделі для прогнозу бажано включити якомога більше факторів. З іншого боку збирання та обробка великої кількості інформації потребують значних витрат, тобто кількість факторів доцільно зменшити.

Для вибору компромісного рішення не існує єдиної процедури.

Тому для побудови “найкращого” рівняння застосовують один із таких підходів.

1. Перебір усіх можливих регресій – історично один із перших методів побудови регресійної моделі – найбільш громіздкий, тому що передбачає побудову регресій, які містять усі можливі комбінації впливових факторів. Іншими словами, якщо розглядається *m* факторів, то досліджується 2^m регресій, які порівнюються між собою за значеннями коефіцієнта детермінації та стандартною похибкою рівняння. Хоча цей метод і дає змогу дослідити усі можливі рівняння, однак при великій кількості факторів він, звичайно, неприйнятний.

2. Підхід, що ґрунтується на виключеннях змінних економніший щодо обчислень і базується на дослідженні часткових *F*-критеріїв, які дають змогу встановлювати статистичну значущість співвідношення між залишками моделі з найбільшою кількістю факторів і залишками моделі з одним вилученим

фактором. Якщо для деякого вилученого фактора таке співвідношення не є значущим (приймається нульова гіпотеза), то він до моделі не повертається. Таке дослідження проводиться також для рівняння з меншою кількістю факторів, але з більшою кількістю ступенів свободи.

3. Підхід, що ґрунтується на включеннях змінних (покроковий регресійний метод) діє у зворотному порядку порівняно з попереднім, тобто до моделі послідовно включаються фактори, що мають найбільший коефіцієнт кореляції із залежною змінною. Модель аналізується за значеннями коефіцієнта детермінації та частковими *F*-критеріями. Фактори, що не задовольняють критерії, з моделі вилучаються. Процес припиняється, якщо жоден з факторів рівняння вилучити не вдається, а новий претендент на включення не відповідає частковому *F*-критерію. На практиці цей метод найпоширеніший.

ТЕМА 4. Мультиколінеарність

План

4.1. Поняття мультиколінеарності. Її вплив на оцінки параметрів моделі.

4.2. Ознаки мультиколінеарності.

4.3. Алгоритм Фаррара-Глобера.

4.4. Методи усунення мультиколінеарності.

4.1 Поняття мультиколінеарності. Її вплив на оцінки параметрів моделі

Поняття мультиколінеарності. Її вплив на оцінки параметрів моделі. Ознаки мультиколінеарності. Алгоритм Фаррара-Глобера. Методи усунення мультиколінеарності.

Однією з умов використання методу найменших квадратів (1МНК) для знаходження параметрів економетричної моделі є те, що незалежні змінні у матриці X мають бути незалежними між собою, тобто $|X'X| \neq 0$ (четверта умова застосування 1 МНК). Проте на практиці можуть мати місце випадки, коли незалежні змінні пов’язані між собою, що стає перешкодою до використання 1МНК.

Явище існування тісної лінійної залежності, або сильної кореляції, між двома або більше незалежними змінними називається *мультиколінеарністю*.

Взагалі, термін «колінеарність» описує лінійний зв’язок між двома незалежними змінними, тоді як «мультиколінеарність» – між більш, ніж двома змінними.

Термін «мультиколінеарність» вперше було впроваджено Рагнаром Фрішем (1934р.).

Мультиколінеарність негативно впливає на кількісні характеристики економетричної моделі або робить неможливою її побудову.

Якщо пояснюючі змінні пов'язані строгим функціональним зв'язком, то говорять про *досконалу мультиколінеарність*. У випадку досконалої мультиколінеарності матриця X^TX виявляється виродженою, тобто її визначник рівний нулю, а значить, не існує оберненої матриці $\left(X^TX\right)^{-1}$, котра використовується в основних співвідношеннях методу найменших квадратів.

Досконала мультиколінеарність являється скоріше теоретичним прикладом. Реальна ж ситуація, коли між пояснюючими змінними існує доволі сильний кореляційний зв'язок, а не строго функціональний. Така залежність має назву *недосконалої мультиколінеарності*. Мультиколінеарність ускладнює розділення впливу пояснюючих змінних на поведінку залежної змінної і робить оцінки коефіцієнтів регресії ненадійними. Проблема простої мультиколінеарності (недосконалої) характерна для часових рядів. Таким чином, основною причиною мультиколінеарності є погана матриця незалежних змінних X.

Серед основних *наслідків*, до яких може призвести мультиколінеарність, можна виділити наступні:

1) при перевірці основної гіпотези про незначущість коефіцієнтів множинної регресії за допомогою t – критерію в більшості випадків вона приймається, проте саме рівняння регресії при перевірці за допомогою F – критерію виявляється значущим, що говорить про завищену величину коефіцієнта множинної кореляції;

2) отримані оцінки коефіцієнтів рівняння множинної регресії в основному невиправдано завищені або мають неправильні знаки;

3) додавання або виключення з вихідних даних одного-двох спостережень робить сильний вплив на оцінки коефіцієнтів моделі;

4) наявність мультиколінеарності в моделі множинної регресії може зробити її непридатною для подальшого застосування (наприклад, для побудови прогнозів).

Для виявлення мультиколінеарності не існує ніяких точних критеріїв, а застосовується ряд емпіричних методів.

4.2 Ознаки мультиколінеарності

Точних кількісних критеріїв для визначення наявності (відсутності) мультиколінеарності не існує. Тим паче, існують деякі рекомендації по виявленню мультиколінеарності.

1) В першу чергу аналізують матрицю парних коефіцієнтів кореляції

$$\widehat{R} = \begin{bmatrix} 1 & r_{y1} & r_{y2} & r_{y3} & \dots & r_{ym} \\ r_{1y} & 1 & r_{12} & r_{13} & \dots & r_{1m} \\ r_{2y} & r_{21} & 1 & r_{23} & \dots & r_{2m} \\ r_{3y} & r_{31} & r_{32} & 1 & \dots & r_{3m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{my} & r_{m1} & r_{m2} & r_{m3} & \dots & 1 \end{bmatrix}, \qquad (4.1)$$

точніше ту її частину, яка відноситься до пояснюючих змінних:

$$\widehat{R} = \begin{bmatrix} 1 & r_{12} & r_{13} & \dots & r_{1m} \\ r_{21} & 1 & r_{23} & \dots & r_{2m} \\ r_{31} & r_{32} & 1 & \dots & r_{3m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & r_{m3} & \dots & 1 \end{bmatrix}, \qquad (4.2)$$

де r_{ij} – парний коефіцієнт кореляції між змінними:

r_{yj} – парний коефіцієнт кореляції між Y і X.

Вважається, що присутність коефіцієнтів r_{ij}, які перевищують значення 0,75–0,8, свідчить про наявність мультиколінеарності.

2) Якщо визначник матриці $R=X^TX$ близький до нуля (наприклад, одного порядку з накопичуваними похибками обчислень), то це свідчить про наявність мультиколінеарності.

3) Коефіцієнт детермінації R^2 достатньо великий, але деякі з коефіцієнтів регресії статистично не значимі, тобто вони мають низькі t-статистики.

4) Високі часткові коефіцієнти кореляції свідчать про наявність мультиколінеарності. При вивченні багатомірних зв’язків необхідно вимірювати дійсну силу лінійного зв’язку між двома змінними, очищену від впливу на досліджувану пару змінних інших факторів. Часткові коефіцієнти кореляції

визначають силу лінійного зв'язку між двома змінними без врахування впливу на них інших змінних.

4.3 Алгоритм Фаррара-Глобера

Найповніше дослідити мультиколінеарність дає змогу алгоритм Фаррара-Глобера, який застосовує три види статистичних критеріїв для виявлення мультиколінеарності:

- усього масиву незалежних змінних (критерій χ^2);
- кожної незалежної змінної з усіма іншими (*F*-критерій);
- кожної пари незалежних змінних (*t*-критерій).

Порівнявши ці критерії з їх критичними значеннями, можна зробити конкретні висновки щодо наявності чи відсутності мультиколінеарності незалежних змінних. Опишемо цей алгоритм.

Складемо покроковий алгоритм Фаррара-Глобера.

1-ий крок: Нормалізуємо значення факторів:

$$x_{ij}^{*} = \frac{\left(x_{ij} - \bar{x}_j\right)}{\sqrt{n\sigma_{x_j}^2}} \tag{4.3}$$

де n – кількість спостережень (i = 1, 2, ..., n);

m – кількість незалежних змінних (j=1, m);

$\bar{x}_j$ – середня арифметична j-ої незалежної змінної;

$\sigma_{x_j}^2$ – дисперсія j-ої незалежної змінної.

2-ий крок: Знаходимо кореляційну матрицю:

$$R = \left(X^{*}\right)^T X^{*} = \begin{pmatrix} 1 & r_{12} & \ldots & r_{1m} \\ r_{21} & 1 & \ldots & r_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ r_{m1} & r_{m2} & \ldots & 1 \end{pmatrix}, \tag{4.4}$$

де $\left(X^{*}\right)^T$ – транспонована матриця X^{*} (елементи матриці R характеризують щільність зв'язку однієї незалежної змінної з іншою);

$r_{ij} = r_{x_i x_j}$ – парні коефіцієнти кореляції.

Однак на основі цієї залежності не можна стверджувати, що отриманий зв'язок є явищем мультиколінеарності. Якщо діагональні елементи матриці *R* не дорівнюють одиниці, то на діагоналі цієї матриці ми проставляємо одиниці, а до решти елементів додаємо різницю між одиницею й значенням діагонального елемента.

3-ий крок: Визначаємо значення χ^2

$$\chi^2 = -\left[n - 1 - \frac{1}{6}(2m + 5) \right] \cdot \ln|R| \tag{4.5}$$

де $|R|$ – визначник кореляційної матриці *R*;

Порівнюємо χ^2 з табличним значенням при $\frac{1}{2}m(m-1)$ ступенях свободи і рівні значущості α. Якщо $\chi^2 > \chi^2_{табл}$, то в масиві незалежних змінних існує мультиколінеарність.

Для обчислення $\chi^2_{табл}$ використовують функцію ХИ2ОБР(рівень значущості; ступені свободи).

4-ий крок: Визначаємо обернену матрицю (матрицю похибок)

$$\mathrm{C} = \mathrm{R}^{-1} = \left(\left(X^* \right)^T X^* \right)^{-1} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mm} \end{pmatrix} \tag{4.6}$$

5-ий крок: Визначаємо *F*-критерії Фішера

$$F_k = \frac{(c_{kk} - 1)(n - m)}{(m - 1)} \tag{4.7}$$

де c_{kk} – діагональні елементи матриці *C*;

Розраховані значення критеріїв порівнюються з табличними при (*n*-*m*) і (*m*-1) ступенях свободи й рівні значущості α. Якщо $\mathrm{F_k} > F_{табл}$, то відповідна *k*-та незалежна змінна мультиколінеарна з іншими);

6-ий крок: Розраховуємо часткові коефіцієнти кореляції, які характеризують щільність зв'язку між двома змінними за умови, що інші змінні не впливають на цей зв'язок (існування парної мультиколінеарності):

$$r_{kj} = \frac{-c_{kj}}{\sqrt{c_{kk} c_{jj}}} \tag{4.8}$$

де c_{kj} – елементи матриці C, що розміщені в k-ому рядку та j-му стовпці;

c_{kk} і c_{jj} – діагональні елементи матриці C.

7-ий крок: Розрахуємо t-критерії:

$$t_{kj} = \left|r_{kj}\right| \cdot \frac{\sqrt{n-m}}{\sqrt{1-r_{kj}^2}} \tag{4.9}$$

Розраховані значення t-критеріїв t_{kj} порівняти з табличними при (n-m) ступенях свободи та рівні значущості α. Якщо $t_{kj} > t_{табл}$, то між незалежними змінними x_k і x_j існує мультиколінеарність.

Для оцінки параметрів моделя в яку входять мультиколінеарні змінні використовують метод головних компонент.

Приклад 4.1. Розглянемо дослідження впливу на економічний показник y – реальне споживання країни (у млрд. грн) трьох факторів: x_1 – купівлі та оплати товарів і послуг (у млрд. грн), x_2 – усіх заощаджень від загального грошового доходу (у % від загальної суми доходу), x_3 – рівня ставки ПДВ (у %). Необхідно перевірити факти на мультиколінеарність.

№ п/п	y	x_1	x_2	x_3
1	25,75	4,69	11,97	29,23
2	25,34	5,64	13,43	29,35
3	31,26	6,26	12,92	33,4
4	33,5	6,99	14,74	30,97
5	32,3	6,36	14,64	32,92
6	38,9	7,6	17,1	37,27
7	41,58	7,12	15,63	30,97
8	48,02	6,81	15,35	33,58
9	43,3	8,67	15,85	35,62

10	51,78	7,83	18,05	34,99
11	52,14	7,84	17,24	39,34
12	54,94	8,85	20,52	41,5
13	59,18	9,61	19,18	45,58
14	62,22	10,67	19,03	41,08
15	63,62	11,04	21,45	40,54
16	65,01	11,85	22,25	42,75
17	67,78	12,94	24,75	43,89
18	71,45	14,24	25,03	41,95
19	75,24	15,67	27,87	44,06
20	77,38	16,33	30,49	46,77
Середнє	51,03	9,35	18,87	37,79
Дисперсія	273,36	11,35	26,08	31,86

1-ий крок: Нормуємо значення факторів, використовуючи формулу (4.3). В результі отримуємо матрицю X^*.

$$X^* = \begin{pmatrix} -0{,}309 & -0{,}302 & -0{,}339 \\ -0{,}246 & -0{,}238 & -0{,}334 \\ -0{,}205 & -0{,}261 & -0{,}174 \\ -0{,}157 & -0{,}181 & -0{,}270 \\ -0{,}198 & -0{,}185 & -0{,}193 \\ -0{,}116 & -0{,}078 & -0{,}021 \\ -0{,}148 & -0{,}142 & -0{,}270 \\ -0{,}169 & -0{,}154 & -0{,}167 \\ -0{,}045 & -0{,}132 & -0{,}086 \\ -0{,}101 & -0{,}036 & -0{,}111 \\ -0{,}100 & -0{,}072 & 0{,}061 \\ -0{,}033 & 0{,}013 & 0{,}147 \\ 0{,}017 & 0{,}007 & 0{,}309 \\ 0{,}088 & 0{,}113 & 0{,}130 \\ 0{,}112 & 0{,}148 & 0{,}109 \\ 0{,}166 & 0{,}257 & 0{,}197 \\ 0{,}238 & 0{,}270 & 0{,}242 \\ 0{,}324 & 0{,}394 & 0{,}165 \\ 0{,}419 & 0{,}072 & 0{,}248 \\ 0{,}463 & 0{,}509 & 0{,}356 \end{pmatrix}$$

2-ий крок: Знаходимо кореляційну матрицю за формулою (4.4):

$$R = \left(X^*\right)^T X = \begin{pmatrix} 0{,}950 & 0{,}929 & 0{,}820 \\ 0{,}929 & 0{,}950 & 0{,}831 \\ 0{,}820 & 0{,}831 & 0{,}950 \end{pmatrix},$$

3-ий крок: Визначаємо значення χ^2

Розрахуємо визначних матриці R – $|R|$

$|R| = 0{,}009$, тоді

$$\chi^2 = -\left[20 - 1 - \frac{1}{6}(2 \cdot 3 + 5)\right] \cdot 0{,}009 = 85{,}157$$

де $|R|$ – визначник кореляційної матриці *R*;

Порівнюємо χ^2 з табличним значенням при $\frac{1}{2}m(m-1) = 3$ ступенях свободи і рівні значущості $\alpha = 0{,}05$.

$$\chi^2_{табл} = 7{,}814724$$

Оскільки $\chi^2 > \chi^2_{табл}$, то в масиві незалежних змінних існує мультиколінеарність у сукупності.

4-ий крок: Визначаємо обернену матрицю (матрицю похибок) за формулою (4.6):

$$\mathrm{C} = \mathrm{R}^{-1} = \left(\left(X^*\right)^T X^*\right)^{-1} = \begin{pmatrix} 23{,}989 & -22{,}810 & -0{,}736 \\ -22{,}810 & 26{,}180 & -3{,}230 \\ -0{,}736 & -3{,}230 & 4{,}513 \end{pmatrix}$$

5-ий крок: Визначаємо *F*-критерії Фішера за формулою (4.7):

$$F_1 = \frac{(23{,}989 - 1)(20 - 3)}{(3 - 1)} = 195{,}405$$

$$F_2 = \frac{(26{,}180 - 1)(20 - 3)}{(3 - 1)} = 214{,}028$$

$$F_3 = \frac{(4{,}513 - 1)(20 - 3)}{(3 - 1)} = 29{,}864$$

Розраховані значення критеріїв порівнюються з табличними при (*n*-*m*) = 17 і (*m*-1) = 2 ступенях свободи й рівні значущості $\alpha = 0{,}05$.

$$F_{табл} = 19{,}43703$$

Оскільки $\mathrm{F}_1 > F_{табл}$, $\mathrm{F}_2 > F_{табл}$, $\mathrm{F}_3 > F_{табл}$, то робимо висновок, що перша, друга й третя незалежні змінні мультиколінеарні з іншими;

6-ий крок: Розраховуємо часткові коефіцієнти кореляції, які характеризують щільність зв'язку між двома змінними за умови, що інші змінні не впливають на цей зв'язок (існування парної мультиколінеарності) (4.8):

$$r_{12} = \frac{-(-22{,}810)}{\sqrt{23{,}989 \cdot 26{,}180}} = 0{,}910$$

$$r_{13} = \frac{-(-0{,}736)}{\sqrt{23{,}989 \cdot 4{,}513}} = 0{,}071$$

$$r_{23} = \frac{-(-3{,}230)}{\sqrt{26{,}180 \cdot 4{,}513}} = 0{,}297$$

7-ий крок: Розрахуємо *t*-критерії (4.9):

$$t_{12} = |0{,}910| \cdot \frac{\sqrt{20-3}}{\sqrt{1-0{,}910^2}} = 9{,}061$$

$$t_{13} = |0{,}071| \cdot \frac{\sqrt{20-3}}{\sqrt{1-0{,}071^2}} = 0{,}292$$

$$t_{13} = |0{,}297| \cdot \frac{\sqrt{20-3}}{\sqrt{1-0{,}297^2}} = 1{,}283$$

Значення критеріїв порівнюємо з табличним при (n-m) = 17 ступенях свободи й рівні значущості α = 0,05.

$$t_{табл} = 2{,}109818$$

Оскільки $t_{12} > t_{табл}$, $t_{13} < t_{табл}$, $t_{23} < t_{табл}$, то між першою та другою незалежними змінними існує мультиколінеарність.

4.4 Методи усунення мультиколінеарності

Відмітимо, що в ряді випадків мулько лінеарність не являється таким вже серйозним «злом», щоб докладати суттєві зусилля по її виявленню і усуненню. в основному все залежить від цілей дослідження.

Якщо основна задача моделі – прогноз майбутніх значень залежної змінної, то при достатньо великому коефіцієнті детермінації $R^2 (\geq 0{,}9)$, наявність мультиколінеарності зазвичай не впливає на прогнозні якості моделі (якщо в майбутньому між корельованими змінними будуть зберігатися ті ж відношення, що і раніше).

Якщо необхідно визначити ступінь впливу кожної з пояснюючих змінних на залежну змінну, то мультиколінеарність, що призводить до збільшення стандартних похибок, скоріше за все, викривить істинні залежності між змінними. В цій ситуації мультиколінеарність являється серйозною проблемою.

Єдиного методу усунення мультиколінеарності не існує. Це пов’язано з тим, що причини і наслідки мультиколінеарності неоднозначні і багато в чому залежать від результатів вибірки.

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

1) отримання додаткових даних чи нової вибірки;
2) зміна специфікації моделі;
3) виключення змінної з моделі;
4) використання попередньої інформації про деякі параметри;
5) перетворення змінних;
6) метод покрокового включення;
7) гребенева регресія (ridge).

Виключення змінної з моделі

Найпростішим методом усунення мультиколінеарності являється виключення з моделі однієї чи ряду корельованих змінних. При застосуванні даного методу необхідна певна обачливість. В даній ситуації можливі помилки специфікації, тому в прикладних економетричних моделях бажано не виключати пояснюючі змінні до тих пір, доки мультиколінеарність не стане серйозною проблемою.

Отримання додаткових даних чи нової вибірки

Оскільки мультиколінеарність прямо залежить від вибірки, то, можливо, при іншій вибірці мультиколінеарності не буде чи вона не буде такою серйозною. Іноді для зменшення мультиколінеарності достатньо збільшити об’єм вибірки. Наприклад, при використанні щорічних даних можна перейти до щоквартальних даних. Збільшення кількості даних скорочує дисперсії коефіцієнтів регресії і тим самим збільшує їх статистичну значимість. Однак отримання нової вибірки чи розширення старої не завжди можливе чи пов’язане з серйозними витратами. Крім того, такий підхід може посилити автокореляцію. Ці проблеми обмежують можливість використання даного методу.

Зміна специфікації моделі

У ряді випадків проблема мультиколінеарності може бути вирішена шляхом зміни специфікації моделі: або змінюється форма моделі, або додаються пояснюючі змінні, не враховані в початковій моделі, але суттєво впливаючі на залежну змінну. Якщо даний метод має основу, то його використання зменшує суму квадратів відхилень, тим самим, скорочуючи стандартну похибку регресії. Це призводить до зменшення стандартних похибок коефіцієнтів.

Використання попередньої інформації про деякі параметри

Іноді при побудові моделі множинної регресії можна скористатися попередньою інформацією, зокрема відомими значеннями деяких коефіцієнтів регресії. Цілком ймовірно, що значення коефіцієнтів, розрахованих для яких-небуть попередніх (зазвичай простих) моделей чи для аналогічної моделі по раніше отримані вибірці, можуть бути використані для створюваної моделі в даний момент.

Перетворення змінних

Одним із способів усунення мутиколінеарності є перетворення змінних, наприклад замість значення всіх змінних, що включені в модель (і результуючої в тому числі), можна взяти їх логарифми:

$$\ln y = \beta_0 + \beta_1 \ln x_1 + \beta_2 \ln x_2 + u \qquad (4.10)$$

Але і це не гарантує позбавлення від мультиколінеарності.

Якщо ні один з наведених методів не дав бажаних результатів, то необхідно використати зміщені методи оцінки невідомих параметрів моделі.

До зміщених методів оцінки коефіцієнтів регресії можна віднести *гребеневу регресію чи рідж* (ridge). Її використовують в тому випадку, коли ні одну із змінних, що включені в модель, видалити не можна.

Суть гребеневої регресії полягає в тому, що до всіх діагональних елементів матриці (X^TX) додається невелике число τ (тау): $10^{-6} < \tau < 0{,}1$. Невідомі параметри множинного рівняння регресії в данному випадку визначаються по формулі:

$$\tilde{\beta}_{ridge} = \left(X^T X + \tau \cdot E\right)^{-1} X^T Y , \qquad (4.11)$$

де E – одинична матриця.

У результаті застосування ріджа оцінки коефіцієнтів рівняння множинної регресії стабілізуються до певного значення, і їх стандартні похибки зменшуються.

Метод покрокового включення

Метод покрокового включення змінних в модель дозволяє обрати із можливого набору змінних саме ті, які підсилюють якість моделі регресії.

Алгоритм методу покрокового включення:

1) із числа всіх змінних в модель регресії включають ті, котрі мають найбільший модуль парного лінійного коефіцієнта кореляції з результативною змінною;

2) при додаванні в модель нових факторів необхідно перевіряти їх значимість за допомогою *F*-критерію Фішера.

Основна гіпотеза формулюється як недоцільність включення фактору x_k в модель множинної регресії. Альтернативна гіпотеза виходить із оберненого твердження.

$$F_{набл}(\mathrm{x_k}) = \frac{R^2(y, x_1, ..., x_{m+1}) - R^2(y, x_1, ..., x_m)}{R^2(y, x_1, ..., x_m)} \cdot (n - q - 1) \quad (4.12)$$

де q – число вже включених в модель змінних.

Критичне значення *F*-критерію визначається по таблиці розподілу Фішера-Снедекора з рівнем значимості α і числом ступенів свободи: *k1*=1 і *k2*=*n*-1: $F_{табл}(\alpha; \mathrm{k_1}; k_2)$.

Якщо $F_{набл} > F_{табл}$, то включення змінної в модель множинної регресії являється обґрунтованим.

Перевірка факторів на значимість здійснюється до тих пір, доки не знайдеться хоча б одна змінна, для якої не виконується умова $F_{набл} > F_{табл}$.

ТЕМА 5. Гетероскедастичність

План

5.1. Поняття гомо- і гетероскедастичності. Вплив гетероскедастичності на властивості оцінок параметрів.

5.2. Методи виявлення гетероскедастичності.

5.3. Тест рангової кореляції Спірмена.

5.4. Параметричний тест Гольдфельда – Квандта.

5.1 Поняття гомо- і гетероскедастичності. Вплив гетероскедастичності на властивості оцінок параметрів

Поняття гомо- і гетероскедастичності. Вплив гетероскедастичності на властивості оцінок параметрів. Методи визначення гетероскедастичності.

Одним з основних припущень моделі класичної лінійної регресії є припущення про сталість дисперсії кожної випадкової величини u_i (гомоскедастичність).

Формалізовано це припущення записується у вигляді:

$$Du_i = M\left(u_i - Mu_i\right)^2 = \sigma_{u_i}^2 = const, \quad i = \overline{1,n}.$$

Якщо це припущення не задовольняється в деякому окремому випадку, то

$$Du_i = \sigma_{u_i}^2 \neq const,$$

і має місце гетероскедастичність. Звичайно нас цікавить питання про доцільність цього припущення і про те, що відбувається, коли припущення про сталість дисперсії випадкової величини u_i не задовольняється.

Суть припущення гомоскедастичності полягає в тому, що варіація кожної u_i навколо її математичного сподівання не залежить від значення x. Дисперсія кожної u_i. зберігається сталою незалежно від малих чи великих значень факторів: $\sigma_{u_i}^2$ не є функцією x_{ij} тобто $\sigma_{u_i}^2 \neq f\left(x_{1i}, x_{2i}, ..., x_{pi}\right)$.

Графічно випадок гомоскедастичності для простої лінійної регресії показано випадковою дисперсією u_i у межах сталої відстані навколо лінії регресії (див. рис. 5.1).

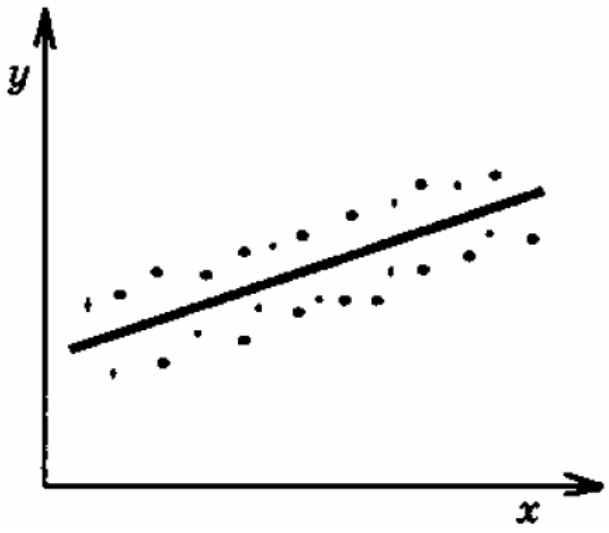


Рисунок 5.1 – Гомоскедастичність

Якщо $\sigma_{u_i}^2$ не є сталою, а її значення залежать від значень x, можемо записати

$$\sigma_{u_i}^2 = f\left(x_{1i}, x_{2i}, ..., x_{pi}\right)$$

У цьому разі маємо справу з гетероскедастичністю.

Графічна форма розкиду спостережень залежить від форми гетероскедастичності, тобто форми зв'язку між $\sigma_{u_i}^2$ та x_i На рисунку 5.2-5.4 показано три різні форми гетероскедастичності.

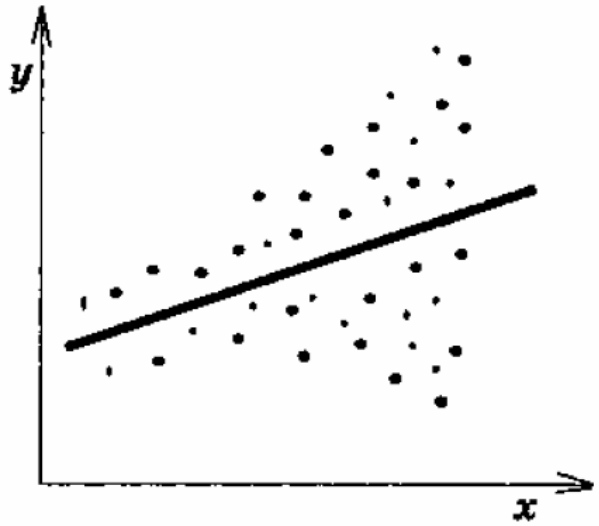


Рисунок 5.2 – Зростання дисперсії u_i

Зокрема на рис. 5.2 показано випадок (монотонно) зростаючої дисперсії $\sigma_{u_i}^2$ (із зростанням x зростає і дисперсія u_i).

Це загальноприйнята форма гетероскедастичності, що допускається в економетричних дослідженнях. На рис. 5.3 показано випадок спадної гетероскедастичності: коли x набуває більших значень, відхилення спостережень від лінії регресії зменшується, таким чином дисперсія випадкової змінної зменшується із зростанням x.

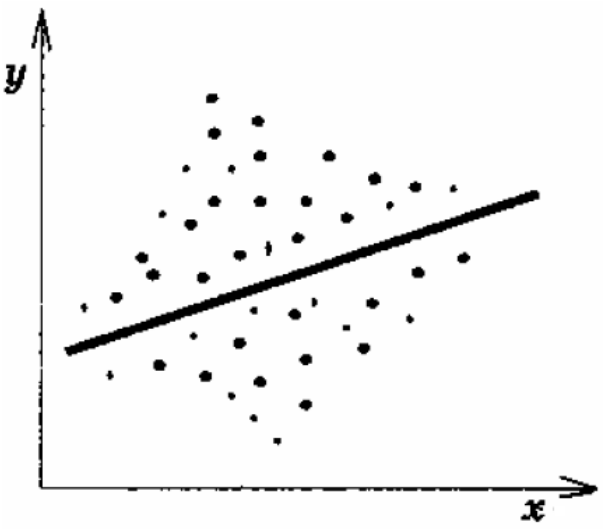


Рисунок 5.3 – Спадання дисперсії u_i

На рис. 5.4 зображено більш складну форму гетероскедастичності: спочатку дисперсія є зменшується із зростанням x, але після певного рівня x^* починає зростати із зростанням x.

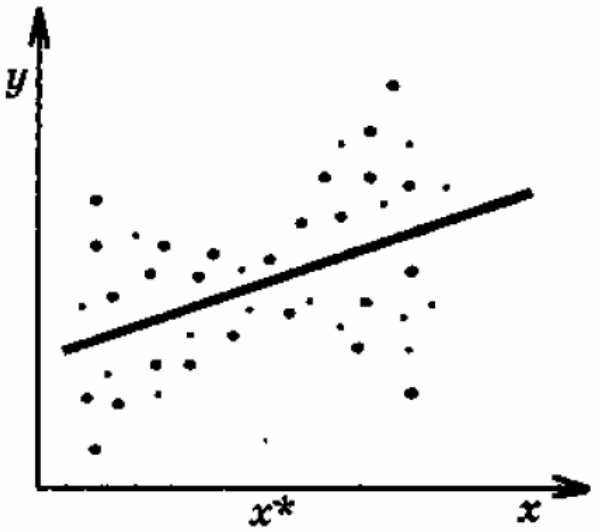


Рисунок 5.4 – Складний випадок

Форма гетероскедастичності залежить від знаків та значень коефіцієнтів у залежності Оскільки u_i – неспостережувана випадкова величина, ми не знаємо справжньої форми гетероскедастичності.

У прикладних дослідженнях, як правило, використовується зручне припущення, що гетероскедастичність (у разі простої лінійної регресії) має форму $\sigma_{u_i}^2 = k^2 x_i^2$, де k — константа, яку потрібно оцінити.

У багатьох економетричних дослідженнях може очікуватись, що припущення про сталу дисперсію випадкової змінної не зберігатиметься. Це можна легко зрозуміти, якщо врахувати фактори, вплив яких абсорбується значенням помилки. Згадаємо, що випадкова величина u_i виражає вплив на залежну змінну помилок в її вимірюванні та неврахованих факторів. У обох випадках є підстави для зміни з часом дисперсії u_i. Помилки вимірювання мають тенденцію до накопичення з плином часу, тому їхня величина збільшується. У такому разі дисперсія u_i збільшується із зростанням значень x. З іншого боку, техніка вибірки та інші методи збору даних постійно вдосконалюються і тому похибки вимірювання можуть зменшуватися. У такому разі $\sigma_{u_i}^2$ також зменшуватиметься. Але важливіше те, що багато із неврахованих змінних можуть змінюватись в однаковому з x напрямку, викликаючи, таким чином, збільшення відхилення спостережень від лінії регресії.

Наслідки порушення припущення про гомоскедастичність:

1) неможливо знайти середньоквадратичне відхилення параметрів σ_{a_i} регресії, а отже, неможливо оцінити значущість параметрів;

2) неможливо побудувати довірчий інтервал для прогнозних значень y_p;

3) отримані за МНК оцінки параметрів регресії не є ефективними (не мають найменшої дисперсії).

Зазначимо, що, якщо незважаючи на гетероскедастичність, ми використовуватимемо звичайні процедури перевірки гіпотез, то висновки можуть бути неправильними. Тому оцінки параметрів доцільно знаходити за узагальненим методом найменших квадратів (інша назва – метод Ейткена).

5.2 Методи виявлення гетероскедастичності

Як і в разі мультиколінеарності, єдиних правил виявлення гетероскедастичності немає, а є різноманітні тести (критерії): аналіз змісту проблеми, графічний аналіз, тест рангової кореляції Спірмена, критерій f, параметричний та непараметричний тести Гольдфельда-Квандта, тест Глейзера тощо. Розглянемо деякі з них.

1. *Аналіз змісту проблеми.*

Інколи при проведенні економетричних досліджень гетероскедастичність вгадується інтуїтивно або висувається як абсолютне припущення. Попередній аналіз проблеми, що вивчається, може навести на думку про наявність гетероскедастичності. Наприклад, при вивченні бюджету сім'ї можна помітити, що дисперсія залишків зростає відповідно до зростання доходу. При зведеному аналізі діяльності різних за розміром фірм також можна очікувати гетероскедастичність. І таких прикладів багато. Отже, перший крок до вияву гетероскедастичності — глибокий аналіз досліджуваної проблеми.

2. Графічний аналіз

Досить наочним та простим методом тестування припущення про наявність гетероскедастичності є графічний метод. Не завжди дослідник володіє необхідним для аналізу проблеми емпіричним матеріалом. Крім того, його висновки щодо наявності або відсутності гетероскедастичності носять суб'єктивний характер, і в цих умовах на допомогу приходять графіки.

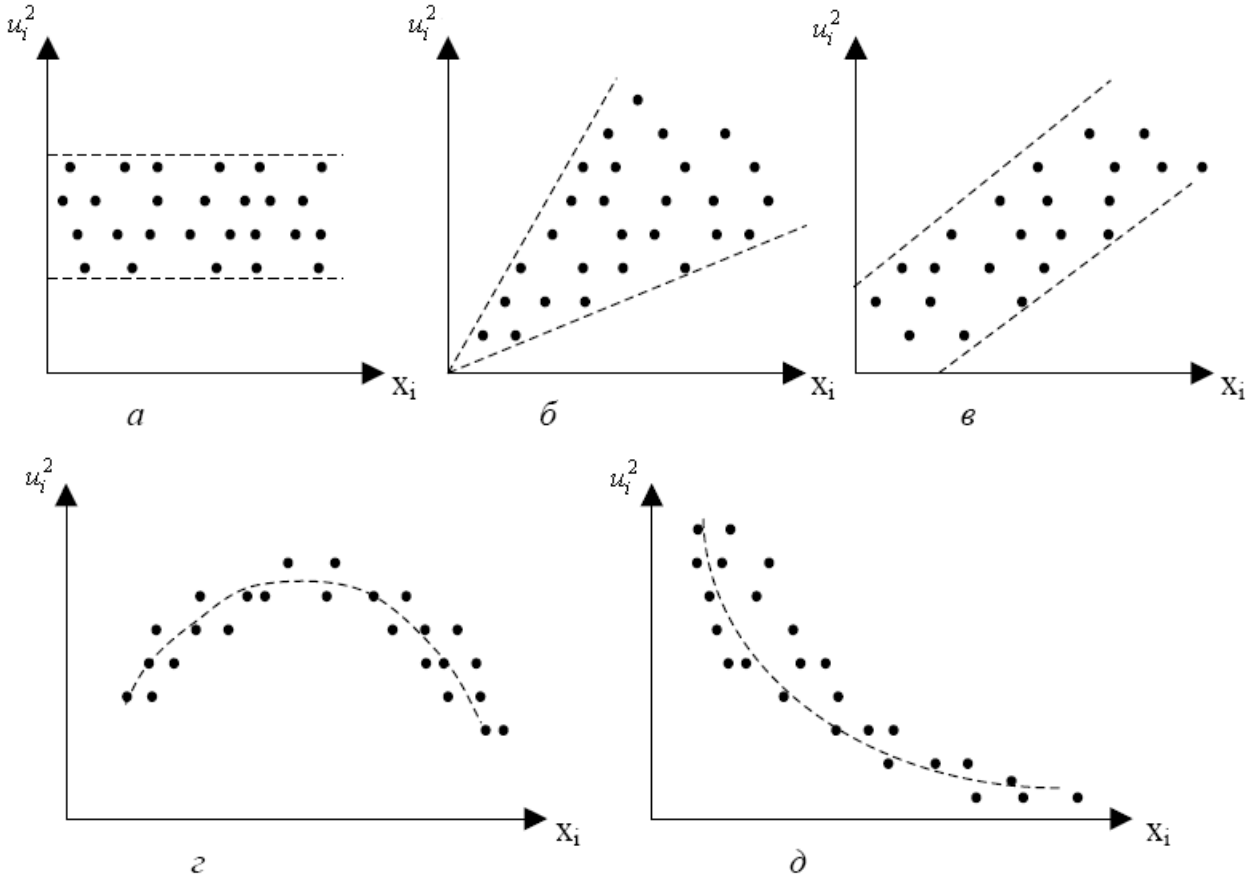


Рисунок 5.5 – Різні типи графіків квадрату залишків

5.3 Тест рангової кореляції Спірмена

Це найпростіший тест, який можна використовувати як до малих, так і до великих вибірок.

Використання цього тесту ґрунтується на припущенні, що дисперсія відхилень u_i буде збільшуватися або зменшуватися при зростанні величини x_i.

Спочатку запишемо коефіцієнт рангової кореляції Спірмана:

$$\rho = 1 - \frac{6\sum d^2}{n\left(n^2 - 1\right)} \tag{5.1}$$

де d – різниця рангів кожної пари значень;

n – кількість спостережень.

Якщо серед рангів зустрічається кілька однакових, то утворюються так звані зв'язні ранги, тобто, однакові середні номери, наприклад, замість однакових по порядку третього і четвертого значень будуть два ранги по 3,5.

У цьому випадку коефіцієнт Спірмена обчислюється за формулою:

$$\rho = 1 - 6\frac{\sum d^2 - A - B}{\sqrt{(n^3 - n - 12A)(n^3 - n - 12B)}},$$

$$A = \frac{1}{12}\sum_{j=1}^{J}(A_j^3 - A_j)$$

$$B = \frac{1}{12}\sum_{k=1}^{K}(B_k^3 - B_k)$$

де J – кількість зв'язок по x_i;

A_j - кількість однакових рангів у зв'язці номер j;

K – кількість зв'язок по u_i;

B_k - кількість однакових рангів у зв'язці номер k.

Наведений коефіцієнт рангової кореляції може використовуватись для визначення гетероскедастичності таким чином.

Припустимо, $y_i = a_0 + a_1 x_i + u_i$.

Етап 1. Побудувати регресію для даних y та x і розрахувати відхилення u_i.

Етап 2. Нехтуючи знаком u_i, тобто беручи абсолютні значення $|u_i|$, ранжуємо $|u_i|$ та x_i у зростаючому чи спадному порядку і підрахувати коефіцієнт рангової кореляції Спірмана.

Етап 3. Перевірити значущість отриманого коефіцієнта рангової кореляції за t–критерієм Ст'юдента. Для цього розрахуємо t-статистику:

$$t = \frac{\rho\sqrt{n-2}}{\sqrt{1-\rho^2}}. \qquad (5.2)$$

При даних ступенях свободи за таблицями Ст'юдента знаходимо $t(\alpha; n-m-1)$ при $m=1$. Якщо розраховане значення перевищує $t_{кр}$ ($t > t_{кр}$) - це підтверджує гіпотезу про гетероскедастичність. Якщо $t \leq t_{кр}$, тоді в регресійній моделі правильним є припущення про гомоскедастичність.

Для прикладу:

	A	B	C	D	E	F	G	H
1				РАНГИ				
2	x	u		x	u			
3	1,5	0,07		5	1			
4	2,2	0,02		3	4			
5	4	0,02		2	4			
6	1,1	0,05		6	2			
7	2,2	0,02		3	4			
8	8	0,05		1	2			
9								
10								
11				розрахунок зв'язок (сортування)			перерахунок рангів	
12				x	u		x	u
13				1	1		1	1
14				2	2		2	2,5
15				3	2		3,5	2,5
16				3	4		3,5	5
17				5	4		5	5
18				6	4		6	5
19								
20				оновлена таблиця рангів				
21				x	u	d^2		
22				5	1	16		
23				3,5	5	2,25		
24				2	5	9		
25				6	2,5	12,25		
26				3,5	5	2,25		
27				1	2,5	2,25		
28					сумма d^2=	44		

	A	B	C	D	E	F	G	H
1				РАНГИ				
2	x	u		x	u			
3	1,5	0,07		=РАНГ(A3;$A$3:$A$8)	=РАНГ(B3;$B$3:$B$8)			
4	2,2	0,02		=РАНГ(A4;$A$3:$A$8)	=РАНГ(B4;$B$3:$B$8)			
5	4	0,02		=РАНГ(A5;$A$3:$A$8)	=РАНГ(B5;$B$3:$B$8)			
6	1,1	0,05		=РАНГ(A6;$A$3:$A$8)	=РАНГ(B6;$B$3:$B$8)			
7	2,2	0,02		=РАНГ(A7;$A$3:$A$8)	=РАНГ(B7;$B$3:$B$8)			
8	8	0,05		=РАНГ(A8;$A$3:$A$8)	=РАНГ(B8;$B$3:$B$8)			
9								
10								
11				розрахунок зв'язок (сортування)			перерахунок рангів	
12				x	u		x	u
13				1	1		1	1
14				2	2		2	=(2+3)/2
15				3	2		=(3+4)/2	=(2+3)/2
16				3	4		=(3+4)/2	=(4+5+6)/3
17				5	4		5	=(4+5+6)/3
18				6	4		6	=(4+5+6)/3
19								
20				оновлена таблиця рангів				
21				x	u	d^2		
22				5	1	=(D22-E22)^2		
23				3,5	5	=(D23-E23)^2		
24				2	5	=(D24-E24)^2		
25				6	2,5	=(D25-E25)^2		
26				3,5	5	=(D26-E26)^2		
27				1	2,5	=(D27-E27)^2		
28					сумма d^2=	=СУММ(F22:F27)		
29								

Розрахуємо:

$$A = \frac{1}{12}(2^3 - 2) = \frac{1}{2};$$

$$B = \frac{1}{12}\left[\left(2^3 - 2\right) + (3^3 - 3)\right] = \frac{1}{12}(6 + 24) = 2{,}5.$$

$$\rho = -0{,}2838$$

ρ= 0,2837598
tрозр= 0,5918472
tкрит= 2,7764451

$t_{розр} < t_{крит} \rightarrow$ гетероскедастичність відсутня

5.4 Параметричний тест Гольдфельда – Квандта

Цей тест застосовується до великих вибірок. Спостережень має бути хоча б удвічі більше, ніж оцінюваних параметрів. Тест припускає нормальний розподіл та незалежність випадкової величини u_i.

1-ий крок: спостереження (вихідні дані) впорядкувати відповідно до величини елементів вектора x_i, який може спричинити зміну дисперсії залишків.

2-ий крок: відкинути c спостережень, які розміщені всередині векторів вихідних даних, де $c = \frac{4n}{15}$, n – кількість елементів вектора $x_{i.}$.

3-ий крок: побудувати дві моделі на основі звичайного МНК за двома створеними сукупностями спостережень обсягом $\frac{n-c}{2}$ за умови, що $\frac{n-c}{2} \geq m$, де m – кількість змінних.

4-ий крок: знайти суму квадратів залишків S_1 і S_2 за першою і другою моделями:

$$S_1 = u_1^T u_1, \quad S_2 = u_2^T u_2, \tag{5.3}$$

де u_1, u_2 – залишки відповідно за першою і другою моделями.

5-ий крок: розрахувати критерій $F^* = \frac{S_2}{S_1}$ (якщо $S_2 > S_1$), який у разі виконання гіпотези про гомоскедастичність відповідатиме F-розподілу з $k_1 = \frac{n-c-2m}{2}$, $k_2 = \frac{n-c-2m}{2}$ ступенями свободи; Значення критерію F^* порівняти з табличним значенням F-критерію при вибраному рівні значущості α і відповідних ступенях свободи. Якщо $F^* \leq F_{табл}$, то гетероскедастичність відсутня.

Чим більше значення F^*, тим більша гетероскедастичність залишків.

Цей тест може бути використаний при припущенні про обернену пропорційність між σ_i і значеннями пояснюючої змінної. При цьому статистика Фішера має вид: $F^* = \frac{S_1}{S_2}$.

ТЕМА 6. Автокореляція в економетричних моделях

План

6.1. Природа і наслідки автокореляції.
6.2. Методи визначення автокореляції. Критерій Дарбіна-Уотсона. Критерій фон Неймана.
6.3. Коефіцієнти автокореляції та їх застосування.
6.4. Моделі з автокорельованими залишками.

6.1 Природа і наслідки автокореляції

Природа і наслідки автокореляції. Методи визначення автокореляції. Критерій Дарбіна-Уотсона. Критерій фон Неймана. Коефіцієнти автокореляції та їх застосування. Моделі з автокорельованими залишками.

Кореляція, що виникає між рівнями досліджуваної змінної, називається автокореляція. Явище автокореляції в основному властиво даним, представленим у вигляді часових рядів.

Часовий ряд - набір послідовних результатів спостереження, виміряних у різні часові проміжки. Приклади часових рядів: змінення курсу валют, середня заробітна плата по країні по місяцях тощо.

Автокореляція залишків регресійної моделі u_i – кореляційна залежність між теперішніми і попередніми значеннями залишків. Величина зсуву між рядами залишків називається часовим лагом. Якщо існує кореляційна залежність між залишками u_n і u_{n-1}, то величина часового лагу дорівнює 1. Значення часового лагу визначає порядок коефіцієнту автокореляції. Дану залежність буде

характеризувати коефіцієнт автокореляції першого порядку між рядами залишків $u_1, ..., u_{n-1}$ і $u_2, ..., u_n$.

Нормальна лінійна модель регресії будується виходячи з припущення, що випадкові помилки рівняння регресії не корелюють між собою, тобто коваріація випадкових помилок будь-яких двох різних спостережень дорівнює нулю:

$$\operatorname{cov}(u_i, u_j) = M(u_i, u_j) = 0, \qquad (6.1)$$

де $i \neq j$.

Явище ***автокореляції залишків*** регресійної моделі порушує цю передумову, що призводить до необхідності усунення коваріаційної залежності між випадковими помилками моделі (взаємозв'язок кожного наступного значення залишків з попереднім).

В економічних задачах значно частіше зустрічається так звана додатна автокореляція ($M(u_i u_j) > 0$), ніж від'ємна автокореляція ($M(u_i u_j) < 0$).

Частіше всього додатна автокореляція викликається направленою взаємодією деяких неврахованих у моделі факторів. Суть автокореляції пояснюється наступним прикладом. Нехай досліджується попит Y на прохолоджуючі напої від доходу X за щомісячними даними. Трендова залежність, що відображає зростання попиту з ростом доходу, може бути представлена лінійною функцією $Y = a_0 + a_1 x$, що зображена на рисунку 6.1

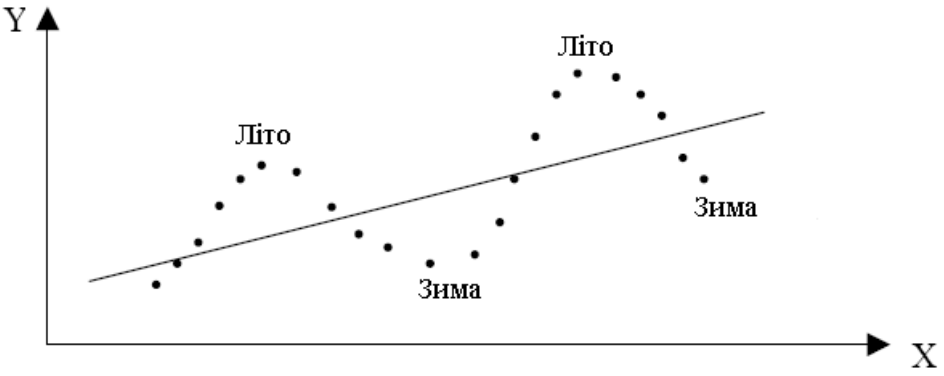


Рисунок 6.1

Однак фактичні точки спостережень зазвичай будуть перевищувати трендову лінію в літні періоди і будуть нижче неї в зимові.

Аналогічна картина має місце в макроекономічному аналізі з урахуванням циклів ділової активності.

Від'ємна автокореляція фактично означає, що за додатним відхиленням має місце від'ємне і навпаки. Можлива схема розсіювання точок в цьому випадку представлена на рисунку 6.2.

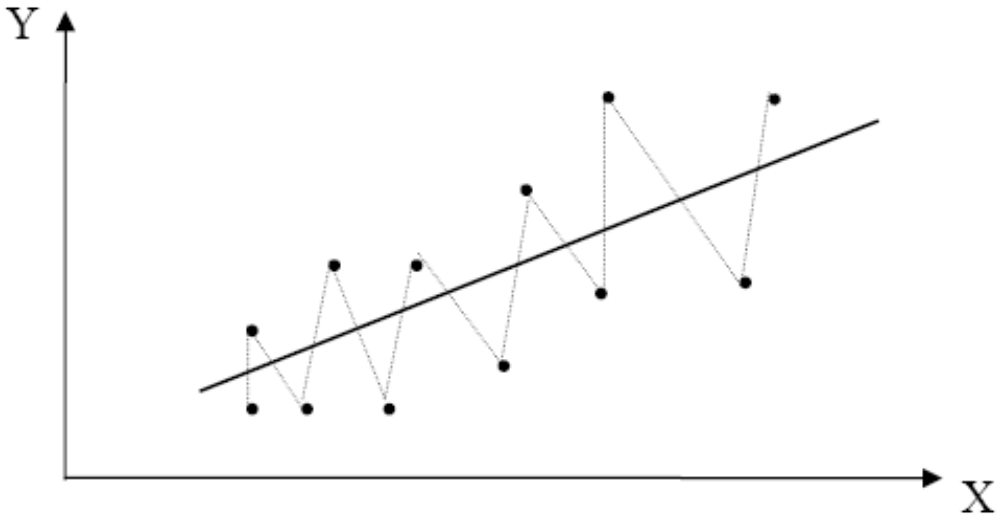


Рисунок 6.2

Така ситуація може мати місце, коли розглядати ту ж залежність між попитом на прохолоджуючі напої і доходами за сезонними даними (зима-літо).

Серед основних причин, що викликають появу автокореляції, можна виділити:

1) помилки специфікації;
2) інерцію в змінах економічних показників;
3) ефект павутини;

4) згладжування даних.

Помилки специфікації. Неврахування в моделі будь-якої важливої пояснюючої змінної чи неправильний вибір форми залежності зазвичай приводить до системних відхилень точок спостережень від лінії регресії, що може призвести до автокореляції.

Інерція. Багато економічних показників (наприклад, інфляція, безробіття, ВНП тощо) мають певну циклічність, що пов'язана з хвилеподібною діловою активністю. Дійсно, економічний підйом призводить до росту зайнятості, скороченню інфляції, збільшенню ВВП тощо. Цей ріст продовжується до тих пір, доки зміни кон'юнктури ринку і ряду економічних характеристик не призведуть до уповільнення росту, потім до зупинки і руху в протилежну сторону зазначених показників. У будь-якому випадку ця інформація приходить не миттєво, а має певну інертність.

Ефект павутини. В багатьох сферах економіки економічні показники реагують на зміни економічних умов із запізненням (часовим лагом). Наприклад, пропозиція сільськогосподарської продукції реагує на зміни ціни із запізненням (рівним періоду дозрівання врожаю). Більша ціна сільськогосподарської продукції в минулому році викличе (скоріш всього) її надвиробництво в поточному році, і, як наслідок, ціна на неї знизиться і т.д.

Згладжування даних. Часто дані за деякий тривалий часовий період отримуються усередненням даних по його підінтервалах. Це може призвести до певного згладжування коливань, які спостерігалися в середині періоду, що, в свою чергу, може стати причиною автокореляції.

Наслідки автокореляції певною мірою схожі з наслідками гетероскедастичності. Серед них при застосуванні МНК зазвичай виділяють наступні:

1. Оцінки параметрів, залишаючись лінійними і незміщеними, перестають

бути ефективними, тобто вибіркові дисперсії вектора оцінок a можуть бути невиправдано великими. Як наслідок, вони перестають бути найкращими лінійними незміщеними оцінками (BLUE–оцінками).

2. Дисперсії оцінок являються зміщеними. Досить часто дисперсії, що розраховуються за стандартними формулами, є заниженими, що призводить до збільшення t-статистик. Це може призвести до визнання статистично значущими пояснюючих змінних, які в дійсності такими не являються.

3. Статистичні критерії t- і F-статистик, які отримані для класичної лінійної моделі, не можуть бути використані для дисперсійного аналізу, тому що їх розрахунок не враховує наявності коваріації залишків.

4. Неефективність оцінок параметрів економетричної моделі, як правило, призводить до неефективних прогнозів, тобто прогнозні значення матимуть велику вибіркову дисперсію.

За наявності автокореляції поширеним методом оцінювання невідомих параметрів є узагальнений метод найменших квадратів, який було розглянуто в попередній лекції. Отримані за допомогою УМНК оцінки є незміщеними та ефективними.

6.2 Методи визначення автокореляції. Критерій Дарбіна-Уотсона. Критерій фон Неймана

Найбільш простим і розповсюдженим методом виявлення автокореляції випадкових залишків регресійної моделі є графічний метод. Є кілька варіантів графічного визначення автокореляції. Один з них, пов’язуючий відхилення u_i з їх порядковими номерами, наведений на рисунку 6.1. Це так звані, послідовно-часові графіки. В цьому випадку по осі абсцис зазвичай відкладають або момент отримання статистичних даних, або порядковий номер спостереження, а по осі ординат відхилення.

Природно припустити, що на рисунку 6.3, *а-г* є певні зв’язки між відхиленнями, тобто автокореляція присутня. Відсутність залежності на рисунку 6.3 *д*, скоріш всього, свідчить про відсутність автокореляції.

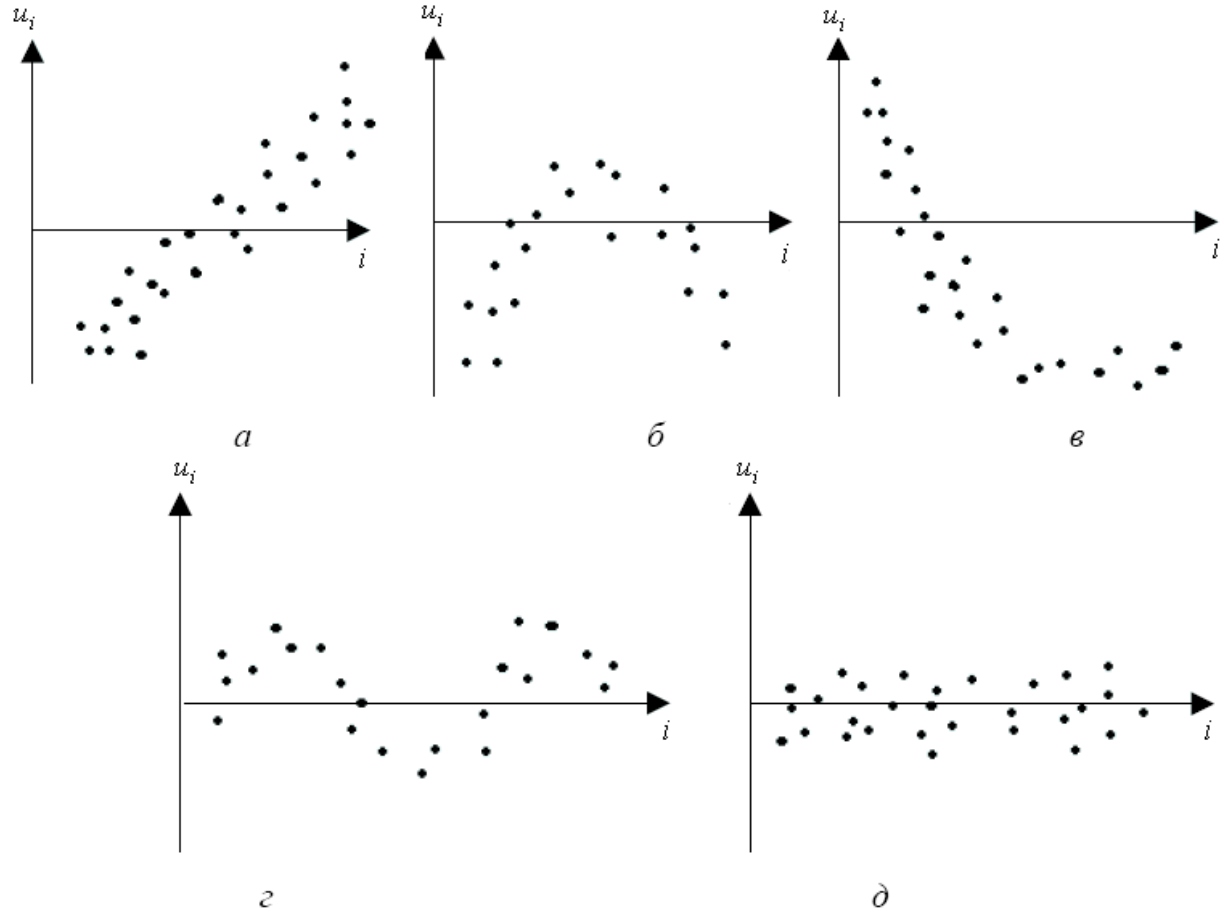


Рисунок 6.3

Наприклад, на рис. 6.3, б відхилення на початку від’ємні, потім додатні, потім знову від’ємні. Це свідчить про наявність між відхиленнями певної залежності. Більше того, можна стверджувати, що в цьому випадку має місце додатна автокореляція залишків. Вона становиться наочною, якщо графік 6.3, б доповнити графіком залежності u_i від u_{i-1}, котрий в цьому випадку орієнтовно буде виглядати так.

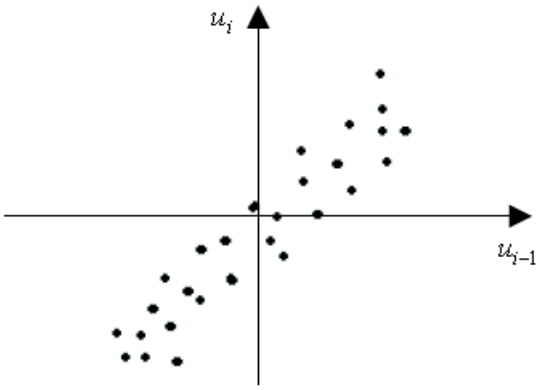


Рисунок 6.4

Переважна більшість точок на цьому графіку розташована в I та III чвертях декартової системи координат, що підтверджує додатну залежність між сусідніми відхиленнями.

Слід сказати, що в сучасних економетричних пакетах аналітичний вираз регресії супроводжується графічним представленням результатів. На графік реальних коливань залежної змінної накладається графік коливань цієї змінної, розрахованих за рівнянням регресії. Співставивши ці два графіки, можна висунути гіпотезу про наявність автокореляції залишків. Якщо ці графіки перетинаються рідко, то можна припустити наявність автокореляції залишків.

Критерій Дарбіна-Уотсона

Це простий критерій наявності автокореляції між сусідніми членами.

Тест Дарбіна-Уотсона базується на простій ідеї: якщо кореляція похибок регресії не дорівнює нулю, то вона присутня в залишках регресії u_i , що отримуються в результаті застосування 1МНК. У тесті Дарбіна-Уотсона для оцінки кореляції використовується статистика виду:

$$DW = d = \frac{\sum_{i=2}^{n}(u_i - u_{i-1})^2}{\sum_{i=1}^{n} u_i^2} \qquad (6.2)$$

Для даного критерію існує два порогових значення $d_н$ і $d_в$, що залежать від кількості спостережень, кількості регресорів і рівня значущості.

Якщо фактично спостережуване значення d:

а) $d_B < d < 4 - d_B$, то автокореляція відсутня;

б) $d_H < d < d_B$ чи $4 - d_B < d < 4 - d_H$, то питання про прийняття гіпотези залишається відкритим (область невизначеності критерію);

в) $0 < d < d_H$, то приймається альтернативна гіпотеза про додатну автокореляцію;

г) $4 - d_H < d < 4$, то приймається альтернативна гіпотеза про від'ємну автокореляцію.

Зобразимо результат Дарбіна-Уотсона графічно:



Рисунок 6.5 – Зони автокореляційного зв'язку за критерієм *DW*

Критичні значення d-статистики (нижня ($d_н$) та верхня ($d_в$) межа) можна знайти в статистичних таблицях для рівнів значущості 0,01; 0,025; 0,05.

При використанні критерію Дарбіна-Уотсона необхідно враховувати наступні обмеження:

1. Критерій *DW* використовується лише для тих моделей, які містять вільний член.

2. Припускається, що випадкові відхилення u_i визначаються пза наступною ітераційною схемою $u_i = \rho u_{i-1} + v_i$, яку називають авторегресійною схемою

першого порядку AR(1). Тут v_i – випадковий член, ρ – коефіцієнт автокореляції $|\rho| < 1$.

3. Статистичні дані повинні мати однакову періодичність (тобто не має бути пропусків у спостереженнях).

4. Критерій *DW* не можна використовувати для регресійних моделей, що містять в складі пояснюючих змінних залежну змінну з часовим лагом в один період, тобто так званих авторегресійних моделей.

Критерії фон Неймана

$$Q = Q_{факт} = \frac{\sum_{i=2}^{n}(u_i - u_{i-1})^2}{\sum_{i=1}^{n} u_i^2} \cdot \frac{n}{n-1}. \qquad (6.3)$$

Звідси $Q_{розр} = DW_{розр} \cdot \frac{n}{n-1}$. Отже, при $n \to \infty$, $Q = DW$.

Фактичне значення критерію фон Неймана порівнюється з табличним при вибраному рівні значущості α і заданій кількості спостережень: $Q_{табл} = Q_{(\alpha,n)}$.

Якщо $Q_{факт} < Q_{табл}$, то існує додатна автокореляція.

6.3 Коефіцієнти автокореляції та їх застосування

Окрім статистик Дарбіна-Уотсона та фон Неймана, для перевірки автокореляції застосовують також *нециклічний коефіцієнт автокореляції r**, який відображає ступінь взаємозв'язку рядів $u_1, u_2 ..., u_{n-1}$ і обчислюється за формулою

$$r^* = \frac{\sum_{i=2}^{n}(u_i u_{i-1}) - \frac{1}{n-1}\cdot\left(\sum_{i=2}^{n} u_i\right)\left(\sum_{i=2}^{n} u_{i-1}\right)}{\sqrt{\left[\sum_{i=2}^{n} u_i^2 - \frac{1}{n-1}\cdot\left(\sum_{i=2}^{n} u_i\right)^2\right]\left[\sum_{i=2}^{n} u_{i-1}^2 - \frac{1}{n-1}\cdot\left(\sum_{i=2}^{n} u_{i-1}\right)^2\right]}} \qquad (6.4)$$

Коефіцієнт r^* може набувати значень в інтервалі (–1;+1). Його від'ємні значення свідчать про від'ємну автокореляцію залишків, а додатні – про додатну автокореляцію. Значення, що лежать в деякій критичній області поблизу нуля, підтверджують нульову гіпотезу про відсутність автокореляції в залишках. Оскільки ймовірнісний розподіл r^* встановити важко, то на практиці замість r^* обчислюють *циклічний коефіцієнт автокореляції* r^0.

$$r^0 = \frac{\sum_{t=2}^{n} u_t\ u_{t-1} + u_n\ u_1 - \frac{1}{n}\left(\sum_{t=1}^{n} u_t\right)^2}{\sum_{t=1}^{n} u_t^2 - \frac{1}{n}\left(\sum_{t=1}^{n} u_t\right)^2}. \qquad (6.5)$$

Фактичне значення цього критерію порівнюється з табличним для вибраного рівня довіри і довжини ряду спостережень n. Якщо $r^0_{факт} \geq r^0_{табл}$, то існує автокореляція. Припускаючи, що середня залишків дорівнює нулю, тобто $\bar{u} = 0$, а отже

$$\sum_{t=2}^{n} u_t \approx \sum_{t=2}^{n} u_{t-1} \approx 0\ , \qquad (6.6)$$

циклічний коефіцієнт автокореляції можна записати так:

$$\rho = r = \frac{\sum_{t=2}^{n} u_t\ u_{t-1}}{\sum_{t=1}^{n} u_t^2}\ . \qquad (6.7)$$

причому $r \in (-1;+1)$. Значення r використовується при оцінюванні параметрів

моделі.

6.4 Моделі з автокорельованими залишками

Зазначимо, що параметри моделі з автокорельованими залишками можна оцінити на основі чотирьох методів:

1) Ейткена (УМНК);

2) перетворення вихідної інформації;

3) Кохрена-Оркатта;

4) Дарбіна.

Перші два методи доцільно застосовувати тоді, коли залишки описуються авторегресійною моделлю першого порядку

$$u_t = \rho u_{t-1} + v_t \tag{6.8}$$

Ітераційні методи Кохрена-Оркатта і Дарбіна можна застосовувати для оцінки параметрів економетричної моделі й тоді, коли залишки описуються авторегресійною моделлю вищого порядку:

$$\begin{aligned} u_t &= \rho_1 u_{t-1} + \rho_2 u_{t-2} + v_t \\ u_t &= \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + v_t \end{aligned} \tag{6.9}$$

Метод оцінювання параметрів Ейткена

Оператор оцінювання УМНК можна записати так:

$$\hat{a} = \left(X^T \Omega^{-1} X\right)^{-1} X^T \Omega^{-1} Y \tag{6.10}$$

або

$$\hat{a} = \left(X^T S^{-1} X\right)^{-1} X^T S^{-1} Y \tag{6.11}$$

де Ω^{-1} – матриця, обернена до дисперсійно-коваріаційної матриці залишків Ω;

S^{-1} – матриця, обернена до матриці $S = \sigma_u^2 \Omega$.

Оскільки в Ω коваріація залишків $\rho^s \to 0$ при $S > 2$, то матриця Ω^{-1} матиме вигляд

$$\Omega^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & 0 & 0 & \cdots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 & \cdots & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix} \tag{6.12}$$

На практиці для розрахунку ρ використовується співвідношення

$$\rho \approx r = \frac{\sum_{i=2}^{n} u_i u_{i-1}}{\sum_{t=1}^{n} u_i^2} \tag{6.13}$$

або

$$\rho \approx r = \frac{\sum_{t=2}^{n} u_t u_{t-1}}{\sum_{t=1}^{n} u_t^2} \cdot \frac{n}{n-1} \tag{6.14}$$

Метод Кохрена–Оркатта

Метод Кохрена-Оркатта є ітераційним методом наближеного пошуку оцінок параметрів моделі з автокорельованими залишками, який базується на МНК.

Алгоритм методу Кохрена-Оркатта реалізується в декілька етапів.

Крок 1. Добровільно вибрати значення параметра ρ, наприклад $\rho = r_1$. Підставивши його у

$$\sum_{t=2}^{n} u_t^2 = \sum_{t=2}^{n} \left[(y_t - \rho y_{t-1}) - a_0(1-\rho) - a_1(x_t - \rho x_{t-1}) \right]^2,$$

обчислити $\hat{a}_0^{(1)}$ і $\hat{a}_1^{(1)}$.

Крок 2. Покласти $\hat{a}_0 = \hat{a}_0^{(1)}$ і $\hat{a}_1 = \hat{a}_1^{(1)}$; підставивши їх у рівняння (6.15), обчислити $\rho = r_1$.

Крок 3. Підставивши в рівняння (6.15) значення $\rho = r_2$, знайти $\hat{a}_0^{(2)}$ і $\hat{a}_1^{(2)}$.

Крок 4. Використати $\hat{a}_0 = \hat{a}_0^{(2)}$ і $\hat{a}_1 = \hat{a}_1^{(2)}$ для мінімізації суми квадратів залишків (6.15) за невідомим параметром $\rho = r_3$. Процедуру повторювати доти, доки наступні значення параметрів $\hat{a}_0$, $\hat{a}_1$ і ρ не відрізнятимуться менш як на задану величину.

Зазначимо, що наведений метод завжди забезпечує:

– знаходження глобального оптимуму;

– порівняно добру збіжність.

ТЕМА 7. Модель Коба-Дугласа

План

7.1. Сутність та призначення функції Кобба-Дугласа..
7.2. Застосування виробничої функції Кобба-Дугласа.

7.1 Сутність та призначення функції Кобба-Дугласа.

Виникнення теорії виробничих функцій прийнято відносити до 1927 р., коли з'явилася стаття американських учених економіста П. Дугласа (P. Douglas) і математика Д. Кобба (D. Cobb) «Теорія виробництва». У цій статті, була зроблена спроба, емпіричним шляхом визначити вплив капіталу, що витрачається, і праці на об'єм продукції, що випускається, в оброблювальній промисловості США.

Як вже було сказано, виробнича функція відображає функціональний зв'язок між об'ємом ефективно використовуваних чинників виробництва (працею і майновим капіталом) і з їх допомогою випуском, що досягається, при існуючому технічному і організаційному знанні.

При субституційній виробничої функції виробництво може бути збільшене за рахунок підвищення кількісної характеристики одного з чинників, тоді як кількісна характеристика іншого чинника залишається без зміни, в іншому варіанті ж виробництво залишається без зміни при різних кількісних комбінаціях чинників праці і майнового капіталу.

На основі умовної введеної субституційних чинників виробництва можна зробити наступні два висновки щодо функціонального взаємозв'язку даних чинників:

При інших рівних збільшення одного з чинників виробництва веде до збільшення випуску – перша похідна позитивна.

Проте гранична продуктивність зростаючого чинника зменшується із збільшенням величини даного чинника – друга похідна негативна.

Рівень організаційних і технічних знань відображається у відповідних формах взаємодій чинників. У даному випадку рівень знань постійний, тобто в даних рамках передбачається відсутність технічного прогресу. Таким чином, субституційна функція виробництва може бути представлена у вигляді наступного зображення, що відображає взаємозв'язок між кількістю праці і випуском при заданій кількості майнового капіталу (Рис. 7.1):

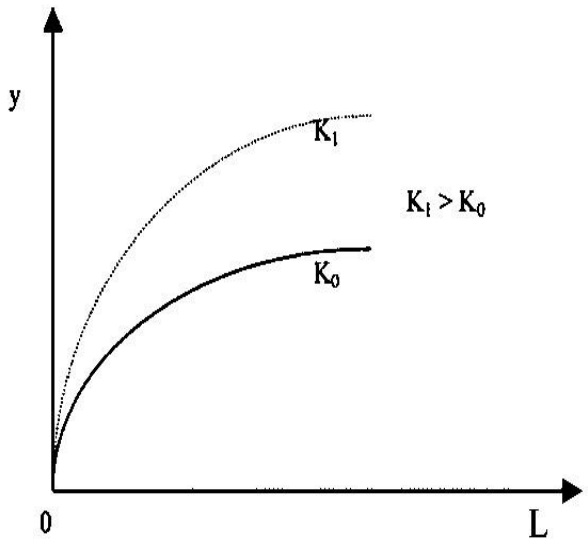


Рисунок 7.1. Зв'язок між виробництвом і виробничою працею

Кожне збільшення кількісного параметра майнового капіталу означає зсув кривої вгору і одночасного збільшення граничної продуктивності праці при заданій кількості робочої сили.

Із збільшенням кількісного параметра майнового капіталу збільшується і середня продуктивність праці, яка є часткою від ділення величини випуску на величину витраченої праці. Проте при цьому зменшується коефіцієнт праці, що визначає середню кількість витраченої праці на кожну одиницю випуску і є таким чином зворотною величиною середньої продуктивності праці.

Величина майнового капіталу приймається в рамках даного короткочасного аналізу як екзогенна задана, тому в моделі і описі не враховується технічний прогрес, а також ефект збільшення виробничих потужностей за рахунок інвестицій.

У 1927 р. П. Дуглас виявив, що якщо сумістити графіки залежності від часу логарифмів показників реального об'єму випуску (у), капітальних витрат (К) і витрат праці (L), то відстані від крапок графіка показників випуску до крапок графіків показників витрат праці і капіталу складатимуть постійну пропорцію. Потім він звернувся до Чарльза Кобба з проханням знайти математичну залежність, що володіє такою особливістю.

Існує ряд проблем по застосуванню такої функції, особливо в тих випадках, коли вона використовується для економіки в цілому. Зокрема, навіть в тих випадках, коли між випуском продукції, виробничим устаткуванням і працею у виробничому процесі існує технологічна залежність, то абсолютно необов'язково, що подібна залежність існує тоді, коли вказані чинники комбінуються в масштабах економіки в цілому. По-друге, навіть якщо така залежність для економіки в цілому існує, то немає ніяких підстав вважати, що вона матиме просту форму.

Отже, будь-яка виробнича система характеризується залежністю між кількістю виробленої в ній продукції та спожитими для цього ресурсами. Причому певні показники цієї залежності мають деякі випадкові коливання. Залежність між ними, формалізовану у відповідний спосіб у вигляді регресійного рівняння, називають виробничою функцією.

Якщо виробнича функція відома, то за кількістю спожитих системою ресурсів можна передбачити кількість виробленої продукції і, навпаки, за заданою кількістю виробленої продукції можна розрахувати необхідну кількість відповідних ресурсів.

У реальних системах неможливо врахувати всі можливі фактори, що впливають на обсяги продукції. Тому розглядають найвизначніші з них і на підставі спостережень за цими факторами та результатом виробничої діяльності будують так звану емпіричну виробничу функцію.

Отже, виробнича функція - це економетрична модель, яка кількісно описує зв'язок основних результативних показників виробничо-господарської діяльності з факторами, що визначають ці показники.

Виробничі функції можуть мати різні галузі застосування, оскільки принцип "витрати — випуск", покладений в основу залежності, може бути реалізований як на мікроекономічному, так і на макроекономічному рівні.

7.2 Застосування виробничої функції Кобба-Дугласа

У 1928 р. ними було запропоновано степеневу виробничу функцію такого вигляду:

$$Q = A \cdot L^{\alpha} K^{1-\alpha},$$

де Q — обчислений або очікуваний індекс виробництва продукції обробної промисловості за деякий характерний інтервал часу;
L — індекс зайнятості в обробній промисловості;
K — індекс постійного капіталу;
A, α — позитивне постійне число, що характеризує технологію виробництва.

Використовуючи статистичні дані за зазначений період, Ч. Кобб і П. Дуглас одержали таку виробничу функцію для обробної промисловості США:

$$Q = 1{,}01 L^{0{,}75} K^{0{,}25} \quad \left(R^2 = 0{,}9402\right).$$

Тут R2 — коефіцієнт множинної детермінації, що показує, яка частина змін залежної змінної (у даному випадку — Q) обумовлена змінами незалежних змінних (L, K). Його значення говорить про те, що зміни Q на 94 % обумовлені змінами L і K. Зауважимо, що значення показників ступеня за незалежних змінних у сумі дорівнює одиниці.

Щоб оцінити внесок кожного фактора в кінцевий продукт, Ч. Кобб і П. Дуглас за допомогою диференціального числення, використовуючи часткові похідні, розрахували

$$MP_L = \frac{\partial Q}{\partial L} = (0{,}75)(1{,}01)L^{-0{,}25}K^{0{,}25} = 0{,}7575L^{-0{,}25}K^{0{,}25};$$

$$MP_K = \frac{\partial Q}{\partial K} = (0{,}25)(1{,}01)L^{0{,}75}K^{-0{,}75} = 0{,}2525L^{0{,}75}K^{-0{,}75}.$$

Середні продукти праці та капіталу були визначені за загальноприйнятими формулами:

$$AP_L = \frac{Q}{L} = \frac{1{,}01L^{0{,}75}K^{0{,}25}}{K} = 1{,}01L^{-0{,}25}K^{0{,}25};$$

$$AP_K = \frac{Q}{K} = \frac{1{,}01L^{0{,}75}K^{0{,}25}}{K} = 1{,}01L^{0{,}75}K^{-0{,}75}.$$

Еластичності випуску за кожним змінним фактором становили:

$$E_{QL} = \frac{MR}{AP_L} = \frac{(0{,}75)(1{,}01)L^{-0{,}25}K^{0{,}25}}{1{,}01L^{-0{,}25}K^{0{,}25}} = 0{,}75;$$

$$E_{QK} = \frac{MP_K}{AP_K} = \frac{(0{,}75)(1{,}01)L^{0{,}75}K^{-0{,}75}}{1{,}01L^{0{,}75}K^{-0{,}75}} = 0{,}25.$$

граничний продукт праці та капіталу:

Отже, у функції (3.15), отриманій Коббом і Дугласом, збільшення обсягу трудовипрат на 1 % веде до збільшення випуску на 0,75 %, а збільшення капіталу на 1 % збільшує випуск на 0,25 %. Оскільки значення EQ,L та EQ,K менше одиниці, випуск продукції відносно нееластичний і за працею, і за капіталом.

Поширення та модифікації. Після опублікування основної праці Ч. Кобба і П. Дугласа численними дослідниками були отримані виробничі функції аналогічного типу для різних видів виробництв, секторів економіки, географічних районів США і навіть для різних країн.

Таблиця 3.3

ОЦІНКА КОЕФІЦІЄНТІВ ФУНКЦІЇ КОББА—ДУГЛАСА [11, т. 1, с. 228]

Галузь промисловості (країна)	Праця, a_1	Капітал, a_2	Сировина, a_3	$a_1 + a_2 + a_3$
Продукти харчування (США)	0,72	0,35	—	1,07
Метали і верстати (США)	0,71	0,26	—	0,97
Бавовна (Індія)	0,84	0,14	—	0,98
Вугілля (Великобританія)	0,51	0,49	—	1,00
Зернові (США, штат Монтана)	0,04	0,50	0,50	1,04
Зернові (США, штат Айова)	0,09	0,15	0,97	1,21
Свійські тварини (США, штат Монтана)	0,08	0,94	—	1,02

Оцінки виробничих функцій, наведені в табл. 3.3, отримано на основі узагальнення

$$Q = AL^{a_1} K^{a_2} M^{a_3}, \qquad (3.14)$$

де Q — обсяг випуску;
L і K — обсяги використаної праці і капіталу;
M — обсяг застосовуваних у виробництві сировини і матеріалів;
A, a_1, a_2, a_3 — коефіцієнти, які характеризують технологію виробництва.

емпіричних даних з використанням функції Кобба—Дугласа такого вигляду:

У другій половині XX ст. степеневі функції почали застосовувати не тільки для аналізу виробничих процесів із відчутними результатами, а й для аналізу різних явищ суспільно-політичного життя. Так, у 1975 р. було розроблено степеневу виробничу функцію для поліції Лос-Анджелеса, що пов'язує кількість арештів за кримінальні злочини з такими змінними факторами, як кількість груп поліцейських-мотоциклістів, кількість поліцейських усередині помешкань, кількість осіб, які недавно звільнені і проживають у місті тощо [24, с. 366].

Виробничу функцію Кобба—Дугласа типу (3.13) часто подають у логарифмічному вигляді:

$$lgY = lgA + a lgL + (1-a) lgK$$

У СРСР Б. М. Михайлевським і Ю. П. Соловйовим на основі опрацювання статистичних даних за 1951—1963 рр. по народному господарству в цілому було отримано таку виробничу функцію:

$$lgYt = -0.524 + 1.11(0.541 lgLt + 0.334 lgKt + 0.1251 lgRt + 0.00518$$

де Yt — отриманий національний дохід;

Lt — кількість відпрацьованих людино-годин;

Kt — вартість капітальних благ;

Rt — оцінка сільськогосподарських земель;

t — рік.

Приклади виробничих функцій розглянутого типу, отриманих для галузей, видів виробництв, підприємств, можна продовжити. Дослідниками було встановлено, що сума коефіцієнтів при змінних може дорівнювати одиниці, але може бути і більше, і менше одиниці; еластичність випуску за змінними факторами була пов'язана з еластичністю масштабу; для багатьох виробництв були визначені мінімально ефективні розміри випуску і т. д. Нами розглядатимуться ці питання в наступних темах.

На теперішній час підходи П. Дугласа і Ч. Кобба розвиваються за такими напрямами:

урахування невизначеності;

урахування лагів запізнення випуску стосовно часу здійснення витрат;

повніше відображення технічного прогресу та якості праці;

поєднання кореляційного (кількісного) аналізу з дисперсійним (якісним) для повнішого врахування впливу факторів на ефективність виробництва тощо.

Ключові положення:

Ізокванта — геометричне місце точок у просторі факторів виробництва, для яких характерний однаковий обсяг випуску продукції.

Ізоквантна варіація показує, як можна комбінувати фактори виробництва, щоб забезпечити визначений обсяг випуску продукції.

Ізокванти можуть бути побудовані на основі даних виробничої сітки або шляхом розрізу "пагорба випуску" горизонтальною площиною, рівнобіжною площині факторів виробництва OKL.

Сукупність ізоквант, кожна з яких відбиває максимально можливий випуск за різних комбінацій факторів виробництва, називається картою ізоквант.

Ізокванти мають такі властивості: вони не перетинаються, опуклі до початку координат; мають негативний нахил; чим далі розміщені від початку координат, тим більший обсяг випуску характеризують.

Виробничі фактори мають властивості заміщеності і доповнюваності. Показником

заміщення виступає MRTS — гранична норма технологічного заміщення одного фактора іншим — та кількість ресурсу і-го виду, що може бути вивільнена в обмін на збільшення застосування ресурсу j-го виду на одиницю так, щоб загальний обсяг продукції залишився незмінним. Кількісні значення MRTS варіюють від 0 (відсутність взаємозамінності факторів) до ∞ (абсолютна, ідеальна взаємозамінність факторів).

Доповнюваність факторів вимірюють коефіцієнтом доповнюваності, що його кількісні значення, як і MRTS, варіюють від 0 до ∞. При цьому зі збільшенням доповнюваності рівень заміщеності знижується, і навпаки.

Ізокванти можуть мати різну конфігурацію, яка визначається особливостями технологічного процесу і, насамперед, ступенем заміщеності та доповнюваності виробничих факторів. У мікроекономічній теорії виробництва найчастіше фігурують лінійна і неперервна ізокванти.

Геометричне місце точок у просторі факторів виробництва, для яких характерний однаковий розмір граничної норми технічного заміщення між факторами, називається ізокліной. Ізокліни можуть бути як прямими, так і кривими, як з позитивним, так і з негативним нахилом до осі абсцис.

Ізокліни у вигляді променів, що виходять із початку координат, характеризують для однорідних виробничих функцій технічно можливі шляхи збільшення обсягів випуску продукції, тобто переходу з нижчої ізокванти на вищу.

Еластичність заміщення факторів виробництва показує, на скільки відсотків (за інших рівних умов) зміниться співвідношення витрат двох факторів, якщо гранична норма їх заміщення (або співвідношення їх граничних продуктів, або співвідношення їх цін) зміниться на 1 %.

Виробнича функція Кобба—Дугласа зв'язує випуск (або іншій кінцевий результат) із розміром виробничих факторів у вигляді добутку обсягів застосування факторів із визначеними степеневими коефіцієнтами. Дослідження показали, що багато явищ виробничого і соціального життя суспільства досить добре апроксимуються залежностями такого типу.

Тема 8. Вибіркові розділи прикладної економетрики

План

8.1. Моделі одновимірних часових рядів
8.2. Поняття часового ряду та його основні компоненти
8.3. Побудова адитивної моделі
8.4. Побудова мультиплікативної моделі

8.1. Моделі одновимірних часових рядів

Моделі часових рядів відіграють вирішальну роль у вилученні значущої інформації з послідовних даних, дозволяючи нам розрізняти закономірності, тенденції та циклічну поведінку, властиву різноманітним сферам.

Основні компоненти часового ряду можуть включати: часовий ряд можна розбити на три основні компоненти: тренд, сезонність і залишки, кожен з яких надає окрему інформацію про базову динаміку даних.

Аналіз трендів це розуміння та точна оцінка компонента тренду є фундаментальним, оскільки він показує довгостроковий рух або напрямок у середніх значеннях часового ряду, допомагаючи у прогнозуванні та прийнятті рішень.

Вплив сезонності, сезонність, що представляє короткострокові циклічні коливання, додає рівень складності моделям часових рядів. Ідентифікація та моделювання цих шаблонів має важливе значення для фіксації регулярної, повторюваної поведінки в даних. Залишки, або нерегулярності, підкреслюють незрозумілу мінливість у часовому ряді. Аналіз залишків має вирішальне значення для оцінки адекватності моделі та розуміння частин даних, які не враховуються трендом і сезонністю.

Адитивні моделі розкладають часовий ряд на компоненти шляхом підсумовування тренду, сезонності та залишкових умов. Ця структура сприяє більш чіткому розумінню індивідуальних внесків у загальну модель даних.

Етапи побудови адитивної моделі: побудова адитивної моделі передбачає оцінку тенденції, виявлення сезонності та аналіз залишків. Кожен крок сприяє вдосконаленню точності моделі в представленні спостережуваних часових рядів.

Структура мультиплікативної моделі представляють альтернативний підхід, виражаючи дані спостереження як добуток його компонентів. Ця структура особливо ефективна, коли вплив сезонності змінюється пропорційно рівню ряду.

Мультиплікативні моделі добре підходять для даних часових рядів зі змінною мінливістю з часом. Їх адаптивність до різноманітних умов робить їх цінними для вловлювання динамічної природи певних явищ.

Вибір між адитивною та мультиплікативною моделями залежить від характеристик наявних часових рядів даних. Розуміння природи даних має вирішальне значення для вибору моделі, яка забезпечує найбільш точне представлення.

Комплексний аналіз при побудові як адитивних, так і мультиплікативних моделей забезпечує повне розуміння динаміки часових рядів, дозволяючи аналітикам досліджувати різні аспекти даних і вибирати моделі, які відповідають конкретним характеристикам часових рядів.

Практичні наслідки використання моделі часових рядів мають практичне значення в різних сферах, включаючи фінанси, економіку, метеорологію та охорону здоров’я, де точні прогнози та обґрунтоване прийняття рішень є найважливішими для успіху.

8.2. Поняття часового ряду та його основні компоненти

Часовий ряд – це послідовне розташування точок даних, зібраних або записаних протягом послідовних інтервалів, що забезпечує хронологічну перспективу еволюції явища чи змінної. Поняття тимчасової структури,

притаманне даним часових рядів, дозволяє ідентифікувати закономірності, тенденції та коливання, які можуть бути неочевидними в перехресних даних, що покращує нашу здатність розуміти динамічні процеси.

Дані часових рядів є універсальними, знаходячи застосування в безлічі галузей, включаючи економіку, фінанси, метеорологію, охорону здоров'я та соціальні науки, що відображає їх адаптивність до різноманітних областей. Основною характеристикою даних часових рядів є їхня часова залежність, де на значення кожного спостереження впливають його попередники, що підкреслює послідовний характер набору даних.

Компоненти часового ряду: аналіз часових рядів включає розбиття даних на три основні компоненти: тренд, сезонність і залишки, кожен з яких надає окрему інформацію про основну динаміку спостережуваного явища. Компонент тенденції представляє довгостроковий рух або напрям у середніх значеннях часового ряду, надаючи розуміння загальної траєкторії явища протягом тривалого періоду. Сезонність створює короткострокові циклічні моделі в часових рядах, що відображають повторювані коливання, які повторюються через регулярні проміжки часу, наприклад щодня, щомісяця або щороку.

Залишки як незрозуміла мінливість: залишки фіксують незрозумілу мінливість або випадкові коливання в даних часових рядів, які не можна віднести до тенденції або сезонності, служачи мірою непередбачуваності, притаманної спостережуваному явищу.

Точний аналіз тенденцій необхідний для розуміння загального напрямку часових рядів, допомоги в прогнозуванні майбутніх значень і створення основи для прийняття рішень у різних сферах.

Сезонність і циклічна поведінка: виявлення та моделювання сезонності має вирішальне значення для розпізнавання циклічної поведінки в даних, дозволяючи аналітикам передбачати регулярні, повторювані закономірності та їхній потенційний вплив на спостережуване явище.

Цілісне розуміння через аналіз компонентів: розкладання часових рядів на основні компоненти дозволяє отримати більш цілісне розуміння даних, дозволяючи аналітикам роз'єднати різні впливи, що формують спостережувані моделі. Концепція часового ряду та його основні компоненти служать основою для побудови ефективних моделей часових рядів, забезпечуючи структурований підхід до аналізу та інтерпретації послідовних даних.

8.3. Побудова адитивної моделі

Адитивна модель служить потужним інструментом для аналізу часових рядів, декомпозуючи спостережувані дані на окремі компоненти — тренд, сезонність і залишки — забезпечуючи чітку структуру, яку можна інтерпретувати.

Математично виражається як $Y(t) = T(t) + S(t) + \varepsilon(t)$, де ($Y(t)$ — спостережуване значення в момент часу (t), $T(t)$ — тренд, $S(t)$ — сезонність, а $\varepsilon(t)$ — залишковий компонент.

Процес побудови адитивної моделі починається з оцінки тенденції, де такі методи, як ковзні середні або експоненціальне згладжування, використовуються для фіксації довгострокового руху або напрямку в часовому ряді.

Ідентифікація сезонності адитивної моделі вимагає ідентифікації та оцінки сезонних закономірностей у даних. Розпізнавання циклічної поведінки, що відбувається через фіксовані проміжки часу, є важливим для точної побудови моделі.

Вивчення залишків є критичним кроком у побудові адитивної моделі. Залишки представляють незрозумілу мінливість у даних і служать засобом перевірки ефективності моделі в охопленні спостережуваних закономірностей.

Адитивна модель покращує інтерпретативність, розбиваючи часовий ряд на компоненти. Це дозволяє аналітикам зрозуміти індивідуальний внесок тенденції, сезонності та залишків у загальне спостережуване явище. Адитивні моделі знаходять застосування в різних сферах, таких як фінанси, економіка та наука про навколишнє середовище, надаючи цінну інформацію про тенденції та сезонні коливання, які впливають на прийняття рішень.

Структура адитивної моделі пропонує гнучкість, що робить її адаптованою до часових рядів із різними характеристиками. Ця адаптивність є перевагою при роботі з даними різного рівня складності. Поетапна побудова адитивної моделі передбачає послідовне оцінювання тенденції, визначення сезонності та аналіз залишків. Такий системний підхід сприяє підвищенню точності моделі.

Адитивні моделі, якщо вони точно побудовані, сприяють підвищенню точності прогнозування. Враховуючи як довгострокові тенденції, так і короткострокові циклічні моделі, ці моделі забезпечують повне розуміння даних. Аналіз залишків у побудові адитивної моделі служить засобом перевірки припущень, що лежать в основі моделі. Розуміння непоясненої мінливості допомагає вдосконалити модель для більш точних прогнозів.

Побудова адитивної моделі – це не одноразовий процес, а скоріше ітеративний. Постійне вдосконалення та вдосконалення на основі спостережуваних закономірностей і продуктивності моделі сприяють його надійності у відловленні тимчасової динаміки.

8.4. Побудова мультиплікативної моделі

Мультиплікативна модель представляє складний підхід до аналізу часових рядів, виражаючи спостережувані дані як добуток його компонентів — тенденції, сезонності та залишків, — забезпечуючи гнучку та динамічну структуру.

Математичний вираз для мультиплікативної моделі такий: $(Y(t) = T(t) \times S(t) \times \varepsilon(t)$, де $Y(t)$ — спостережене значення при часі (t), $T(t)$ — тренд, $S(t)$ — сезонність, а $\varepsilon(t)$ — залишковий компонент.

Подібно до адитивної моделі, побудова мультиплікативної моделі починається з оцінки тенденції. Виявлення та кількісна оцінка довгострокового руху в часовому ряді має вирішальне значення для визначення загального напрямку спостережуваного явища.

Мультиплікативні моделі включають ідентифікацію та моделювання сезонних моделей. На відміну від адитивних моделей, сезонність множиться, підкреслюючи її пропорційний вплив на загальний ряд, що робить її придатною для різних рівнів даних. Аналіз залишків у мультиплікативних моделях передбачає перевірку добутку залишків. Цей підхід розглядає залишковий компонент як фактор, а не суму, що відображає мультиплікативний характер моделі. Мультиплікативні моделі найкращі в ситуаціях, коли мінливість сезонних коливань змінюється з рівнем ряду. Ця адаптивність робить їх особливо корисними при роботі з даними часових рядів, що демонструють динамічну мінливість.

Цілісне розуміння взаємодій відбувається у випадку: це коли побудова мультиплікативної моделі забезпечує цілісне розуміння взаємодії між трендом, сезонністю та залишками. Множення компонентів фіксує складні взаємозв’язки, пропонуючи детальне уявлення про базову динаміку часових рядів.

Мультиплікативні моделі знаходять застосування в різних сферах, включаючи фінанси, економіку та охорону здоров'я. Їхня здатність фіксувати різну величину сезонних впливів робить їх універсальними для розміщення різних характеристик даних часових рядів. Побудова мультиплікативної моделі передбачає оцінку тренду, визначення сезонних закономірностей та аналіз залишків. Ці послідовні кроки сприяють точності моделі в представленні спостережуваних часових рядів.

Побудова мультиплікативної моделі вимагає ретельного розгляду специфічних характеристик даних часових рядів. Розуміння природи даних має вирішальне значення для вибору найбільш прийнятної моделі для підвищення точності прогнозування.

Мультиплікативні моделі чудово фіксують динамічні явища, забезпечуючи гнучку структуру, яка пристосовується до мінливих умов. Це робить їх цінними в сценаріях, де базові шаблони демонструють часову мінливість. Подібно до адитивних моделей, побудова мультиплікативної моделі є ітераційним процесом. Безперервне вдосконалення на основі спостережуваних закономірностей і оцінки продуктивності забезпечує оптимальну адаптацію моделі до еволюції даних часових рядів.

Тема 9. Економетричний аналіз виробничих функцій, інтерпретація результатів

План

9.1. Розуміння ролі виробничих функцій в економіці
9.2. Види виробничих функцій і особливості їх використання
9.3. Інтерпретація результатів виробничих функцій

9.1. Розуміння ролі виробничих функцій в економіці

Виробничі функції є фундаментальними інструментами, які використовують економісти для аналізу та розуміння того, як вхідні ресурси перетворюються на результати в гонитві за економічним зростанням і ефективністю. Буде розглянуто визначення, компоненти та застосування виробничих функцій в контексті економіки.

Визначення виробничої функції: виробнича функція — це математична залежність, яка виражає максимальну кількість продукції, яку можна отримати від даного набору вхідних ресурсів. Простіше кажучи, він описує, як різні

ресурси, такі як праця, капітал і технології, поєднуються для виробництва товарів і послуг. Загальна форма виробничої функції така: Q = f(L, K, T), де Q представляє кількість продукції, L означає затрати праці, K означає вкладення капіталу, а T означає технологію.

Компоненти виробничих функцій:

Праця (L): Представляє людські зусилля та навички, задіяні у виробничому процесі. Збільшення затрат праці часто призводить до підвищення продуктивності, але віддача може зменшуватися, оскільки кількість робочої сили стає занадто великою для наявних ресурсів.

Капітал (K): Охоплює фізичні та фінансові активи, такі як машини, будівлі та інструменти. Капітал підвищує продуктивність, дозволяючи ефективно використовувати інші ресурси.

Технологія (T): Відноситься до знань, методів та інновацій, застосованих у виробничому процесі. Технологічний прогрес може призвести до підвищення продуктивності, зміщуючи виробничу функцію вгору.

Застосування виробничих функцій в економіці:

Аналіз ефективності - виробничі функції допомагають оцінити ефективність використання ресурсів в економіці. Економісти можуть визначити оптимальну комбінацію ресурсів для досягнення максимального результату та мінімізації відходів.

Формулювання політики - уряди та політики використовують виробничі функції для розробки та впровадження економічної політики. Наприклад, розуміння впливу податкових пільг на капіталовкладення та участь у робочій силі.

Прогнозування та планування - підприємства використовують виробничі функції для прогнозування майбутнього виробництва на основі очікуваних змін у вхідних ресурсах. Допомагає в стратегічному плануванні розподілу ресурсів і розширення.

Міжнародна торгівля - виробничі функції відіграють певну роль у поясненні порівняльних переваг, впливаючи на структуру торгівлі між країнами. Країни можуть спеціалізуватися на виробництві товарів, де вони мають порівняльну перевагу.

Виробничі функції служать безцінним інструментом у вивченні та управлінні економічними процесами. Розуміючи, як вхідні ресурси поєднуються для створення результатів, економісти, підприємства та політики можуть приймати обґрунтовані рішення щодо підвищення продуктивності, економічного зростання та покращення загального добробуту.

9.2. Види виробничих функцій і особливості їх використання

Розглянемо 10 різних типів виробничих функцій, які економісти використовують для моделювання та аналізу виробничих процесів.

1. Лінійна виробнича функція:

Визначення: $Q = aL + bK + c$

Характеристики - прямо пропорційне співвідношення між витратами та виходами, що передбачає постійну віддачу від масштабу.

Використання - підходить для ситуацій, коли існує прямий, пропорційний зв’язок між входами та виходами. Часто використовується у вступних економічних моделях через свою простоту.

Застосування - допомагає зрозуміти базові сценарії виробництва, де подвоєння ресурсів призводить до пропорційного збільшення випуску. Обмеження в охопленні складності реальних виробничих процесів.

2. Виробнича функція Кобба-Дугласа:

Визначення: $Q = AL^{\alpha} \times K^{\beta}$

Характеристики - нелінійна функція, що дозволяє отримати різний ступінь віддачі від масштабу, широко використовується через свою гнучкість.

Використання - широко використовується для моделювання виробничих процесів у різних галузях промисловості. Пропонує гнучкість, дозволяючи нелінійні зв'язки між входами та виходами.

Застосування - корисно для аналізу впливу змін праці та капіталу на загальне виробництво. Часто застосовується в теорії зростання та емпіричних дослідженнях для оцінки еластичності виробництва.

3. Виробнича функція CES (постійна еластичність заміщення):

Визначення: $Q = (aL^{\rho} + bK^{\rho})^{((1/\rho))}$

Характеристики - узагальнена форма, яка враховує як взаємодоповнюваність, так і замінність між входами.

Використання - підходить для фіксації ситуацій, коли вхідні дані можуть бути замінниками або доповненнями. Забезпечує більш гнучку структуру, ніж функція Кобба-Дугласа.

Застосування - корисно для галузей промисловості, де є суміш додаткових і замінних ресурсів. Включає параметри еластичності для кількісного визначення ступеня заміни між вхідними ресурсами.

4. Виробнича функція Леонтьєва:

Визначення: $Q = \min(AL, BK)$

Характеристики - фіксовані пропорції, припускаючи, що ресурси використовуються у фіксованих співвідношеннях, що представляє певні галузі з жорсткими співвідношеннями витрат.

Використання - підходить для галузей, де сировина використовується у фіксованих пропорціях. Представляє спрощене уявлення про виробничі процеси зі строгими коефіцієнтами витрат.

Застосування - зазвичай використовується в моделях витрати-випуск для аналізу взаємозалежностей між різними секторами економіки. Корисно для розуміння галузей із жорсткими вхідними відносинами, таких як сільське господарство.

5. Функція виробництва змінних пропорцій:

Визначення - нелінійні виробничі функції, які враховують зміну вхідних пропорцій із різними рівнями випуску.

Характеристики - відображає динамічний характер реальних виробничих процесів.

Використання - застосовується, коли вхідна суміш змінюється разом із варіаціями рівня виходу. Представляє більш реалістичний погляд на виробничі процеси.

Застосування - корисно для моделювання галузей, де технологія або методи виробництва можуть адаптуватися до змін у вхідних кількостях. Передає динамічний характер багатьох сценаріїв реального виробництва.

6. Функція трансцендентного виробництва:

Визначення: $Q = aL^b \times K^c$

Характеристики - використовує трансцендентні функції, що забезпечує більшу гнучкість у моделюванні складних виробничих відносин.

Використання - забезпечує більшу гнучкість у моделюванні складних взаємозв'язків між входами та виходами. Дозволяє включати різні математичні функції.

Застосування - застосовується в сценаріях, де традиційні функціональні форми можуть неадекватно відображати тонкощі виробничого процесу. Корисно для зйомки унікальних виробничих технологій або процесів.

7. Квадратична виробнича функція:

Визначення: $Q = aL + bK + cL^2 + dK^2$

Характеристики - вводить квадратичні терміни, фіксуючи нелінійні зв’язки між входами та виходами.

Використання - вводить нелінійні терміни для охоплення більш складних зв’язків між входами та виходами. Дозволяє розглянути можливість зменшення або збільшення віддачі від масштабу.

Застосування - корисно для галузей, де гранична продуктивність ресурсів змінюється нелінійно з їх кількістю. Застосовується в ситуаціях, коли традиційні лінійні функції або функції Кобба-Дугласа не працюють.

8. Узагальнена виробнича функція Леонтьєва:

Визначення: $Q = \min(aL, bK) + cL + dK$

Характеристики - розширення моделі Леонтьєва з додаванням додаткових лінійних термінів для підвищення гнучкості.

Використання - розширює основну модель Леонтьєва шляхом введення додаткових лінійних термінів для підвищення гнучкості. Дозволяє більш деталізоване представлення вхідних зв'язків.

Застосування - корисно, коли необхідно враховувати додаткові фактори або обмеження у виробництві, окрім фіксованих коефіцієнтів витрат. Застосовується в моделях витрат-виходу з більш складними виробничими структурами.

9. Виробнича функція CES із еластичною заміною вхідних даних:

Визначення: $Q = aL^{\rho} + bK^{\rho} + c(L/K)^{\sigma}$

Характеристики - включає еластичне заміщення витрат для врахування зміни відносин між працею та капіталом.

Використання - поєднує характеристики CES з еластичним вхідним заміщенням, дозволяючи різні ступені заміщення. Підходить для ситуацій, коли зв’язок між входами є динамічним.

Застосування - корисно для галузей, де взаємозамінність праці та капіталу змінюється з рівнем виробництва. Включає параметри еластичності для кількісної оцінки гнучкості заміщення вхідних даних.

10. Функція багатопрофільного виробництва:

Визначення: $Q1 = f(L, K)$, $Q2 = g(L, K)$, ...

Характеристики - представляє ситуації, коли фірма виробляє кілька товарів або послуг одночасно, що відображає складність сучасної економіки.

Використання - застосовується, коли фірма виробляє кілька товарів або послуг одночасно. Представляє більш складний сценарій виробництва.

Застосування - корисно для галузей промисловості з різноманітними товарними лініями. Дозволяє аналізувати виробничий процес, коли є кілька виходів, кожен із власним набором вхідних даних і виробничих функцій.

Наведені 10 типів виробничих функцій відкриває багатий інструментарій, який визначає взаємозв'язки між витратами та результатами в економічних системах. Кожен тип служить унікальній меті, пропонуючи економістам і особам, які приймають рішення, цінні інструменти для аналізу та оптимізації виробничих процесів. Детальне розуміння цих виробничих функцій дасть змогу орієнтуватися в різноманітних концепціях економічної теорії та практики.

Таким чином, вибір виробничої функції залежить від конкретних характеристик галузі або економічної системи, що розглядається. Економісти та дослідники ретельно обирають найбільш відповідну функцію, виходячи з характеру ресурсів, технологій виробництва та бажаного рівня. Кожен тип виробничої функції служить цінним інструментом для розуміння динаміки виробництва та прийняття обґрунтованих рішень в економічному плануванні та політиці.

9.3. Інтерпретація результатів виробничих функцій

В попередніх підрозділах були досліджені різні типи виробничих функцій, розуміння того, як отримати значущу інформацію з результатів цих функцій, має важливе значення для осіб, які приймають рішення, економістів і всіх, хто прагне зрозуміти динаміку виробничих процесів. Тепер доцільно розглянути інтерпретацію отриманих результатів.

І. Розуміння еластичності випуску:

Визначення - еластичність випуску вимірює відсоткову зміну випуску в результаті одновідсоткової зміни конкретного входу, зберігаючи при цьому інші входи постійними.

Інтерпретація - позитивна еластичність випуску вказує на те, що збільшення витрат веде до збільшення випуску, і навпаки. Величина має значення: вища еластичність означає більш значний вплив на випуск.

II. Повернення до масштабу:

Визначення - віддача від масштабу стосується швидкості, з якою випуск змінюється у відповідь на пропорційну зміну всіх входів.

Інтерпретація - підвищення віддачі від масштабу відбувається, коли пропорційне збільшення витрат призводить до більш ніж пропорційного збільшення випуску. Постійне повернення до масштабу відбувається, коли збільшення витрат пропорційно відповідає збільшенню випуску.

Зменшення віддачі від масштабу відбувається, коли пропорційне збільшення витрат призводить до непропорційного збільшення випуску.

III. Замінність і взаємодоповнюваність:

Визначення – у таких функціях, як Кобб-Дуглас або CES, коефіцієнти представляють ступінь замінності або комплементарності між входами.

Інтерпретація - коефіцієнт більше 1 вказує на взаємозамінність, коли входи можуть взаємозамінюватися. Коефіцієнт менше 1 вказує на взаємодоповнюваність, коли входи працюють разом і їх важко замінити.

IV. Гранична продуктивність:

Визначення - гранична продуктивність ресурсу — це додатковий випуск продукції, отриманий шляхом збільшення цього ресурсу на одну одиницю, а всі інші витрати залишаються незмінними.

Інтерпретація - зменшення граничної продуктивності означає, що зі збільшенням кількості вхідних ресурсів додатковий отриманий випуск зменшується.

Від’ємна гранична продуктивність вказує на те, що подальше збільшення витрат призводить до зменшення випуску.

V. Оптимальна комбінація вхідних даних:

Визначення - оптимальна комбінація вхідних ресурсів – це суміш, яка максимізує випуск продукції за даного набору цін на вхідні ресурси.

Інтерпретація - досягнення оптимальної комбінації ресурсів передбачає збалансування граничного продукту кожного ресурсу з його вартістю.

Відхилення від оптимального поєднання можуть свідчити про неефективність або неоптимальне використання ресурсів.

VI. Еластичність заміщення:

Визначення - у таких функціях, як CES, еластичність заміщення вимірює, наскільки легко один ресурс можна замінити іншим.

Інтерпретація - вища еластичність передбачає більшу здатність замінювати вхідні ресурси.

Нижча еластичність вказує на нижчий ступінь замінності, маючи на увазі, що входи менш взаємозамінні.

Підсумовуючи, інтерпретацію результатів виробничих функцій, що є складним завданням, яке потребує глибокого розуміння економічних принципів. Завжди потрібно враховувати контекст досліджуваної галузі чи економічної системи та пам’ятати, що інтерпретація результатів не є однозначною справою.

СПИСОК РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

Базова

1. Диха М. В. Економетрія: навчальний посібник / М. В. Диха, В. С. Мороз – К. : «Центр учбової літератури», 2019. – 206 с.
2. Козьменко О.В. Економіко-математичні методи та моделі (економетрика) : навчальний посібник / О.В. Козьменко, О.В. Кузьменко. - Суми : Університетська книга, 2020. - 406 с.
3. Кузьмичов А.І. Економетрія: посібник / А.І. Кузьмичов. – Київ : Ліра, 2020. – 212 с.
4. Лугінін, О.Є. Економетрія [Текст]: навч. посіб., 2-ге вид., перероб. та доп. / О.Є. Лугінін. – К.: ЦУЛ, 2008. – 278 с.
5. Прикладна економетрика : навч. посіб. : у двох частинах. Частина 1 : [Електронне видання] / Л. С. Гур'янова, Т. С. Клебанова, С. В. Прокопович та ін. – Харків : ХНЕУ ім. С. Кузнеця, 2016. – 235 с.

Допоміжна

1. Perkhun L. A study of the effect of various types of tax receipts on general revenues of Ukraine’s regions // Scientific letters of academic society of Michal Baludansky, Volume 4, No 4/2016 – P. 25-28.
2. Perkhun L. Typology of Ukraine region on the basis of the financial independence level // Economics, management, law: socio-economic aspects of development: Collection of scientific articles. Vol 1. – Edizioni Magi, Italy, 2016. – C. 326-329.
3. Економетрія (вікіпедія) [Електронний ресурс]. – Режим доступу: https://uk.wikipedia.org/wiki/%D0%95%D0%BA%D0%BE%D0%BD%D0%BE%D0%BC%D0%B5%D1%82%D1%80%D1%96%D1%8F
4. Назаренко, О.М. Основи економетрики : підручник, 2-е вид., випр. / О.М. Назаренко. – К.: ЦНЛ, 2005. – 392с.
5. Перхун Л.П. Вплив податкової реформи в Україні на конкурентоспроможність регіонів // Детермінанти соціально-економічного розвитку підприємства: монографія. Вип 3. / за наук. ред. Строченко Н.І., Пилипенко В.В., Ковальової О.М. – Суми, Сумський національний аграрний університет, 2016, С. 278-284.
6. Присенко Г. В. Прогнозування соціально-економічних процесів : навч. посібн. / Г. В. Присенко, Є. І. Равікович. – К. : КНЕУ, 2005. – 378 с.
7. Черняк О.І., Ставицький А.В. Динамічна економетрика. – К.: КВІЦ, 2000. – 120 с.
8. Кухаренко В. М. , Перхун Л. П. , Товмаченко Н. М. Методика комплексного оцінювання якості тестів. Частина 1 // Статистика України. 2018. № 3. С 40–48.

Кухаренко В. М., Перхун Л. П., Товмаченко Н. М. Методика комплексного оцінювання якості тестів. Частина 2 // Статистика України. 2018. № 4. С. 72–79

Інформаційні ресурси

1. Офіційний сайт Державної служби статистики [Електронний ресурс]. - Режим доступу: http://www.ukrstat.gov.ua/
2. Світовий банк даних [Електронний ресурс]. - Режим доступу: https://worlddata.ai/
3. Офіційний сайт Євростату [Електронний ресурс]. - Режим доступу: https://ec.europa.eu/eurostat/home?