Міністерство освіти і науки України
ДЕРЖАВНИЙ БІОТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ
Факультет економічних відносин та фінансів
Кафедра економіки та бізнесу

О.Д. Тімченко

ЕКОНОМЕТРИКА

КУРС ЛЕКЦІЙ

для здобувачів першого (бакалаврського) рівня
вищої освіти спеціальності
051 Економіка
292 Міжнародні економічні відносини
071 Облік і оподаткування
072 Фінанси, банківська справа, страхування та фондовий ринок

Харків
2024

Міністерство освіти і науки України
ДЕРЖАВНИЙ БІОТЕХНОЛОГІЧНИЙ УНІВЕРСИТЕТ
Факультет економічних відносин та фінансів
Кафедра економіки та бізнесу

О.Д. Тімченко

ЕКОНОМЕТРИКА

КУРС ЛЕКЦІЙ

**для здобувачів першого (бакалаврського) рівня
вищої освіти спеціальності
051 Економіка
292 Міжнародні економічні відносини
071 Облік і оподаткування
072 Фінанси, банківська справа, страхування та фондовий ринок**

Затверджено рішенням науково –
методичної комісії факультету
економічних відносин та фінансів
Протокол № 6 від 23.02.2024 р.

Харків
2024

УДК 330.43(042.4)

Схвалено
на засіданні кафедри економіки та бізнесу,
протокол № 9 від 15.02.2024 р.

Рецензенти:

Т.О. Ставерська, кандидат економічних наук, доцент, завідувач кафедри фінансів, банківської справи та страхування Державного біотехнологічного університету
Н.Б. Кащена, доктор економічних наук, професор, завідувач кафедри обліку, аудиту та оподаткування Державного біотехнологічного університету

Економетрика: курс лекцій для здобувачів першого (бакалаврського) рівня вищої освіти спеціальності 051 Економіка, 292 Міжнародні економічні відносини, 071 Облік і оподаткування, 072 Фінанси, банківська справа, страхування та фондовий ринок / укладач: О.Д. Тімченко; ДБТУ.- Харків:, 2024. – 90 с.

Визначено роль і місце навчальної дисципліни, її зміст за розділами і темами, надано розгорнутий перелік рекомендованої базової та допоміжної літератури. Курс лекцій надає допомогу здобувачам в опануванні навчальної дисципліни як під керівництвом викладача, так й під час виконання самостійної роботи.

Курс лекцій призначений здобувачам першого (бакалаврського) рівня вищої освіти денної та заочної форм навчання спеціальності 051 Економіка, 292 Міжнародні економічні відносини, 071 Облік і оподаткування, 072 Фінанси, банківська справа, страхування та фондовий ринок

УДК 330.43(042.4)

Відповідальний за випуск: О.Д. Тімченко, доц. кафедри економіки та бізнесу

ЗМІСТ

ПЕРЕДМОВА

Ефективність прийнятих рішень у підприємництві, комерції, бізнесі та інших сферах діяльності залежить від того, наскільки особа, котра приймає ці рішення, використовує інформацію, що характеризує кількісний зв’язок між економічними показниками.

Економетрика - розділ економічної науки, в якому вивчаються методи кількісного вимірювання взаємозв'язків між економічними показниками.

Економетрика - одна з фундаментальних дисциплін у підготовці бакалаврів з економіки для всіх спеціальностей, побудована на основі математичних та економічних знань. Для засвоєння дисципліни потрібна ґрунтовна математична база, особливо з матричної алгебри, диференціального числення, теорії ймовірностей та математичної статистики. Важливо також мати підготовку з економічної теорії, макро- та мікроекономіки, статистики, економічного аналізу. Зауважимо, що економетрика з огляду на громіздкість обчислень та вимоги до точності результатів вивчається за комп'ютерної підтримки.

Знання, здобуті студентами під час вивчення економетрики, широко застосовуються в менеджменті, маркетингу, фінансовій справі, податковому менеджменті і т. ін.

Мета вивчення дисципліни “Економетрика” полягає в тому, щоб навчити студентів кількісно оцінювати взаємозв'язки економічних показників для різних масивів економічної інформації, вдаючись до тестування останньої стосовно відповідності її певним передумовам, а також до визначення методів кількісного вимірювання зв'язків, які доцільно застосовувати в кожному конкретному випадку згідно з особливостям і економічної інформації.

Під час вивчення цієї дисципліни студенті мають:

- опанувати методи побудови та реалізації економетричних моделей за допомогою персонального комп'ютера;
- набути практичних навичок кількісного вимірювання взаємозв'язків між економічними показниками;
- поглибити теоретичні знання в галузі математичного моделювання економічних процесів і явищ;
- здобути знання про застосування економетричних моделей в економічних дослідженнях.

Економетрика надає додаткові можливості оволодіти обчислювальною технікою, розвиває аналітичні навички та є основою економічних досліджень.

Досягненню цілей курсу підпорядкована логіка його викладання та вивчення.

Тема 1. ПРИРОДА ТА СУТНІСТЬ ЕКОНОМЕТРИКИ

1.1 Природа та сутність економетрики. Предмет і метод курсу

Процес прийняття науково обґрунтованих рішень в економіці тісно пов'язаний з визначенням кількісних співвідношень між економічними показниками. Ефективність прийнятих рішень у підприємництві, комерції, бізнесі й інших сферах діяльності залежить від того, наскільки людина, що приймає ці рішення, використовує інформацію, що характеризує кількісний зв'язок між економічними показниками.

Економетрика – це галузь економічної теорії, яка вивчає моделі економічних систем у формі, що уможливлює перевірку цих моделей на адекватність засобами математичної статистики.

Мета економетрики – здійснювати емпіричну перевірку положень економічної теорії, підтверджуючи чи відхиляючи останні. Цим економетрика відрізняється від математичної економіки, зміст якої полягає виключно у застосуванні математики, і теоретичні положення якої не обов'язково потребують емпіричного підтвердження. Економетрика є результатом синтезу економічної теорії, математичної статистики та економічної статистики.

Економетрика - одна з фундаментальних дисциплін у підготовці бакалаврів з економіки для всіх спеціальностей, яка побудована на основі математичних і економічних знань. Для засвоєння дисципліни потрібна ґрунтовна математична база, особливо матричної алгебри, диференціального числення, теорії імовірностей і математичної статистики. Важливо також мати підготовку по економічній теорії, макро- і мікроекономіки, статистики, економічного аналізу. Звідси очевидно, що економетрику студенти можуть вивчати лише тоді, коли засвоїли основні розділи математики й одержали загальноекономічні знання.

У буквальному перекладі з латинської мови ***економетрія*** означає "вимірювання економіки". Але поняття економетрії є набагато ширшим, хоча вимірювання залишається однією з її складових частин.

Економетрика – область науки, ціль якої полягає в тому, щоб представити кількісну міру економічним відносинам.

Економетрика є інструментом, що дозволяє перейти від якісного рівня аналізу до рівня, що використовує кількісні статистичні значення досліджуваних величин.

Можливості економетрики залежать не тільки від якості тих моделей, що повинні відображати закономірності економічних процесів, а і значною мірою, від якості самої економетричної технології, що сьогодні є достатньо розвиненою.

Економетрика є синтезованою дисципліною, вона поєднує в собі економічну теорію, математичну економіку, економічну і математичну статистику. Економічна теорія пропонує твердження або гіпотези, які за своєю сутності є переважно якісними. Наприклад, мікроекономічна теорія стверджує, що зниження ціни товару буде сприяти зростанню попиту на цей

товар. Але сама теорія не приводить жодного кількісного виміру взаємозв'язків цих двох показників, тобто вона не показує, наскільки зросте або зменшиться кількість товару у результаті певної зміни ціни товару. Таким чином, економічна теорія приймає без доказів обернену залежність між ціною і попитом на товар. Завдання ж економетрики полягає в обчисленні відповідних кількісних оцінок. Інакше кажучи, економетрика забезпечує кількісну сторону економічної теорії.

На відміну від чистої математичної економіки, що виражає економічну теорію в математичній формі без мети вимірювання, економетрика зацікавлена в емпіричному підтвердженні економічної теорії. Економетрист повинен використовувати математичні рівняння, запропоновані математиком - економістом, але перетворює їх у форму, найбільш придатну для емпіричного тестування.

Відмінність економетрики від економічної статистики також дуже наочна. Економічна статистика в основному стосується збирання, обробки і зображення економічних даних у формі діаграм і таблиць. У цьому складається робота економіста-статиста. Зібрані дані складають основу для роботи економетриста.

Економетрика поділяється на теоретичну та прикладну.

Теоретична економетрика розглядає методи вимірювання економічних зв'язків, визначених економетричними моделями. У цьому аспекті економетрика базується на математичній статистиці. Наприклад, одним зі способів, що найбільше часто застосовується в економетриці, є метод найменших квадратів. Використання цього методу ставить перед теоретичною економетрикою задачу: детально розглянути припущення, властивості методу найменших квадратів і т.д.

У прикладній економетриці безпосередньо використовуються методи теоретичної економетрики, наприклад, для вивчення функцій споживання, попиту та пропозиції і т.д.

Виходячи з того, що головне призначення економетрики складається в кількісному вимірюванні зв'язків в економічних системах, прикладна економетрика дозволяє:

- побудувати економічно і статистично обґрунтовані моделі розвитку різних економічних явищ або процесів, на їхній основі визначити і дослідити кількісні внутрішні і зовнішні причинно-наслідкові зв'язки між економічними показниками. За побудованими моделями розрахувати прогнози розвитку економічних показників і системи в цілому;
- виконати імітаційні розрахунки, які дозволяють провести глибокий економічний аналіз можливих варіантів розвитку економічного об'єкта, здійснити вибір ефективної економічної стратегії;
- отримати детальну розрахункову інформацію, необхідну для прийняття більш обґрунтованих рішень;
- дослідити динаміку економічного об'єкта або процесу.

До основних задач економетрики можна віднести наступне:

- побудова економетричних моделей, тобто представлення економічних моделей у математичній формі, зручної для проведення емпіричного аналізу. Дану проблему прийнято називати проблемою *специфікації*. Найчастіше вона може бути вирішена декількома способами;
- оцінка параметрів побудованої моделі, що роблять обрану модель найбільш адекватним реальним даним. Це так званий етап *параметризації*;
- перевірка якості знайдених параметрів моделі і самої моделі в цілому. Іноді цей етап аналізу називають етапом *верифікації*;
- використання побудованих моделей для пояснення поводження досліджуваних економічних показників, прогнозування і пророкування, а також для осмисленого проведення економічної політики.

? Запитання для самоперевірки знань

1. Дайте визначення предмета курсу економетрики.
2. Альтернативні підходи до визначення предмета економетрики.
3. Зв’язок економетрики з іншими дисциплінами.
4. Наведіть етапи розвитку економетрики як економічної науки.
5. Які задачі економетричного дослідження?
6. Характеристика структури економетричних досліджень.
7. Що розуміється під специфікацією моделі?

Тема 2. МАТЕМАТИЧНЕ МОДЕЛЮВАННЯ ЕКОНОМІЧНИХ ЯВИЩ І ПРОЦЕСІВ

2.1 Поняття системи. Основні характеристики економічних систем

Системний аналіз – це методологія дослідження об’єктів з метою визначення найбільш ефективних методів управління ними. Системи є сукупністю взаємозалежних об’єктів і процесів, що змінюються в часі.

Під *системою* розуміють будь-який комплекс взаємозалежних елементів, що динамічно взаємодіють.

Системи розрізняють по різними ознаками. Якщо спрямована дія призводить до певних і точно передбачених результатів, то такі системи називаються *детермінованими*. Якщо ж результат дії на систему точно передбачити неможливо, то система є *вірогідною*.

Головною властивістю економічної системи є її складність. Складність економіки визначається величезною кількістю складових цієї системи, зв’язків між ними, якісними особливостями економічних явищ і процесів. Перебіг цих процесів у кожний момент істотно залежить від їхнього попереднього стану. Динамічність і стійкість економічних процесів впливають не тільки на якісну структуру господарських зв’язків, але й відчутно змінюють їх кількісні

характеристики. У зв'язку з цим зростає кількість елементів системи, що підлягають дослідженню і визначенню.

Специфічною особливістю економічної системи є її належність до класу керованих систем. Разом з тим у ній самій можуть відбуватися процеси, які розвиваються за принципом саморегулювання. Це властивість тісна пов'язана з динамічністю і стійкістю економічних процесів. Функціонування більшості економічних систем має стохастичний характер, що визначається впливом великої кількості факторів. Це зумовлює необхідність застосування системного підходу до вивчення економіки.

Структурний аналіз передбачає кілька *етапів*. На першому експерти формулюють мету, уточнюють область дослідження. На другому етапі здійснюють первинну структуризацію системи – окреслюють межі системи, що досліджується, її зовнішнє середовище, прогнозують вплив системи на середовище і навпаки. Якщо система мало залежить від зовнішнього середовища, вона вважається *замкненою*. Система, яка залежить від зовнішнього середовища, але сама на нього впливає мало, є *відкритою*. На третьому етапі формулюють математичну (статичну) модель системи, що досліджується.

Усі ці етапи не піддаються формалізації, тому процес побудови моделі є процесом інтелектуальним та творчим.

Класифікація системи - це найпростіший акт моделювання. Як і будь-яка модель, класифікація носить цільовий характер і значною мірою є умовною. Тому можливі різноманітні класифікації систем (рис. 2.1).

Реальним економічним системам властиві наступні властивості:

- складність структури системи;
- цілісність системи;
- складність інформаційних процесів;
- безліч цілей і багатомірність критеріїв ефективності системи;
- динамічність процесів, що відбуваються в системах;
- безліч суб'єктивних факторів, що впливають на функціонування економічної системи;
- тісні зв'язки між суспільною й економічною системами.

Усі ці якості роблять економічні системи складними для вивчення, описи і моделювання. Слід зазначити, що при описі економічних систем важко обмежитися умовами детермінації, особливо при рішенні задач планування і прогнозування, тому при їхньому дослідженні необхідно використовувати стохастичні моделі.

Рис. 2.1 – Класифікація систем за способом управління

2.2 Економічна модель. Складові елементи економічної моделі

Економетричний аналіз у широкому розумінні має справа з перевіркою гіпотез про економічні системи за результатами спостережень. У зв'язку з цим до першочергових завдань у дослідженнях такого роду повинна відноситися *специфікація* (побудова) моделі, яка відповідно до наших уявлень адекватно відтворює зв'язки між змінними. Таким чином, головним моментом економетрії, а також економетричного аналізу, є *побудова* економічної моделі.

Економічна модель – це логічний опис того, що саме економічна теорія вважає особливо важливим під час дослідження даної проблеми. Зокрема, модель має форму системи рівнянь, яка характеризує взаємозалежності між більш-менш взаємодіючими змінними. Така система поєднує визначення, теоретичні припущення відносно економічної поведінки й умов рівноваги, що у сукупності дає можливість отримувати відповіді на поставлені економічні питання. Модель у змозі відповідати на такі питання, які відносяться до пояснення того або іншого явища, чи його прогнозу.

Побудова моделі – це творчий процес, який має дві основні стадії.

1 стадія – виключення тих елементів, відносно яких можливо припустити, що їхній вплив на досліджуваний показник невеликий.

2 стадія – вилучення у наявному виді всіх залежностей між економічними змінними і встановлення їх логічної структури.

Економічна модель складається зі змінних і параметрів. *Змінні* – це економічні величини, які приймають різні значення з деякої множини. *Параметри* – це постійні коефіцієнти, які не завжди відомі. Дослідник виходить з того, що вони повинні мати фіксовані значення за будь-якої ситуації, де можливе спостереження. Параметри також можна назвати "незмінними", що зв'язують змінні в рівняннях. Ці рівняння, а також параметри, формують структуру моделі: вони вказують на характер співвідношень між змінними.

Параметри класифікуються відповідно до економічної природи рівняння, в якому вони з'являються. Виділяють три типи параметрів: поведінки, технологічні і тотожності.

Наприклад, параметри споживчої функції $C = C(Y_d)$ носять характер поведінки, параметри рівняння $Y_d = Y\text{-}T$ належать до параметрів тотожності. Параметри рівняння, що зв'язують масштаби виготовленого продукту з витратами праці, капіталу і землі, можна розглядати як технологічні.

У будь-якій економічній моделі існують також розрізнення між змінними. Деякі змінні визначаються в рамках самої моделі, у той час як інші входять у модель із зовнішнього середовища.

Змінні, які визначаються поза моделлю, називаються *екзогенними* змінними. Таким чином, вони дуже схожі на параметри. Але є суттєві відмінності. Передбачається, що параметри залишаються незмінними протягом усього періоду спостереження, а екзогенні змінні безумовно повинні змінюватися за часом. Саме зміна екзогенних змінних приводить у рух модель і викликає перехід моделі до нового стану.

Інший тип змінних, які називаються *ендогенними*, визначається таким чином: їх значення одержують у результаті рішення всіх рівнянь моделі за заданими значеннями екзогенних змінних. Екзогенні змінні представляють ті сили, які характеризують зовнішнє середовище і, які у рамках даної структури моделі, впливають на ендогенні змінні.

2.3 Основні поняття економіко-математичного моделювання. Етапи проведення економетричного аналізу

Основними методами дослідження систем є методи математичного моделювання, що дозволяють скласти опис структури і поводження соціальних, економічних, технічних і інших систем. До таких методів відносяться методи теорії імовірності, математичної статистики, математичного аналізу, факторного аналізу, аналізу динамічних рядів і ін.

Особливістю математичного моделювання є те, що воно дає можливість вивчення, прогнозування і управління реальними економічними системами,

для яких фізичний (натуральний) експеримент утруднений чи економічно невигідний, а іноді і небезпечний, оскільки може привести до великих матеріальних і інших витрат.

Економетричне моделювання припускає виконання певної кількості етапів. У розгорнутому вигляді послідовність основних етапів економетричного моделювання може бути представлена наступними етапами:

Етап 1. **Окреслення задачі дослідження**. Насамперед, необхідно чітко окреслити економічні задачі дослідження, вивчити економічну теорію і різноманітні точки зору на те, які причинно-наслідкові зв'язки існують між аналізованими показниками. На цьому етапі здійснюється попередній теоретичний аналіз економічної системи, яка моделюється, і тенденцій її розвитку. Ці знання є основою для розробки специфікації рівнянь моделі, яка здійснюється на наступному етапі.

Етап 2. **Розробка специфікації моделі в загальному вигляді.** На цьому етапі обирається тип моделі, визначається список змінних, які входять у модель, їхній розподіл на екзогенні й ендогенні. Далі розробляється безпосередньо специфікація економетричних зв'язків у вигляді загальних рівнянь.

У процесі оцінювання рівнянь і перевірки статистичних гіпотез вигляд рівняння може бути змінений зі збереженням тих же регрессоров, або, навпаки, може залишитися незмінною форма рівняння, а специфікація піддана коригуванню. Проте, на даному етапі у загальному виді записується вихідний варіант специфікації кожного рівняння моделі.

Етап 3. **Формування статистичної бази** припускає збирання і дослідження статистичної інформації, необхідної для коректного оцінювання параметрів моделі, специфікованої в загальному вигляді при виконанні попереднього етапу. Під час цьому, в першу чергу, здійснюється аналіз рядів даних на економічну відповідність, що припускає перевірку на адекватність значень статистичних даних відносно економічних показників, включених у модель.

Статистичним даним, що використовуються в економетричному моделюванні, часто властиві роз'єднаність, непорівнянність, а іноді і відсутність окремих значень у числовій послідовності. Тому на етапі формування статистичної бази необхідно забезпечувати порівнянність даних за одиницями виміру, за періодичністю, за формою вираження, за типом ціни і т.п.

Етап 4. **Оцінювання параметрів рівнянь моделі.** На даному етапі, насамперед, з'ясовується, чи відповідає специфікація моделі, що розробляється, передумовам класичної моделі лінійної регресії. Попереднє уявлення про це можна одержати відразу на підставі апріорної інформації, що є в результаті виконання перших трьох етапів, а потім ці уявлення варто обов'язково перевірити за допомогою різноманітних статистичних тестів (п'ятий етап).

Переважна більшість реальних моделей не виконує ряд передумов класичної моделі лінійної регресії. У такому випадку робиться послаблення

(узагальнення) цих передумов, що призводить до узагальненої економетричної моделі. Це є архіважливим моментом у всьому процесі економетричного моделювання, оскільки саме цей факт визначає вибір конкретного методу оцінювання регресійних коефіцієнтів рівняння.

Якщо метод оцінювання обрано некоректно, то про якість результатів моделювання (прогнозу й імітації) не можна говорити, тому що рівняння моделі будуть описувати що завгодно, тільки не досліджуваний об'єкт. Тому даній проблемі приділяється велика увага під час вивчення економетрії.

Етап 5. **Обчислення критеріальних характеристик і перевірка статистичних гіпотез.** За відповідними формулами обчислюються критеріальні характеристики, такі як коефіцієнт детермінації, стандартні помилки параметрів і рівняння, t- і F-статистики для перевірки гіпотез про дійсні значення окремих регресійних коефіцієнтів і їхніх груп, d-статистика, що визначає наявність автокореляції залишків у рівнянні, коефіцієнт міри загальної мультиколінеарності. На підставі цих характеристик здійснюється вибір відповідного рівняння моделі, що адекватно описує досліджуваний економічний показник.

Проте варто пам'ятати, що вибір алгебраїчного вираження для змінної передбачає рішення як статистичних проблем, так і логічних аспектів. Існує така рекомендація: вибір змінних в економетричних моделях повинен спиратися не тільки на можливості статистики, але і, насамперед, на економіко-теоретичні висновки. Тому остаточний вибір функціональної залежності для досліджуваної змінної повинен здійснюватися з позицій логіко-статистичного підходу, що забезпечує врахування економічних аспектів на стадії застосування статистичних методів.

Етап 6. **Перевірка моделі на адекватність**. Розраховані статистики порівнюються з їх критеріальними значеннями. Якщо результати порівняння будуть задовільними, то економетрична модель вважається адекватною досліджуваному об'єкту (процесу). В іншому випадку необхідно вносити зміни в специфікацію моделі, тобто повторити всі етапи, починаючи з другого.

Етап 7. **Рішення моделі** може бути здійснене в двох режимах: прогнозному й імітаційному.

Рішення економетричної моделі в режимі прогнозування дозволяє відповісти на запитання: як буде розвиватися той чи інший економічний показник, якщо тенденція розвитку досліджуваного об'єкта не зміниться? Тобто, у цьому випадку передбачається, що кожен економічний показник буде в період прогнозування розвиватися так, як він розвивався до нього і динаміка його розвитку адекватно описується відповідним рівнянням моделі. Під час цього в якості майбутніх значень екзогенних змінних і деяких параметрів можуть використовуватися експертні оцінки.

Рішення імітаційних задач зводиться до аналізу впливу тих чи інших конкретно-історичних умов на розвиток економіки. Здійснюється це шляхом попередньої оцінки альтернативної траєкторії показників чи параметрів, що дозволила б дати відповідь на питання: як буде розвиватися економіка протягом деякого періоду, якщо не буде яких-небудь умов, або навпаки, якщо

вони з’являться. Ці умови вводяться в модель і її рішення дає відповідь на поставлене питання.

Етап 8. **Економічний аналіз результатів моделювання і їхнє використання при прийнятті управлінських рішень** цілком спирається на економічну теорію і за своїм змістом тяжіє до суміжних дисциплін, чим безпосередньо до самої економетрії.

Послідовність основних етапів економетричних досліджень може бути представлена наступною схемою:

Рис. 2.2 – Послідовність основних етапів економетричних досліджень

2.4 Основні типи економетричних моделей і принципи їхньої класифікації

Існують різноманітні типи економетричних моделей. Їх необхідно знати, оскільки вид моделі є одним з найважливіших факторів, що визначають вибір конкретних методів оцінки характеристик моделі і її рішення. Назвемо деякі основні типи економетричних моделей і принципи їхньої класифікації.

1. *За кількістю рівнянь*, що входять у модель, розрізняють:
   - прості (з одного рівняння);

- комплексні (із двох і більше рівнянь).

У моделях, що представляють собою систему рівнянь, стає складним розподіл змінних на залежні і незалежні (такий розподіл стосується лише простих моделей). У комплексних моделях розрізняють взаємозалежні змінні, які розраховуються в моделі за допомогою рівнянь, і змінні, визначені поза рамками моделі, і які входять в неї ззовні.

2. *За зв'язком між рівняннями* комплексні моделі бувають:
   - одночасні (система одночасних рівнянь);
   - рекурсивні;
   - блочно-рекурсивні.

Взаємозалежні економетричні моделі (системи одночасних рівнянь) характеризуються тим, що складаються з рівнянь, регресори яких описуються власними рівняннями, що містять у якості пояснюючих змінних регресанти інших рівнянь, що знаходяться нижче. Це найбільш розповсюджений у практичних дослідженнях тип економетричних моделей, найбільш складний з погляду оцінки параметрів і рішення моделі.

Якщо в моделі має місце односпрямований зв'язок між рівняннями, то мова йде про рекурсивні моделі. Характерним для цих моделей є те, що каждое рівняння в якості пояснюючих перемінних може включати регресанти попередніх (уже вирішених) рівнянь, але не з наступним. Система таких рівнянь утворить причинний ланцюг. Такі моделі не викликають розрахункових ускладнень під час оцінки параметрів і рішення моделі.

Але рекурсивності в реальних моделях домогтися важко, тому часто приводять економетричну модель до блочно-рекурсивного вигляду, в якому причинний ланцюг утворюють не окремі рівняння, а їхні блоки, а група рівнянь усередині блоку може представляти собою систему одночасних рівнянь. Зусилля, пов'язані з уявленням економетричної моделі в блочно-рекурсивному вигляді, виправдовуються спрощенням і підвищенням ефективності процедур оцінки параметрів рівнянь моделі і їх рішення як у прогнозному, так і в імітаційному режимах.

3. Залежно від того, включені в модель *змінні, стосовні попередніх тимчасових періодів*, або їх немає, економетричні моделі поділять на: статичні; динамічні.

Статичні моделі використовують змінні, стосовні тільки одного тимчасового періоду t, тобто в моделі робиться як би тимчасовий зріз.

Якщо ж у модель включена хоча б одна змінна попереднього (t-1) періоду, то модель здобуває динамічного характеру. Реальні економетричні моделі завжди є динамічними.

4. За видом *функціонального зв'язку* розрізняють економетричні моделі: лінійні; нелінійні.

Під час наявності нелінійності в моделях виникають значні методологічні і розрахункові ускладнення. Тому на практиці нелінійні моделі за допомогою ряду перетворень призводять до лінійного виду та працюють далі, як з лінійними.

5. *За ступенем ідентифікованості* (ототожності з реальним об'єктом) розрізняють моделі: неідентифіковані; точно ідентифіковані; понадідентифіковані.

Проблема ідентифікації виникає під час оцінювання одночасних рівнянь і має велике практичне значення, оскільки:

- якщо рівняння недоідентифіковане, то, у принципі, не існує засобів одержання його оцінки, тобто параметри рівняння не можуть бути оцінені;
- якщо рівняння точне ідентифіковано, то воно може бути коректно оцінено і вирішено будь-якими методами;
- при понадідентифікації рівняння існує більш, ніж один засіб одержання його оцінок і ускладнення полягає в їхньому виборі.

Ступінь ідентифікації моделі відіграє важливу роль також під час вибору методу рішення моделі. Перевірка на ідентифікацію є обов'язковим етапом проведення побудови економетричної моделі.

? Запитання для самоперевірки знань

1. Дайте визначення поняттю система.
2. Які основні класифікаційні признаки систем?
3. Які основні властивості економічних систем?
4. Назвіть основні методи дослідження систем?
5. Назвіть етапи моделювання складних систем.
6. З яких елементів складається математична модель?
7. Що розуміється під специфікацією моделі?
8. Назвіть типи математичних моделей. Чим вони відрізняються між собою?
9. До якого типу належить економетрична модель?
10. Які особливості має економетрична модель?
11. Які типи даних використовуються в економетричному дослідженні? Які проблеми виникають під час роботи з даними?
12. Наведіть основні типи економетричних моделей і принципи їхньої класифікації.
13. Чому виникає проблема ідентифікації та необхідність її рішення?
14. Як треба розуміти сукупність спостережень та її однорідність?
15. Чим забезпечується порівнянність даних у просторі й часі?

Тема 3. ПРОСТА ЛІНІЙНА РЕГРЕСІЯ ТА КОРЕЛЯЦІЯ В ЕКОНОМЕТРИЧНИХ ДОСЛІДЖЕННЯХ

3.1 Специфікація моделі

В економетриці широко використовуються методи статистики. Ставлячи мету дати кількісний опис взаємозв'язків між економічними змінними, економетрія насамперед зв'язана з методами регресії і кореляції.

Однієї з головних перешкод застосування системного підходу у всіх сферах економічного аналізу є проблема невизначеності. Одним з можливих підходів до рішення цієї проблеми є регресійний аналіз.

Регресією називають стохастичну залежність однієї випадкової величини від іншої (чи декількох інших) випадкових величин.

Стохастична залежність виражається за допомогою функції, що називається ***функцією регресії***. Принциповою відмінністю між строгою функціональною залежністю і функцією регресії є те, що в першому випадку незалежна змінна (X) цілком визначає значення функції і ця функція зворотна (наприклад $Y=5X$ та $X=Y/5$). В другому випадку цього сказати неможливо.

У залежності від кількості факторів, які входять у рівняння регресії, прийнято розрізняти просту (парну) і множинну регресії.

Проста регресія являє собою регресію між двома змінними - y и x, тобто модель виду:

$$y = f(x), \tag{3.1}$$

де y – залежна змінна (результативна ознака);

x – незалежна, або пояснююча, змінна (ознака-фактор).

Множинна регресія відповідно являє собою регресію результативної ознаки з двома і більшою кількістю факторів, тобто модель виду:

$$y = f(x_1, x_2, ..., x_k). \tag{3.2}$$

Будь-яке економетричні дослідження починається зі специфікації моделі, тобто з формулювання виду моделі, виходячи з відповідної теорії зв’язку між змінними. Іншими словами, дослідження починається з теорії, що встановлює зв'язок між явищами.

Насамперед з кола факторів, що впливають на результативну ознаку, необхідно виділити найбільш впливові. Парна регресія достатня, якщо мається домінуючий фактор, який і використовується в якості пояснюючої змінної.

Практично в кожнім окремому випадку величина y складається з двох доданків:

$$y = y_x + e, \tag{3.3}$$

де y – фактичне значення результативної ознаки;

y_x – теоретичне значення результативної ознаки, знайдене виходячи з відповідної математичної функції зв’язку y та x, тобто з рівняння регресії;

e – випадкова величина, що характеризує відхилення реального значення результативної ознаки від теоретичного, знайденого з рівняння регресії.

Випадкова величина e включає вплив не врахованих у моделі факторів, випадкових помилок і особливостей виміру. Її присутність у моделі покликана

трьома джерелами: специфікацією моделі, вибірковим характером вихідних даних, особливостями виміру змінних.

Від правильно обраної специфікації моделі залежить величина випадкових помилок: вони тим менше, ніж у більшій мері теоретичні значення результативної ознаки $\hat{y}_x$ підходять до фактичним даним y.

До ***помилок специфікації*** будуть відноситися не тільки невірний вибір тієї або іншої математичної функції для $\hat{y}_x$, але і недооблік у рівнянні регресії якого-небудь істотного фактора, тобто використання парної регресії замість множинної.

Поряд з помилками специфікації мають місце ***помилки вибірки***, оскільки дослідник більш за все має справу з вибірковими даними при встановленні закономірного зв'язку між ознаками. Помилки вибірки мають місце й у силу неоднорідності даних у вихідній статистичній сукупності, що, як правило, буває при вивченні економічних процесів. Якщо сукупність неоднорідна, то рівняння регресії не має практичного змісту. Для одержання якісного результату виключають із сукупності одиниці з аномальними значеннями досліджуваних ознак. І в цьому випадку результати регресії являють собою вибіркові характеристики.

Найбільшу небезпеку в практичному використанні методів регресії представляють ***помилки виміру***. Якщо помилки специфікації можна зменшити, змінюючи форму моделі (вигляд математичної формули), а помилки вибірки - збільшуючи обсяг вихідних даних, то помилки виміру практично руйнують усі зусилля по кількісній оцінці зв'язку між ознаками. Особливо велика роль помилок виміру при дослідженні на макрорівні. Так, у дослідженнях попиту і споживання в якості пояснюючої змінної широко використовується "доход на душу населення". Разом з тим статистичний вимір величини доходу пов'язаний з труднощами і не позбавлений можливих помилок, наприклад у результаті наявності прихованих доходів.

Припускаючи, що помилки виміру зведені до мінімуму, основна увага в економетричних дослідженнях приділяється помилкам специфікації моделі.

У парній регресії вибір виду математичної функції $\hat{y}_x = f(x)$ може бути здійснений трьома методами:

- графічним;
- аналітичним, тобто виходячи з теорії досліджуваного взаємозв'язку;
- експериментальним.

При вивченні залежності між двома ознаками ***графічний метод*** підбору виду рівняння регресії досить наочний. Він заснований на полі кореляції. Основні типи кривих, використовувані при кількісній оцінці зв'язків, представлені на рис. 3.1.

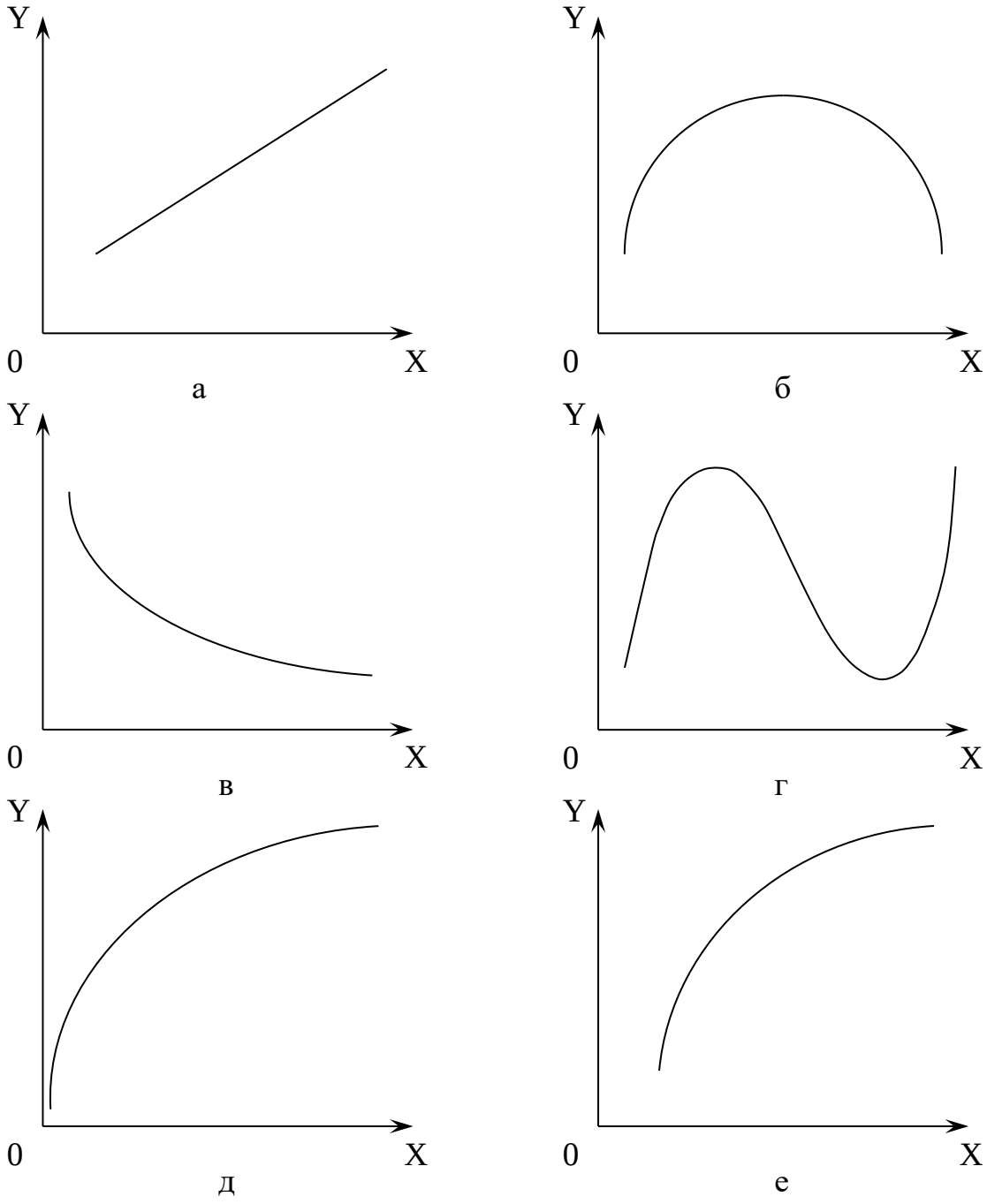


Рис. 3.1 – Основні типи кривих, використовувані при кількісній оцінці зв'язків між двома змінними:

$$а - \hat{y}_x = a + b \cdot x;$$

$$б - \hat{y}_x = a + b \cdot x + c \cdot x^2;$$

$$в - \hat{y}_x = a + \frac{b}{x};$$

$$г - \hat{y}_x = a + b \cdot x + c \cdot x^2 + d \cdot x^3;$$

$$д - \hat{y}_x = a \cdot x^b;$$

$$е - \hat{y}_x = a \cdot b^x.$$

Значний інтерес представляє ***аналітичний метод*** вибору типу рівняння регресії. Він заснований на вивченні матеріальної природи зв'язку досліджуваних ознак.

При обробці інформації на комп'ютері вибір виду рівняння регресії здійснюється ***експериментальним методом***, тобто шляхом порівняння величини залишкової дисперсії $D_{зал}$, розрахованої при різних моделях.

Якщо рівняння регресії проходить через усі крапки кореляційного поля, що можливо тільки при функціональному зв'язку, коли всі крапки лежать на лінії регресії $\hat{y}_x = f(x)$, то фактичні значення результативної ознаки збігаються з теоретичними $y = \hat{y}_x$, тобто вони цілком обумовлені впливом фактора x. У цьому випадку залишкова дисперсія $D_{зал} = 0$. У практичних дослідженнях, як правило, має місце деяке розсіювання крапок щодо лінії регресії. Воно обумовлено впливом інших факторів, що не враховуються в рівнянні регресії. Іншими словами, мають місце відхилення фактичних даних від теоретичних $(y - \hat{y}_x)$. Величина цих відхилень і лежить в основі розрахунку залишкової дисперсії:

$$D_{зал} = \frac{1}{n}\sum (y - \hat{y}_x)^2. \qquad (3.4)$$

Чим менше величина залишкової дисперсії, тим у меншій мері спостерігається вплив інших факторів, що не враховуються в рівнянні регресії, тим краще рівняння регресії підходить до вихідних даних. При обробці статистичних даних на комп'ютері перебираються різні математичні функції в автоматичному режимі і з них обирається та, для якої залишкова дисперсія є найменшою.

Якщо залишкова дисперсія виявляється приблизно однаковою для декількох функцій, то на практиці перевага віддається більш простим видам функцій, тому що вони в більшому ступені піддаються інтерпретації і вимагають меншого обсягу спостережень.

3.2 Загальне поняття про вибіркову лінійну регресію. Оцінювання параметрів лінійної регресії

Лінійна регресія знаходить широке застосування в економетрії у вигляді чіткої економічної інтерпретації її параметрів. Лінійна регресія зводиться до побудови рівняння виду

$$\hat{y}_x = a + bx \quad або \quad y = a + bx + e. \qquad (3.5)$$

Рівняння виду $\hat{y}_x = a + bx$ дозволяє за заданими значеннями фактора x отримати теоретичні значення результативної ознаки, підставляючи в рівняння фактичні значення фактора x.

Побудова лінійної регресії зводиться до оцінки її параметрів - *a* і *b*. Оцінки параметрів лінійної регресії можуть бути знайдені різними методами. Можна звернутися до поля кореляції і, вибравши на графіку дві крапки, провести через них пряму лінію. Далі за графіком можна визначити значення параметрів. Параметр *a* визначимо як крапку перетинання лінії регресії з віссю *OY*, а параметр *b* оцінимо, виходячи з кута нахилу лінії регресії, як

$$dy/dx,$$

де *dy* – збільшення результату *y*,

dx – збільшення фактора *x*.

Класичний підхід до оцінювання параметрів лінійної регресії заснований на ***методі найменших квадратів (МНК)***.

МНК дозволяє одержати такі оцінки параметрів *a* і *b*, при яких сума квадратів відхилень фактичних значень результативної ознаки *y* від розрахункових (теоретичних) $\hat{y}_x$ мінімальна:

$$\sum_i (y_i - \hat{y}_{x_i})^2 \to \min \ . \qquad (3.6)$$

Щоб знайти мінімум даної функції, треба обчислити похідні частки за кожним параметром *a* і *b* і дорівняти них до нуля.

Позначимо $\sum e^2$ через *S*, тоді:

$$S = \sum (y_i - \bar{y}_x)^2 = \sum (y - a - b \cdot x)^2$$
$$\frac{dS}{da} = -2\sum y + 2 \cdot n \cdot a + 2 \cdot b \sum x = 0$$
$$\frac{dS}{db} = -2\sum y \cdot x + 2 \cdot a \sum x + 2 \cdot b \sum x^2 = 0.$$

Перетворюючи формулу, одержимо наступну систему нормальних рівнянь для оцінки параметрів *a* та *b*:

$$\begin{cases} n \cdot a + b \sum x = \sum y, \\ a \sum x + b \sum x^2 = \sum y \cdot x. \end{cases}$$

Вирішуючи систему нормальних рівнянь знайдемо оцінки параметрів *a* і *b*. Можна скористатися наступними готовими формулами:

$$b = \frac{\text{cov}(x, y)}{\sigma_x^2}, \tag{3.7}$$

де *cov (x, y)* – коваріація ознак;

σ^2_x – дисперсія ознаки *x*.

Через те, що $\text{cov} = \overline{xy} - \bar{y} \cdot \bar{x}, \sigma_x^2 = \overline{x^2} - \bar{x}^2$, одержимо наступну формулу розрахунку оцінки параметра *b*:

$$b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - (\bar{x})^2}$$
$$a = \bar{y} - b \cdot \bar{x} \ . \tag{3.8}$$

Параметр *b* називається ***коефіцієнтом регресії***. Його величина показує середню зміну результату зі зміною фактора на одну одиницю. Так, якщо у функції витрат $\hat{y}_x = 725 + 1.5x$ (*y* - витрати (тис. грн.), *x* - кількість одиниць продукції), то зі збільшенням обсягу продукції (*x*) на 1 одиницю витрати виробництва зростають у середньому на 1,5 тис. грн., тобто додатковий приріст продукції на 1 од. викликає збільшення витрат у середньому на 1,5 тис. грн.

Можливість чіткої економічної інтерпретації коефіцієнта регресії зробила лінійне рівняння регресії досить розповсюдженим у економетричних дослідженнях.

Параметр *a* не має чіткого економічного змісту у відмінності від параметра *b*. Інтерпретувати можна лише знак при параметрі *a*. Якщо *a > 0*, то відносна зміна результату відбувається повільніше, ніж зміна фактора. Іншими словами, варіація результату менше варіації фактора - коефіцієнт варіації по фактору *x* вище коефіцієнта варіації для результату *y*: *Vx > Vy*.

3.3 Поняття тісноти зв'язку, оцінка коефіцієнтів кореляції та детермінації

Рівняння регресії завжди доповнюється показником тісноти зв'язку. При використанні лінійної регресії як такий показник виступає лінійний коефіцієнт кореляції r_{xy}. Існують різні модифікації формули лінійного коефіцієнта кореляції. Деякі з них приведені нижче:

$$r_{xy} = b\frac{\sigma_x}{\sigma_y} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \sigma_y} \ . \tag{3.9}$$

Як відомо, лінійний коефіцієнт кореляції знаходиться в границях

$$-1 < r_{xy} < 1.$$

Якщо коефіцієнт регресії $b > 0$, то $0 < r_{xy} < 1$, і, навпаки, при $b < 0$, $-1 < r_{xy} < 0$.

Варто мати на увазі, що величина лінійного коефіцієнта кореляції оцінює тісноту зв'язку розглянутих ознак у її лінійній формі. Тому близькість абсолютної величини лінійного коефіцієнта кореляції до нуля ще не означає відсутність зв'язку між ознаками. При іншій специфікації моделі зв'язок між ознаками може виявитися досить тісною.

Поряд з коефіцієнтом кореляції використовується ще один критерій, за допомогою якого також вимірюється щільність зв'язку між двома або більше показниками та перевіряється адекватність (відповідність) побудованої регресійної моделі реальній дійсності. Тобто дається відповідь на запитання, чи справді зміна значення y лінійно залежить саме від зміни значення x, а не відбувається під впливом різних випадкових факторів. Таким критерієм є ***коефіцієнт детермінації***. Перед тим, як розглянути, що саме являє собою коефіцієнт детермінації та як він пов'язаний з коефіцієнтом кореляції, розглянемо питання про ***декомпозицію дисперсій***.

Відхилення фактичних значень від середніх значень можна записати у вигляді:

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i). \qquad (3.10)$$

Різницю $(y_i - \bar{y})$ прийнято називати *загальним відхиленням*. Різницю $(\hat{y}_i - \bar{y})$ називають *відхиленням, яке можна пояснити, виходячи з регресійної прямої*. Справді, якщо x_i змінюється, то ми можемо завжди знайти значення цього відхилення, маючи тільки регресійну пряму, бо $\bar{y}$ завжди залишається незмінною величиною. Різницю $(y_i - \hat{y}_i)$ називають відхиленням, яке не можна пояснити, виходячи з регресійної прямої, або *непояснювальним відхиленням*. Справді, якщо x_i змінюється, то змінюються обидві величини y_i і $\hat{y}_i$ тому, виходячи тільки з регресійної прямої, неможливо пояснити це відхилення.

Піднесемо обидві частини (3.10) до квадрату та підсумуємо за всіма індексами, отримаємо:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 + \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2, \qquad (3.11)$$

де $\sum_{i=1}^{n}(y_i - \bar{y}_i)^2$ – загальна сума квадратів, яка позначається, як правило, через SST;

$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ – сума квадратів помилок, яка позначається через SSE;

$\sum_{i=1}^{n}(\hat{y}_i - \bar{y}_i)^2$ – сума квадратів, що пояснює регресію та позначається через SSR.

Отже, вираз (3.11) у скороченому вигляді можна записати так:

$$SST = SSE + SSR.$$

Якщо поділити (3.11) на *n*, то отримаємо вираз для дисперсій:

$$\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n} = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n} + \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{n}, \qquad (3.12)$$

де $\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}$ – загальна дисперсія, яку позначимо $\sigma^2_{заг}$;

$\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}$ – дисперсія помилок, яку позначимо $\sigma^2_{пом}$;

$\frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{n}$ – дисперсія, яку прийнято називати дисперсією, що пояснює регресію, позначимо її через $\sigma^2_{регр}$.

Таким чином, ми розклали загальну дисперсію на дві частини: дисперсію, що пояснює регресію, та дисперсію помилок. Умовно це можна записати у вигляді:

$$\sigma^2_{заг} = \sigma^2_{пом} + \sigma^2_{регр}. \qquad (3.13)$$

Поділивши обидві частини (3.13) на $\sigma^2_{заг}$ отримаємо:

$$1 = \frac{\sigma^2_{пом}}{\sigma^2_{заг}} + \frac{\sigma^2_{регр}}{\sigma^2_{заг}}. \qquad (3.14)$$

Як можна побачити з виразу (3.14), перша частина ($\frac{\sigma^2_{пом}}{\sigma^2_{заг}}$) є пропорцією дисперсії помилок у загальній дисперсії, тобто є частиною дисперсії, яку не

можна пояснити через регресійний зв'язок. Друга частина ($\frac{\sigma^2_{регр}}{\sigma^2_{заг}}$) – частина дисперсії, яку можна пояснити, виходячи з регресії.

Частина дисперсії, що пояснює регресію, називається ***коефіцієнтом детермінації*** та позначається R^2 . Коефіцієнт детермінації використовується як критерій адекватності моделі, бо є мірою пояснювальної сили незалежної змінної *x*.

Таким чином, коефіцієнт детермінації можна записати у вигляді:

$$R^2 = \frac{SSR}{SST}. \qquad (3.15)$$

Коефіцієнт детермінації завжди додатний і знаходиться в межах від нуля до одиниці ($0 \le R^2 \le 1$).

Розглянемо зв'язок між коефіцієнтом кореляції та нахилом регресійної лінії, тобто параметром *b*:

$$r_{xy} = b\frac{\sigma_x}{\sigma_y} \qquad \text{або} \qquad R^2 = (r_{xy})^2. \qquad (3.16)$$

Наприклад, якщо величина R^2 = 0,96, те це означає, що рівнянням регресії пояснюється 96% дисперсії результативної ознаки, а на долю інших факторів приходиться лише 4 % її дисперсії (тобто залишкова дисперсія). Величина коефіцієнта детермінації служить одним із критеріїв оцінки якості лінійної моделі. Чим більше частка поясненої варіації, тим відповідно менше роль інших факторів, і, отже, лінійна модель добре апроксимує вихідні дані і нею можна скористатися для прогнозу значень результативної ознаки.

3.4 Оцінка якості лінійного рівняння регресії

Оцінка якості лінійного рівняння регресії (адекватність, значимість) простої лінійної регресійної моделі можна перевірити за допомогою коефіцієнта детермінації. Якщо його значення близьке до одиниці, то можна вважати, що модель адекватна. Якщо його значення близьке до нуля, то модель неадекватна, тобто немає лінійного зв'язку між залежною та незалежною змінними. Але який висновок можна зробити, якщо значення коефіцієнта детермінації має не явно виражене граничне значення, тобто знаходиться в середині інтервалу від 0 до 1?

Зрозуміло, що в таких випадках важко зробити однозначний висновок про наявність зв'язку, тобто про адекватність моделі. Потрібен інший критерій, який би однозначно давав відповідь на запитання про адекватність побудованої моделі. Найбільш поширеним із таких критеріїв є ***критерій***

Фішера. При цьому висувається нульова гіпотеза, за якою коефіцієнт регресії дорівнює нулю, тобто $b=0$, та фактор x не впливає на результат y.

Безпосередньому розрахунку F-критерію Фішера відбувається аналіз дисперсії, де загальна сума квадратів відхилень розкладується на факторну та залишкову:

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2. \qquad (3.17)$$

Кожна сума квадратів пов'язана з числом, яке називають її "*ступенем вільності*". Це число показує, скільки незалежних елементів інформації, що утворились з елементів $y_1, y_2,\ldots,y_n$, потрібно для розрахунку даної суми квадратів.

У статистиці ***кількістю ступеня вільності*** певної величини часто називають різницю між кількістю різних дослідів та кількістю констант, знайдених завдяки цим дослідам незалежно один від одного. Окреме застосування цього поняття відноситься до суми квадратів.

Розглянемо, скільки ступенів вільності має кожна вивчена нами сума квадратів.

Для утворення SST потрібно $(n-1)$ незалежних чисел, тому що з чисел $\{(Y_1-\bar{Y}),(Y_2-\bar{Y}),\ldots(Y_n-\bar{Y})\}$ незалежні тільки $(n-1)$ завдяки властивості:

$$\sum_{i=1}^{n}(y_i-\bar{y})=0.$$

Суму квадратів, що пояснює регресію (SSR), отримують, використовуючи тільки одну незалежну одиницю інформації, яка утворюється з $y_1, y_2,\ldots. y_n$, а саме b.

Отже, суму квадратів, що пояснює просту лінійну регресію, можна утворити, використовуючи тільки одну одиницю незалежної інформації, а саме b. Звідси SSR має один ступінь вільності.

Сума квадратів помилок (SSE) має $(n-2)$ ступенів вільності. Ця сума базується на кількості ступенів вільності, яка дорівнює різниці між кількістю спостережень і кількістю параметрів, що оцінюються. У разі простої лінійної регресії оцінюються два параметри a та b. Якщо позначити кількість спостережень через n, то для SSE маємо $(n-2)$ ступенів вільності.

Ступені вільності прийнято позначати через *DF* або *df*.

У разі простої лінійної регресії ступені вільності, як і суми квадратів, можна розкласти таким чином:

$$n-1=1+(n-2). \qquad (3.18)$$

Поділивши кожну суму квадратів на відповідне їй число ступенів

вільності, отримаємо середній квадрат відхилень, або дисперсію на одну ступень вільності:

$$D_{заг} = \frac{\sum (y - \bar{y})^2}{n-1}; D_{регр} = \frac{\sum (\hat{y} - \bar{y})^2}{1}; D_{пом} = \frac{\sum (y - \hat{y})^2}{n-2}. \quad (3.19)$$

Відношення факторної дисперсії до залишкової на одну ступень вільності дає розрахунок ***F-критерію Фішера***:

$$F = \frac{D_{регр}}{D_{пом}} \quad (3.20)$$

Якщо нульова гіпотеза справедлива, то факторна та залишкова дисперсії не відрізняються одна від одної. Визначене значення F-критерію буде достовірним, якщо воно більш табличного. У цьому випадку нульова гіпотеза про відсутність зв'язку між показниками відхиляється та робиться висновок щодо значимості цього зв'язку: $\boldsymbol{F_{факт} > F_{табл}}$.

Якщо ж величина виявиться менш табличної $\boldsymbol{F_{факт} < F_{табл}}$, то імовірність нульової гіпотези вище заданого рівня (наприклад, 0,05) і вона не може бути відхилена без ризику зробити неправильний висновок щодо наявності зв'язку. У цьому випадку рівняння регресії вважається статистично не значимим, нульова гіпотеза не відхиляється.

Оцінка значимості рівняння регресії зазвичай виконується у вигляді таблиці дисперсійного аналізу.

Таблиця 3.1 – ANOVA - таблиця

Джерело варіації	Кількість ступенів вільності	Сума квадратів відхилень	Дисперсія на одну ступень вільності	F- значення	
				фактичне	табличне
Загальна, **SST**	*n-1*	$\sum_{i=1}^{n} (y_i - \bar{y})^2$	-	-	-
Регресійна (факторна), **SSR**	*1*	$\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$	$\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$	$\frac{\sum (\hat{y}_i - \hat{y}_i)^2}{\sum (y_i - \hat{y}_i)^2}{n-2}$	зі стат. таблиці
Помилкова (залишкова), **SSE**	*n-2*	$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$	$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 / n-2$		

Величина *F*- критерію пов'язана з коефіцієнтом детермінації R^2:

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

3.5 Оцінка значущості параметрів лінійної регресії та кореляції. Побудова інтервалів довіри

У випадку простої лінійної регресії оцінюється не тільки значимість рівняння регресії в цілому, а і окремих його параметрів. З цією метою по кожному з параметрів визначається його стандартна помилка: m_b та m_a.

Стандартна помилка коефіцієнта регресії визначається за формулою:

$$m_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n-2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S^2}{\sum (x - \bar{x})^2}}, \qquad (3.21)$$

де S^2 – помилкова (залишкова) дисперсія на одну ступень вільності.

Величина стандартної помилки разом з *t*-розподілом Ст'юдента при *(n-2)* ступеня вільності застосовується для перевірки значущості коефіцієнта регресії та для побудови його інтервалів довіри.

Для оцінки значущості коефіцієнта регресії його величина порівнюється з його стандартною помилкою, тобто визначається фактичне значення *t*-критерія Ст'юдента:

$$t_b = \frac{b}{m_b}, \qquad (3.22)$$

яке потім порівнюється з табличним значенням при певному рівні значимості α та кількості ступенів вільності *(n-2)*.

Між *t*-критерієм Ст'юдента та *F*-критерієм Фішера існує певна залежність, яка виражається формулою:

$$t_b = \sqrt{F}. \qquad (3.23)$$

Інтервал довіри для коефіцієнта регресії визначається як

$$b \pm t_{табл} \cdot m_b \ \ або \quad b - t_{табл} \cdot m_b \langle b \langle b + t_{табл} \cdot m_b. \qquad (3.24)$$

Оскільки коефіцієнт регресії в економетричних дослідженнях має чітку економічну інтерпретацію, то межі інтервалу довіри для коефіцієнта регресії не повинні мати суперечливих результатів, наприклад, нижня межа інтервалу має від'ємне значення, а верхня – додатне. Такий інтервал довіри містить в собі значення нуля, а це означає, що коефіцієнт регресії *b=0*, фактор *x* не впливає на результат *y*, коефіцієнт регресії є статистично не значимий.

Стандартна помилка для параметра *a* визначається за формулою:

$$m_a = \sqrt{\frac{\sum (y - \hat{y}_x)^2 \cdot \sum x^2}{(n-2) \cdot n \sum (x - \bar{x})^2}} = \sqrt{S^2 \cdot \frac{\sum x^2}{n \cdot \sum (x - \bar{x})^2}} . \quad (3.25)$$

Процедура оцінювання значущості даного параметра не відрізняється від оцінки значимості коефіцієнта регресії; розраховується *t*-критерій Ст'юдента для параметра *a*:

$$t_a = \frac{a}{m_a}, \quad (3.26)$$

його величина порівнюється з табличним значенням при *df = n-2* ступенів вільності.

Значимість лінійного коефіцієнта кореляції перевіряється на основі величині помилки коефіцієнта кореляції m_r:

$$m_r = \sqrt{\frac{1 - r^2}{n-2}} . \quad (3.27)$$

Фактичне значення *t*-критерія Ст'юдента визначається як

$$t_r = \frac{r}{\sqrt{1 - r^2}} \cdot \sqrt{n-2} . \quad (3.28)$$

Дана формула свідчить проте, що у випадку простої лінійної регресії $t_r^2 = F$. Таким чином,

$$t_r^2 = t_b^2 \quad (3.29)$$

Таким чином, перевірка гіпотез про значимість коефіцієнта регресії, параметра *a* та коефіцієнта кореляції рівносильна перевірки гіпотези про значимість та адекватність лінійного рівняння регресії.

Для оцінки якості рівняння регресії застосовується також показник, який характеризує відхилення фактичних даних від теоретичних, які отримують з рівняння моделі. Фактичні значення результативної ознаки відрізняються від теоретичних, розрахованих по рівнянню регресії, тобто *y* і $\hat{y}_x$. Чим менше ця відмінність, тим ближче теоретичні значення підходять до емпіричних даних, краще якість моделі. Величина відхилень фактичних і розрахункових значень результативної ознаки (y- $\hat{y}_x$) за кожним спостереженням являє собою ***помилку апроксимації***. Їхнє число відповідає обсягу сукупності. Оскільки (y- $\hat{y}_x$) може бути як величиною позитивною, так і негативною, то помилки апроксимації

для кожного спостереження прийнято визначати у відсотках по модулю.

Відхилення (y- $\hat{y}_x$) можна розглядати як абсолютну помилку апроксимації, а

$$\left|\frac{y-\hat{y}_x}{y}\right|\cdot 100$$

- як відносну помилку апроксимації. Щоб мати загальне поняття про якість моделі з відносних відхилень за кожним спостереженням, визначають ***середню помилку апроксимації*** як середню арифметичну просту:

$$\overline{A}=\frac{1}{n}\cdot\sum\left|\frac{y-\hat{y}_x}{y}\right|\cdot 100. \qquad (3.30)$$

Якщо значення середньою помилки апроксимації знаходиться в інтервалі від 5 % до 8 %, то це свідчить про високу якість побудованого рівняння моделі.

3.6 Побудова інтервалів прогнозу за лінійним рівнянням регресії

У прогнозних розрахунках по рівнянню регресії визначається прогнозне ($y_{пр}$) значення шляхом підстановки в рівняння регресії $\hat{y}_x = a + bx$ відповідного значення x. Однак крапковий прогноз явно не реальний. Тому він доповнюється розрахунком стандартної помилки $\hat{y}_x$, тобто $m_{\hat{y}_x}$, і відповідно інтервальною оцінкою прогнозного значення (y^*)

$$\hat{y}_x - m_{\hat{y}x} \le y^* \le \hat{y}_x + m_{\hat{y}x} \quad . \qquad (3.31)$$

З теорії вибірки відомо, що $m_{\bar{y}}^{\ 2}=\frac{\sigma^2}{n}$. Використовуючи в якості оцінки σ^2 залишкову дисперсію на одну ступінь вільності S^2, одержимо формулу розрахунку помилки середнього значення змінної y:

$$m_{\bar{y}}^{\ 2}=\frac{S^2}{n}.$$

Помилка коефіцієнта регресії визначається за формулою:

$$m_b^{\ 2}=\frac{S^2}{\sum(x-\bar{x})^2}. \qquad (3.32)$$

Вважаючи, що прогнозне значення фактора $x_{пр}= x_к$, одержимо наступну формулу розрахунку стандартної помилки прогнозного значення по лінії регресії:

$$m^2_{\hat{y}x} = \frac{S^2}{n} \cdot \frac{S^2}{\sum(x-\bar{x})^2} \cdot (x_k - \bar{x})^2 = S^2 \cdot \left(\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum(x-\bar{x})^2} \right) \quad (3.33)$$

Відповідно $m_{\hat{y}x}$ має вираз:

$$m_{\hat{y}x} = S \cdot \sqrt{\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum(x-\bar{x})^2}} \quad (3.34)$$

Розглянута формула стандартної помилки прогнозного значення y при заданому значенні $x_к$ характеризує помилку положення лінії регресії.

Для прогнозованого значення $\hat{y}_x$ інтервали довіри при заданому $x_к$ визначаються як

$$\hat{y}_{xk} \pm t_\alpha \cdot m_{\hat{y}x}. \quad (3.35)$$

? Запитання для самоперевірки знань

1. Що є функцією регресії?
2. Чим регресійна модель відрізняється від функції регресії?
3. Назвіть причини наявності в регресійної моделі випадкового відхилення?
4. Назвіть основні етапи регресійного аналізу.
5. В чому полягають помилки специфікації моделі?
6. У чому різниця між теоретичним та емпіричним рівнянням регресії?
7. У чому суть методу найменших квадратів?
8. Які основні передумови МНК?
9. Поясните зміст коефіцієнта регресії, назвіть способи його оцінювання.
10. У чому сутність статистичної значущості коефіцієнтів регресії?
11. Які висновки можна зробити об оцінках коефіцієнтів регресії та випадкового відхилення, що знайдені по МНК?
12. Проінтерпретуйте коефіцієнти емпіричного парного лінійного рівняння регресії.
13. Поясніть суть коефіцієнту кореляції.
14. У яких межах змінюється коефіцієнт кореляції?
15. Поясніть суть коефіцієнту детермінації?
16. У яких межах змінюється коефіцієнт детермінації?

17. Як визначається дисперсія залишків, загальна дисперсія і дисперсія регресії? Який між ними зв'язок?

18. Що таке число ступенів волі і як воно визначається для факторної і залишкової сум квадратів?

19. Як визначається F-критерій? Для чого він застосовується?

20. Як оцінити достовірність коефіцієнта кореляції?

21. Як визначити довірчі інтервали для параметрів моделі?

22. У чому відмінність стандартної помилки положення лінії регресії від середньої помилки прогнозованого індивідуального значення результативної ознаки при заданому значенні фактора?

23. У чому зміст середньої помилки апроксимації і як вона визначається?

Тема 4. МНОЖИННА ЛІНІЙНА РЕГРЕСІЯ ТА КОРЕЛЯЦІЯ

4.1 Поняття класичної багатофакторної регресії. Специфікація моделі

Множинна регресія – один з найбільш розповсюджених методів в економетрії. Основна мета множинної регресії – побудувати модель з великим числом факторів, визначивши при цьому вплив кожного з них окремо, а також сукупний їхній вплив на показник, який досліджується.

Парна регресія може дати гарний результат при моделюванні, якщо впливом інших факторів, що впливають на об'єкт дослідження, можна зневажити. Разом з тим дослідник ніколи не може бути упевнений у справедливості даного припущення. Для того щоб мати правильне представлення про вплив одного фактора на інший, необхідно вивчити їхню кореляцію при незмінному рівні інших факторів. Іншими словами, необхідно спробувати виявити вплив інших факторів, увівши їх у модель, тобто побудувати рівняння множинної регресії:

$$y = a + b_1x_1 + b_2x_2 + \ldots + b_p x_p + e \qquad (4.1)$$

де y – залежна змінна;

a, b_1, b_2, b_p – невідомі параметри рівняння регресії;

x_1, x_2, x_p – незалежні змінні (фактори);

e – випадкова величина.

Побудова рівняння множинної регресії починається з рішення питання про специфікацію моделі. Суть проблеми специфікації містить у собі два кола питань: добір факторів і вибір виду рівняння регресії. Їхнє рішення при побудові моделі множинної регресії має деяку специфіку.

Включення в рівняння множинної регресії того чи іншого набору факторів зв'язано насамперед із представленням дослідника про природу взаємозв'язку показника з іншими економічними явищами.

Фактори, що включаються в множинну регресію, повинні відповідати наступним вимогам:

1. Вони повинні бути кількісно виміряні. Якщо необхідно включити в модель якісний фактор, що не має кількісного виміру, то йому необхідно додати кількісну визначеність.

2. Фактори не повинні бути інтеркорельовані і тим більше знаходитися в точному функціональному зв'язку.

Включення в модель факторів з високою інтеркорреляцією, коли $R_{yx1} < R_{x1x2}$ для залежності $y = a + b_1x_1 + b_2x_2 + e$ може привести до небажаних наслідків – система нормальних рівнянь може виявитися погано обумовленої і спричинити за собою нестійкість і ненадійність оцінок коефіцієнтів регресії.

Якщо між факторами існує висока кореляція, то не можна визначити їхній окремий вплив на результативний показник і параметри рівняння регресії виявляються неінтерпретируємими. Так, у рівнянні $y = a + b_1x_1 + b_2x_2 + e$ передбачається, що фактори x_1 і x_2 незалежні друг від друга, тобто $r_{x1x2} = 0$. Тоді можна говорити, що параметр b_1 вимірює силу впливу x_1 на результат y при незмінному значенні фактора x_2. Якщо ж $r_{x1x2} = 1$, то зі зміною фактора x_1 фактор x_2 не може залишатися незмінним. Звідси b_1 і b_2 не можна інтерпретувати як показники роздільного впливу x_1 і x_2 на y.

3. Фактори, які входять до моделі, повинні пояснити варіацію незалежної змінної. Якщо будується модель з набором p факторів, то для неї розраховується показник детермінації R^2, що фіксує частку поясненої варіації результативної ознаки за рахунок розглянутих у регресії p факторів. Вплив інших факторів, що не ввійшли до моделі, оцінюється як $1–R^2$ з відповідною залишковою дисперсією S^2.

При додатковому включенні в регресію $p+1$ фактора коефіцієнт детермінації повинний зростати, а залишкова дисперсія зменшуватися:

$$R^2_{p+1} \geq R^2_p; S^2_{p+1} \leq S^2_p \qquad (4.2)$$

Якщо ж цього не відбуваються і дані показники практично мало відрізняються друг від друга, то фактор x_{p+1}, що включається до аналізу, не поліпшує модель і практично є зайвим чинником.

Насичення моделі зайвими факторами не тільки не знижує величину залишкової дисперсії і не збільшує показник детермінації, але і приводить до статистичної незначимості параметрів регресії по t-критерію Ст'юдента.

Добір факторів звичайно здійснюється в двох стадіях: на першій підбираються фактори, виходячи із сутності проблеми; на другій – на основі матриці показників кореляції визначають t-статистики для параметрів регресії.

Коефіцієнти інтеркореляції (тобто кореляції між пояснюючими змінними) дозволяють виключати з моделі дублюючі фактори. Вважається, що дві змінні явно колінеарні, тобто знаходяться між собою в лінійній залежності, якщо $r_{x_i\ x_j} \geq 0{,}7$.

Однією з умов, які висуваються до факторів моделі, є перевірка однорідності вихідної інформації.

Критерієм однорідності інформації є *середньоквадратичне відхилення* і *коефіцієнт варіації*, що розраховуються для кожного факторного і результативного показника.

Середньоквадратичне відхилення показує абсолютне відхилення індивідуальних значень від середньоарифметичного:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \ . \qquad (4.3)$$

Коефіцієнт варіації показує відносну міру відхилення окремих значень від середньоарифметичного:

$$V = \frac{\sigma}{\bar{x}} \cdot 100 \ . \qquad (4.4)$$

Чим більше *V*, тим більше розкид варіаційного ряду:

- якщо *V*<10% – розкид варіаційного ряду незначна;
- якщо 10%<*V*<20% – середній розкид варіаційного ряду;
- якщо 20%<*V*<33% – велика розкид варіаційного ряду;
- якщо *V*>33% – інформація неоднорідна і її необхідно чи виключити відкинути нетипові спостереження, що звичайно знаходяться в перших і останніх рядах вибірки.

На підставі найвищого показника варіації визначається необхідний обсяг вибірки даних

$$N \geq \frac{V^2 \cdot t^2}{m^2} \ , \qquad (4.5)$$

де *V* – варіація, %;

t – показник надійності зв'язку, що при рівні імовірності P=0,05 дорівнює 1,96;

m – показник точності розрахунків, %. Для економічних розрахунків припустима помилка 5–8 %.

Перевірка відповідності даних нормальному закону розподілу здійснюється розрахунком асиметрії *A*, ексцесу *E* и їхніх помилок ε_a, ε_e.

$$A = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n \cdot \sigma^3}; \qquad \varepsilon_a = \sqrt{\frac{6}{n}}; \tag{4.6}$$

$$E = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n \cdot \sigma^4}; \qquad \varepsilon_e = \sqrt{\frac{24}{n}}. \tag{4.7}$$

Границя допустимості відхилення від нормального закону розподілу визначається умовами:

$$\frac{A}{\varepsilon_a} < 3; \qquad \frac{E}{\varepsilon_e} < 3. \tag{4.8}$$

4.2 Оцінка параметрів багатофакторної регресії

Параметри рівняння множинної регресії оцінюються, як і в простої регресії, методом найменших квадратів (МНК). При його застосуванні будується система нормальних рівнянь, рішення якої і дозволяє одержати оцінки параметрів регресії.

Так, для рівняння $y=a+b_1 \cdot x_1+b_2 \cdot x_2+....+b_p \cdot x_p+\varepsilon$ система нормальних рівнянь складе:

$$\begin{cases} \Sigma y = n \cdot a + b_1 \cdot \Sigma x_1 + b_2 \cdot \Sigma x_2 + ... + b_p \cdot \Sigma x_p \\ \Sigma y \cdot x_1 = a \cdot \Sigma x_1 + b_1 \cdot \Sigma x^2{}_1 + b_2 \cdot \Sigma x_1 \cdot x_2 + ... + b_p \cdot \Sigma x_p \cdot x_1, \\ ..... \\ \Sigma y \cdot x_p = a \cdot \Sigma x_p + b_1 \cdot \Sigma x_1 \cdot x_p + b_2 \cdot \Sigma x_2 \cdot x_p + ... + b_p \cdot \Sigma \cdot x^2{}_p. \end{cases} \tag{4.9}$$

Її рішення може бути здійснено методом визначників:

$$a = \frac{\Delta a}{\Delta}; \quad b_1 = \frac{\Delta b_1}{\Delta}; ....; b_p = \frac{\Delta b_p}{\Delta}, \tag{4.10}$$

де Δ – визначник системи;

$\Delta a, \Delta b_1, ..., \Delta b_p$ – приватні визначники.

Загальний вираз для розрахунку параметрів моделі множинної регресії має досить громіздкий вигляд, тому ми не будемо його наводити. Зазначимо тільки, що, як і у випадку простої регресії, параметр *a* розраховується за допомогою середніх значень:

$$a = \bar{y} - b_1 \cdot \bar{x}_1 - ... - b_p \cdot \bar{x}_p. \tag{4.11}$$

У випадку лінійної множинної регресії $y=a+b_1 \cdot x_1+b_2 \cdot x_2+....+b_p \cdot x_p+\varepsilon$ параметри при x називаються ***коефіцієнтами "чистої" регресії***. Вони характеризують середню зміну результату зі зміною відповідного фактору на одну одиницю при незмінному значенні інших факторів, закріплених на середньому рівні.

Зважаючи на те, що коефіцієнти "чистої" регресії мають чітке економічне пояснення, лінійна множинна регресія отримала найбільше поширення в економетричних дослідженнях.

Можливий і інший підхід до визначення параметрів множинної регресії, коли на основі матриці парних коефіцієнтів кореляції будується рівняння регресії в стандартизованому масштабі:

$$t_y = \beta_1 . tx_1 + \beta_2 . tx_2 + ... + \beta_p . t_{xp} + \varepsilon, \qquad (4.12)$$

де t_y, t_{x1},... , t_{xp} - стандартизовані змінні:

$$t_y = \frac{y - \bar{y}}{\sigma_y}, t_{x_i} = \frac{x_i - \bar{x}}{\sigma_{x_i}},$$

для яких середнє значення дорівнює нулю: $\bar{t}_y = \bar{t}_{x_i} = 0,$, а середнє квадратичне відхилення дорівнює одиниці $\sigma_{t_y} = \sigma_{t_x} = 1;$;

β - стандартизовані коефіцієнти регресії.

Застосовуючи МНК до рівняння множинної регресії в стандартизованому масштабі, після відповідних перетворень одержимо систему нормальних рівнянь вигляду:

$$\begin{cases} R_{xy1} = \beta_1 + \beta_2 . R_{x2x1} + \beta_3 . R_{x3x1} + ... + \beta_p . R_{xpx1}, \\ R_{xy2} = \beta_1 . R_{x2x1} + \beta_2 + \beta_3 . R_{x3x2} + ... + \beta_p . R_{xpx2}, \\ ..... \\ R_{yxp} = \beta_1 . R_{xpx1} + \beta_2 . R_{x3xp} + ... + \beta_p . \end{cases}$$

Вирішуючи її методом визначників, знайдемо параметри - стандартизовані коефіцієнти регресії (β-коефіцієнти).

Стандартизовані коефіцієнти регресії показують, на скільки сигм зміниться в середньому результат, якщо відповідний фактор x_i зміниться на одну сигму при незмінному середньому рівні інших факторів. В силу того, що всі змінні задано як центровані і нормовані, стандартизовані коефіцієнти регресії β_i порівнянні між собою. Порівнюючи їхній один з одним, можна розташувати фактори за силою їхнього впливу на результат. У цьому основна вада стандартизованих коефіцієнтів регресії на відміну від коефіцієнтів "чистої" регресії, що непорівнянні між собою.

4.3 Коефіцієнти множинної кореляції та детермінації

Практична значимість рівняння множинної регресії оцінюється за допомогою показника множинної кореляції і його квадрата – коефіцієнта детермінації.

Показник множинної кореляції характеризує тісноту зв'язку розглянутого набору факторів з досліджуваною ознакою, чи, інакше, оцінює тісноту спільного впливу факторів на результат.

Незалежно від форми зв'язку показник множинної кореляції може бути знайдений як *індекс множинної кореляції*:

$$R_{yx_1x_2\ldots x_P} = \sqrt{1-\frac{\sigma^2_{зал}}{\sigma^2_y}},\ \sqrt{1-\frac{\sum(y-\hat{y})^2}{\sum(y-\bar{y})^2}} \qquad (4.13)$$

де σ^2_y – загальна дисперсія результативної ознаки;

$\sigma^2_{зал}$ – залишкова дисперсія для рівняння $y = f(x_1, x_2, ..., x_p)$.

Методика побудови індексу множинної кореляції аналогічна побудові індексу кореляції для парної залежності. Границі його зміни ті ж: від 0 до 1. Чим ближче його значення до 1, тим тісніше зв'язок результативної ознаки з усім набором досліджуваних факторів. Величина індексу множинної кореляції повинна бути більше або дорівнювати максимальному парному індексу кореляції:

$$R_{yx_1x_2\ldots x_p} \geq R_{yx_i(\max)} \left(i = \overline{1, p}\right).$$

При правильному включенні факторів у регресійний аналіз величина індексу множинної кореляції буде істотно відрізнятися від індексу кореляції парної залежності. Якщо ж додатково включені в рівняння множинної регресії фактори третьорядні, то індекс множинної кореляції може практично збігатися з індексом парної кореляції (розходження в третьому, четвертому знаках). Звідси ясно, що, порівнюючи індекси множинної і парної кореляції, можна зробити висновок про доцільність включення в рівняння регресії того чи іншого фактора. Так, якщо *y* розглядається як функція *x* й *z* і отриманий індекс множинної кореляції $R_{yzx} = 0{,}85$, а індекси парної кореляції при цьому були $R_{yx} = 0{,}82$ і $R_{yz} = 0{,}75$, то зовсім ясно, що рівняння парної регресії $y = f(x)$ охоплювало 67,2% коливання результативної ознаки під впливом фактора *x*, а додаткове включення в аналіз фактора *z* збільшило частку поясненої варіації до 72,3%, тобто зменшилася частка залишкової варіації на 5,1 відсотки (з 32,8 до 27,7%).

У розглянутих показниках множинної кореляції (індекс і коефіцієнт) використовується залишкова дисперсія, що має систематичну помилку убік зменшення, тим більше значну, чим більше параметрів визначається в рівнянні

регресії при заданому обсязі спостережень *n*. Якщо число параметрів при x_j дорівнює *m* і наближається до обсягу спостережень, то залишкова дисперсія буде близька до нуля і коефіцієнт (індекс) кореляції наблизиться до одиниці навіть при слабкому зв'язку факторів з результатом. Для того, щоб не допустити можливого перебільшення тісноти зв'язку, використовується *скорегований індекс (коефіцієнт) множинної кореляції.*

Скорегований індекс множинної кореляції містить виправлення на число ступенів волі, а саме залишкова сума квадратів $\sum\left(y-\hat{y}_{x_1x_2\ldots x_p}\right)^2$ поділяється на число ступенів волі залишкової варіації $(n-m-1)$, а загальна сума квадратів відхилень $\sum\left(y-\bar{y}\right)^2$ – на число ступенів волі в цілому по сукупності $(n-1)$.

Формула скоректованого індексу множинної детермінації має вид:

$$\overline{R^2}=1-\frac{\sum(y-\hat{y})^2:(n-m-1)}{\sum(y-\bar{y})^2:(n-1)}\ , \qquad (4.14)$$

де *m* – число параметрів при змінних *x*;

n – число спостережень.

Оскільки $\sum(y-\hat{y})^2/\sum(y-\bar{y})^2=1-R^2$, то величину скорегованого індексу детермінації можна представити у вигляді

$$\overline{R^2}=1-\left(1-R^2\right)\cdot\frac{(n-1)}{(n-m-1)}\ . \qquad (4.15)$$

Чим більше величина *m*, тим сильніше розходження $\overline{R^2}$ і R^2.

Для лінійної залежності ознак скорегований коефіцієнт множинної кореляції визначається по тій же формулі, що й індекс множинної кореляції, тобто як корінь квадратний з $\overline{R^2}$. Відмінність складається лише в тім, що в лінійній залежності під *m* мається на увазі число факторів, включених у регресійну модель, а в криволінійній залежності *m* – число параметрів при *x* і їхніх перетвореннях ($x^2, \ln x$ і ін.), що може бути більше числа факторів як економічних змінних. Так, якщо $y=f(x_1,x_2)$, то для лінійної регресії *m=2*, а для регресії виду

$$y=a+b_1\cdot x_1+b_{12}\cdot x_1^2+b_2\cdot x_2+b_{22}\cdot x_2^2+\varepsilon$$

число параметрів при *x* дорівнює 4, тобто *m=4*. При заданому обсязі спостережень за інших рівних умов зі збільшенням числа незалежних змінних (параметрів) скорегований коефіцієнт множинної детермінації зменшується. Його величина може стать і негативною при слабких зв'язках результату з факторами. У цьому випадку він повинний вважатися рівним нулю. При не великому числі спостережень скорегована величина коефіцієнта множинної

детермінації R^2 має тенденцію переоцінювати частку варіації результативної ознаки, зв'язану з впливом факторів, включених у регресійну модель.

Приклад. Припустимо, що при *n=30* для лінійного рівняння регресії з чотирма факторами $R^2 = 0{,}7$, а з урахуванням коректування на число ступенів волі

$$\overline{R^2} = 1-(1-0{,}7)\cdot\frac{(30-1)}{(30-4-1)} = 0{,}652\ .$$

Чим більше обсяг сукупності, по якій обчислена регресія, тим менше розрізняються показники $\overline{R}^2$ і R^2. Так, уже при *n=50* при тім же значенні R^2 і *m* величина $\overline{R}^2$ складе 0,673.

У статистичних пакетах прикладних програм у процедурі множинної регресії звичайно приводиться скорегований коефіцієнт (індекс) множинної кореляції (детермінації). Величина коефіцієнта множинної детермінації використовується для оцінки якості регресійної моделі. Низьке значення коефіцієнта (індексу) множинної кореляції означає, що в регресійну модель не включені істотні фактори – з одного боку, а з іншого боку – розглянута форма зв'язку не відбиває реальні співвідношення між змінними, включеними до моделі. Вимагаються подальші дослідження з поліпшення якості моделі і збільшенню її практичної значимості.

4.4 Частні рівняння множинної регресії

На основі лінійного рівняння множинної регресії

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + ... + b_p \cdot x_p + \varepsilon$$

можуть бути знайдені ***частні рівняння регресії:***

$$\begin{cases} y_{x_1 \cdot x_2, x_3, ..., x_p} = f(x_1), \\ y_{x_2 \cdot x_1, x_3, ..., x_p} = f(x_2), \\ ....., \\ y_{x_p \cdot x_1, x_2, ..., x_{p-1}} = f(x_p), \end{cases} \qquad (4.16)$$

тобто рівняння регресії, які пов'язують результативну ознаку з відповідними факторами *x* при закріпленні інших факторів множинної регресії на середньому рівні. Частні рівняння регресії мають наступний вигляд:

$$y_{x_1 \cdot x_2, x_3, ..., x_p} = a + b_1 \cdot x_1 + b_2 \cdot \bar{x}_2 + b_3 \cdot \bar{x}_3 + ... + b_p \cdot \bar{x}_p + e;$$

$$y_{x_2 \cdot x_1, x_3, ..., x_p} = a + b_1 \cdot \bar{x}_1 + b_2 \cdot x_2 + b_3 \cdot \bar{x}_3 + ... + b_p \cdot \bar{x}_p + e;$$

……

$$y_{x_p \cdot x_{1,} x_2 ,..., x_{p-1}} = a + b_1 \cdot \bar{x}_1 + b_2 \cdot \bar{x}_2 + ... + b_{p-1} \cdot \bar{x}_{p-1} + b_p \cdot x_p + e. \quad (4.17)$$

Під час підстановки в ці рівняння середніх значень відповідних факторів вони приймають вигляд парних рівнянь лінійної регресії, тобто маємо:

$$\begin{cases} \hat{y}_{x_1 \cdot x_{2,} x_3 ,..., x_p} = A_1 + b_1 \cdot x_1, \\ \hat{y}_{x_2 \cdot x_{1,} x_3 ,..., x_p} = A_2 + b_2 \cdot x_2, \\ ....., \\ y_{x_p \cdot x_{1,} x_2 ,..., x_{p-1}} = A_p + b_p \cdot x_p, \end{cases}$$

де

$$\begin{cases} A_1 = a + b_2 \cdot \bar{x}_2 + b_3 \cdot \bar{x}_3 + ... + b_p \cdot \bar{x}_p, \\ A_2 = a + b_1 \cdot \bar{x}_1 + b_3 \cdot \bar{x}_3 + ... + b_p \cdot \bar{x}_p, \\ ....., \\ A_p = a + b_1 \cdot \bar{x}_1 + b_2 \cdot \bar{x}_2 + ... + b_{p-1} \cdot \bar{x}_{p-1}. \end{cases} \quad (4.18)$$

На відміну від простої регресії *частні рівняння регресії* характеризують ізольований вплив фактора на результат, тому що інші фактори закріплені на незмінному рівні. Ефекти впливу інших факторів поєднані у них до вільного члену рівняння множинної регресії. Це дозволяє на основі часткових рівнянь регресії визначати ***частні коефіцієнти еластичності***:

$$E_{y_{x_i}} = b_i \cdot \frac{x_i}{\hat{y}_{x_i \cdot x_1 x_2 ..... x_{i-1} x_{i+1} ... x_p}},$$

де b_i – коефіцієнти регресії для фактора x_i в рівнянні множинної регресії;

$\hat{y}_{x_i \cdot x_1 x_2 ... x_{i-1} x_{i+1} ... x_p}$ – частне рівняння регресії.

Частні коефіцієнти еластичності показують, на скільки відсотків в середньому зміниться результат *y*, якщо відповідний фактор x_i зміниться на 1 %, при цьому інші фактори залишатимуться незмінними.

4.5 Оцінка надійності результатів множинної регресії та кореляції

Значимість рівняння множинної регресії в цілому, так само як і в парній регресії, оцінюється за допомогою F-критерію Фішера:

$$F = \frac{D_{факт}}{D_{зал}} = \frac{R^2}{1 - R^2} \times \frac{n - m - 1}{m}, \quad (4.19)$$

де $D_{факт}$ – факторна сума квадратів на одну ступінь вільності;

$D_{зал}$ – залишкова сума квадратів на одну ступінь вільності;

R^2 – коефіцієнт (індекс) множинної детермінації;

m – число параметрів при змінних x;

n – число спостережень.

Оцінюється значимість не тільки рівняння в цілому, але і фактора, додатково включеного в регресійну модель. Необхідність такої оцінки пов'язана з тим, що не кожен фактор, що ввійшов у модель, може істотно збільшувати частку поясненої варіації результативної ознаки. Крім того, при наявності в моделі декількох факторів вони можуть вводитися в модель у різній послідовності. Через кореляцію між факторами значимість того самого фактора може бути різною в залежності від послідовності його введення в модель. Мірою для оцінки включення фактора в модель служить ***частковий F-критерій, тобто*** F_{x_i}**.**

Припустимо, що оцінюємо значимість впливу x_1 як додатково включеного в модель фактора. Використовуємо наступну формулу:

$$F_{x_1} = \frac{R^2_{yx_1x_2\ldots x_p} - R^2_{yx_2\ldots x_p}}{1 - R^2_{yx_1x_2\ldots x_p}} \times \frac{n-m-1}{1}, \qquad (4.20)$$

де $R^2_{yx_1x_2\ldots x_p}$ – коефіцієнт множинної детермінації для моделі з повним набором факторів;

$R^2_{yx_2\ldots x_p}$ – той же показник, але без включення в модель фактора x_1;

n – число спостережень;

m – число параметрів у моделі (без вільного члена).

Якщо оцінюємо значимість впливу фактора x_p після включення в модель факторів x_1, x_2,…, x_{p-1}, то формула частного F-критерію прийме вигляд:

$$F_{x_p} = \frac{R^2_{yx_1x_2\ldots x_p} - R^2_{yx_1x_2\ldots x_{p-1}}}{1 - R^2_{yx_1x_2\ldots x_p}} \times \frac{n-m-1}{1}. \qquad (4.21)$$

У загальному виді для фактора x_i частний F- критерій визначиться як

$$F_{x_i} = \frac{R^2_{yx_1\ldots x_i\ldots x_p} - R^2_{yx_1\ldots x_{i-1}x_{i+1}\ldots x_p}}{1 - R^2_{yx_1\ldots x_i\ldots x_p}} \times \frac{n-m-1}{1}. \qquad (4.22)$$

Фактичне значення частного F-критерію порівнюється з табличним при 5%-му або 1%-му рівні значимості і кількості ступенів вільності: 1 та $n-m-1$. Якщо фактичне значення F_{x_i} перевищує $F_{табл}$ (α, df_1, df_2), то додаткове включення фактора x_i у модель статистично виправдано і коефіцієнт чистої регресії b_i при факторі x_i статистично значимий. Якщо ж фактичне значення

F_{x_i} менше табличного, то додаткове включення в модель фактора x_i не збільшує істотно частку поясненої варіації ознаки *y*, отже, недоцільно його включення в модель. Коефіцієнт регресії при даному факторі в цьому випадку статистично не значимий.

За допомогою частного *F*-критерію можна перевірити значимість усіх коефіцієнтів регресії в припущенні, що кожен відповідний фактор x_i вводився в рівняння множинної регресії останнім.

Частний *F*-критерій оцінює значимість коефіцієнтів чистої регресії. Знаючи величину F_{x_i}, можна визначити і *t*-критерій для коефіцієнта регресії при *i*-том факторі, t_{b_i}, а саме:

$$t_{b_i} = \sqrt{F_{x_i}} \,. \tag{4.23}$$

Якщо розглядається рівняння

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + \varepsilon \,,$$

то визначаються послідовно *F*-критерій для рівняння з одним фактором x_1, далі *F*-критерій для додаткового включення в модель фактора x_2, тобто для переходу від однофакторного рівняння регресії до двохфакторного, і, нарешті, *F*-критерій для додаткового включення в модель фактора x_3, тобто дається оцінка значимості фактора x_3 після включення в модель факторів x_1 та x_2. У цьому випадку *F*-критерій для додаткового включення фактора x_2 після x_1 є ***послідовним*** на відміну від *F*-критерію для додаткового включення в модель фактора x_3, що є ***частним*** *F*-критерієм, тому що оцінює значимість фактора в припущенні, що він включений у модель останнім. З *t*-критерієм Ст'юдента пов'язаний саме частний *F*-критерій. Послідовний *F*-критерій може цікавити дослідника на стадії формування моделі.

Оцінка значимості коефіцієнтів чистої регресії по *t-критерію Ст'юдента* може бути проведена і без розрахунку частних *F*-критеріїв. У цьому випадку, як і в парній регресії, для кожного фактора використовується формула

$$t_{b_i} = \frac{b_i}{m_{b_i}} \,, \tag{4.24}$$

де b_i – коефіцієнт чистої регресії при факторі x_i;

m_{b_i} – стандартна помилка коефіцієнта регресії b_i.

Для рівняння множинної регресії

$$\hat{y} = a + b_1 \cdot x_1 + b_2 \cdot x_2 + \ldots b_p \cdot x_p$$

стандартна помилка коефіцієнта регресії може бути визначена по наступній формулі:

$$m_{b_i} = \frac{\sigma_y \cdot \sqrt{1 - R^2_{yx_1 \ldots x_p}}}{\sigma_{x_i} \cdot \sqrt{1 - R^2_{x_i x_1 \ldots x_p}}} \cdot \frac{1}{\sqrt{n - m - 1}}, \qquad (4.25)$$

де σ_y – середньоквадратичне відхилення для ознаки *y*;

σ_{x_i} – середньоквадратичне відхилення для ознаки x_i;

$R^2_{yx_1 \ldots x_p}$ – коефіцієнт детермінації для рівняння множинної регресії;

$R^2_{x_i x_1 \ldots x_p}$ – коефіцієнт детермінації для залежності фактора x_i з всіма іншими факторами рівняння множинної регресії;

$n - m - 1$ – число ступенів вільності для залишкової суми квадратів відхилень.

Як бачимо, щоб скористатися даною формулою, необхідна матриця міжфакторної кореляції і розрахунок по ній відповідних коефіцієнтів детермінації $R^2_{x_i x_1 \ldots x_p}$. Так, для рівняння

$$y = a + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3 + \varepsilon$$

оцінка значимості коефіцієнтів регресії b_1, b_2, b_3 припускає розрахунок трьох міжфакторних коефіцієнтів детермінації, а саме: $R^2_{x_1 \cdot x_2 x_3}$, $R^2_{x_2 \cdot x_1 x_3}$, $R^2_{x_3 \cdot x_1 x_2}$.

Разом з тим, якщо врахувати, що

$$b_i = \frac{\sigma_y}{\sigma_{x_i}} \cdot \sqrt{\frac{R^2_{yx_1 \ldots x_p} - R^2_{yx_1 \ldots x_{i-1} x_{i+1} \ldots x_p}}{1 - R^2_{x_i x_1 \ldots x_p}}}, \qquad (4.26)$$

то можна переконатися, що

$$t_{b_i} = \frac{b_i}{m_{b_i}} = \sqrt{F_{x_i}}. \qquad (4.27)$$

? Запитання для самоперевірки знань

1. Основні етапи множинного кореляційного аналізу.
2. Як визначається модель множинної лінійної регресії?

3. Назвіть, у чому полягає специфікація моделі множинної регресії?

4. Сформулюйте вимоги, які висуваються до факторів для включення їх в модель множинної регресії?

5. Перечисліть передумови МНК. Які наслідки, якщо вони не виконуються?

6. Що характеризують коефіцієнти регресії? Як вони інтерпретуються?

7. У чому суть МНК для побудови множинного лінійного рівняння регресії?

8. Опишіть алгоритм визначення коефіцієнтів множинної лінійної регресії.

9. Як визначається статистична значущість коефіцієнтів регресії?

10. Які коефіцієнти використовуються для оцінки порівняльної сили впливу факторів на результат?

11. У чому сутність коефіцієнта детермінації?

12. Чим скоригований коефіцієнт детермінації відрізняється від звичайного?

13. Від чого залежить величина скорегованого коефіцієнта детермінації?

14. Як здійснити аналіз статистичної значущості коефіцієнта детермінації?

15. Як використовується *F*-статистика в регресійному аналізі.

16. Що таке частний *F*-критерій та чим він відрізняється від послідовного *F*-критерію?

17. Як пов'язані між собою *t*-критерій Ст'юдента для оцінки значимості b_i та частний *F*-критерій?

Тема 5. НЕЛІНІЙНА РЕГРЕСІЯ

5.1 Специфікація нелінійної моделі

Якщо між економічними явищами існують нелінійні співвідношення, то вони виражаються за допомогою відповідних нелінійних функцій.

Розрізняють ***два класи нелінійних регресій***:

1. регресії, нелінійні щодо включених в аналіз пояснюючих змінних, але лінійні за параметрами;

2. регресії, які нелінійні за оцінюваними параметрами.

Прикладом нелінійної регресії першого класу можуть служити наступні функції:

гіперболи - $y = a + \frac{b}{x}$;

параболи другого ступеня $y = a + bx + cx^2$;

поліноми різних ступенів та інші.

До нелінійних регресій за оцінюваними параметрами відносяться функції:

ступенева — $y = a \times x^b$;

показова — $y = a \times b^x$;

експонентна — $y = e^{a+bx}$ та інші.

Нелінійна регресія за включеними змінними не має будь-яких складностей в оцінці її параметрів. Вони визначаються, як і в лінійній регресії, методом найменших квадратів (МНК), тому що ці функції лінійні за параметрами. Так, в параболі другого ступеня $y = a + bx + cx^2$,

заміняючи змінні $x = x_1$, $x^2 = x_2$, одержимо двухфакторне рівняння лінійної регресії:

$y = a + bx_1 + cx_2$, для оцінки параметрів якого використовується МНК.

Відповідно для полінома третього порядку

$y = a + bx + cx^2 + dx^3$ при заміні $x = x_1$, $x^2 = x_2$, $x^3 = x_3$, одержимо трьохфакторну модель лінійної регресії.

Отже, поліном будь-якого порядку зводиться до лінійної регресії з її методами оцінювання параметрів і перевірки гіпотез. Як показує досвід більшості дослідників, серед нелінійної поліноміальної регресії найчастіше використовується парабола другого ступеня; в окремих випадках - поліном третього порядку. Обмеження у використанні поліномів більш високих ступенів пов'язані з вимогою однорідності досліджуваної сукупності: чим вище порядок полінома, тим більше вигинів має крива і відповідно менш однорідна сукупність за результативною ознакою.

Серед класу нелінійних функцій, параметри яких без особливих ускладнень оцінюються МНК, варто назвати добре відому в економіці функцію гіперболи $y = a + \frac{b}{x}$, класичним прикладом якої є крива Філіпса, яка характеризує нелінійне співвідношення між нормою безробіття x і відсотком приросту заробітної плати y.

5.2 Коефіцієнти еластичності для математичних функцій

Серед нелінійних функцій, які можуть бути приведені до лінійного виду, в економетричних дослідженнях широко використовується ступенева функція - $y = a \times x^b$. Пов'язано це з тим, що параметр b в ній має чітке економічне тлумачення та називається ***коефіцієнтом еластичності***. Величина коефіцієнта b показує, на скільки відсотків зміниться в середньому результат, якщо фактор зміниться на 1 %. Так, якщо залежність попиту від цін характеризується рівнянням виду $\hat{y}_x = 15{,}37\, x^{-0.14}$, то зі збільшенням цін на 1 % попит знижується в середньому на 0,14 %. Про правомірність подібного тлумачення параметра b для ступеневої функції $y = a \times x^b$ можна судити, якщо розглянути формулу розрахунку коефіцієнта еластичності:

$$E = \mathrm{f}(x)\frac{x}{y}\ , \qquad (5.1)$$

де *f'(x)* - перша похідна, яка характеризує співвідношення приросту результату і фактора для відповідної форми зв'язку.

Коефіцієнт еластичності можна визначати і при наявності інших форм зв'язку, але тільки для ступеневої функції він являє собою постійну величину, рівну параметру *b*. В інших функціях коефіцієнт еластичності залежить від значень фактора *x*.

У зв'язку з тим, що коефіцієнт еластичності не є величиною постійною, а залежить від відповідного значення *x*, то розраховується середній показник еластичності за формулою:

$$\overline{E} = b \times \frac{\overline{x}}{\overline{y}} \qquad (5.2)$$

Оскільки коефіцієнти еластичності становлять економічний інтерес, а види моделей не обмежуються тільки ступеневою функцією, наведемо формули розрахунку коефіцієнтів еластичності для найбільш розповсюджених типів рівнянь регресії.

Таблиця 5.1 - **Коефіцієнти еластичності для математичних функцій**

Вид функції	Коефіцієнт еластичності
Лінійна $y = a + bx$	$E = \frac{b \cdot x}{a + b \cdot x}$
Парабола другого ступеня $y = a + bx + cx^2$	$E = \frac{(b + 2 \cdot c \cdot x) \cdot x}{a + bx + cx^2}$
Гіпербола $y = a + \frac{b}{x}$	$E = \frac{-b}{a \cdot x + b}$
Показова $y = a \times b^x$	$E = x \cdot \ln b$
Ступенева $y = a \times x^b$	$E = b$
Логарифмічна $y = a + b \ln x$	$E = \frac{c \cdot x}{\frac{1}{b} \cdot e^{cx} + 1}$

5.3 Кореляція для нелінійної регресії

Рівняння нелінійної регресії, так само як і в лінійній залежності, доповнюється показником кореляції, а саме ***індексом кореляції (R)***:

$$R = \left(1 - \frac{\sigma^2_{зал}}{\sigma^2_y}\right)^{\frac{1}{2}}, \qquad (5.3)$$

де σ^2_y - загальна дисперсія результативної ознаки *y*;

$\sigma^2_{зал}$ - залишкова дисперсія.

Тому що $\sigma_y^2 = \frac{1}{n}\sum(y-\bar{y})^2$, а $\sigma_{зал}^2 = \frac{1}{n}\sum(y-\hat{y})^2$, то індекс кореляції можна виразити як

$$R = \sqrt{1 - \frac{\sum(y - \hat{y}_x)^2}{\sum(y - \bar{y})^2}} \qquad (5.4)$$

Величина даного показника знаходиться в межах $0 \le R \le 1$: чим ближче до одиниці, тим тісніше зв'язок розглянутих ознак, тим більш надійно знайдене рівняння регресії. Оцінка істотності індексу кореляції проводиться так само, як і оцінка надійності коефіцієнта кореляції.

Оскільки в розрахунку індексу кореляції використовується співвідношення факторної і загальної суми квадратів відхилень, то R^2 має той же зміст, що і коефіцієнт детермінації. У спеціальних дослідженнях величину R^2 для нелінійних зв'язків називають ***індексом детермінації***.

Індекс детермінації використовується для перевірки істотності в цілому рівняння нелінійної регресії за F-критерієм Фішера:

$$F_{факт} = \frac{R^2}{1-R^2} \times \frac{n-m-1}{m}, \qquad (5.5)$$

де R^2 - індекс детермінації;

n - число спостережень;

m - число параметрів при змінних x.

Величина m характеризує число ступенів вільності для факторної суми квадратів, а $(n-m-1)$ - число ступенів вільності для залишкової суми квадратів.

Для ступеневої функції $y = a \times x^b$ $m = 1$ і формула F-критерію прийме той же вигляд, що і при лінійній залежності:

$$F_{факт} = \frac{R^2}{1-R^2} \times (n-2)$$

Для параболи другого ступеня $y = a + bx + cx^2$ формула F-критерію

$$F_{факт} = \frac{R^2}{1-R^2} \times \frac{n-3}{2}$$

Індекс детермінації R^2 можна порівнювати з коефіцієнтом детермінації r^2 для обґрунтування можливості застосування лінійної функції. Чим більше кривизна лінії регресії, тим величина коефіцієнта детермінації r^2 менше індексу детермінації R^2. Близькість цих показників означає, що немає необхідності ускладнювати форму рівняння регресії і можна використовувати лінійну функцію. Практично якщо величина $(R^2 - r^2)$ не перевищує 0,1, то припущення про лінійну форму зв'язку вважається виправданим. В іншому випадку проводиться оцінка істотності розходження R^2, обчислених за тими самими вихідним даним, через *t-критерій Ст'юдента*:

$$t = \frac{R_{yx}^2 - r_{yx}^2}{m_{|R-r|}}, \qquad (5.6)$$

де $m_{|R-r|}$ - помилка різниці між R^2 і r^2, яка визначається за формулою:

$$m_{|R-r|} = 2 \cdot \sqrt{\frac{(R^2 - r^2) - (R^2 - r^2)^2 \cdot (2 - (R^2 + r^2))}{n}}. \qquad (5.7)$$

Якщо $t_{факт} > t_{табл}$, то розходження між розглянутими показниками кореляції істотне і заміна нелінійної регресії рівнянням лінійної функції неможлива. Практично, якщо величина $t < 2$, то розходження між R^2 і r^2 несуттєві і можливе застосування лінійної регресії, навіть якщо є припущення про деяку нелінійність розглянутих співвідношень ознак фактора і результату.

? **Запитання для самоперевірки знань**

1. Назвіть, які існують два класи нелінійних регресій? Наведіть приклади нелінійних функцій, які відносяться до першого та другого класу.
2. За допомогою яких методів нелінійна регресія перетворюється до лінійного вигляду?
3. Як називається процес перетворення нелінійних регресій в лінійну форму?
4. Якою нелінійною функцією може бути замінена парабола другого ступеня, якщо не спостерігається зміна спрямованості зв’язку ознак?
5. Запишіть усі види моделей, нелінійних відносно: включених змінних; оцінюваних параметрів.
6. У чому відмінність застосування МНК до моделей, нелінійним щодо змінних і оцінюваних параметрів?
7. Як визначаються коефіцієнти еластичності за різними видами регресійних моделей?
8. Назвіть показники кореляції, які використовуються при нелінійних співвідношеннях розглянутих ознак.

Тема 6. МУЛЬТИКОЛІНЕАРНІСТЬ

6.1 Суть мультиколінеарності

При побудові множинної лінійної регресії за МНК серйозною проблемою є мультиколінеарність - лінійний взаємозв'язок двох або декількох пояснюючих змінних. Якщо пояснюючі змінні зв'язані строгою функціональною залежністю, то має місце досконала мультиколінеарність. На практиці можна зштовхнутися з дуже високої (або близької до неї) мультиколінеарністю - сильною кореляційною залежністю між пояснюючими змінними.

Мультиколінеарність може бути проблемою лише у випадку множинної регресії. Пояснимо це на прикладі досконалої мультиколінеарності.

Нехай рівняння регресії має вигляд

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon. \qquad (6.1)$$

Нехай також між пояснюючими змінними існує строга лінійна залежність

$$X_2 = \gamma_0 + \gamma_1 X_1. \qquad (6.2)$$

Підставивши (6.2) в (6.1), одержимо

$$Y = \beta_0 + \beta_1 X_1 + \beta_2(\gamma_0 + \gamma_1 X_1) + \varepsilon$$

або

$$Y = (\beta_0 + \beta_2\gamma_{0)} + (\beta_1 + \beta_2\gamma_{1)}\, X_1 + \varepsilon.$$

Позначивши $\beta_0 + \beta_2\gamma_0 = a$, $\beta_1 + \beta_2\gamma_1 = b$, одержуємо рівняння парної лінійної регресії:

$$Y = a + b_1 + \varepsilon. \qquad (6.3)$$

За МНК визначаємо коефіцієнти *a* й *b*. Тоді одержимо систему двох рівнянь:

$$\begin{cases} \beta_0 + \beta_2\gamma_0 = a, \\ \beta_1 + \beta_2\gamma_1 = b. \end{cases} \qquad (6.\ 4)$$

У систему (6.4) входять три невідомі величини β_0, β_1, β_2 (коефіцієнти γ_0 і γ_1 визначені в (6.2)). Така система в багатьох випадків має нескінченно багато рішень. Таким чином, досконала мультиколінеарність не дозволяє однозначно визначити коефіцієнти регресії рівняння (6.1) і розділити внески пояснюючих змінних X_1 і X_2 у їхньому впливі на залежну змінну Y. У цьому випадку неможливо зробити обґрунтовані статистичні висновки про ці коефіцієнти. Отже, у випадку мультиколінеарності висновки за коефіцієнтами й за рівнянням регресії будуть ненадійними.

Досконала мультиколінеарність є скоріше теоретичним прикладом. Реальна ж ситуація, коли між пояснюючими змінними існує досить сильна кореляційна залежність, а не струга функціональна. Така залежність називається недосконалої мультиколінеарністю. Вона характеризується високим коефіцієнтом кореляції ρ між відповідними пояснюючими змінними. Причому, якщо значення ρ за абсолютною величиною близько до одиниці, то говорять про майже досконалу мультиколінеарність. У кожному разі мультиколінеарність утрудняє розподіл впливу пояснюючих факторів на поводження залежної змінної й робить оцінки коефіцієнтів регресії ненадійними. Даний висновок наочно підтверджується за допомогою діаграми Венна (рис. 6.1)

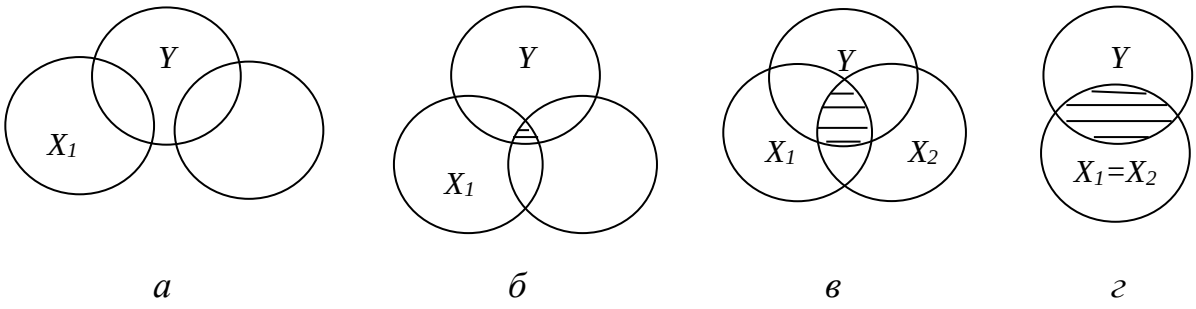


Рис. 6.1

На рис. 6.1, *а* корельованість між пояснюючими змінними X_1 і X_2 відсутня і вплив кожної з них на *Y* знаходить відображення в накладенні кіл X_1 і X_2 на коло *Y*. У міру посилення лінійної залежності між X_1 і X_2 відповідні кола усе більше накладаються один на одного. Заштрихована область відображає співпадаючі частини впливу X_1 і X_2 на *Y*. На рис. 6.1, *г* при досконалої мультиколінеарності неможливо розмежувати ступені індивідуального впливу пояснюючих змінних X_1 і X_2 на залежну змінну *Y*.

6.2 Наслідки мультиколінеарності

При виконанні певних передумов МНК дає найкращі лінійні незміщені оцінки (BLUE-оцінки). Причому властивість незміщеності й ефективності оцінок залишається в чинності, навіть якщо кілька коефіцієнтів регресії виявляються статистично незначущими. Однак незміщеність фактично означає лише те, що при багаторазовому повторенні спостережень (при постійних обсягах вибірок) за досліджуваними величинами середні значення оцінок прагнуть до їхніх дійсних значень. Повторювати спостереження в однакових умовах в економіці практично неможливо. Тому ця властивість нічого не гарантує в кожному конкретному випадку. Найменша можлива дисперсія зовсім не означає, що дисперсія оцінок буде мала в порівнянні із самими оцінками. У ряді випадків така дисперсія досить велика, щоб оцінки коефіцієнтів стали статистично незначущими.

Звичайно виділяються наступні наслідки мультиколінеарності:

1. Великі дисперсії (стандартні помилки) оцінок. Це утрудняє знаходження дійсних значень визначаємих величин і розширює інтервальні оцінки, погіршуючи їхню точність.
2. Зменшуються *t*-статистики коефіцієнтів, що може привести до невиправданого висновку про істотність впливу відповідної пояснюючої змінної на залежну змінну.
3. Оцінки коефіцієнтів за МНК і їхні стандартні помилки стають дуже чутливими до найменших змін даних, тобто вони стають нестійкими.

4. Утрудняється визначення внеску кожної з пояснюючих змінних у дисперсію залежної змінної, що пояснюється рівнянням регресії.

5. Можливе одержання невірного знака у коефіцієнта регресії.

6.3 Визначення мультиколінеарності та методи її усунення

Існує кілька ознак, за якими може бути встановлена наявність мультиколінеарності.

1. Коефіцієнт детермінації R^2 досить високий, але деякі з коефіцієнтів регресії статистично незначимі, тобто вони мають низькі t-статистики.

2. Парна кореляція між малозначимими пояснюючими змінними досить висока.

Однак дана ознака буде надійною лише у випадку двох пояснюючих змінних. При більшій їхній кількості більш доцільним є використання часткових коефіцієнтів кореляції.

3. Високі часткові коефіцієнти кореляції.

Часткові коефіцієнти кореляції визначають силу лінійної залежності між двома змінними без урахування впливу на них інших змінних. Однак при вивченні багатомірних зв'язків у ряді випадків парні коефіцієнти кореляції можуть давати зовсім невірні уявлення про характер зв'язку між двома змінними. Наприклад, між двома змінними X і Y може бути високий позитивний коефіцієнт кореляції не тому, що одна з них стимулює зміну іншої, а тому, що обидві ці змінні змінюються в одному напрямку під впливом інших змінних, як врахованих у моделі, так і, можливо, неврахованих. Тому необхідно вимірювати дійсну силу лінійного зв'язку між двома змінними, очищену від впливу на розглянуту пару змінних інших факторів. Коефіцієнт кореляції між двома змінними, очищений від впливу інших змінних, називається *частний коефіцієнтом кореляції.*

Наприклад, при трьох пояснюючих змінних X_1, X_2, X_3 частковий коефіцієнт кореляції між X_1 і X_2 розраховується по формулі

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}} \ . \qquad (6.5)$$

У загальному випадку вибірковий частковий коефіцієнт кореляції між змінними X_i і Xj ($1 \le i < j \le m$), очищений від впливу інших (m - 2) пояснюючих змінних, символічно позначається

$$r_{yj.12\ldots(j-1)(j+1)\ldots(j-1)(j+1)\ldots m} = \frac{-c_{ij}^*}{\sqrt{c_{ii}^* c_{jj}^*}}, \qquad (6.6)$$

де C^* – зворотна матриця до матриці R (матриця емпіричних парних коефіцієнтів кореляції між усілякими парами $X_1, X_2, \ldots X_m$).

Із загальної формули (6.6) легко знаходяться часткові формули для трьох змінних (формула (6.5)) і для чотирьох змінних:

$$r_{ij.kl} = \frac{r_{ij/k} - r_{il.k} \cdot r_{jl.k}}{\sqrt{(1 - r_{ij.k}^2)(1 - r_{jl.k}^2)}} \ . \qquad (6.7)$$

Нехай $r_j = r_{yj.12...(j-1)(j+1)...m}$ – частковий коефіцієнт кореляції між залежної змінної Y і змінної X_j, очищений від впливу всіх інших пояснюючих змінних. Тоді r_j^2 – частковий коефіцієнт детермінації, що визначає відсоток дисперсії змінної Y, який пояснюється впливом тільки змінної X_j. Інакше кажучи, r_j^2, j = =1, 2,..., m, дозволяє оцінити частку впливу кожної змінної X_j на розсіювання змінної Y.

4. Сильна допоміжна (додаткова) регресія.

Мультиколінеарність може мати місце внаслідок того, що яка-небудь із пояснюючих змінних є лінійною (або близької до лінійного) комбінацією інших пояснюючих змінних.. Для аналізу будуються рівняння регресії кожної з пояснюючих змінних X_j, j = 1, 2, ... , m, на пояснюючі змінні, які залишилися. Обчислюються відповідні коефіцієнти детермінації R_j^2 й розраховується їх статистична значимість на основі F-статистики:

$$F_j = \frac{R_j^2}{1 - R_j^2} \cdot \frac{n - m}{m - 1} \ , \qquad (6.\ 8)$$

де n – число спостережень,

m – число пояснюючих змінних у первісному рівнянні регресії.

Статистика F має розподіл Фішера с $v_1 = m - 1$ $v_2 = n - m$ ступенями волі. Якщо коефіцієнт R_j^2 статистично не значимий, то X_j не є лінійною комбінацією інших змінних і її можна залишити в рівнянні регресії. У противному випадку є підстави вважати, що X_j істотно залежить від інших пояснюючих змінних і має місце мультиколінеарність.

5. Визначення рівня мультиколінеарності. Для цього розраховують величину дисперсійно-інфляційного фактора для кожної змінної:

$$VIF_i = \frac{1}{1 - R_i^2} \ . \qquad (6.\ 9)$$

Якщо $VIF_i \le 10$, то можна стверджувати, що зв'язок між і-м фактором і всіма іншими недостатня, тобто мультиколінеарність відсутня.

Існує і ряд інших методів визначення мультиколінеарності.

У ряді випадків мультиколінеарність не є дуже серйозним «злом», щоб докладати істотних зусиль по її виявленню й усуненню. В основному все

залежить від цілей дослідження.

Якщо основне завдання моделі — прогноз майбутніх значень залежної змінної, то при досить великому коефіцієнті детермінації R^2 (> 0,9) наявність мультиколінеарності звичайно не позначається на прогнозних якостях моделі. Хоча це ствердження буде обґрунтованим лише в тому випадку, якщо і у майбутньому між корельованими змінними будуть зберігатися ті ж відносини, що й раніше.

Якщо ж метою дослідження є визначення ступеня впливу кожної з пояснюючих змінних на залежну змінну, то наявність мультиколінеарності, що приводить до збільшення стандартних помилок, швидше за все, спотворить дійсні залежності між змінними. У цій ситуації мультиколінеарність є серйозною проблемою.

Єдиного методу усунення мультиколінеарності, придатного в кожному разі, не існує. Це пов'язане з тим, що причини й наслідки мультиколінеарності неоднозначні й багато в чому залежать від результатів вибірки.

1. Виключення змінної(их) з моделі

Найпростішим методом усунення мультиколінеарності є виключення з моделі однієї або ряду корельованих змінних.

Однак необхідна певна обачність при застосуванні даного методу. У цій ситуації можливі помилки специфікації. Наприклад, при дослідженні попиту на деяке благо в якості пояснюючих змінних можна використати ціну даного блага й ціни замінників даного блага, які найчастіше корелюють один з одним. Виключивши з моделі ціни замінників, ми, швидше за все, припустимося помилки специфікації. Внаслідок цього можна одержати зміщені оцінки й зробити необґрунтовані висновки.

Таким чином, у прикладних економетричних моделях бажано не виключати пояснюючі змінні доти, поки колінеарність не стане серйозною проблемою.

2. Одержання додаткових даних або нової вибірки

Оскільки мультиколінеарність прямо залежить від вибірки, то, можливо, при іншій вибірці мультиколінеарності не буде або вона не буде настільки серйозною.

Іноді для зменшення мультиколінеарності досить збільшити обсяг вибірки. Наприклад, при використанні щорічних даних можна перейти до поквартальних даних. Збільшення кількості даних скорочує дисперсії коефіцієнтів регресії й тим самим збільшує їхню статистичну значимість.

Однак одержання нової вибірки або розширення старої не завжди можливо або пов'язане із серйозними витратами.

Крім того, такий підхід може підсилити автокореляцію. Ці проблеми обмежують можливість використання даного методу.

3. 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 первісній моделі, але які істотно впливають на залежну змінну.

Якщо даний метод має підстави, то його використання зменшує суму квадратів відхилень, тим самим скорочуючи стандартну помилку регресії. Це приводить до зменшення стандартних помилок коефіцієнтів.

4. Використання попередньої інформації про деякі параметри

Іноді при побудові моделі множинної регресії можна скористатися попередньою інформацією, зокрема відомими значеннями деяких коефіцієнтів регресії. Цілком імовірно, що значення коефіцієнтів, розраховані для яких-небудь попередніх (звичайно більше простих) моделей або для аналогічної моделі за раніше отриманій вибірці, можуть бути використані для розроблювальної в цей момент моделі.

На приклад будується регресія виду (6.1). Припустимо, що змінні X_1 і X_2 корельовані. Для раніше побудованої моделі парної регресії $Y = \gamma_0 + \gamma_1 X_1 + v$ був визначений статистично значимий коефіцієнт γ_1 (для визначеності нехай $\gamma_1 = 0{,}8$), що зв'язує Y з X_1. Якщо є підстави вважати, що зв'язок між Y і X_1 залишиться незмінним, то можна припустити $\gamma_1 = \beta_1 = 0{,}8$. Тоді (6.1) прийме вид:

$$Y = \beta_0 + 0{,}8X_1 + \beta_2 X_2 + \varepsilon, \Rightarrow$$

$$Y - 0{,}8X_1 = \beta_0 + \beta_2 X_2 + \varepsilon. \qquad (6.10)$$

Рівняння (6.10) фактично є рівнянням парної регресії, для якого проблема мультиколінеарності не існує.

Обмеженість використання даного методу обумовлена тим, що, по-перше, одержання попередньої інформації найчастіше важко, а по-друге, імовірність того, що виділений коефіцієнт регресії буде тим самим для різних моделей, невисока.

5. Перетворення змінних

У ряді випадків мінімізувати або взагалі усунути проблему мультиколінеарності можна за допомогою перетворення змінних.

Наприклад, нехай емпіричне рівняння регресії має вигляд

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2, \qquad (6.11)$$

причому X_1 і X_2 – корельовані змінні. У цій ситуації можна спробувати визначати регресійні залежності відносних величин:

$$\frac{\hat{Y}}{X_1} = b_0 + b_1 \frac{X_2}{X_1}, \qquad \frac{\hat{Y}}{X_2} = b_0 + b_1 \frac{X_1}{X_2}. \qquad (6.12)$$

Цілком імовірно, що в моделях, аналогічних (6.12), проблема

мультиколінеарності буде відсутня.

Можливі й інші перетворення, близькі за своєю суттю до вищеописаного. Наприклад, якщо в рівнянні розглядаються взаємозв'язки номінальних економічних показників, то для зниження мультиколінеарності можна спробувати перейти до реальних показників і т.п.

? Запитання для самоперевірки знань

1. Поясніть значення термінів «колінеарність» і «мультиколінеарність» .

2. У чому розходження між досконалою й недосконалої мультиколінеарністю?

3. Які основні наслідки мультиколінеарності?

4. Як можна виявити мультиколінеарність?

5. Як оцінюється корельованість між двома пояснюючими змінними?

6. Перелічите основні методи усунення мультиколінеарності.

7. Які з наступних стверджень дійсні, помилкові або не визначені? Відповідь поясните.

а) При наявності високої мультиколінеарності неможливо оцінити статистичну значимість коефіцієнтів регресії при корельованих змінних.

б) Наявність мультиколінеарності не є перешкодою для одержання за МНК BLUE-оцінок.

в) Мультиколінеарність не є істотною проблемою, якщо основне завдання побудованої регресійної моделі складається в прогнозуванні майбутніх значень залежної змінної.

г) Високі значення коефіцієнтів парної кореляції між пояснюючими змінними не завжди є ознаками мультиколінеарності.

д) Тому що X^2 є строгою функцією від X, то при використанні обох змінних у якості пояснюючих виникає проблема мультиколінеарності.

е) При наявності мультиколінеарності оцінки коефіцієнтів залишаються незміщеними, але їхні t-статистики будуть занадто низькими.

ж) Коефіцієнт детермінації R^2 не може бути статистично значимим, якщо всі коефіцієнти регресії статистично незначимі (мають низькі t-статистики).

з) Мультиколінеарність не приводить до одержання зміщених оцінок коефіцієнтів, але веде до одержання зміщених оцінок для дисперсій коефіцієнтів.

і) У регресійній моделі $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ наявність мультиколінеарності можна виявити, якщо обчислити коефіцієнт кореляції між X_1 і X_2.

8. Нехай за МНК оцінюється рівняння регресії $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$. Для більшості вибірок спостерігається висока корельованість між X_1 і X_2. Нехай корельованість між цими змінними не спостерігається. Коефіцієнти регресії оцінюються за даною вибіркою. Чи будуть у цьому випадку оцінки

незміщеними? Чи будуть незміщеними оцінки дисперсій знайдених емпіричних коефіцієнтів регресії?

9. Поясніть логіку відкидання пояснюючої змінної з метою усунення проблеми мультиколінеарності.

10. Нехай в рівнянні регресії $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ змінні X_1 й X_2 сильно корельовані. Будується рівняння регресії X_2 на X_1, випадкові відхилення від якої позначимо через υ. Будується нове рівняння регресії с залежної змінною Y і двома пояснюючими змінними – X_2 и υ. Чи буде вирішена таким чином проблема мультиколінеарності?

Тема 7. ГЕТЕРОСКЕДАСТИЧНІСТЬ

7.1 Сутність гетероскедастичності та її наслідки

При практичному проведенні регресійного аналізу за допомогою МНК варто звернути серйозну увагу на проблеми, пов'язані з виконанням властивостей випадкових відхилень моделей. Властивості оцінок коефіцієнтів регресії прямо залежать від властивостей випадкового члена в рівнянні регресії. Для одержання якісних оцінок необхідно стежити за виконанням передумов МНК (умов Гаусса — Маркова), тому що за їхніх порушень МНК може давати оцінки з поганими статистичними властивостями. При цьому існують інші методи визначення більше точних оцінок. Однієї із ключових передумов МНК є наступна умова: *дисперсія випадкових відхилень ε_i постійна.*

Виконання даної передумови називається *гомоскедастичністю* (сталістю дисперсії відхилень). Це можна записати як $\mathrm{var}(\varepsilon_i) = \sigma_\varepsilon^2 = const.$

Нездійсненність даної передумови називається *гетероскедастичністю* (мінливістю дисперсій відхилень). Це можна записати як $\mathrm{var}(\varepsilon_i) = \sigma_\varepsilon^2 \neq const.$

Сутність припущення гомоскедастичності полягає в тому, що σ_ε^2 не є функцією x_{ij}, тобто $\sigma_\varepsilon^2 \neq f(x_{1i}, x_{2i}, \ldots, x_{pi})$. Графічно випадок гомоскедастичності зображений на рис. 7.1.

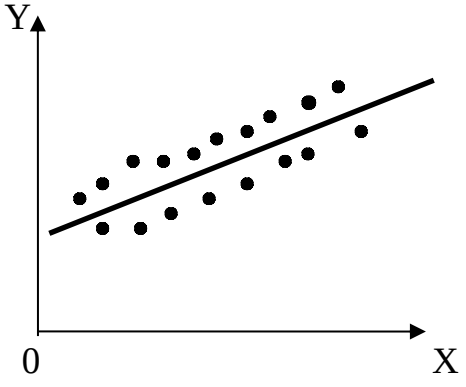


Рис 7.1. Гомоскедастичність

Якщо σ_{ε}^{2} не є постійної, її значення залежать від значень x_{ij}, то можна записати $\sigma_{\varepsilon}^{2} = f(x_{1i,}x_{2i},...,x_{pi})$. У цьому випадку має місце гетероскедастичність. Один з випадків гетероскедастичності є випадок монотонно зростаючої дисперсії ε_i (зі зростанням x зростає й дисперсія ε_i).

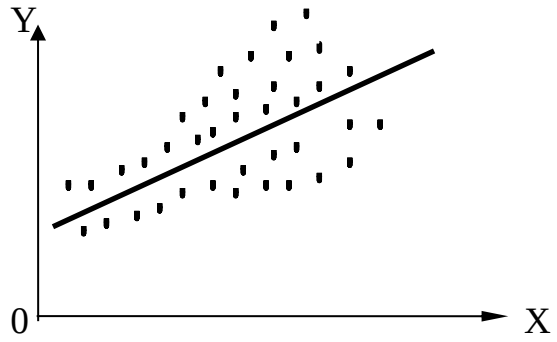


Рис 7.2. Гетероскедастичність (зростання дисперсії ε_i).

Проблема гетероскедастичності характерна для перехресних даних і досить рідко зустрічається при розгляді часових рядів. Це можна пояснити наступним чином. При перехресних даних ураховуються економічні суб'єкти (споживачі, домогосподарства, фірми, галузі, країни й ін.), що мають різні доходи, розміри, потреби й т. ін. У цьому випадку можливі проблеми, пов'язані з ефектом масштабу. У часових рядах звичайно розглядаються ті ж самі показники в різні моменти часу (наприклад, ВНП, чистий експорт, темпи інфляції й т. ін. у певному регіоні за певний період часу). Однак при збільшенні (зменшенні) розглянутих показників із часом може виникнути проблема гетероскедастичності.

При розгляді класичної лінійної регресійної моделі МНК дає найкращі лінійні незміщені оцінки (BLUE-оцінки) лише при виконанні ряду передумов, однієї з яких є сталість дисперсії відхилень (гомоскедастичність): $\mathrm{var}(\varepsilon_i) = \sigma_{\varepsilon}^{2} = const$ для всіх спостережень *i, i = 1, 2, ..., n.*

При нездійсненності даної передумови (при гетероскедастичності) наслідку застосування МНК будуть наступними.

1. Оцінки коефіцієнтів як і раніше залишаться незміщеними й лінійними.
2. Оцінки не будуть ефективними (тобто вони не будуть мати найменшу дисперсію в порівнянні з іншими оцінками даного параметра). Вони не будуть навіть асимптотично ефективними. Збільшення дисперсії оцінок знижує ймовірність отримання максимально точних оцінок.
3. Дисперсії оцінок будуть розраховуватися зі зсувом. Зміщеність

з'являється внаслідок того, що непояснена рівнянням регресії дисперсія $\sigma^2 = \frac{\sum e_i^2}{n-m-1}$ (m – число пояснюючих змінних), що використається при обчисленні оцінок дисперсій всіх коефіцієнтів, не є більше незміщеною.

4. Внаслідок вищесказаного всі висновки, одержані на основі відповідних *t*- і *F*-статистик, а також інтервальні оцінки будуть ненадійними. Отже, статистичні виводи, одержані при стандартних перевірках якості оцінок, можуть бути помилковими й приводити до невірних висновків за побудованою моделлю. Цілком імовірно, що стандартні помилки коефіцієнтів будуть занижені, а отже, *t*-статистики будуть завищені. Це може привести до визнання статистично значимими коефіцієнтів, які такими насправді не є.

7.2 Виявлення гетероскедастичності. Методи пом'якшення проблеми гетероскедастичності

У ряді випадків, знаючи характер даних, поява проблеми гетероскедастичність можна передбачати й спробувати усунути цей недолік ще на етапі специфікації. Однак значно частіше цю проблему доводиться вирішувати після побудови рівняння регресії.

Виявлення гетероскедастичності в кожному конкретному випадку є досить складним завданням, тому що для знання дисперсій відхилень $\sigma^2(e_i)$ необхідно знати розподіл ВВ (випадкових величин) *Y*, що відповідає обраному значенню x_i ВВ *X*. На практиці часто для кожного конкретного значення x_i визначається єдине значення y_i, що не дозволяє оцінити дисперсію ВВ *Y* для даного x_i.

Не існує якого-небудь однозначного методу визначення гетероскедастичності. Однак до теперішнього часу для такої перевірки розроблене досить велике число тестів і критеріїв для них.

Розглянемо найбільш популярні й наочні:

1. Графічний аналіз відхилень,
2. Тест рангової кореляції Спірмана,
3. Тест Парка,
4. Тест Глейзера,
5. Тест Голдфелда-Квандта.

1. *Графічний аналіз залишків.*

Використання графічного подання відхилень дозволяє визначитися з наявністю гетероскедастичності. У цьому випадку по осі абсцис відкладаються значення (x_i) пояснюючої змінної *X* (або лінійної комбінації пояснюючих змінних $Y = b_0 + b_iX_i + ... + b_mX_m$), а по осі ординат або відхилення e_i, або їхні квадрати e_i^2, $i = 1, 2, \ldots, n$. Приклади таких графіків наведені на рис. 7.3.

На рис. 7.3, *а* всі відхилення e_i^2 перебувають усередині напівсмуги постійної ширини, паралельної осі абсцис. Це говорить про незалежність

дисперсій e_i^2 від значень змінної X й їхній сталості, тобто в цьому випадку виконуються умови гомоскедастичності.

На рис. 7.3, *б-д* спостерігаються деякі систематичні зміни в співвідношеннях між значеннями x_i змінної X і квадратами відхилень e_i^2. На рис. 7.3, *б* и *в* відбита лінійна, *г* – квадратична, *д* – гіперболічна залежності між квадратами відхилень і значеннями пояснюючої змінної X. Інакше кажучи, ситуації, представлені на рис. 7.3, *б — д,* відбивають більшу ймовірність наявності гетероскедастичності для розглянутих статистичних даних.

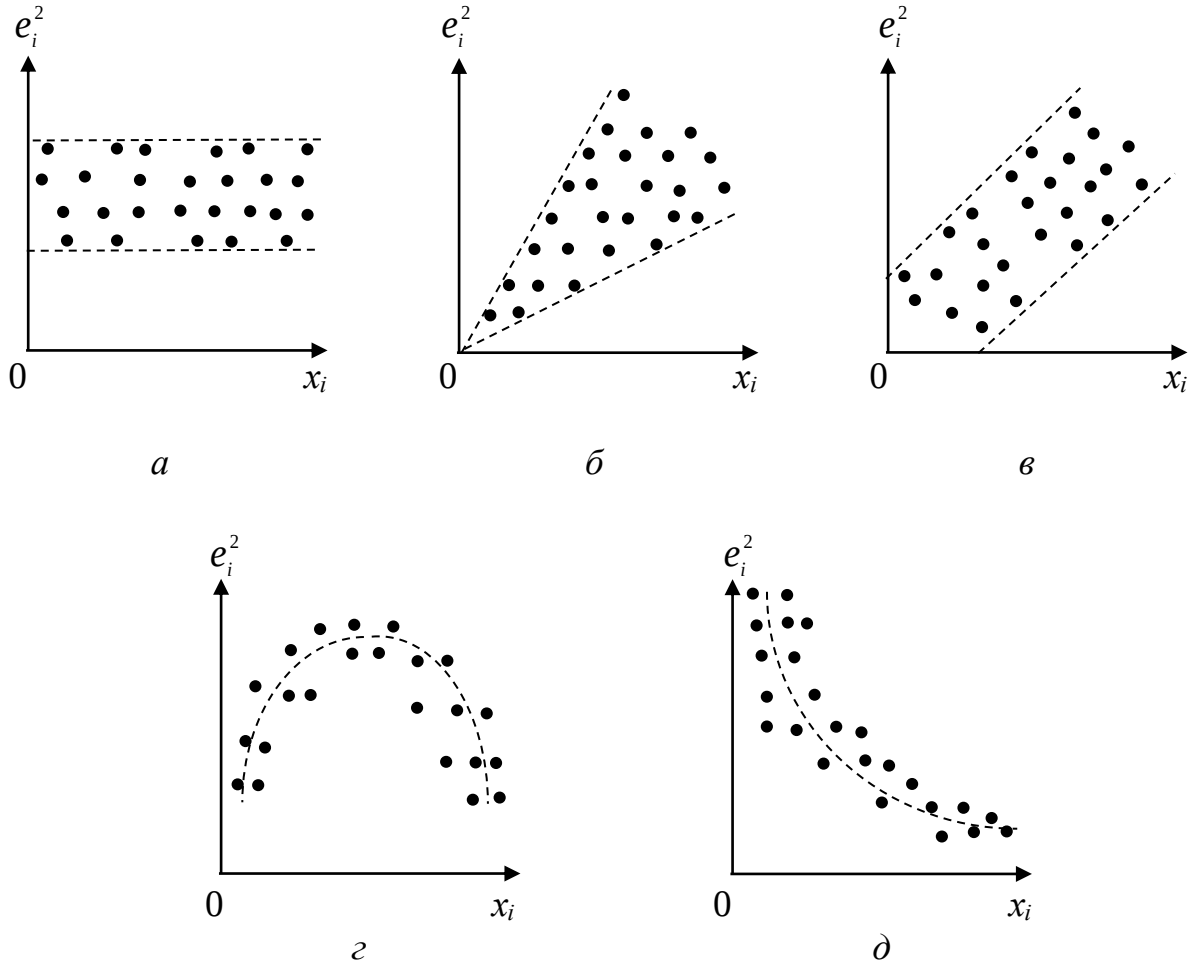


Рис. 7.3

Графічний аналіз відхилень є зручним і досить надійним у випадку парної регресії. При множинній регресії графічний аналіз можливий для кожної з пояснюючих змінних X_j , $j = 1, 2, ..., m$, окремо. Частіше ж замість пояснюючих змінних X_j по осі абсцис відкладають значення $\hat{y}_i$, $i = 1, 2, ..., n$, одержані з емпіричного рівняння регресії. Оскільки за рівнянням множинної лінійної регресії $\hat{y}_i$ є лінійною комбінацією x_{ij}, $j = 1, 2, ... , m$, $i = 1, 2, ..., n$, то графік, що відбиває залежність e_i^2 від $\hat{y}_i$, може вказати на наявність гетероскедастичності аналогічно ситуаціям на рис. 7.3, *б-д.* Такий аналіз найбільш доцільний при великій кількості пояснюючих

змінних.

2. *Тест рангової кореляції Спірмана.*

При використанні даного тесту передбачається, що дисперсія відхилення буде або збільшуватися, або зменшуватися зі збільшенням значень *X*. Тому для регресії, побудованої за МНК, абсолютні величини відхилень e_i і значення x_i СВ *X* будуть корельовані.

Тест проводиться в наступній послідовності:

1). Побудувати регресійну модель *y* и *x* і розрахувати відхилення e_i.

2). Ранжуються абсолютні значення e_i і x_i у зростаючому або убиваючому порядку й підраховується коефіцієнт рангової кореляції Спірмана:

$$r_s = 1 - 6\left[\frac{\sum_{i=1}^{n} d_i}{n(n^2 - 1)}\right], \qquad (7.1)$$

де d_i – різниця між рангами e_i і x_i, $i = 1, 2, \ldots, n$, які приписуються двом характеристикам *i-го* об'єкта;

n – кількість об'єктів, що ранжуються.

Наприклад, якщо x_{20} є 25-м за величиною серед всіх спостережень *X*, а e_{20} є 32-м, то $d_i = 25 - 32 = -7$.

3). Перевіряється значимість отриманого коефіцієнта за *t*-критерієм Ст'юдента із числом ступенів волі $\upsilon = n - 2$:

$$t = \frac{r_{x,e}\sqrt{n-2}}{\sqrt{1 - r_{x,e}^2}}, \qquad (7.2)$$

де *n* – кількість спостережень.

Якщо спостережуване значення *t*-статистики, обчислене за формулою, перевищує $t_{кр} = t_{a/2,n-2}$ (обумовлене за таблицею критичних крапок розподілу Ст'юдента), це підтверджує гіпотезу про гетероскедастичність. У противному випадку гіпотеза про відсутність гетероскедастичності приймається.

Якщо в моделі регресії більше чим одна пояснююча змінна, то перевірка гіпотези може здійснюватися за допомогою *t*-статистики для кожної з них окремо.

3. *Тест Парка.*

Р. Парк запропонував критерій визначення гетероскедастичності, що доповнює графічний метод деякими формальними залежностями. Передбачається, що дисперсія $\sigma_i^2 = \sigma^2(e_i)$ є функцією *i-го* значення x_i пояснюючої змінної. Парк запропонував наступну функціональну залежність:

$$\sigma_i^2 = \sigma^2 x_i^{\beta} e^{\upsilon_i}. \qquad (7.3)$$

Прологарифмірував (7.3), одержимо:

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln x_i + \upsilon_i. \qquad (7.4)$$

Тому що дисперсії σ_i^2 звичайно невідомі, то їх заміняють оцінками квадратів відхилень e_i^2.

Критерій Парка включає наступні етапи:

1. Будується рівняння регресії $y_i = b_0 + b_1 x_i + e_i$.
2. Для кожного спостереження визначаються $\ln e_i^2 = \ln(y_i - \hat{y}_i)^2$.
3. Будується регресія

$$\ln \sigma_i^2 = \alpha + \beta \ln x_i + \upsilon_i. \qquad (7.5)$$

де $\alpha = \ln\sigma^{2.}$

У випадку множинної регресії залежність (7.5) будується для кожної пояснюючої змінної.

4. Перевіряється статистична значимість коефіцієнта β рівняння (7.5) на основі t-статистики. Якщо коефіцієнт β статистично значимий, то це означає наявність зв'язку між $\ln e_i^2$ й $\ln x_i$, тобто гетероскедастичності в статистичних даних.

Використання в критерій Парка конкретної функціональної залежності (7.5) може привести до необґрунтованих висновків (наприклад, коефіцієнт β статистично незначимий, а гетероскедастичність має місце). Можлива ще одна проблема. Для випадкового відхилення υ_i, у свою чергу може мати місце гетероскедастичність. Тому критерій Парку доповнюється іншими тестами.

4. *Тест Глейзера.*

Тест Глейзера за своєю суттю аналогічний тесту Парка й доповнює його аналізом інших (можливо, більше підходящих) залежностей між дисперсіями відхилень σ_i; і значеннями змінної x_i. За даним методом оцінюється регресійна залежність модулів відхилень $|\varepsilon_i|$ (тісно зв'язаних с σ_i^2) від x $|e_i|$. При цьому розглянута залежність моделюється наступним рівнянням регресії:

$$|e_i| = \alpha + \beta x_i^k + \upsilon_i. \qquad (7.6)$$

Змінюючи значення k, можна побудувати різні регресії. Звичайно k = ..., -1, -0,5, 0,5, 1, ... Статистична значимість коефіцієнта β у кожному конкретному випадку фактично означає наявність гетероскедастичності.

Якщо для декількох регресій (7.6) коефіцієнт β виявляється статистично значимим, то при визначенні характеру залежності звичайно орієнтуються на кращу з них.

Так само, як й у тесті Парка, у тесті Глейзера для відхилень υ_i може порушуватися умова гомоскедастичності. Однак у багатьох випадках запропоновані моделі є досить гарними для визначення гетероскедастичності.

5. *Тест Голдфелда-Квандта*.

У цьому випадку передбачається, що стандартне відхилення $\sigma^2 = \sigma(\varepsilon_i)$ пропорційно значенню x_i змінної X у цьому спостереженні, тобто $\sigma_i^2 = \sigma^2 x_i^2$, $i = 1, 2, ..., n$. Передбачається, що ε_i має нормальний розподіл і відсутня автокореляція залишків.

Тест Голдфелда-Квандта полягає в наступному:

1). Ранжуються спостереження змінної X у порядку зростання або убування.

2). Задається величина C – кількість центральних спостережень за незалежними змінними X, які ми будемо виключати з подальшого аналізу. Оптимальною кількістю центральних спостережень є приблизно четверта частина всіх спостережень. Залишок $(n - C)$ спостережень ділиться на дві підвибірки однакові за розміром, одна з яких включає маленькі значення x, інша – більші.

3). Оцінюються окремі регресії для кожної підвибірки, розраховуються суми квадратів відхилень із $((n - C/2 - k)$ – ступенями волі (k – загальна кількість оцінюваних параметрів у моделі).

Якщо припущення про пропорційність дисперсій відхилень значенням X вірно, то дисперсія регресії за першою підвибіркою (сума квадратів відхилень) буде істотно менше дисперсії регресії за другою підвибіркою.

4). Для порівняння відповідних дисперсій будується наступне F-відношення:

$$F = \frac{\sum e_2^2 / \left[\{(n-c)/2\} - k\right]}{\sum e_1^2 / \left[\{(n-c)/2\} - k\right]} = \frac{\sum e_2^2}{\sum e_1^2}, \qquad (7.7)$$

де $\sum e_1^2$ – сума квадратів підвибірки з малими значеннями х,

$\sum e_2^2$ – відхилення від підвибірки з більшими значеннями х.

Якщо розраховані значення F-відношення більше табличного значення, то приймається гіпотеза про наявність гетероскедастичності, навпаки – про гомоскедастичності.

Природним є питання: якими повинні бути розміри підвибірок для прийняття обґрунтованих рішень? Для парної регресії Голдфелд і Квандт пропонують наступні пропорції: $n = 30$, $k = 11$; $n = 60$, $k = 22$.

Для множинної регресії даний тест звичайно проводиться для тієї пояснюючої змінної, котра найбільшою мірою пов'язана з σ_i. При цьому k повинне бути більше, ніж $(m + 1)$. Якщо немає впевненості щодо вибору змінної X_j, то даний тест може здійснюватися для кожної з пояснюючих змінних.

Цей же тест може бути використаний при припущенні про обернену пропорційність між σ_i і значеннями пояснюючої змінної. При цьому статистика Фішера прийме вид: $F = \dfrac{\sum e_1^2}{\sum e_2^2}$.

Гетероскедастичність приводить до неефективності оцінок, незважаючи на їх незміщеність. Це може обумовити необґрунтовані висновки про якість моделі. Тому при встановленні гетероскедастичності виникає необхідність перетворення моделі з метою усунення даного недоліку. Вид перетворення залежить від того, відомі чи ні дисперсії σ_i^2 відхилень ε_i, $i = 1, 2, ..., n$.

1. Метод зважених найменших квадратів (ЗНК)

Даний метод застосовується при відомих для кожного спостереження значеннях σ_i^2. У цьому випадку можна усунути гетероскедастичність, розділив кожне спостережуване значення на відповідне йому значення дисперсії. У цьому суть методу зважених найменших квадратів.

Для простоти викладу опишемо ЗНК на прикладі парної регресії:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i. \qquad (7.8)$$

Розділимо обидві частини (7.8) на відоме $\sigma_i = \sqrt{\sigma_i^2}$:

$$\frac{y_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{x_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i}. \qquad (7.9)$$

Поклавши $\dfrac{y_i}{\sigma_i} = y_i^*$, $\dfrac{x_i}{\sigma_i} = x_i^*$, $\dfrac{\varepsilon_i}{\sigma_i} = \upsilon_i$, $\dfrac{1}{\sigma_i} = z_i$, одержимо рівняння регресії без вільного члена, але з додаткової пояснюючої змінної Z і з «перетвореним» відхиленням v:

$$y_i^* = \beta_0 z_i + \beta_1 x_1^* + v_i \qquad (7.10)$$

При цьому для v_i виконується умова гомоскедастичності.

Отже, для перетвореної моделі (7.10) виконуються передумови 1^0 — 5^0 МНК. У цьому випадку оцінки, отримані за МНК, будуть найкращими лінійними незміщеними оцінками.

Таким чином, ЗНК включає наступні етапи:

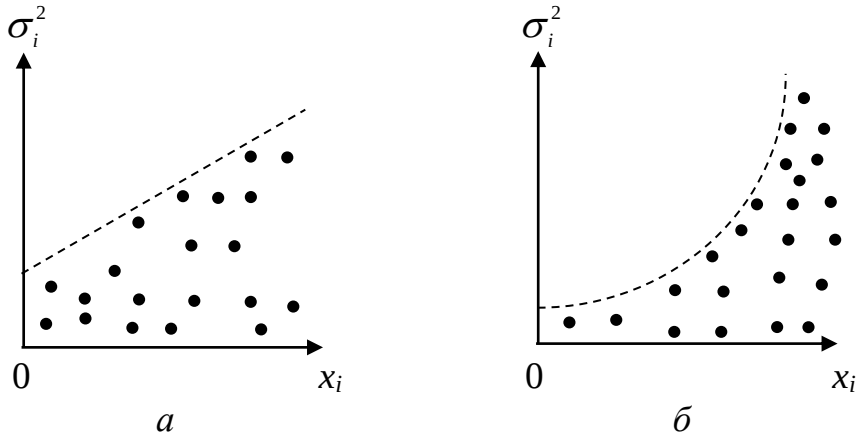
1. Значення кожної пари спостережень (x_i , y_i) ділять на відому величину σ_i. Тим самим спостереженням з найменшими дисперсіями надаються найбільші «ваги», а з максимальними дисперсіями - найменші «ваги». Дійсно, спостереження з меншими дисперсіями відхилень будуть більше значимими при оцінці коефіцієнтів регресії, чим спостереження з більшими дисперсіями. Урахування цього факту збільшує ймовірність одержання більше точних оцінок.

2. За МНК для перетворених значень $\left(\frac{1}{\sigma_i}, \frac{x_i}{\sigma_i}, \frac{y_i}{\sigma_i}\right)$ будується рівняння регресії без вільного члена з гарантованими якостями оцінок.

2. Дисперсії відхилень невідомі

Для застосування ЗНК необхідно знати фактичні значення дисперсій σ_i^2 відхилень. На практиці такі значення відомі вкрай рідко. Отже, щоб застосувати ЗНК, необхідно зробити реалістичні припущення про значення σ_i^2.

Наприклад, може виявитися доцільним припустити, що дисперсії σ_i^2 відхилень ε_i пропорційні значенням x_i (рис. 7.4, *а*) або значенням x_i^2 (рис. 7.4, *б*).



а *б*

Рис. 7.4.

1. Дисперсії σ_i^2 пропорційні x_i (рис. 7.4, *а*):

$\sigma_i^2 = \sigma^2 x_i$ (σ^2 – коефіцієнт пропорційності).

Тоді рівняння (7.8) перетвориться діленням його лівої й правої частин на $\sqrt{x_i}$:

$$\frac{y_i}{\sqrt{x_i}} = \frac{\beta_0}{\sqrt{x_i}} + \beta_1 \frac{x_i}{\sqrt{x_i}} + \frac{\varepsilon_i}{\sqrt{x_i}} \Rightarrow \frac{y_i}{\sqrt{x_i}} = \beta_0 \frac{1}{\sqrt{x_i}} + \beta_1 \sqrt{x_i} + v_i. \qquad (7.11)$$

Для випадкових відхилень $v_i \frac{\varepsilon_i}{\sqrt{x_i}}$ виконується умова гомоскедастичності. Отже, для регресії (7.11) застосуємо звичайний МНК.

Таким чином, оцінивши для (7.11) за МНК коефіцієнти β_0 й β_1, потім повертаються до вихідного рівняння регресії (7.8).

Якщо в рівнянні регресії присутні декілька пояснюючих змінних, можна надійти у такий спосіб. Замість конкретної пояснюючої змінної X_j використовується вихідне рівняння множинної лінійної регресії $\hat{Y} = b_0 + b_1 X_1 + ... + b_m X_m$, тобто фактично лінійна комбінація пояснюючих змінних. У цьому випадку отримують наступну регресію:

$$\frac{y_i}{\sqrt{\hat{y}_i}} = \beta_0 \frac{1}{\sqrt{\hat{y}_i}} + \beta_1 \frac{x_{i1}}{\sqrt{\hat{y}_i}} + ... + \beta_m \frac{x_{im}}{\sqrt{\hat{y}_i}} + \frac{\varepsilon_i}{\sqrt{\hat{y}_i}}. \qquad (7.12)$$

Іноді із всіх пояснюючих змінних вибирається найбільш підходяща виходячи із графічного подання (рис. 7.4).

2. Дисперсії σ_i^2 пропорційні x_i^2 (рис. 7.4, *б).*

У випадку, якщо залежність σ_i^2 від x_i доцільніше виразити не лінійною функцією, а квадратичної, то відповідним перетворенням буде ділення рівняння регресії (7.8) на x_i:

$$\frac{y_i}{x_i} = \beta_0 \frac{1}{x_i} + \beta_1 + \frac{\varepsilon_i}{x_i} \Rightarrow \frac{y_i}{x_i} = \beta_0 \frac{1}{x_i} + \beta_1 + v_i, \qquad (7.13)$$

де $v_i = \frac{\varepsilon_i}{x_i}$.

За аналогією з вищевикладеним для відхилень v_i буде виконуватися умова гомоскедастичності. Після визначення за МНК оцінок коефіцієнтів β_0 й β_1 для рівняння (7.13) повертаються до вихідного рівняння (7.8).

Для застосування описаних вище перетворень досить значимі знання про правдиві значення дисперсій відхилень σ_i^2, або припущення, якими ці дисперсії можуть бути. У багатьох випадках дисперсії відхилень залежать не від включених у рівняння регресії пояснюючих змінних, а від тих, які не включені в модель, але відіграють істотну роль у досліджуваній залежності. У цьому випадку вони повинні бути включені в модель. У ряді випадків для усунення гетероскедастичності необхідно змінити специфікацію моделі (наприклад, лінійну на лог-лінійну, мультиплікативну на адитивну і т. ін.).

На практиці має сенс застосувати кілька методів визначення гетероскедастичності й способів її коректування (перетворень, що стабілізують дисперсію).

? Запитання для самоперевірки знань

1. У чому суть гетероскедастичності?
2. Яке з наступних тверджень вірно, помилково або не визначено:

а) внаслідок гетероскедастичності оцінки перестають бути ефективними й спроможними;

б) оцінки й дисперсії оцінок залишаються незміщеними;

в) виводи по *t*- і *F*-статистиках є ненадійними;

г) при наявності гетероскедастичності стандартні помилки оцінок будуть заниженими;

д) гетероскедастичність проявляється через низьке значення статистики Дарбина-Уотсона *DW*;

е) не існує загального тесту для аналізу гетероскедастичності;

ж) тест рангової кореляції Спірмана заснований на використанні *t*-статистики;

з) тест Парка є частковим випадком тесту Глейзера;

і) використання методу зважених найменших квадратів носить обмежений характер, тому що для його використання необхідно знати дисперсії відхилень;

к) якщо в парній регресії дисперсія випадкових відхилень пропорційна величині пояснюючої змінної (*x)*, то для одержання ефективних оцінок необхідно всі спостережувані значення поділити на *x*?

3. Приведіть аргументи на користь графічного тесту, тесту Парка й тесту Глейзера.
4. Приведіть схему тесту Голдфелда-Квандта.
5. У чому суть методу зважених найменших квадратів (ЗНК)?
6. Чому при наявності гетероскедастичності ЗНК дозволяє одержати більше ефективні оцінки, чим звичайний МНК.
7. Є підстава вважати, що в регресії, побудованої за квартальним даними, випадкові відхилення в перші квартали більше, ніж відхилення в інших кварталах. Як це можна перевірити?

Тема 8. АВТОКОРЕЛЯЦІЯ В ЕКОНОМЕТРИЧНИХ МОДЕЛЯХ

8.1 Сутність і причини автокореляції

Важливою передумовою побудови якісної регресійної моделі за МНК є незалежність значень випадкових відхилень ε_i від значень відхилень у всіх інших спостереженнях. Відсутність залежності гарантує відсутність корельованості між будь-якими відхиленнями ($\sigma(\varepsilon_i, \varepsilon_j) = \text{cov}(e_i, \varepsilon_j) = 0$ при $i \neq j$) і, зокрема, між сусідніми відхиленнями ($\sigma(\varepsilon_{i-1}, \varepsilon_i) = 0$), $i = 2, 3, ..., n$.

Автокореляція (послідовна кореляція) визначається як кореляція між спостережуваними показниками, упорядкованими в часі (часові ряди) або в

просторі (перехресні дані). Автокореляція залишків (відхилень) звичайно зустрічається в регресійному аналізі при використанні даних часових рядів. При використанні перехресних даних наявність автокореляції (просторової кореляції) украй рідко. У силу цього в подальшому замість символу *i* порядкового номера спостереження будемо використати символ *t,* що відбиває момент спостереження. Обсяг вибірки при цьому будемо позначати символом *T* замість *n.* В економічних задачах значно частіше зустрічається так називана *позитивна автокореляція* ($\sigma(\varepsilon_{t-1}, \varepsilon_t) > 0$), ніж *негативна автокореляція* ($\sigma(\varepsilon_{t-1}, \varepsilon_t) < 0$).

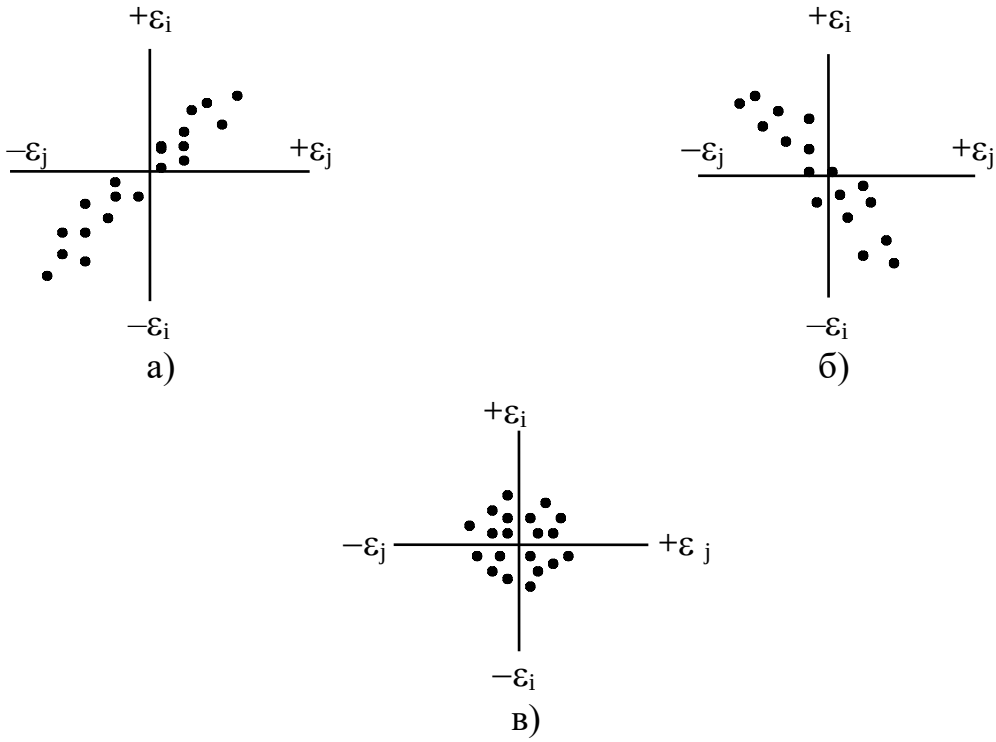


Рис. 8.1.

На рис. 8.1 а) зображена наявність позитивної кореляції між випадковими величинами ε (позитивне значення ε_i супроводжується позитивним $-\varepsilon_j$ і навпаки). У більшості випадків позитивна автокореляція викликається спрямованим постійним впливом деяких неврахованих у моделі факторів. Негативна автокореляція фактично означає, що за позитивним відхиленням наступає негативне й навпаки. Можлива схема розсіювання крапок у цьому випадку представлена на рис. 8.1 б) негативне значення ε_i супроводжується позитивним $-\varepsilon_j$ і навпаки. На рис. 8.1 в) зображений класичний приклад відсутності кореляції між випадковими величинами, тобто немає систематичності в розміщенні випадкових значень ε, тому коваріація між ними дорівнює нулю.

Серед основних причин, що викликають появу автокореляції, можна виділити помилки специфікації, інерцію в зміні економічних показників, ефект павутини, згладжування даних.

Помилки специфікації. Неврахування в моделі якої-небудь важливої пояснюючої змінної або неправильний вибір форми залежності звичайно приводить до системних відхилень крапок спостережень від лінії регресії, що може обумовити автокореляцію.

Інерція. Багато економічних показників (наприклад, інфляція, безробіття, ВНП і т. ін.) мають певну циклічність, пов'язаної з хвилеподібністю ділової активності. Дійсно, економічний підйом приводить до росту зайнятості, скороченню інфляції, збільшенню ВНП і т. ін. Цей ріст триває доти, поки зміна кон'юнктури ринку й ряду економічних характеристик не приведе до зросту, потім зупинці й руху назад розглянутих показників. У кожному разі ця трансформація відбувається не миттєво, а має певну інертність.

Ефект павутини. У багатьох виробничих й інших сферах економічні показники реагують на зміну економічних умов із запізнюванням (часовим лагом). Наприклад, пропозиція сільськогосподарської продукції реагує на зміну ціни із запізнюванням (рівним періоду дозрівання врожаю). Більша ціна сільськогосподарської продукції в минулому році викличе (швидше за все) її надвиробництво цього року, а отже, ціна на неї знизиться й т. ін.

Згладжування даних. Найчастіше дані по деякому тривалому часовому періоді одержують усередненням даних за підінтервалами, які його складають. Це може привести до певного згладжування коливань, які були усередині розглянутого періоду, що у свою чергу може послужити причиною автокореляції.

8.2 Наслідки автокореляції

Наслідки автокореляції деякою мірою подібні з наслідками гетероскедастичності. Серед них при застосуванні МНК звичайно виділяються наступні.

1. Оцінки параметрів, залишаючись лінійні і незміщеними, перестають бути ефективними. Отже, вони перестають мати властивості найкращих лінійних незміщених оцінок (BLUE-оцінок).

2. Дисперсії оцінок є зміщеними. Часто дисперсії, що обчислюють за стандартними формулами, є заниженими, що спричиняє збільшення *t*-статистик. Це може привести до визнання статистично значимими пояснюючі змінні, які в дійсності такими можуть і не бути.

3. Оцінка дисперсії регресії $\sigma^2 = \sum_{t=1}^{T} \frac{e_t^2}{T-m-1}$ є зміщеною оцінкою дійсного значення σ^2, у багатьох випадках занижуючи його.

4. У силу вищесказаного виводи за *t*- і *F*-статистиками, що визначають значимість коефіцієнтів регресії й коефіцієнта детермінації,

можливо, будуть невірними. Внаслідок цього погіршуються прогнозні якості моделі.

8.3 Виявлення автокореляції

В силу невідомості значень параметрів рівняння регресії невідомими будуть також і дійсні значення відхилень ε_t, $t = 1, 2, ..., T$. Тому висновки о їх незалежності здійснюються на основі оцінок e_t, $t = 1, 2, ..., T$, отриманих із емпіричного рівняння регресії. Існують декілька можливих методів визначення автокореляції.

1. Графічний метод.

Існує кілька варіантів графічного визначення автокореляції. Один з них, що погоджує відхилення e_t з моментами t їхнього одержання (їхніми порядковими номерами i), наведений на рис. 8.2. Це так названі послідовно-часові графіки. У цьому випадку по осі абсцис звичайно відкладаються або час (момент) одержання статистичних даних, або порядковий номер спостереження, а по осі ординат – відхилення ε_t (або оцінки відхилень e_t).

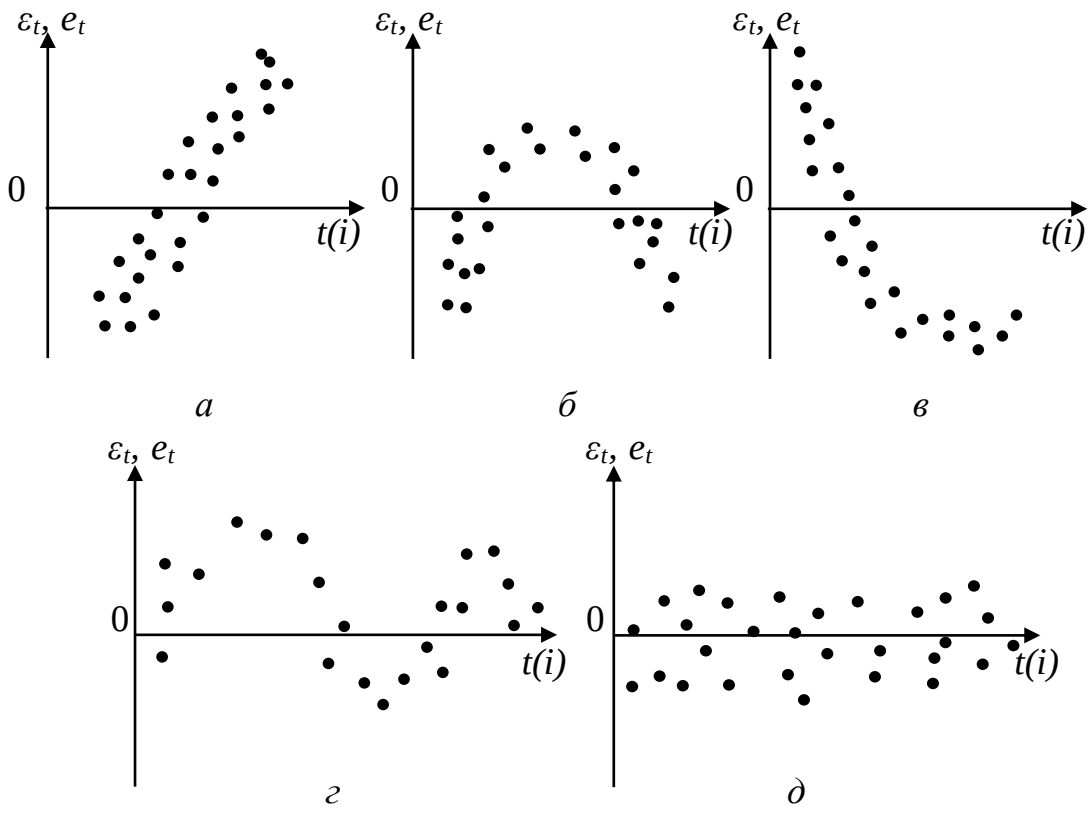


Рис. 8.2

Природно припустити, що на рис. 8.2, *а-г* є певні зв'язки між відхиленнями, тобто автокореляція має місце. Відсутність залежності на рис. 8.2, *д* швидше за все свідчить про відсутність автокореляції.

На рис. 8.2, *б* відхилення спочатку в основному негативні, потім позитивні, потім знову негативні. Це свідчить про наявність між відхиленнями певної залежності. Більше того, можна стверджувати, що в

цьому випадку має місце позитивна автокореляція залишків. Вона стає досить наочною, якщо графік 8.2, *б* доповнити графіком залежності e_t від e_{t-1} (рис. 8.3).

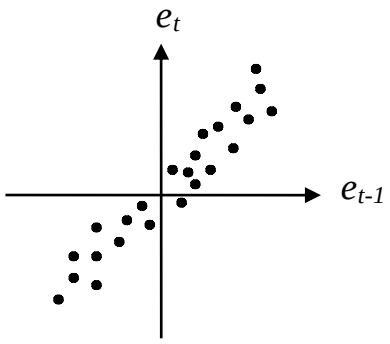


Рис. 8.3

Переважна більшість крапок на цьому графіку розташоване в *I* й *III* чвертях декартової системи координат, підтверджуючи позитивну залежність між сусідніми відхиленнями.

У сучасних комп'ютерних прикладних програмах для рішення завдань економетрії аналітичне вираження регресії доповнюється графічним поданням результатів. На графік реальних коливань залежної змінної накладається графік коливань змінної за рівнянням регресії. Зіставивши ці дві графіки, можна висунути гіпотезу про наявність автокореляції залишків. Якщо ці графіки перетинаються рідко, то можна припустити наявність позитивної автокореляції залишків.

2. Метод рядів.

Цей метод досить простий: послідовно визначаються знаки відхилень e_t, $t = 1, 2, \ldots, T$. Наприклад,

(- - - - -)(+ + + + + + +)(- - -)(+ + + +)(-),

тобто 5 «-», 7 «+», 3 «-», 4 «+», 1 «-» при 20 спостереженнях.

Ряд визначається як безперервна послідовність однакових знаків. Кількість знаків у ряді називається *довжиною ряду*.

Візуальний розподіл знаків свідчить про невипадковий характер зв'язків між відхиленнями. Якщо рядів занадто мало в порівнянні з кількістю спостережень *n,* то цілком імовірна позитивна автокореляція. Якщо ж рядів занадто багато, то ймовірно негативну автокореляцію. Для більше детального аналізу пропонується наступна процедура. Нехай

n – обсяг вибірки;

n_1 – загальна кількість знаків «+» при n спостереженнях (кількість позитивних відхилень e_t);

n_2 – загальна кількість знаків «-» при n спостереженнях (кількість негативних відхилень e_t);

k – кількість рядів.

При досить великій кількості спостережень ($n_1 > 10$, $n_2 > 10$) і відсутності автокореляції ВВ k має асимптотично нормальний розподіл з

$$M(k) = \frac{2n_1n_2}{n_1 + n_2} + 1;$$

$$D(k) = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}.$$

Тоді, якщо $M(k) - e_{a/2}D(k) < k < M(k) + e_{a/2}D(k)$, то гіпотеза про відсутність автокореляції не відхиляється.

Для невеликого числа спостережень ($n_1 < 20$, $n_2 < 20$) за таблицями критичних значень кількості рядів при n спостереженнях (додаток) на перетинанні рядка n_1 і стовпця n_2 визначаються нижнє k_1 і верхнє k_2 значення при рівні значимості $\alpha = 0.05$.

Якщо $k_1 < k < k_2$, то говорять про відсутність автокореляції.

Якщо $k \leq k_1$, то говорять про позитивну автокореляцію залишків.

Якщо $k \geq k_2$, то говорять про негативну автокореляцію залишків.

У нашому прикладі $n = 20$, $n_1 = 11$, $n_2 = 9$, $k = 5$. За таблицями (додаток) визначаємо $k_1 = 6$, $k_2 = 16$. Оскільки $k = 5 < 6 = k_1$, то приймається припущення про наявність позитивної автокореляції при рівні значимості $\alpha = 0{,}05$.

3. Критерій Дарбіна-Уотсона.

Найбільш відомим критерієм виявлення автокореляції першого порядку є критерій Дарбіна–Уотсона. Статистика *DW* Дарбіна–Уотсона приводиться у всіх спеціальних прикладних комп'ютерних програмах як найважливіша характеристика якості регресійної моделі.

Суть методу полягає в тому, що на основі обчисленої статистики *DW* Дарбіна–Уотсона робиться висновок про автокореляцію.

$$DW = \frac{\sum_{t=2}^{T}(e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} \quad . \tag{8.1}$$

Статистика Дарбіна–Уотсона тісно пов'язана з вибірковим коефіцієнтом кореляції $r_{e_t e_{t-1}}$:

$$DW \approx \frac{2\left(\sum e_t^2 - \sum e_t e_{t-1}\right)}{\sum e_t^2} = 2(1 - r_{e_t e_{t-1}}). \tag{8.2}$$

Якщо $e_t = e_{t-1}$, то $r_{e_t e_{t-1}} = 1$ й DW = 0. Якщо $e_t = -e_t$, то $r_{e_t e_{t-1}} = -1$ й DW = 4. У всіх інших випадках $0 \leq DW \leq 4$ і її значення можуть указати на наявність або відсутність автокореляції. Дійсно, якщо $r_{e_t e_{t-1}} \approx 0$

(автокореляція відсутня), то $DW \approx 2$. Якщо $r_{e_t e_{t-1}} \approx 1$ (позитивна автокореляція), то $DW \approx 0$. Якщо $r_{e_t e_{t-1}} \approx -1$ (негативна автокореляція), то $DW \approx 4$.

Для більш точного визначення, яке значення DW свідчить про відсутність автокореляції, а яке — про її наявність, була побудована таблиця критичних крапок розподілу Дарбіна—Уотсона. За неї для заданого рівня значимості α, числа спостережень n і кількості пояснюючих змінних m визначаються два значення: d_l – нижня межа й d_u – верхня межа.

Порядок тестування за критерієм Дарбіна–Уотсона:

1. За побудованим емпіричним рівнянням регресії

$$\hat{y}_t = b_0 + b_1 x_{t1} + ... + b_m x_{tm}$$

визначаються значення відхилень $e_t = y_t - \hat{y}_t$ для кожного спостереження t, $t = 1, 2, ..., T$.

2. За формулою (8.1) розраховується статистика DW.

3. За таблицею критичних крапок Дарбіна–Уотсона визначаються два числа d_l й d_u і здійснюють висновки за правилом:

$0 \le DW < d_l$ — існує позитивна автокореляція;

$d_u \le DW < d_u$ — висновок про наявність автокореляції невизначений;

$d_u \le DW < 4 - d_u$ — автокореляція відсутня;

$4 - d_u \le DW < 4 - d_l$ — висновок про наявність автокореляції невизначений;

$4 - d_l \le DW \le 4$ — існує негативна автокореляція.

Всі випадки зображені на рис (8.4).

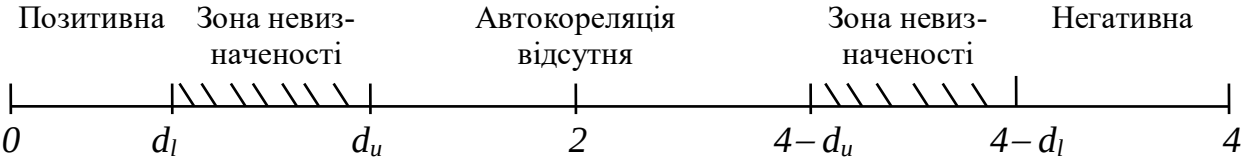


Рис. 8.4.

При використанні критерію Дарбіна-Уотсона необхідно враховувати наступні обмеження.

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

2. Передбачається, що випадкові відхилення ε_t визначаються за ітераційною схемою: $\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t$, називаної авторегресійною схемою першого порядку AR(1). Тут ν_t – випадковий член.

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

4. Критерій Дарбіна–Уотсона не застосуємо для регресійних моделей, що містять у складі пояснюючих змінних залежну змінну з часовим лагом в один період, тобто для так званих *авторегресійних моделей* виду:

$$y = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + ... + \beta_k x_{tk} + \gamma y_{t-1} + \varepsilon_t . \qquad (8.3)$$

Причину четвертого обмеження пояснюється наступним. Нехай рівняння регресії має вигляд:

$$y_t = \beta_0 + \beta_1 x_t + \gamma y_{t-1} + \varepsilon_t . \qquad (8.4)$$

Нехай випадкове відхилення ε_t піддано впливу авторегресії першого порядку:

$$\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t . \qquad (8.5)$$

Тоді рівняння регресії (8.4) можна представити в наступному виді:

$$y_t = \beta_0 + \beta_1 x_t + \gamma y_{t-1} + \rho\varepsilon_{t-1} + \nu_t . \qquad (8.6)$$

Але y_{t-1} залежить від ε_{t-1}, тому що якщо (8.4) вірно для t, то воно вірно й для t–1. Отже, є систематичний зв'язок між однією з пояснюючих змінних й одним з компонентів випадкового члена, тобто не виконується одна з основних передумов МНК (передумова 4^0) — пояснюючі змінні не повинні бути випадковими (не мати випадкової складової). Значення будь-якої пояснюючої змінної повинне бути екзогенним, повністю визначеним. У противному випадку оцінки будуть зміщеними навіть при більших обсягах вибірок.

Для авторегресійних моделей розроблені спеціальні тести виявлення автокореляції, зокрема h-статистика Дарбіна, що визначається за формулою

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n\,\mathrm{var}(g)}}, \qquad (8.7)$$

де $\hat{\rho}$ – оцінка ρ авторегресії першого порядку (8.5);

var(g) – вибіркова дисперсія коефіцієнта при лаговій змінній y_{t-1},

n – число спостережень.

При великому обсязі вибірки n і справедливості нульової гіпотези H_0: $\rho = 0$ статистика h має стандартизований нормальний розподіл ($h \sim N(0, 1)$). Тому за заданим рівнем значимості α визначається критична крапка $u_{\alpha/2}$ з умови $\Phi(u_{\alpha/2}) = (1-\alpha)/2$ і рівняється h з $u_{\alpha/2}$. Якщо $h > u_{\alpha/2}$, то нульова гіпотеза про відсутність автокореляції повинна бути відхилена. У противному випадку вона не відхиляється.

Відзначимо, що звичайно значення $\hat{\rho}$ розраховується за формулою $\hat{\rho} = 1 - 0.5DW$, а $var(\text{g})$ дорівнює квадрату стандартної помилки S_g оцінки g коефіцієнта γ. Тому h легко обчислюється на основі даних оціненої регресії.

У такий спосіб можна записати

$$h = \left(1 - \frac{1}{2}DW\right)\sqrt{\frac{n}{1 - n[\text{var}(g)]}}.$$

Якщо

а) h> 1.96, то є позитивна автокореляція;

б) h< −1.96, то є негативна автокореляція;

в) −1.96< h< 1.96, то автокореляція відсутня.

Основна проблема при використанні цього тесту полягає в неможливості обчислення h при $nvar(\text{g}) > 1$.

8.4 Методи усунення автокореляції

Основною причиною наявності випадкового члена в моделі є недосконалі знання про причини й взаємозв'язки, що визначають то або інше значення залежної змінної. Тому властивості випадкових відхилень, у тому числі й автокореляція, у першу чергу залежать від вибору формули залежності й состава пояснюючих змінних. Тому що автокореляція найчастіше викликається неправильною специфікацією моделі, то необхідно насамперед скорегувати саму модель. Можливо, автокореляція викликана відсутністю в моделі деякої важливої пояснюючої змінної. Варто спробувати визначити даний фактор і врахувати його в рівнянні регресії. Також можна спробувати змінити формулу залежності (наприклад, лінійну на лог-лінійну, лінійну на гіперболічну й т. ін.).

Однак якщо всі розумні процедури зміни специфікації моделі вичерпані, а автокореляція має місце, то можна припустити, що вона обумовлена якимись внутрішніми властивостями ряду $\{e_t\}$. У цьому випадку можна скористатися авторегресійним перетворенням. У лінійній регресійній моделі або в моделях, що зводяться до лінійного, найбільш доцільним і простим перетворенням є *авторегресійна схема першого порядку AR(1).*

Для простоти викладу $AR(1)$ розглянемо модель парної лінійної регресії

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i. \qquad (8.8)$$

Тоді спостереженням t й $(t\text{-}1)$ відповідають формули:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t. \qquad (8.9)$$

$$y_{t-1} = \beta_0 + \beta_1 x_{t-1} + \varepsilon_{t-1}\,. \qquad (8.10)$$

Нехай випадкові відхилення піддаються впливу авторегресії першого порядку (8.5):

$$\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t\,.$$

де ν_t, t = 2, 3, ... , T, — випадкові відхилення, що задовольняють всім передумовам МНК, а коефіцієнт ρ відомий.

Віднімемо з (8.9) співвідношення (8.10), помножене на ρ:

$$н_e - \rho н_{e-1} = \beta_0 (1 - \rho) + \beta_1(ч_e - \rho\, ч_{e-1}) + (\varepsilon_e - \rho\varepsilon_{e-1})\,. \qquad (8.11)$$

Поклавши $y_t^* = y_t - \rho y_{t-1}$, $x_t^* = x_t - \rho x_{t-1}$, $\beta_0^* = \beta_0(1 - \rho)$, і з урахуванням (8.5) одержимо:

$$y_t^* = \beta_0^* + \beta_1 x_t^* + \nu_t\,. \qquad (8.12)$$

Тому що по припущенню коефіцієнт ρ відомий, то очевидно, y_t^*, x_t* , ν_t обчислюються досить просто. У силу того що випадкові відхилення ν_t задовольняють передумовам МНК, оцінки β_0^* й β_1 будуть мати властивості найкращих лінійних незміщених оцінок.

Однак спосіб обчислення y_t^* , x_t^* приводить до втрати першого спостереження (якщо ми не володіємо попереднім йому спостереженням). Число ступенів волі зменшиться на одиницю, що при більших вибірках не так істотно, але при малих вибірках може привести до втрати ефективності. Ця проблема звичайно переборюється за допомогою *виправлення Прайса—Вінстена:*

$$x_1^* = \sqrt{1-\rho^2}\cdot x_1,$$
$$y_1^* = \sqrt{1-\rho^2}\cdot y_1\,. \qquad (8.13)$$

Авторегресійне перетворення може бути узагальнене на довільне число пояснюючих змінних, тобто використано для рівняння множинної регресії.

Авторегресійне перетворення першого порядку *AR*(1) може бути узагальнене на перетворення більш високих порядків *AR*(2), *AR*(3) і т. ін.:

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \nu_t,$$
$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \rho_3 \varepsilon_{t-3} + \nu_t\,. \qquad (8.14)$$

Однак на практиці значення коефіцієнта ρ звичайно невідомо і його необхідно оцінювати. Існує кілька методів оцінювання. Розглянемо найбільш уживані.

1. Визначення ρ на основі статистики Дарбіна-Уотсона.

Оскільки статистика Дарбіна-Уотсона тісно пов'язана з коефіцієнтом кореляції між сусідніми відхиленнями через співвідношення (8.2):

$$DW \approx \frac{2\left(\sum e_t^2 - \sum e_t e_{t-1}\right)}{\sum e_t^2} \approx 2(1 - r_{e_t e_{t-1}})$$

Тоді як оцінку коефіцієнта ρ може бути взятий коефіцієнт $r = r_{e_t e_{t-1}}$. З (8.2) маємо:

$$r \approx 1 - \frac{DW}{2}. \qquad (8.15)$$

Цей метод оцінювання досить непоганий при великій кількості спостережень. У цьому випадку оцінка r параметра ρ буде досить точною.

2. Метод Кохрана-Оркатта.

Іншим можливим методом оцінювання ρ є ітеративний процес, називаний методом Кохрана-Оркатта. Розглянемо даний метод на прикладі парної регресії (8.8):

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i.$$

і авторегресійної схеми (8.5) першого порядку

$$\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t.$$

а) Оцінюється за МНК регресія (8.8) і для неї визначаються оцінки e_t відхилень e_t, $t = 1, 2, \ldots, T$.

б) З використанням схем $AR(1)$ оцінюється регресійна залежність

$$e_t = \hat{\rho} e_{t-1} + \nu_t, \qquad (8.16)$$

де $\hat{\rho}$ — оцінка коефіцієнта ρ.

в) На основі даної оцінки будується рівняння:

$$(y_t - \hat{\rho} y_{t-1}) = \alpha(1 - \hat{\rho}) + \beta(x_t - \hat{\rho} x_{t-1}) + (\varepsilon_t - \hat{\rho}\varepsilon_{t-1}). \qquad (8.17)$$

за допомогою якого оцінюються коефіцієнти α і ρ (у цьому випадку

значення $\hat{\rho}$ відомо).

г) Значення $\beta_0 = \alpha\ (1 - \hat{\rho})$ і $\beta_1 = \beta$ підставляються в (8.8). Знову обчислюються оцінки e_t відхилень і процес повертається до етапу б).

Чергування етапів здійснюється доти, поки не буде досягнута необхідна точність, тобто поки різниця між попередньою й наступною оцінками ρ не стане менше кожного наперед заданого числа.

3. Метод Хілдрета-Лу.

За даним методом регресія (8.11) оцінюється для кожного можливого значення ρ з відрізка [-1, 1] з будь-яким кроком (наприклад, 0,001; 0,01 і т. ін.). Величина $\hat{\rho}$, що дає найменшу стандартну помилку регресії, приймається як оцінка коефіцієнта ρ. І значення β_0^* й β_1 оцінюються з рівняння регресії (8.11) саме з даним значенням $\hat{\rho}$.

Цей ітераційний метод широко використається в пакетах прикладних програм.

4. Метод перших разностей.

У випадку, коли є підстава вважати, що автокореляція відхилень дуже велика, можна використати метод перших разностей.

Для часових рядів характерна позитивна автокореляція залишків. Тому при високій автокореляції думають $\rho = 1$, і, отже, рівняння (8.11) приймає вид

$$y_t - y_{t-1} = \beta_1(x_t - x_{t-1}) + (\varepsilon_t - \varepsilon_{t-1})$$

або (8.18)

$$y_t - y_{t-1} = \beta_1(x_t - x_{t-1}) + v_t .$$

Позначивши $\Delta y_t = y_t - y_{t-1}$, $\Delta x_t = x_t - x_{t-1}$, з (6.18) одержимо

$$\Delta y_t = \beta_1 \Delta x_t + v_t . \qquad (8.19)$$

З рівняння (8.19) за МНК оцінюється коефіцієнт β_1. Коефіцієнт β_0 у цьому випадку не визначається безпосередньо. Однак із МНК відомо, що $\beta_0 = \bar{y} - \beta_1 \bar{x}$.

У випадку $\rho = -1$, склавши (8.9) і (8.10) з урахуванням (8.5), можна одержати наступне рівняння регресії:

$$y_t + y_{t-1} = 2\beta_0 + \beta_1(x_t + x_{t-1}) + v_t .$$

або (8.20)

$$\frac{y_t - y_{t-1}}{2} = \beta_0 + \beta_1 \frac{x_t - x_{t-1}}{2} + v_t .$$

Однак метод перших разностей припускає занадто сильне спрощення ($\rho = \pm 1$). Тому більше кращими є наведені вище ітераційні методи.

8.5 Моделі розподіленого лагу

При аналізі багатьох економічних показників (особливо в макроекономіці) часто використають щорічні, щоквартальні, щомісячні, щоденні дані. Наприклад, це можуть бути річні дані по ВНП, ВВП, обсягу експорту, інфляції й т.д., місячні дані за обсягами продажу продукції, щоденні обсяги випуску якої-небудь фірми. Для раціонального аналізу необхідно систематизувати моменти одержання відповідних статистичних даних.

У цьому випадку варто впорядкувати дані за часом їхнього одержання й побудувати так звані *часові ряди.*

Нехай досліджується показник *Y*. Його значення в сучасний момент (період) часу *t* позначають y_t; значення *Y* у наступні моменти позначаються y_{t+1}, y_{t+2} ... , y_{t+k}, ...; значення *Y* у попередні моменти позначаються y_{t-1}, y_{t-2}, ..., y_{t-k}, ...

Неважко зрозуміти, що при вивченні залежностей між такими показниками або при аналізі їхнього розвитку в часі в якості пояснюючих змінних використаються не тільки поточні значення змінних, але й деякі попередні за часом значення, а також сам час *T*. Моделі даного типу називають *динамічними.*

У свою чергу змінні, вплив яких характеризується певним запізнюванням, називаються *лаговими змінними*.

Звичайно динамічні моделі підрозділяють на два класи.

1. *Моделі з лагами (моделі з розподіленими лагами)* — це моделі, що містять у якості лагових змінних лише незалежні (пояснюючі) змінні. Прикладом є модель

$$y = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + ... + \beta_k x_{t-k} + \varepsilon_t \qquad (8.21)$$

2. *Авторегресійні моделі* — це моделі, рівняння яких у якості лагових пояснюючих змінних включають значення залежних змінних. Прикладом є модель

$$y_t = \alpha + \beta x_t + \gamma y_{t-1} + \varepsilon_t \qquad (8.22)$$

В економетричному аналізі динамічні моделі використаються досить широко. Це цілком природно, тому що в багатьох випадках вплив одних економічних факторів на інші здійснюється не миттєво, а з деяким тимчасовим запізнюванням - лагом. Причин наявності лагів в економіці досить багато, і серед них можна виділити наступні.

Психологічні причини, які звичайно виражаються через інерцію в поводженні людей. Наприклад, люди витрачають свій дохід поступово, а не миттєво. Звичка до певного способу життя приводить до того, що люди здобувають ті ж блага протягом деякого часу навіть після падіння реального доходу.

Технологічні причини. Наприклад, винахід персональних комп'ютерів не привело до миттєвого витиснення ними більших ЕОМ у силу необхідності заміни відповідного програмного забезпечення, що потребувало тривалого часу.

Інституціональні причини. Наприклад, ті, хто зберігає гроші на довгострокових рахунках, фактично зв'язані, хоча на грошовому ринку можуть бути найбільш вигідні умови.

Механізми формування економічних показників. Наприклад, інфляція багато в чому є інерційним процесом; грошовий мультиплікатор (створення грошей у банківській системі) також проявляє себе на певному тимчасовому інтервалі й т.д.

Оскільки моделі з розподіленими лагами відіграють важливу роль в економіці, встає питання як оцінити невідомі параметри α й β_i. Це можна зробити двома способами:

1. Метод послідовного оцінювання запропонований Ф. Альтом і Дж. Тінбергеном. За даним методом рівняння (8.21) рекомендується оцінювати з послідовно збільшуючи кількість лагів (спочатку будують регресію y_t від x_t й оцінюють параметри, потім y_t від x_t і x_{t-1}, потім y_t від x_t і x_{t-1} і x_{t-2} і так далі). Ця послідовна процедура припиняється, коли параметри при лагові змінних x_t починають бути статистично незалежними або коефіцієнт хоча б одної змінної змінює свій знак.

Однак застосування цього методу досить обмежено в силу постійного зменшення числа ступенів волі, що супроводжується збільшенням стандартних помилок і погіршенням якості оцінок, а також можливості мультиколінеарності. Крім цього, при неправильному визначенні кількості лагів можливі помилки специфікації.

2. Підхід Койка. Койк припустив, що коефіцієнти β_i при лагових значеннях пояснюючої змінної змінюються в геометричній прогресії:

$$\beta_k = \beta_0 \lambda^k, \qquad k = 0, 1, ... \tag{8.23}$$

де $0 < \lambda < 1$ характеризує швидкість убування коефіцієнтів зі збільшенням лага.

Зазначене співвідношення показує, що кожен наступний коефіцієнт β менший, чим попередній, тобто з кожним наступним кроком у минуле вплив лага на y_t поступово зменшується.

У даному випадку рівняння 8.21 перетворюється у

$$y = \alpha + \beta_0 x_t + \beta_0 \lambda x_{t-1} + \beta_0 \lambda^2 x_{t-2} + ... + \varepsilon_t \quad . \tag{8.24}$$

Віднімаючи з рівняння (8.24) таке ж рівняння, але помножене на λ и обчислене для попереднього періоду часу *t-1*

$$\lambda y_{t-1} = \lambda\alpha + \beta_0\lambda x_t + \beta_0\lambda x_{t-1} + \beta_0\lambda^2 x_{t-2} + ... + \lambda\varepsilon_{t-1} \quad , \qquad (8.25)$$

одержимо наступне рівняння

$$y_t - \lambda y_{t-1} = (1-\lambda)\alpha + \beta_0 x_t + (\varepsilon_t - \lambda\varepsilon_{t-1}) \Rightarrow$$

$$y_t = \alpha(1-\lambda) + \beta_0 x_t + \lambda y_{t-1} + \nu_t \,, \qquad (8.26)$$

де $\nu_t = \varepsilon_t - \lambda\varepsilon_{t-1}$.

Ця процедура називається *перетворенням Койка*, що переводить модель з безкінечним числом лагів в авторегресійну.

Порівнюючи первісну модель

$$y = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + ... + \beta_k x_{t-k} + \varepsilon_t$$

с отриманої (8.26) бачимо, що тепер необхідно оцінити тільки α, β_0 й λ.

Крім цього це знімає одну з гострих проблем моделей з лагами – проблему мультиколінеарності.

При застосуванні перетворення Койка можливі наступні проблеми:

• Серед пояснюючих змінних з'являється змінна y_{t-1}, що, у принципі, носить випадковий характер, що порушує одну з передумов МНК. Крім того, дана пояснююча змінна, швидше за все, корелює з випадковим відхиленням ν_t.

• Якщо для випадкових відхилень e_t, e_{t-1} вихідної моделі виконується передумова 3° МНК, то для випадкових відхилень ν_t, мабуть, має місце автокореляція. Для її аналізу замість звичайної статистики DW Дарбіна–Уотсона необхідно використати *h*-статистику Дарбина.

• При зазначених вище проблемах оцінки, отримані за МНК, є зміщеними й неспроможними.

Модель Койка має дві модифікації:

а) модель адаптивних очікувань

$$y_t = \gamma\alpha + \gamma\beta x_t + (1-\gamma) y_{t-1} + \nu_t \,, \qquad (8.27)$$

де γ – коефіцієнт очікування;

$$\nu_t = \varepsilon_t - (1-\gamma)\varepsilon_{t-1} .$$

б) модель часткового коригування

$$y_t = \lambda\alpha + \lambda\beta x_t + (1-\lambda)y_{t-1} + \lambda\varepsilon_t, \qquad (8.28)$$

λ – коефіцієнт коригування.

Таким чином, ми розглянули три моделі, які в загальному виді можна записати як

$$y_t = \alpha_0 + \alpha_1 x_t + \alpha_2 y_{t-1} + \nu_t$$

Встає проблема оцінювання невідомих параметрів цих моделей, оскільки до них не можна прямо застосовувати метод найменших квадратів. Причина неможливості застосування методу найменших квадратів полягає в тому, що змінна y_{t-1} корелює з помилкою ν_t. Якщо якимсь образом усунути цю кореляцію, то можна використати метод найменших квадратів.

Припустимо, що ми знайшли "замінник" для y_{t-1}, що з ним сильно корелює, але не корелює з помилкою ν_t. Такий замінник називається допоміжної змінної. Якщо ввести x_{t-1} як допоміжну змінну для y_{t-1}, то параметри регресії можна одержати, вирішивши систему нормальних рівнянь

$$\begin{aligned}
\sum y_t &= Na_0 + a_1\sum x_t + a_2\sum y_{t-1} \\
\sum y_t x_t &= a_0\sum x_t + a_1\sum x_t^2 + a_2\sum y_{t-1}x_t \\
\sum y_t x_{t-1} &= a_0\sum x_{t-1} + a_1\sum x_t x_{t-1} + a_2\sum y_{t-1}x_{t-1}
\end{aligned}$$

? Запитання для самоперевірки знань

1. Що таке автокореляція?
2. Назвіть основні причини автокореляції.
3. Що може викликати негативну автокореляцію?
4. Яка передумова МНК порушується при автокореляції?
5. Які наслідки автокореляції?
6. Перелічите основні методи виявлення автокореляції.
7. Опишіть схему використання статистики *DW* Дарбіна–Уотсона.
8. Перелічите обмеження використання статистики *DW* Дарбіна–Уотсона.
9. Яка статистика використається для виявлення автокореляції в авторегресійних моделях?
10. Опишіть авторегресійну схему першого порядку *AR*(1).
11. У чому зміст виправлення Прайса-Вінстена?
12. Опишіть способи визначення коефіцієнта автокореляції ρ в авторегресійній схемі першого порядку *AR*(1).
13. Вірні або помилкові наступні твердження? Відповіді поясните.

а) Автокореляція характерна в основному для часових рядів.

б) При наявності автокореляції оцінки, отримані по МНК, є зміщеними.

в) Статистика *DW* Дарбіна–Уотсона не використається в авторегресійних моделях.

г) Статистика *DW* Дарбіна–Уотсона лежить у межах від 0 до 4.

д) Для використання статистики *DW* статистичні дані повинні мати однакову періодичність.

е) Авторегресійна схема першого порядку AR(1) усуває автокореляцію тільки у випадку, коли коефіцієнт автокореляції $\rho = 1$.

ж) При наявності автокореляції значення коефіцієнта детермінації R^2 буде завжди істотно нижче одиниці.

з) Автокореляція завжди є наслідком неправильної специфікації моделі.

14. У чому суть часового ряду?

15. У чому складається розходження між моделями з розподіленими лагами й авторегресійними моделями?

16. Які основні причини лагів в економетричних моделях?

17. Перелічите основні способи визначення оцінок для моделей з розподіленими лагами.

18. У чому суть перетворення Койка?

19. У чому суть моделі адаптивних очікувань?

20. У чому складається відмінність моделі адаптивних очікувань від моделі часткового коректування?

ТЕМА 9. ПОБУДОВА ЕКОНОМЕТРИЧНОЇ МОДЕЛІ НА ОСНОВІ ОДНОЧАСНИХ СТРУКТУРНИХ РІВНЯНЬ

9.1 Поняття економетричних систем рівнянь. Структурна та зведена форма моделі

При вивченні складних економічних явищ економетричний аналіз може базуватися на системі рівнянь, причому деякі змінні y_t, x_t можуть входити більш ніж в одне рівняння, кожне з яких описує змінну та її взаємозв'язок з певними чинниками.

Такі економетричні моделі називаються *системами одночасних структурних рівнянь.*

Система одночасних рівнянь містить *ендогенні* та *екзогенні* змінні.

Ендогенними є ті змінні, які визначаються внутрішньою структурою досліджуваного економічного явища, тобто їх значення вивчаються на основі економетричної моделі. Вони визначені в системі одночасних рівнянь як *y*. Це залежні змінні, число яких дорівнює числу рівнянь в системі.

Екзогенні змінні не залежать від внутрішньої структури економічного явища і їх значення задаються поза моделлю (пояснювальні змінні). Зазвичай вони позначаються як *x*. Вони впливають на ендогенні змінні, але не залежать від них.

Одночасні моделі мають дві форми: структурну і зведену.

Структурна форма моделі створюється у процесі формування моделі, спрямованому на відбиття причинно-наслідкового механізму, що реально існує. Вона дає змогу простежити вплив змін екзогенних змінних моделі на значення ендогенних змінних. Структурна форма моделі має вигляд:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + e_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + e_2, \end{cases} \quad (9.1)$$

де y – ендогенні змінні;

x – екзогенні змінні.

Класифікація змінних на ендогенні та екзогенні залежить від теоретичної концепції прийнятої моделі. Економічні змінні можуть виступати в одних моделях як ендогенні, в інших – як екзогенні. В якості екзогенних змінних можуть розглядатися значення ендогенних за попередній період часу (лагові змінні).

Структурна форма моделі (СФМ) дозволяє визначити вплив зміни екзогенної змінної на значення ендогенної змінної. Необхідно в якості екзогенних змінних вибирати такі змінні, які можуть бути об'єктом регулювання. Заміняючи їх та керуючи ними, можна заздалегідь знати значення ендогенних змінних.

Структурна форма моделі в правій частині містить при ендогенних та екзогенних змінних коефіцієнти b_i та a_j (b_i – коефіцієнт при ендогенній змінній, a_j - коефіцієнт при екзогенній змінній), які називаються ***структурними коефіцієнтами моделі***. Всі змінні у моделі виражені у відхиленнях від середнього рівня, тобто під x розуміється $x - \bar{x}$, а під y – відповідно $y - \bar{y}$. Тому вільний член у кожному рівнянні системи відсутній.

Використання МНК для оцінювання структурних коефіцієнтів моделі надає зміщенні оцінки цих коефіцієнтів. Тому зазвичай для визначення структурних коефіцієнтів моделі структурна форма моделі перетворюється у ***зведену форму моделі***.

Зведена форма моделі (ЗФМ) представляє собою систему лінійних функцій ендогенних змінних від екзогенних:

$$\begin{cases} \hat{y}_1 = c_{11} \cdot x_1 + c_{12} \cdot x_2 + ... + c_{1m} \cdot x_m, \\ \hat{y}_2 = c_{21} \cdot x_1 + c_{22} \cdot x_2 + ... + c_{2m} \cdot x_m, \\ ..... \\ \hat{y}_n = c_{n1} \cdot x_1 + c_{n2} \cdot x_2 + ... + c_{nm} \cdot x_m, \end{cases} \quad (9.2)$$

де c_i – коефіцієнти зведеної форми моделі.

9.2 Проблема ідентифікації. Оцінювання параметрів структурної моделі

Під час перетворення зведеної форми моделі до структурної дослідник стає перед проблемою ідентифікації. Ідентифікація – це єдність відповідності між зведеною та структурною формами моделі.

З позиції ідентифікації структурні моделі можливо поділити на три види:

- точно ідентифіковане;
- не ідентифіковане;
- понад ідентифіковане.

Модель точно ідентифікована, якщо всі її структурні коефіцієнти визначаються одночасно, за коефіцієнтами зведеної форми моделі, тобто якщо число параметрів структурної моделі дорівнює числу параметрів зведеної форми моделі. В такому випадку структурні коефіцієнти моделі оцінюються через параметри зведеної форми моделі та модель точно ідентифікована. Структурна модель

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2, \\ y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 \end{cases} \quad (9.3)$$

з двома ендогенними та трьома екзогенними змінними є точно ідентифікована.

Модель не ідентифікована, якщо число зведених коефіцієнтів менш числа структурних коефіцієнтів, в результаті структурні коефіцієнти не можуть бути оцінені через коефіцієнти зведеної форми моделі. Структурна модель у повному вигляді

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1, \\ y_2 = b_{21}y_1 + a_{22}x_2, \end{cases} \quad (9.4)$$

яка складається з n ендогенних та m екзогенних змінних у кожному рівнянні системи, завжди не ідентифікована.

Модель понад ідентифікована, якщо число зведених коефіцієнтів більш числа структурних коефіцієнтів. В цьому випадку на основі коефіцієнтів зведеної форми моделі можна отримати два або більш значень одного структурного коефіцієнта. В цій моделі число структурних коефіцієнтів менш числа коефіцієнтів зведеної форми. Так, якщо в структурній моделі повного вигляду

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1, \\ y_2 = b_{21}y_1 + a_{22}x_2, \end{cases} \quad (9.5)$$

припустити нульові значення не тільки коефіцієнтів a_{13} та $a_{21,}$ як в моделі

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2 + a_{13}x_3, \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + a_{23}x_3, \end{cases} \quad (9.6)$$

а і $a_{22} = 0$, то система рівнянь стане понад ідентифікована:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2, \\ y_2 = b_{21}y_1 + a_{23}x_{3..} \end{cases} \quad (9.7)$$

Структурну форму системи одночасних економетричних рівнянь можна подати й у матричному вигляді:

$$AY_t + BX_t = u_t, \quad (9.8)$$

де A – невироджена матриця невідомих параметрів при ендогенних змінних розмірності (k x k);

B – матриця ендогенних змінних розмірності (k x 1);

Y_t – вектор екзогенних змінних розмірності (m x 1);

X_t - вектор випадкових залишків розмірності (k x 1).

До структурної форми системи одночасних рівнянь можуть увійти також балансові рівняння або тотожності, які відбивають балансові зв'язки між деякими змінними та об'єднують регресійні рівняння в систему. Характерною ознакою цих моделей є те, що ендогенна змінна, будучи залежною в одному з рівнянь системи, може відігравати роль незалежної, тобто пояснювальної, змінної в іншому рівнянні.

Систему (9.8) можна розв'язати відносно ендогенних змінних Y_{1t}, Y_{2t}, ..., Y_{kt} (припускаємо, що ранг системи дорівнює t). Тоді дістанемо зведену форму:

$$\begin{aligned} Y_{1t} &= r_{11}X_{1t} + r_{12}X_{2t} + ... + r_{1k}X_{kt} + \upsilon_{1t}; \\ Y_{2t} &= r_{2t}X_{1t} + r_{22}X2t + ... + r_{2k}X_{kt} + \upsilon_{2t}; \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ Y_{kt} &= r_{k1}X_{1t} + r_{k2}X_{2t} + ... + r_{kk}X_{kt} + \upsilon_{kt}. \end{aligned} \quad (9.9)$$

Кожна ендогенна змінна у зведеній формі міститься лише в одному рівнянні і залежить (коли не брати до уваги залишків) тільки від екзогенних змінних. Залишки V_t тут є лінійними функціями залишків u_t, а коефіцієнти r_{ij} - лінійними функціями коефіцієнтів структурної форми.

Зведену форму (9.9) у матричному вигляді можна подати так:

$$Y_t \cdot A^{-1*} + B^*X_t = A^{-1*}u_t, \quad (9.10)$$

або

$$Y_t = RX_t + \upsilon_t, \quad (9.11)$$

де $R = -A^{-1}B$, $\upsilon_t = A^{-1}u_t$;

R – *матриця коефіцієнтів зведеної форми розмірності (m x n);*

υ_t – вектор-стовпець, складений з лінійних комбінацій випадкових змінних u_t, присутніх у структурній формі рівняння.

Частинним випадком системи одночасних рівнянь є *рекурсивні системи*, в яких матриця A параметрів ендогенних змінних має трикутний вигляд, а випадкові залишки не корелюють між собою.

Не завжди всі елементи матриць параметрів *A* та *B* можуть бути оцінені. Можливість їх оцінювання пов'язана з проблемою ідентифікації.

Коли система така, що визначення будь-якого структурного рівняння в ній неможливе, то це рівняння є не ідентифікованим і не може бути оцінене взагалі. Якщо умов, які допускають оцінювання, достатньо, але й не більш того, рівняння може бути строго ідентифіковане. І якщо умов більше, ніж потрібно для оцінювання рівняння, маємо його надідентифікацію.

Для оцінювання параметрів системи одночасних структурних рівнянь застосовують спеціальні методи (метод 1 МНК, який оцінює окремо кожне рівняння системи, дуже часто приводить до необґрунтованих оцінок). Найпоширенішими є дво- та трикроковий метод найменших квадратів.

Надідентифіковані рівняння оцінюються за допомогою двокрокового методу найменших квадратів (2 МНК).

Двокроковий метод найменших квадратів (2 МНК) дістав свою назву на тій підставі, що 1МНК тут застосовується на двох етапах.

На першому етапі за допомогою 1 МНК оцінюються параметри кожного регресійного рівняння, зокрема у зведеній формі, тобто оцінюється залежність між певною ендогенною змінною та всіма екзогенними змінними.

На другому етапі 1 МНК застосовується для оцінювання параметрів регресійних рівнянь у структурній формі, але у праву частину таких рівнянь включаються розрахункові значення ендогенних змінних, що дозволяє звільнитися від залежності пояснюючих змінних та стохастичної складової. Якщо такої залежності немає, то для оцінювання параметрів моделі доцільно застосовувати метод 1 МНК.

Формалізовано процедуру оцінювання параметрів 2 МНК опишемо у вигляді алгоритму.

1. Перевіряється кожне рівняння моделі на ідентифікованість. Якщо рівняння точно ідентифіковані або надідентифіковані, то для оцінювання параметрів кожного з них можна використати оператор оцінювання:

$$\begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} Y_1'X(X'X)^{-1}X'Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X \end{bmatrix}^{-1} \begin{bmatrix} Y_1'X(X'X)^{-1}X'Y \\ X_1'Y \end{bmatrix}. \qquad (9.12)$$

2. Знаходження здобутку матриць поточних ендогенних змінних, які містяться у правій частині моделі, то матриці всіх екзогенних змінних моделі, тобто Y'_1X.

3. Обчислення матриці $X' X$ і знаходження оберненої матриці $(X' X)^{-1}$.

4. Визначення добутку $X'Y_1$ матриць усіх екзогенних і всіх ендогенних змінних у правій частині моделі.

5. Знаходження здобутку Y'_1X $(X'X)$ $X'Y_1$матриць, які здобуто на кроках 2, 3, 4.

6. Визначення добутку Y'_1X матриць ендогенних змінних у правій частині моделі і екзогенних змінних, які внесені до даного рівняння.

7. Знаходження добутку XY_1 матриць екзогенних змінних, які входять у дане рівняння, і ендогенних змінних правої частини системи рівнянь.

8. Визначення добутку Y'_1X_1 матриць екзогенних змінних даного рівняння.

9. Знаходження матриці, оберненої до блочної:

$$Q_S^{-1} = \begin{bmatrix} Y_1'X(X'X)^{-1}X'Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X \end{bmatrix}^{-1} \tag{9.13}$$

10. Визначення добутку матриць $X'Y_1$, де X' – матриця всіх ендогенних змінних моделі, Y_1 – вектор залежної ендогенної змінної лівої частини рівняння.

11. Знаходження добутку матриць:

$$q_s = Y_1' X_1' X_1 (X'X)^{-1} X'Y_1. \tag{9.14}$$

12. Визначення параметрів моделі:

$$\begin{bmatrix} \hat{b}_s \\ \hat{a}_s \end{bmatrix} = Q_s^{-1} q_s \,. \tag{9.15}$$

13. Обчислення s-ї залежної ендогенної змінної на основі знайдених параметрів a_s і b_s:

$$\hat{Y}_s = a_s Y_1 + b_s X_1. \tag{9.16}$$

14. Обчислення вектора залишків в s-му рівнянні системи:

$$u_s = Y_s - \hat{Y}_s. \tag{9.17}$$

15. Визначення дисперсії залишків для кожного рівняння:

$$\sigma^2_{u_s} = \frac{1}{n-k-m+1} u_s' u_s \,. \tag{9.18}$$

16. Знаходження матриці коваріацій для параметрів кожного рівняння:

$$\text{asy} \quad \text{var}\begin{bmatrix} \hat{b}_s \\ \hat{a}_s \end{bmatrix} = \sigma^2_{u_s} Q_s^{-1}. \tag{9.19}$$

17. Знаходження стандартної похибки параметрів і визначення довірчих інтервалів:

$$S\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \sqrt{\sigma^2_{u_s} Q_s^{-1}}; \tag{9.20}$$

$$\hat{a} - t_{(\alpha)} S \le a \le \hat{a} + t_{(\alpha)} S; \tag{9.21}$$

$$\hat{b} - t_{(\alpha)} S \le b \le \hat{b} + t_{(\alpha)} S. \tag{9.22}$$

? Запитання для самоперевірки знань

1. Які економетричні моделі називаються системою одночасних структурних рівнянь?
2. Дайте поняття ендогенним та екзогенним змінним. В чому полягає їх основна різниця?
3. Які форми може приймати система одночасних рівнянь? Надайте характеристику кожній із них.
4. Запишіть в загальному вигляді структурну форму моделі на основі одночасних рівнянь.
5. Що означає зведена форма моделі? Як її одержати?
6. В якому випадку системи одночасних рівнянь є рекурсивними системами?
7. Дайте визначення рекурсивних систем і запишіть модель на основі рекурсивної системи.
8. За яких умов рівняння можуть бути строго ідентифіковане та понадідентифіковане?
9. Яка система рівнянь називається точно ідентифікованою?
10. Яка система рівнянь називається надідентифікованою?
11. Яка умова ідентифікованості системи рівнянь?
12. Застосування якого методу використовують при оцінюванні параметрів системи одночасних структурних рівнянь?
13. Охарактеризуйте алгоритм процедури оцінювання параметрів системи одночасних структурних рівнянь.

14. На основі якого методу можна оцінити параметри моделі, якщо вона складається із системи рекурсивних рівнянь?

15. Який метод оцінки параметрів можна застосувати, коли всі рівняння моделі є точно ідентифікованими?

16. На основі якого методу можна оцінити параметри моделі, якщо вона має надідентифіковані рівняння?

17. Чи можна виконувати оцінку параметрів моделі окремо для групи точно ідентифікованих і над ідентифікованих рівнянь?

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

1. Березька К.М. Економетрика: основи теорії та комп'ютерний практикум. Тернопіль: ЗУНУ, 2022. 152 с.

2. Григорків, В.С. Моделювання економіки: підручник. Чернівці : ЧНУ ім. Ю. Федьковича, 2019. 360 с.

3. Диха М. В., Мороз В. С. Економетрія: Навчальний посібник. К.: Центр навчальної літератури (ЦУЛ), 2019. 206 с.

4. Економіко-математичне моделювання: навчальний посібник / за ред. О. Т. Іващука. – Тернопіль : ТНЕУ Економічна думка, 2008. – 704 с.

5. Економетрика : навч. посіб. / О. Є. Лугінін та ін. Херсон : ОЛДІ ПЛЮС, 2016. 320 с.

6. Економетрика : підручник / О. І. Черняк, А. В. Ставицький, О. В. Баженова та ін.; за ред. О. І. Черняка. – 2-ге вид., перероб. та доп. – Миколаїв: МНАУ, 2014. – 414 с.

7. Єрьоменко В., Алілуйко А., Березька К., Мартинюк О. Економетрика : навчальний посібник. Тернопіль: Підручники і посібники, 2023. 168 с.

8. К.Ю. Величко, О.Д. Тімченко. Моделювання та прогнозування міжнародних економічних відносин: методичні вказівки для самостійного вивчення дисципліни для здобувачів першого (бакалаврського) рівня вищої освіти спеціальності 292 Міжнародні економічні відносини ОПП Міжнародна економіка [Електронний ресурс] – Х.: ДБТУ, 2024. – 82 с.

9. Ковальчук О. Я. Математичне моделювання та прогнозування в міжнародних відносинах: Підручник. Тернопіль: ТНЕУ, 2019. 412 с.

10. Козьменко О. В., Кузьменко О. В. Економіко-математичні методи та моделі (економетрика): Навч. посібник. Суми: Університетська книга, 2018. 406 с.

11. Лукяненко І. Г. Економетрика : підручник / І. Г. Лукяненко, Л. І. Краснікова – К. : Товариство «Знання», КОО, 1998. – 494 с.

12. Назаренко О. М. Основи економетрики : підручник вид. 2-ге, допов. та перероб. / О. М. Назаренко. – Київ : Центр навчальної літератури, 2005. – 395 с.

13. Руська Р. В. Економетрика: навчальний посібник. видання 2-е перероб. доп. Тернопіль: ЗУНУ, 2022. 224 с.

14. Тімченко О.Д., Филипенко О.М. Економетрія: Конспект лекцій / Харк. держ. університет харчування та торгівлі.– Харків, 2006. – 113 с.

Навчальне видання

ЕКОНОМЕТРИКА

КУРС ЛЕКЦІЙ

для здобувачів першого (бакалаврського) рівня
вищої освіти спеціальності
051 Економіка
292 Міжнародні економічні відносини
071 Облік і оподаткування
072 Фінанси, банківська справа, страхування та фондовий ринок

ТІМЧЕНКО Ольга Дмитрівна

Формат 60х84/16. Гарнітура Times New Roman
Папір для цифрового друку. Друк ризографічний.
Ум. друк. арк._.
Наклад___пр.
ДБТУ
61002, м. Харків, вул. Алчевських, 44