



МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

Харківський державний університет

харчування та торгівлі

Кафедра економіки та управління

О.М. Филипенко, О.Д. Тімченко

ЕКОНОМЕТРИКА

Навчальний посібник

Харків
ХДУХТ
2021

УДК 330.43(075.8)
ББК 65.051
Е45

Рецензенти:

д-р. екон. наук, проф. В.А. Гросул,
д-р. екон. наук, проф. Н.С. Краснокутська

Рекомендовано до видання

Редакційно-видавничою радою ХДУХТ, протокол № 15 від 17 лютого 2021 р.
і вченою радою ХДУХТ, протокол № 9 від 18 лютого 2021 р.

Е-45 **Филипенко О.М.** Економетрика : навч. посібник / О.М. Филипенко,
О.Д. Тімченко. – Харків : ХДУХТ, 2021. – 170 с.

Навчальний посібник «Економетрика» розраховано на здобувачів економічних спеціальностей. Матеріал посібника спрямовано на активізацію роботи здобувачів під час лекцій і самостійного опрацювання лекційного матеріалу, підвищення ефективності організації самостійної роботи здобувачів, систематизацію знань і формування в студентів цілісного уявлення про предмет курсу.

Навчальний посібник є інструментом вивчення дисципліни «Економетрика», формування у здобувачів системи знань із питань побудови, дослідження та використання економетричних моделей.

УДК 330.43(075.8)
ББК 65.051

© Филипенко О. М., Тімченко О.Д., 2021
© Харківський державний університет
харчування та торгівлі, 2021

ЗМІСТ

Вступ.....	4
Програма курсу.....	6
Тема 1. Природа та сутність економетрики.....	10
Тема 2. Економіка як об'єкт моделювання.....	17
Тема 3. Концептуальні аспекти математичного моделювання економіки	26
Тема 4. Парна лінійна регресія та кореляція в економетричних дослідженнях.....	41
Тема 5. Множинна лінійна регресія та кореляція.....	74
Тема 6. Нелінійна регресія.....	100
Тема 7. Мультиколінеарність як особливий випадок множинної.....	104
Тема 8. Сутність гомо- та гетероскедастичності.....	112
Тема 9. Автокореляція в економетричних моделях.....	123
Тема 10. Побудова економетричної моделі на основі одночасних структурних рівнянь.....	139
Перелік рекомендованої літератури.....	147

ВСТУП

Реформування економіки України створює умови і робить необхідним перехід до нової ідеології управління. Процес прийняття науково обґрунтованих рішень в економіці тісно пов'язаний з визначенням кількісних співвідношень між економічними показниками. Ефективність прийнятих рішень у підприємстві, 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. Звідси очевидно, що економетрію студенти можуть вивчати лише тоді, коли вже засвоїли основні розділи математики для економістів та здобули загальноекономічні знання.

Зауважимо, що економетрія з огляду на громіздкість обчислень та вимоги до точності результатів вивчається за комп'ютерної підтримки.

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

Мета вивчення дисципліни «Економетрика» полягає в тому, щоб навчити здобувачів вищої освіти методам побудови економетричних моделей, які кількісно описують взаємозв'язки між економічними показниками; оцінюванню коректності їх використання; аналізу та інтерпретації результатів економетричних досліджень; здійснення прогнозних розрахунків у майбутньому та прийняття управлінських рішень.

Предметом курсу є економетричні методи та моделі, які дозволяють визначати і вивчати кількісні взаємозв'язки між соціально- економічними явищами.

Завдання вивчення дисципліни: теоретична та практична підготовка здобувачів вищої освіти економічних спеціальностей, яка повинна допомагати майбутнім керівникам самостійно ставити та вирішувати задачі, що пов'язані з виробленням господарських рішень, а також прогнозуванню в майбутньому з урахування того соціально-економічного середовища, в якому функціонують підприємства.

Навчальний посібник за курсом «Економетрика» має за мету активізацію роботи студентів під час лекцій та самостійного опрацювання лекційного матеріалу, підвищення ефективності організації самостійної роботи, систематизацію знань та формування в студентів цілісного уявлення

про предмет курсу.

Навчальний посібник сформований за основними питаннями в межах тем робочої програми курсу «Економетрика», який викладається для студентів економічних спеціальностей і є основою організації цілеспрямованого засвоєння учбового матеріалу, реалізації вимог щодо формування системи певних знань і навичок у студентів.

Посібник за кожною темою курсу супроводжується основними питаннями, які пропонуються до розгляду під час лекцій, питаннями для самоперевірки, тестовими завданнями та завданнями для розв'язання, відповіді на які сприятимуть формуванню в студентів навичок самоаналізу та самоконтролю, а також більш якісному засвоєнню лекційного матеріалу.

Робота студентів із навчальним посібником за кожною темою починається безпосередньо на лекції під час оголошення викладачем основних проблемних питань лекції, і має продовження у вигляді самостійного опрацювання лекційного матеріалу індивідуально кожним здобувачем в позааудиторний час.

ПРОГРАМА КУРСУ

Розділ 1. Основні поняття теорії та методів моделювання

Тема 1. Природа та сутність економетрики

Природа та сутність економетрики як науки. Предмет і метод дисципліни.

Зміст, мета і основні задачі курсу «Економетрика». Взаємозв'язок курсу з попередніми дисциплінами навчального плану, його місце і роль в підготовці бакалаврів. Предмет курсу. Завдання для індивідуальної та самостійної роботи. Форми контролю знань студентів. Теоретичні та прикладні задачі економетрики.

Тема 2. Економіка як об'єкт моделювання

Характеристика економіки, її структури як об'єкта моделювання.

Проблеми методології макроекономічного аналізу.

Ризик, невизначеність та конфліктність розвитку соціально-економічних процесів.

Тема 3. Концептуальні аспекти математичного моделювання економіки

Поняття системи. Основні характеристики економічних систем.

Визначення системи у вузькому та широкому розумінні. Моделювання як метод досліджень економічних систем. Основні методи та етапи дослідження систем. Основні типи економічних систем.

Поняття моделі. Економетрична модель та її структура. Класифікація змінних та параметрів в економетричних моделях. Основні типи економетричних моделей і принципи їхньої класифікації.

Принципи класифікації економетричних моделей, їх типи та види.

Основні поняття економіко-математичного моделювання. Етапи проведення економетричного моделювання

Методи моделювання економічних процесів, етапи проведення математичного моделювання. Етапи економетричного моделювання, його цілі та задачі.

Розділ 2. Основи лінійного моделювання

Тема 4. Парна лінійна регресія та кореляція в економетричних дослідженнях

Специфікація моделі

Загальне поняття про специфікацію економетричної моделі. Випадкова величина та джерела її виникнення. Види помилок. Методи вибору математичної функції.

Загальне поняття про вибірку лінійну регресію. Оцінювання параметрів лінійної регресії

Основи регресійного аналізу. Поняття регресії, її види. Загальне поняття про лінійну регресію. Складові парної лінійної регресії. Оцінка параметрів лінійної регресії за методом найменших квадратів. Властивості простої вибіркової лінійної регресії. Оцінка коефіцієнта регресії.

Поняття тісноти зв'язку, оцінка коефіцієнтів кореляції та детермінації

Поняття про коефіцієнт кореляції. Декомпозиція дисперсій. Поняття про коефіцієнт детермінації. Зв'язок між коефіцієнтом кореляції і коефіцієнтом детермінації.

Оцінка якості лінійного рівняння регресії

Поняття про ступені вільності. Аналіз дисперсій у лінійній регресії. Перевірка простої регресійної моделі на адекватність за допомогою F -критерію Фішера. Середня помилка апроксимації.

Оцінка значущості параметрів лінійної регресії та кореляції. Побудова інтервалів довіри

Перевірка нуль-гіпотези за допомогою t -тесту Стьюдента. Оцінка значущості параметрів лінійної регресії та кореляції. Побудова інтервалів довіри для параметрів a та b .

Побудова інтервалів прогнозу за лінійним рівнянням регресії.

Виконання прогнозу за лінійним рівнянням регресії та побудова його інтервалів.

Тема 5. Множинна лінійна регресія та кореляція

Поняття класичної багатофакторної регресії. Специфікація моделі. Загальне поняття про специфікацію багатофакторної моделі. Вибір чинників при побудові багатофакторної регресії. Вибір форми рівняння багатофакторної регресії. Показники перевірки однорідності вихідної інформації для побудови множинної регресії

Оцінка параметрів багатофакторної регресії

Класична багатофакторна модель. Основні припущення у багатофакторному регресійному аналізі. Оцінка невідомих параметрів багатофакторної регресії за МНК. Оцінка коефіцієнтів регресії у випадку множинної регресії. Стандартизовані коефіцієнти регресії.

Коефіцієнти множинної кореляції та детермінації

Коефіцієнт множинної кореляції та детермінації. Коефіцієнт детермінації та оцінений коефіцієнт детермінації. ANOVA-дисперсійний аналіз.

Частні рівняння множинної регресії

Поняття частних рівнянь. Розклад лінійного рівняння множинної регресії на частні рівняння, їх оцінка.

Оцінка надійності результатів множинної регресії та кореляції. Перевірка моделі на адекватність за допомогою F -критерію Фішера. Оцінка значущості коефіцієнтів регресії та кореляції за допомогою t -тесту Стьюдента.

Розділ 3. Нелінійне моделювання та випадки множинної регресії

Тема 6. Нелінійна регресія

Загальне поняття про нелінійну регресію. Класи нелінійних моделей.

Загальне поняття про нелінійну регресію. Різні види нелінійних моделей.
Два класи нелінійних регресій.

Побудова та оцінка параметрів логарифмічної функції $\ln Y = a + b \ln X$.
Зведення логарифмічної функції до лінійного виду.

Застосування та оцінка параметрів напівлогарифмічної функції $Y = a + b \ln X$ або $\ln Y = a + bX$. Зведення напівлогарифмічних моделей до лінійного виду.

Поняття про нелінійну модель виду $Y = a + b \cdot \frac{1}{X} + e$. Зведення зворотної моделі до лінійного виду шляхом заміни змінних.

Застосування та використання нелінійних моделей в економіці.
Поліноміальна функція виду $Y = a + b_1X + b_2X^2 + \dots + b_mX^m + \varepsilon$. Застосування та використання поліноміальної моделі в економіці. Лінеаризація поліноміальної моделі.

Показова функція $Y = a \cdot e^{bX}$, приклади застосування, оцінювання параметрів та зведення до лінійного виду.

Проблеми та підходи до удосконалення специфікації моделей

Критерії вибору форми моделі. Ознаки «якісної» моделі. Помилки специфікації та їх тестування. Послідовні процедури схеми аналізу залежностей. Проблеми та підходи до удосконалення специфікації моделей

Тема 7. Мультиколінеарність як особливий випадок множинної регресії

Сутність мультиколінеарності

Поняття мультиколінеарності. Умови виникнення мультиколінеарності.

Наслідки мультиколінеарності

Вплив мультиколінеарності та труднощі, пов'язані з оцінкою параметрів.

Визначення мультиколінеарності та методи її усунення

Методи визначення мультиколінеарності та способи її усунення.

Алгоритм Фаррара-Глобера для дослідження мультиколінеарності – лінійної залежності або кореляції між двома і більш незалежними змінними.

Тема 8. Сутність гомо- та гетероскедастичності

Сутність гомо- та гетероскедастичності, їх наслідки. Поняття гомо- та гетероскедастичності. Вплив гетероскедастичності на властивості оцінок параметрів. Алгоритм перевірки на гетероскедастичність: критерій μ , параметричний тест Гольдфельда-Квандта, непараметричний тест Гольдфельда-Квандта, тест Глейсера.

Виявлення гетероскедастичності. Методи пом'якшення проблеми гетероскедастичності

Тести та критерії визначення гетероскедастичності (графічний аналіз відхилень, тест рангової кореляції Спірмана та ін.). Методи пом'якшення гетероскедастичності – перетворення моделі за допомогою методу зважених найменших квадратів (ЗНК).

Тема 9. Автокореляція в економетричних моделях

Сутність і причини автокореляції

Поняття автокореляції. Поняття позитивної та негативної автокореляції залишків. Основні причини виникнення автокореляції (помилки специфікації, інерція тощо).

Наслідки автокореляції

Наслідки автокореляції: неефективні оцінки параметрів, статистично не значимі критерії за t - і F -статистиками.

Виявлення автокореляції

Методи визначення автокореляції – графічний, метод рядів, критерій Дарбіна-Уотсона.

Методи усунення автокореляції

Основні методи усунення автокореляції – авторегресійна схема першого порядку AR(1), виправлення Прайса-Вінстена. Методи оцінювання параметрів за допомогою методу Кохрана-Оркатта, методу Хілдрета-Лу та ін. Поняття динамічного ряду. Типи динамічних рядів. Причини наявності лагів в економіці. Методи оцінки невідомих параметрів в лагових моделях – метод послідовного оцінювання, метод Койка.

Тема 10. Побудова економетричної моделі на основі одночасних структурних рівнянь

Поняття економетричних систем рівнянь. Структурна та зведена форма моделі

Загальне поняття про системи одночасних структурних рівняннях. Визначення ендогенних та екзогенних змінних. Дві форми одночасних моделей - структурна та зведена.

Проблема ідентифікації. Оцінювання параметрів структурної моделі

Ідентифікація як відповідність між зведеною та структурною формами моделі. Три форми моделі з позиції ідентифікації. Алгоритм процедури оцінювання параметрів структурної моделі.

Розділ 1. Основні поняття теорії та методів моделювання

Тема 1. ПРИРОДА ТА СУТНІСТЬ ЕКОНОМЕТРИКИ



Основні питання:

1.1 Природа та сутність економетрики.

1.2 Предмет, методи і завдання економетрики



Ключові слова та поняття: економетрика, об'єкт, предмет, завдання, етапи розвитку, кількісні закономірності, кількісне вимірювання, взаємозв'язки, економічні об'єкти і процеси

1.1 Природа та сутність економетрики

Реформування економіки України створює умови і робить необхідним перехід до нової ідеології управління. Процес прийняття науково обґрунтованих рішень в економіці тісно пов'язаний з визначенням кількісних співвідношень між економічними показниками. Ефективність прийнятих рішень у підприємстві, 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 – розділ економічної науки, у якому вивчаються методи кількісного виміру взаємозв'язків між економічними показниками.

Економетрика – одна з фундаментальних дисциплін у підготовці бакалаврів по економіки для всіх спеціальностей, 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 економетрію студенти можуть вивчати лише тоді, коли засвоїли основні розділи математики й одержали загальноекономічні знання.

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

Термін “економетрика” вперше був наведений бухгалтером П. Цьемпой. Він вважав, що якщо до даних бухгалтерського обліку застосувати методи алгебри і геометрії, то буде отримане нове, більш якісне уявлення про результати господарської діяльності. Ця концепція не знайшла подальшого використання, але термін “економетрика” виявився вдалим для визначення нового напрямку в економічній науці.

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

Визначення предмета економетрії в різних виданнях тлумачиться по-

різному. Сьогодні можна умовно розрізнити кілька підходів до визначення предмета економетрії, що характерні для закордонної економічної літератури. Так, Л. Клейн визначає економетрію як науку, що вивчає вимір зв'язків відповідно економічному аналізу. М. Тинтнер ототожнює економетрію з економічною статистикою, тоді як Г. Хансен під економетрією розуміє застосування математичних і статистичних методів в економіці.

За іншими визначеннями економетрія є синтезом економічної теорії і математики. Деякі автори під економетрією розуміють економічну теорію, математику і статистику.

В журналі “Економетрика”, що був заснований у 1933 р. Р. Фрішем, він дав наступне визначення економетрії: “Економетрика – це не теж саме, що економічна статистика. Вона не ідентична і тому, що ми називаємо економічною теорією, хоча значна частина цієї теорії носить кількісний характер. Економетрія не є синонімом доповнень математики до економіки. Як показує досвід, кожна з трьох відправних крапок – статистика, економічна теорія і математика – необхідна, але не достатня, умова для розуміння кількісних співвідношень у сучасному економічному житті. Це об'єднання всіх трьох складових. І це об'єднання формує економетрику”.

За класичним визначенням, *економетрія – це наука, що вивчає кількісні закономірності і взаємозв'язки економічних об'єктів і процесів за допомогою математико – статистичних методів.*

Але не можна стверджувати, що визначено однозначне ставлення до поняття економетрія. Так, Е. Маленво дотримувався більш широкого розуміння цього поняття, інтерпретуючи економетрію “як будь-яке застосування математики або статистичних методів до вивчення економічних явищ”. О. Ланге писав, що економетрія займається визначенням конкретних кількісних закономірностей економічного життя, застосовуючи до цієї цілі статистичні методи. Таким чином, статистичний підхід до економетричних вимірювань став домінуючим.

Економетрія 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.

Економетрія 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 економетрії полягає в обчисленні відповідних кількісних оцінок. Інакше кажучи, економетрія забезпечує кількісну сторону економічної теорії.

На відміну від чистої математичної економіки, що виражає економічну теорію в математичній формі без мети вимірювання, економетрія 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 економетрія базується на математичній статистиці. Наприклад, одним зі способів, що найбільше часто застосовується в економетрії, є метод найменших квадратів. Використання цього методу ставить перед теоретичною економетрією задачу: детально розглянути припущення, властивості методу найменших квадратів і т.д.

У прикладній економетриці безпосередньо використовуються методи теоретичної економетрії, 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 економетрія 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.

1.2 Предмет, методи і завдання економетрики

Об'єктом економетрики є економічні системи різного рівня складності: від окремого підприємства чи фірми до економіки галузей, регіонів, держави

й світу загалом.

Предмет економетрики – це методи побудови та дослідження математико-статистичних моделей економіки, проведення кількісних досліджень економічних явищ, пояснення та прогнозування розвитку економічних процесів.

До основних задач економетрії 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 називаний етап *параметризації*;

– перевірка якості знайдених параметрів моделі і самої моделі в цілому. Іноді цей етап аналізу називають етапом *верифікації*;

– використання побудованих моделей для пояснення поведінки досліджуваних економічних показників, прогнозування і пророкування, а також для осмисленого проведення економічної політики.

Роль економетрики в економічній науці неухильно зростає. Фактично економетрика відіграє роль основного методологічного інструмента в економіці. З її допомогою підтверджують або відхиляють економічні теорії, а також встановлюють межі їх застосування. У якості кінцевих прикладних цілей економетричного дослідження виділяють дві основні:

– прогноз економічних показників, що характеризують стан і розвиток аналізованої системи (ВВП, інфляція, прибуток, зарплата, обсяги виробництва, рівноважна ціна тощо);

– імітація різноманітних можливих сценаріїв соціально-економічного розвитку аналізованої системи, коли статистично виявлені взаємозв'язки між її різними характеристиками використовуються для відстеження того, як можливі зміни тих чи інших параметрів впливають на значення характеристик, що нас цікавлять.



Запитання для перевірки засвоєних знань

1. Дайте визначення предмета курсу економетрика.
2. Які існують альтернативні підходи до визначення економетрики?
3. Зв'язок економетрики з іншими дисциплінами.
4. Наведіть етапи розвитку економетрики як економічної науки.
5. Що є об'єктом та предметом економетрики?
6. Які задачі економетричного дослідження?
7. Дайте характеристику структури економетричних досліджень.
8. У чому полягає призначення економічної моделі?
9. Які основні завдання прикладної економетрики?
10. Що розуміється під специфікацією моделі?

11. Що розуміється під верифікацією моделі?



Завдання для самостійного розв'язання



Тестовий контроль для перевірки знань

1. У рамках кількісного економічного аналізу розглядаються гіпотези, які
 - а) пов'язані з перевіркою практичної цінності економетричної моделі;
 - б) містять спробу пояснення в самому загальному виді реальних ситуацій;
 - в) пов'язані з організацією вихідних даних;

2. Економетрика - це
 - а) наука, що за допомогою статистичних методів намагається встановити кількісні взаємозв'язки між економічними змінними;
 - б) наука, що зв'язана згодом емпіричних даних з тими значеннями, які можна одержати в результаті побудови моделі;
 - в) наука, що пов'язана з обробкою, описом вихідної економічної інформації.

3. До прикладних задач економетрики не відносяться:
 - а) організація даних;
 - б) перевірка практичної цінності економетричної моделі;
 - в) формування моделі;
 - г) методика статистичного дослідження.

4. Виберіть найбільш точне визначення економічної моделі:
 - а) логічний опис того, що вважається особливо важливим при дослідженні даної проблеми;
 - б) логічне (звичайно математичне) опис того, що виходячи з попереднього аналізу, економічна теорія вважає особливо важливим при дослідженні даної проблеми;
 - в) система рівнянь, що характеризує виділені дослідником взаємозалежності.

5. Призначення економічної моделі полягає
 - а) у поясненні того або іншого економічного економічного явища;
 - б) у виділенні взаємозалежності між більш-менш добре взаємодіючими змінними;
 - в) у поясненні того або іншого економічного явища або його прогнозу.
6. Виберіть ту форму, за допомогою якої не можна представити економічну модель:

- а) система рівнянь;
- б) графік;
- в) таблиця.

7. Макроекономічна модель вирішує наступні проблеми:

- а) характеристика розвитку окремих товарних ринків;
- б) характеристика поведіння індивідуальних учасників господарського процесу;
- в) характеристика поведіння окремих учасників господарського процесу на рівні всього народного господарства.

8. Специфікація моделі в загальному виді являє собою:

- а) відбір домінуючому факторі в моделі;
- б) побудова моделі;
- в) формулювання математичної функції.

9. Верифікації моделі в загальному виді являє собою:

- а) відбір домінуючому факторі в моделі;
- б) перевірка якості знайдених параметрів моделі і самої моделі;
- в) формулювання математичної функції.

10. Найбільш наочним методом вибору математичної функції є:

- а) графічний;
- б) аналітичний;
- в) експериментальний.



Завдання для розв'язання

Завдання 1.

Встановіть відповідність між термінами та їх визначеннями:

	Термін		Визначення
1	Кількісний економічний аналіз	<i>A</i>	Логічний опис того, що виходячи з попереднього аналізу економічна теорія вважає особливо важливим при дослідженні
2	Емпіричне твердження	<i>B</i>	Розгляд гіпотез, які містять спробу пояснення в самому загальному виді реальних ситуацій
3	Економетрика	<i>B</i>	Статистичне дослідження, що має справу з вибірковими даними й використовує певні

	Термін		Визначення
--	--------	--	------------

4	Теоретичні задачі економетрії	<i>Г</i>	Формування моделі, організація даних, перевірка практичної цінності економетричної моделі
5	Задачі прикладної економетрії	<i>Д</i>	Спеціальна область знань, що охоплює питання застосування статистико-математичних методів до теоретичних моделей, що описують реальні економічні процеси
6	Економічна модель	<i>Є</i>	Модель, у якій змінні формуються шляхом підсумовування поведження окремих
7	Макроекономічна модель	<i>Ж</i>	Твердження, що може бути зіставлено (з метою підтвердження або спростування) зі спостережуваними фактами
9	Мікроекономічна модель	<i>І</i>	Модель, що описує розвиток окремих товарних ринків і поведження індивідуальних учасників на них

Тема 2. ЕКОНОМІКА ЯК ОБ'ЄКТ МОДЕЛЮВАННЯ



Основні питання:

2.1 Характеристика економіки, її структури як об'єкта моделювання.

2.2 Проблеми методології макроекономічного аналізу

2.3 Ризик, невизначеність та конфліктність розвитку соціально-економічних процесів



Ключові слова та поняття: модель, економічний аналіз, інструменти, соціально-економічна система, метод моделювання, адекватність, макроекономічний аналіз невизначеність, ризик

2.1 Характеристика економіки та її структури як об'єкта моделювання

Необхідно зауважити, що термін «економіка» в українській мові має два значення: економіка може означати як науку, так і господарську систему. Основне призначення *економіки як господарської системи* – забезпечення суспільства предметами споживання та послугами, котрі створюють умови для життя і безпеки людини, суспільства, країни. *Економіка як економічна теорія* – це суспільна наука, що вивчає аспекти прийняття рішень людьми, які намагаються задовольнити свої матеріальні бажання, використовуючи обмежені ресурси.

Слід наголосити, що економічна теорія у цілому та макро- і мікроекономіка зокрема, ґрунтується на органічному поєднанні емпіричних спостережень з *моделями*.

Модель – це уявний або реальний об'єкт, який у процесі свого вивчення замінює об'єкт-оригінал.

З точки зору використання моделей економічна теорія мало чим відрізняється від інших наук. Причини використання модельного підходу різноманітні, проте, мабуть, одними з найголовніших є *недоступність об'єкта-оригіналу* або *небезпечність безпосереднього експерименту з цим об'єктом*.

Оскільки економіка – це наука про людей, то експериментувати потрібно над людьми, що м'яко кажучи, не завжди морально та етично. Такі явища, як, наприклад, екологічні катастрофи тощо, краще описувати і передбачати на модельному рівні. Тому й створюються моделі катастроф, щоб вони ніколи не сталися. Економіст має справу не лише з моделями того, що вже відбулося чи відбудеться, а й того, що може не відбутися ніколи.

Зауважимо, що, спектр різновидів моделей, котрі використовують економісти, досить широкий. *Модель є інструментом економічного аналізу*. І чим різноманітніший інструментарій має у своєму арсеналі економіст, тим

більша ймовірність отримання вірогідних результатів.

Тисячолітня історія науки свідчить, що без теорії немає науки. І економічна наука не може претендувати у цьому розумінні на якусь винятковість. Економіст «утопить» в океані фактів і чисел, якщо не буде озброєний методами їх систематизації, оброблення, пояснення, передбачення. А це й є теорія. Теорія ж неможлива без моделей. *Немає науки без теорії, а теорії немає без моделей.*

Нагадаємо, що економіко-математичне моделювання разом з іншими економічними дисциплінами мають спільний об'єкт дослідження – *економіку як соціально-економічну систему.*

Соціально-економічна система – це складна ймовірнісна динамічна система, що охоплює процеси виробництва, обміну, розподілу й споживання матеріальних та інших благ. Соціально-економічні системи належать до класу *кібернетичних, тобто керованих, систем.*

Але слід зазначити, що економіко-математичні методи й моделі мають свій напрям дослідження економіки (соціально-економічних систем), свій власний предмет – передусім за все, це кількісні та логічні взаємозв'язки й закономірності поточного функціонування і розвитку соціально-економічних систем.

Центральним поняттям кібернетики є поняття «система». Єдиного означення цього поняття поки що не існує. Можливим є таке формулювання:

Системою називають комплекс взаємопов'язаних елементів разом із відношеннями (зв'язками) між цими елементами та між їхніми атрибутами, які спільно реалізують певні цілі.

Основним методом дослідження систем є *метод моделювання*, тобто спосіб теоретичних і практичних дій, спрямованих на створення та використання моделей. Метод моделювання ґрунтується на принципі аналогії, тобто можливостях вивчення реального об'єкта не безпосередньо, а шляхом дослідження подібного йому й більш доступного дослідженню об'єкта – його моделі.

Одним із важливих аспектів у економіко-математичному моделюванні є поняття адекватності моделі, тобто відповідності моделі модельованому об'єктові чи процесові. Адекватність моделі – дещо умовне поняття, оскільки повної відповідності моделі реальному об'єктові не може бути. Йдеться не просто про адекватність, а про відповідність тим властивостям, які вважаються суттєвими для дослідника, відповідають меті дослідження та усталеній системі гіпотез. Зазначимо, що перевірка адекватності економіко-математичних моделей є непростюю. Вона обтяжена складністю вимірювання економічних величин. Але без такої перевірки застосування результатів моделювання в аналізі та управлінських рішеннях може не лише виявитися малокорисним, а й призвести до негативних наслідків.

Слід наголосити, що соціально-економічні системи належать до так званих складних систем, яким притаманна низка властивостей, котрі потрібно враховувати під час моделювання. Інакше неможливо твердити про

адекватність побудованої моделі. Серед цих властивостей зазначимо такі:

- *емерджентність* як прояв у найяскравішій формі властивості цілісності системи, тобто наявність у економічній системі таких властивостей, які не є притаманними жодному з її елементів, котрий розглядається окремо, поза системою;

- *динамічність* економічних процесів, що полягає в зміні у часі параметрів і структури економічних систем під впливом як внутрішніх, так і зовнішніх чинників (навколишнього середовища);

- *невизначеність* щодо розвитку економічних явищ (процесів);

- *неможливість* ізолювати процеси, що здійснюються в економічних системах, від процесів у навколишньому середовищі з тим щоб спостерігати та досліджувати їх незалежно;

- *активна* реакція на нові чинники, що з'являються.

Використовуючи *системний підхід* у дослідженні економіки, вирізняють, зокрема, макро- та мікроекономічні моделі. Перші відображають функціонування та розвиток усієї економічної системи чи її великих підсистем, другі – господарських одиниць та їх об'єднань.

2.2 Проблеми методології макроекономічного аналізу

Зауважимо, що розвиток сучасної макроекономічної теорії відбувається значною мірою на іншій методологічній основі порівняно з викладеною у підручниках для вищих економічних навчальних закладів теоретичною базою. Проблема полягає у значних семантичних розбіжностях між «канонічним» викладенням макроекономічних моделей і сучасним станом макроекономічного моделювання.

У сучасному макроекономічному моделюванні домінує підхід, що ґрунтується на динамічних моделях раціональної поведінки типового агента на ефективному ринку. Така поведінка формує раціональні очікування в умовах невизначеності, ризику та відсутності арбітражу. Моделювання макроекономічної поведінки у багатьох випадках уявляється як задача оптимального управління, зокрема динамічного програмування Р. Беллмана, що формулюється для детермінованих чи стохастичних процесів.

Наголосимо, що у макроекономічному моделюванні вирішальну роль відіграє адекватне формулювання якісної гіпотези, що характеризує, зокрема, межі використання певної макроекономічної моделі.

Різні макроекономічні процеси, наприклад, позики держави на вільному ринку, емісія грошей, інфляція тощо, можуть бути змодельовані випадковими процесами, що описуються стохастичними диференціальними рівняннями. Однак паралельно стохастичній інтерпретації макроекономічних процесів останнім часом інтенсивно розвивається альтернативний, у певному сенсі, підхід до аналізу макроекономічної динаміки, який вбачає джерело і природу невизначеності у принциповій *нелінійності* економічних процесів. Виявлення множинності точок рівноваги, граничних циклів, бифуркацій і хаосу, асиметрії інформації змушує економістів по-новому аналізувати макроекономічні

системи. Теорія сучасної нелінійної динаміки, наприклад методи обчислення розмірності фрактала, дає змогу обчислити для таких процесів, зокрема, автокореляцію, хоча природа таких залежностей зовсім інша, ніж стохастичних процесів.

Незважаючи на сучасний бурхливий розвиток наукового підходу – «синергетичної економіки», що спирається на суттєву нелінійність економічних процесів, багатьох науковців у науковій та навчальній літературі для спрощення обмежуються здебільшого лінійними моделями.

У математичному аспекті важливим є поняття лінійності, котре означає, що будь-яка лінійна комбінація розв'язків також буде розв'язком. Використовуючи цей принцип, можна побудувати розв'язок для більш загальної ситуації, відшукавши розв'язки в часткових випадках. Тому про якісні властивості у загальному випадку можна судити виходячи з властивостей часткових. У лінійних моделях відгук (реакція) об'єкта на зміну умов є пропорційним величині цих змін.

Для явищ, математичні моделі котрих є нелінійними, знання про поведінку частини об'єкта не гарантують знань про поведінку об'єкта в цілому, а його відгук на зміну умов може якісно залежати від кількісної величини (обсягів) цих змін. Більшість реальних процесів і відповідних (адекватних) їм математичних моделей є нелінійними. Лінійні ж моделі відповідають частковим випадкам і, як правило, послугують лише як перше наближення до реальності.

2.3 Ризик, невизначеність та конфліктність розвитку соціально-економічних процесів

З погляду теорії систем економіку слід віднести до класу динамічних, слабо структурованих систем великої складності. Крім того, економіка має яскраво виражену ієрархічну, багаторівневу структуру, за якої більш вищий рівень ієрархії інтегрує за певними правилами (алгоритмами) інформаційні сигнали (потoki) нижчих рівнів ієрархії та оперує інформаційними агрегатами.

Економічна кібернетика розглядає економіку, її структурні та функціональні блоки як системи, в яких відбуваються процеси регулювання й управління, що реалізуються рухом і перетворенням інформації. Наголосимо, що інформація, з одного боку, є генератором розвитку соціально-економічної системи, і водночас, з іншого – джерелом невизначеності та породженого цим ризику в економіці. Тут принципово важливим є інноваційний характер інформації, яка, за означенням, не може бути повністю передбачуваною.

Невизначеність – це те, що не піддається однозначній оцінці. Наголосимо, що економічний ризик породжується невизначеністю (невідомістю, недостовірністю, неоднозначністю), конфліктністю, наявністю альтернатив щодо рішень, планів, проектів і внутрішньо притаманний економіці та бізнесу.

У низці випадків з огляду на обрані цілі доводиться приймати рішення

на підставі побудови системи гіпотез, зокрема, через відсутність вичерпної, достовірної інформації. А це призводить до ризику відхилення від цілей, недоотримання очікуваних результатів, можливих збитків, які можуть виникнути через недостатню обґрунтованість (помилковість, неадекватність) тих чи інших гіпотез (припущень).

Ступінь ризику залежить також і від ставлення суб'єкта прийняття рішень (управлінської команди) до нього: схильності, несхильності, байдужості. Тому всі чинники невизначеності, конфліктності та зумовленого ними ризику розділяють на об'єктивні та суб'єктивні.

Отже, можна дати таке означення економічного ризику, що базується на принципах системного аналізу.

***Ризик**, яким обтяжена економічна діяльність, – це економічна категорія, що відбиває особливості сприйняття заінтересованими суб'єктами економічних відносин об'єктивно існуючих невизначеності та конфліктності, іманентних процесам цілепокладання, управління, прийняття рішень, оцінювання.*

Існує ціла система показників кількісної оцінки ступеня ризику. Але від жодного кількісного показника (кількісної оцінки) ступеня ризику не слід очікувати, що він показуватиме адекватні результати за будь-яких обставин. Тобто встановлення певного єдиного показника, як кількісної міри (ступеня) ризику, є спробою подолати невизначеність, характеризуючи випадкову величину (ефективність, збитки) одним числом. На нашу думку, кількісна оцінка ризику є багатовимірною величиною (вектором), компоненти якої відбивають різні грані ризику й формуються залежно від мети дослідження, прийнятої системи гіпотез, суб'єктивного чинника, що характеризує ставлення суб'єкта ризику до невизначеності та ризику, тощо.



Запитання для перевірки засвоєних знань

1. Яке основне призначення економіки як господарської системи?
2. Що Ви розумієте під терміном “модель”?
3. Що є причинами використання модельного підходу?
4. У чьому полягає сутність понять система та соціально-економічна система?
5. Який основний метод дослідження систем? На якому принципі він ґрунтується?
6. Охарактеризуйте властивості соціально-економічних систем.
7. У чому полягає сутність та роль макроекономічного моделювання?
8. У чому сутність лінійності та нелінійності економічних процесів?
9. У чому сутність невизначеності та конфліктності розвитку соціально-економічних процесів?
10. У чому сутність ризику розвитку соціально-економічних процесів?



Завдання для самостійного розв'язання

Тестовий контроль для перевірки знань

1. Основне призначення економіки як господарської системи:
 - а) забезпечення суспільства послугами, котрі створюють умови для життя і безпеки людини, суспільства;
 - б) забезпечення суспільства предметами споживання та послугами, котрі створюють умови для життя і безпеки людини, суспільства;
 - в) забезпечення суспільства ресурсами для того, щоб задовольнити свої матеріальні бажання;
 - г) жодної правильної відповіді.

2. Модель – це:
 - а) складна ймовірнісна динамічна система, що охоплює процеси виробництва, обміну, розподілу й споживання матеріальних та інших благ;
 - б) це уявний або реальний об'єкт, який у процесі свого вивчення замінює об'єкт-оригінал;
 - в) це абстракція реальної дійсності, в якій відношення між реальними елементами, а саме ті, що цікавлять дослідника, замінені відношеннями між математичними категоріями;
 - г) жодної правильної відповіді.

3. Яка властивість соціально-економічної системи означає – наявність у економічній системі таких властивостей, які не є притаманними жодному з її елементів, котрий розглядається окремо, поза системою:
 - а) емерджентність;
 - б) динамічність;
 - в) невизначеність;
 - г) активна реакція на нові чинники.

4. Яка властивість соціально-економічної системи означає – зміні у часі параметрів і структури економічних систем під впливом навколишнього середовища:
 - а) емерджентність;
 - б) невизначеність;
 - в) динамічність;
 - г) активна реакція на нові чинники.

5. Поняття, що означає певний вид рівнянь, що містять шукані величини в ступенях більше одиниці або коефіцієнти, що залежать від властивостей середовища – це:
 - а) нелінійність;

- б) лінійність;
- в) синергія;
- г) жодної правильної відповіді.

6. До причин не лінійності економічних процесів відносяться:

- а) множинність точок рівноваги;
- б) множинність граничних циклів;
- в) асиметрії інформації;
- г) всі відповіді вірні.

7. Можлива небезпека втрат, небезпечні події, які можуть відбутися і в результаті яких виникають збитки, майнова втрата в учасників підприємницької діяльності – це:

- а) невизначеність;
- б) синергія;
- в) ризик;
- г) всі відповіді вірні.

8. Ступінь ризику залежить від:

- а) схильності суб'єкта прийняття рішень;
- б) несхильності суб'єкта прийняття рішень;
- в) байдужості суб'єкта прийняття рішень;
- г) всі відповіді вірні.

9. До причин виникнення ризику не відноситься:

- а) конфліктність;
- б) неоднозначність;
- в) емерджентність;
- г) інформація.

10. Наука, що розглядає економіку а також її структурні і функціональні ланки як системи, в яких відбуваються процеси регулювання й управління, що реалізуються рухом і перетворенням інформації – це:

- а) економетрика;
- б) економічна кібернетика;
- в) інформатика;
- г) макроекономіка.



Завдання для розв'язання

Завдання 1.

Встановіть відповідність між термінами та їх визначеннями:

	Термін		Визначення
1	Економіка як економічна теорія	<i>Г</i>	суспільна наука, що вивчає аспекти прийняття рішень людьми, які задовольняють свої матеріальні бажання, використовуючи обмежені ресурси
2	Економіко-математичні методи	<i>Є</i>	математичні методи і моделі в аналізі, плануванні, прогнозуванні й управлінні економічними об'єктами та процесами
3	Модель	<i>К</i>	це уявний або реальний об'єкт, який у процесі свого вивчення замінює об'єкт-оригінал
4	Соціально-економічна система	<i>Д</i>	складна ймовірнісна динамічна система, що охоплює процеси виробництва, обміну, розподілу й споживання матеріальних та інших благ
5	Система	<i>А</i>	комплекс взаємопов'язаних елементів разом із відношеннями (зв'язками) між цими елементами та між їхніми атрибутами, які спільно реалізують певні цілі
6	Метод моделювання	<i>З</i>	спосіб теоретичних і практичних дій, спрямованих на створення та використання моделей
7	Адекватність моделі	<i>Л</i>	відповідність моделі модельованому об'єктові чи процесові
8	Емерджентність	<i>І</i>	прояв у найяскравішій формі властивості цілісності системи
9	Динамічність економічних процесів	<i>Ж</i>	зміна у часі параметрів і структури економічних систем під впливом навколишнього середовища
10	Невизначеність	<i>Б</i>	це неповноцінність чи неточність інформації про умови підготовки та реалізації господарських рішень
11	Ризик	<i>В</i>	економічна категорія, що відбиває особливості сприйняття заінтересованими суб'єктами економічних відносин об'єктивно існуючих невизначеності та конфліктності

Завдання 2.

Обведіть правильну відповідь, визначену літерою під запитанням:

1. Твердження «На ідеї моделювання по суті базується будь-який метод наукового пізнання як теоретичний, так і експериментальний»

- А) правильне Б) неправильне

2. Економіко-математичне моделювання – це комплекс економічних і математичних дисциплін.

- А) так Б) ні

3. Система – це модель прийняття управлінських рішень.

- А) так Б) ні

4. Модель – це оригінал, яким управляємо або повинні управляти.

- А) так Б) ні

5. Твердження «низько структуровані проблеми, вони пов'язані з розробкою довгострокових напрямів дій, які висвітлюють багато аспектів у діяльності підприємств. У цьому разі важко або майже неможливо описати математичні зв'язки. Ці проблеми вирішуються переважно з використанням методології системного аналізу».

- А) правильне Б) неправильне

Тема 3. КОНЦЕПТУАЛЬНІ АСПЕКТИ МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ ЕКОНОМІКИ



Основні питання:

- 3.1 Сутність моделювання як методу наукового пізнання.
- 3.2 Особливості та принципи математичного моделювання.
- 3.3 Основні дефініції економіко-математичного моделювання.
- 3.4 Особливості економічних спостережень і вимірів.
- 3.5 Етапи економіко-математичного моделювання.
- 3.6 Елементи класифікації економіко-математичних моделей.
- 3.7 Роль прикладних економіко-математичних досліджень.



Ключові слова та поняття: математичне моделювання, математична модель, форма, об'єкти, методи, принципи, дефініції, етапи моделювання, класифікація

3.1 Сутність моделювання як методу наукового пізнання

Модель від лат. («modulus» – зразок, норма, міра.) – це об'єкт, що заміщує оригінал і відбиває його найважливіші риси й властивості для даного дослідження, даної мети дослідження за обраної системи гіпотез.

Математична модель – це абстракція реальної дійсності (світу), в якій відношення між реальними елементами, а саме ті, що цікавлять дослідника, замінені відношеннями між математичними категоріями. Ці відношення зазвичай подаються у формі рівнянь і/чи нерівностей, відношеннями формальної логіки між показниками (змінними), які характеризують функціонування реальної системи, що моделюється.

Сутність методології математичного моделювання полягає в заміні досліджуваного об'єкта його «образом» – математичною моделлю – і подальшим вивченням (дослідженням) моделі на підставі аналітичних методів та обчислювально-логічних алгоритмів, які реалізуються за допомогою комп'ютерних програм. Робота не із самим об'єктом (явищем, процесом), а з його моделлю дає можливість відносно швидко і безболісно досліджувати його основні (суттєві) властивості та поведінку за будь-яких імовірних ситуацій (це переваги теорії). Водночас обчислювальні (комп'ютерні, симулятивні, імітаційні) експерименти з моделями об'єктів дозволяють ретельно та досить глибоко вивчати об'єкт, що недоступно суто теоретичним підходам (це перевага експерименту).

Вже сама постановка питання щодо математичного моделювання будь-якого об'єкта породжує чіткий план дій, який умовно можна поділити на три етапи: модель – алгоритм – програма (рис.3.1).

На *першому етапі* обирається (чи будується) «еквівалент» об'єкта, що відображає в математичній формі найважливіші (ключові) його властивості – закони, яким він підпорядковується, зв'язки, що притаманні складовим його

частинам, тощо. Математична модель (чи її фрагменти) досліджуються теоретичними методами, що дає змогу отримати важливі (концептуального характеру) нові знання про об'єкт.

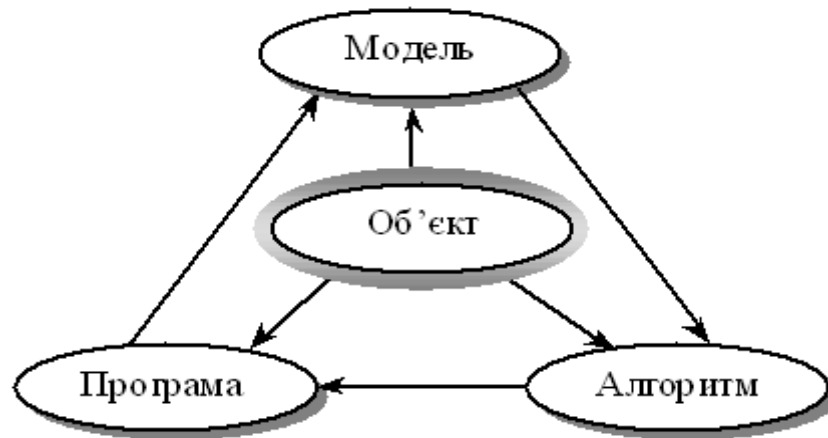


Рисунок 3.1 – Узагальнена схема математичного моделювання

Другий етап – вибір (чи розроблення) алгоритму для реалізації моделі на комп'ютері. Модель подається у формі, зручній для застосування числових методів, визначається послідовність обчислювальних і логічних операцій, котрі необхідно здійснити, щоб отримати шукані величини із заданою точністю.

На *третьому етапі* створюються програми, що «переносять» модель і алгоритм на доступну комп'ютерну мову. Їх можна назвати «електронним» еквівалентом досліджуваного об'єкта, що є придатним для безпосереднього експериментування на комп'ютері.

Створивши *тріаду*: «модель – алгоритм – програма», дослідник (системний аналітик) отримує універсальний, гнучкий і відносно дешевий інструмент, який тестується в «пробних» обчислювальних експериментах. Після того як *адекватність* (достатній рівень відповідності, зважаючи на цілі та прийняту систему гіпотез) тріади щодо досліджуваного об'єкта засвідчена, з моделлю проводять різноманітні та детальні «досліди», які дають нову інформацію про необхідні якісні та кількісні властивості й характеристики об'єкта. Процес моделювання супроводжується поліпшенням та уточненням, за необхідності, всіх складових (ланок) тріади.

3.2 Особливості і принципи математичного моделювання

Головна особливість моделювання полягає у тому, що це метод опосередкованого пізнання за допомогою об'єктів-замішувачів. Саме ця особливість моделювання визначає специфічні форми використання абстракцій, аналогій, гіпотез, інших категорій і методів пізнання.

Сформулюємо принципи, які визначають ті загальні вимоги, яким повинна задовольняти правильно побудована математична модель деякого об'єкта (системи).

Принцип 1. Полярність діалектичної пари «модель – об’єкт». Ця діалектична пара завжди полярна, має два полюси – «модель» і «об’єкт».

Принцип 2. Первинність об’єкта. З двох взаємно пов’язаних полюсів діалектичної пари «модель – об’єкт» один із них (об’єкт) є первинним, інший (модель) – похідним від нього.

Принцип 3. Зумовленість моделі об’єктом. Наявність полюсу «модель» зумовлює необхідність наявності полюсу «об’єкт».

Принцип 4. Множинність моделей щодо об’єкта дослідження. Як «модель» для об’єкта, так і «об’єкт» для даної «моделі» семантично та інтерпретаційно багатозначні: «об’єкт» описується не однією, а багатьма «моделями», «модель» віддзеркалює властивості не одного, а багатьох «об’єктів».

Принцип 5. Адекватність. Цей принцип передбачає відповідність моделі меті дослідження, прийнятій системі гіпотез за рівнем складності й організації, а також відповідність реальній системі (об’єкту). Доки не вирішено питання, чи правильно відображає модель досліджувану систему (об’єкт), цінність моделі незначна.

Принцип 6. Спрощення за умови збереження суттєвих (ключових) властивостей об’єкта (системи). Модель повинна бути в деяких аспектах суттєво простішою від прототипу — в цьому власне й полягає сенс моделювання, тобто модель ігнорує несуттєві властивості об’єкта. Цей принцип може бути названий принципом абстрагування від другорядних деталей.

Практичні рекомендації щодо зменшення складності моделі:

- зменшення кількості змінних за допомогою виключення несуттєвих змінних або їх об’єднання. Процес перетворення (редукції) моделі в модель з меншою кількістю змінних і обмежень називають *агрегуванням*;

- зміна природи змінних величин й параметрів. Змінні величини й параметри наближено розглядаються як постійні, дискретні – як неперервні тощо;

- зміна функціональної залежності між змінними. Нелінійна залежність замінюється зазвичай лінійною, дискретна функція розподілу ймовірностей – неперервною тощо;

- зміна обмежень (збільшення, виключення чи модифікація). Після зняття обмежень одержуємо оптимістичне рішення, після введення – песимістичне. Варіюючи обмеженнями, можна знайти можливі граничні значення ефекту чи ефективності. Такий спосіб часто застосовують для знаходження попередніх оцінок ефективності рішень на етапі постановки задач;

- обмеження точності моделі. Точність результатів моделі не може бути вищою за точність вхідних даних.

Принцип 7. Блочна побудова. За дотримання цього принципу блочної побудови полегшується розроблення складних моделей і з’являється можливість використання накопиченого досвіду та адаптації готових блоків із

мінімально необхідними зв'язками між ними. Виокремлення блоків відбувається з урахуванням розподілення моделі за етапами й режимами функціонування об'єкта (системи).

Складні об'єкти (системи) потребують розроблення цілої ієрархії моделей. Виокремлюють такі рівні, як вся система, підсистеми, підсистеми керування тощо.

Наголосимо, що використання математичних методів в економічному аналізі жодною мірою не зводиться до підбору прийнятих формул, підстановки в них певних чисел та певного чаклування, в результаті чого виходить «відповідь».

Нагадаємо рекомендації відомого американського вченого Р.Хемінга: «Мета обчислень – розуміння, а не числа»; «перш ніж розв'язувати задачу, подумай, що робити з її розв'язком».

3.3 Основні дефініції економіко-математичного моделювання

Якщо йдеться про математичну модель, що описує механізм функціонування певної гіпотетичної економічної чи соціально-економічної системи, то таку модель називають економіко-математичною чи просто економічною.

Під економіко-математичною моделлю розуміють концентроване вираження найсуттєвіших економічних взаємозв'язків досліджуваних об'єктів (процесів) у вигляді математичних функцій, нерівностей і рівнянь.

Математична модель – це об'єкт, котрий створюється системним аналітиком для отримання нових знань про об'єкт-оригінал і відбиває лише суттєві (з погляду системного аналітика) властивості об'єкта-оригіналу.

Модель вважається адекватною об'єкту-оригіналу, якщо вона з достатнім ступенем наближення, на рівні розуміння системним аналітиком модельованого процесу відбиває закономірності процесу функціонування реальної економічної системи у зовнішньому середовищі.

Як було зазначено, під моделюванням розуміють процес побудови, вивчення й використання моделей.

Процес моделювання включає три системотвірних елементи:

- суб'єкт дослідження (системний аналітик);
- об'єкт дослідження;
- модель, яка опосередковує відносини між об'єктом, який вивчається, та суб'єктом, який пізнає (системним аналітиком).

У загальних рисах можна виокремити чотири основні етапи процесу математичного моделювання економічних систем і подати їх за такою узагальненою схемою (рис. 3.2).

3.4 Особливості економічних спостережень і вимірів

Головним гальмом для практичного застосування математичного моделювання в економіці є проблема наповнення розроблених моделей конкретною та якісною інформацією. Точність і повнота первинної

інформації, реальні можливості її збору й опрацювання справляють визначальний вплив на вибір типів прикладних моделей. З іншого боку, завдання моделювання економіки висувають нові вимоги до системи інформації.



Рисунок 3.2 – Узагальнена схема процесу економіко-математичного моделювання

Залежно від модельованих об'єктів і призначення моделей використовується в них вхідна інформація має суттєво відмінний характер і походження. Вона може бути розподіленою на дві категорії: щодо минулого розвитку та сучасного стану об'єктів (економічне спостереження й опрацювання); про майбутній розвиток об'єктів, яка включає дані про очікувані зміни, внутрішні параметри та зовнішні умови (прогнози). Інша категорія інформації є результатом самостійних досліджень, які також можуть проводитися за допомогою моделювання.

Методи економічних спостережень і використання їхніх результатів розробляються економічною статистикою. З огляду на це варто визначити лише специфічні проблеми економічних спостережень, які стосуються

моделювання економічних процесів. В економіці чимало процесів є масовими: вони характеризуються закономірностями, що не проявляються на підставі лише одного чи кількох спостережень. Тому моделювання в економіці має спиратися на масові спостереження.

Інша проблема породжується динамічністю економічних процесів, мінливістю їхніх параметрів і структурних відношень. Унаслідок цього доводиться постійно вивчати економічні процеси, здійснювати їх моніторинг. Оскільки спостереження за цими процесами й опрацювання емпіричних даних зазвичай забирають досить багато часу, то, будуючи економіко-математичні моделі, необхідно коригувати вхідну інформацію з урахуванням її надходження із деяким запізненням у часі.

Дослідження кількісних відношень економічних процесів і явищ спирається на економічні виміри. Точність проведення вимірювань значною мірою впливає на точність кінцевих результатів кількісного аналізу. Тому застосування математичного моделювання загострило проблему вимірювання та кількісного зіставлення різних аспектів і явищ соціально-економічного розвитку та повноти одержуваних даних, захисту їх від навмисних і технічних викривлень (деформації).

3.5 Етапи економіко-математичного моделювання

В різних галузях знань, зокрема в економіці, етапи моделювання набувають специфічних рис. Проаналізуємо послідовність і зміст етапів одного циклу економіко-математичного моделювання.

1. Постановка економічної проблеми та розроблення концептуальної моделі. Головне на цьому етапі – чітко сформулювати сутність проблеми (цілі дослідження), припущення, що приймаються, і ті питання, на які необхідно одержати відповіді. З урахуванням цілей дослідження проводиться якісний аналіз об'єкта; виокремлюються, абстрагуючись від другорядних, найважливіші риси і властивості об'єкта, що моделюється. З позиції системного підходу вивчаються структура об'єкта й головні взаємозв'язки між його елементами (підсистемами). Обираються та обґрунтовуються основні показники й система гіпотез, що пояснюють поведінку та розвиток об'єкта і на основі яких буде відбуватись подальша формалізація.

На цьому етапі моделювання широко застосовуються якісні методи описання систем, знакові та мовні моделі. Таке попереднє, наближене зображення системи називають концептуальною моделлю.

2. Розроблення математичних моделей. Це етап формалізації економічної проблеми, вираження її у вигляді конкретних математичних залежностей і відношень (функцій, рівнянь, нерівностей тощо). На цьому етапі проводиться теоретичне (аналітичне) дослідження моделі, обираються методи дослідження й розв'язку.

Метою теоретичного (аналітичного) дослідження є з'ясування загальних властивостей моделі. Найважливіший момент – доведення існування розв'язку

для моделі. Знання загальних властивостей моделі настільки важливе, що часто задля доведення подібних властивостей дослідники свідомо йдуть на ідеалізацію первинної моделі. У тому разі, коли аналітичними методами не вдається з'ясувати загальні властивості моделі, а спрощення моделі спричиняється до недопустимих (неадекватних) результатів, переходять до числових методів дослідження.

3. Реалізація моделі у вигляді пакету прикладних програм (ППП) та проведення розрахунків. Цей етап включає розробку алгоритмів для числового розв'язування задачі, складання програм на ЕОМ (можливе використання існуючих ППП з відповідною адаптацією) і безпосереднє проведення розрахунків. Труднощі цього етапу зумовлені передусім великою розмірністю економічних задач, необхідністю опрацювання значних масивів інформації. Завдяки високій швидкодії сучасних ЕОМ вдається проводити числові «модельні» експерименти, вивчаючи «поведінку» моделі за різних значень деяких умов. Дослідження, що проводяться за допомогою числових методів, можуть стати суттєвим доповненням до результатів аналітичного дослідження. Клас економічних задач, які можна розв'язувати числовими методами, значно ширший, ніж клас задач, доступних аналітичному дослідженню.

4. Перевірка адекватності моделі. Вимога адекватності є суперечною вимозі простоти, і це слід враховувати, перевіряючи модель на адекватність. Початковий варіант моделі попередньо перевіряється за такими основними аспектами: чи всі суттєві параметри включені в модель; чи містить модель несуттєві параметри; чи правильно відображені функціональні зв'язки між параметрами; чи правильно визначені обмеження на значення параметрів тощо.

Для встановлення відповідності створюваної моделі оригіналу використовують такі методи:

- порівняння результатів моделювання з окремими експериментальними результатами, одержаними за однакових (подібних) умов;
- використання інших схожих моделей;
- порівняння структури і функціонування моделі з прототипом.

Головним шляхом перевірки адекватності моделі досліджуваного об'єкта виступає практика. Але вона потребує накопичення статистики, котра не завжди буває достатньою для отримання надійних даних. Для багатьох моделей перші два методи виявляються менш прийнятними. Тоді залишається лише один шлях: висновок про подібність моделі та прототипу робити на підставі порівняння їхніх структур і виконуваних функцій. Такі висновки не мають формального характеру, оскільки ґрунтуються на досвіді та інтуїції дослідника.

Згідно з результатами перевірки моделі на адекватність приймається рішення про можливість її практичного використання чи проведення коригування.

5. Аналіз числових результатів та прийняття відповідних рішень.

Результати досліджень подаються у вигляді, зручному для огляду, і на основі обробки отриманих результатів проводиться аналіз матеріалів дослідження моделі. На цьому, завершальному, етапі виникає питання про правильність і повноту результатів моделювання, про можливість практичного застосування останніх, і, найголовніше, про досягнення цілей дослідження.

Звернімо увагу на зворотні зв'язки етапів, які виникають унаслідок того, що в процесі дослідження виявляються недоліки попередніх етапів моделювання. Недоліки, які не вдається виправити на проміжних етапах моделювання, усуваються в наступних циклах. Але результати кожного циклу мають і цілком самостійне значення. Розпочавши дослідження від побудови простої моделі, можна швидко одержати корисні результати, а потім перейти до створення досконалішої моделі.

3.6 Елементи класифікації економіко-математичних моделей

Існують різні форми зображення математичної моделі – інваріантна, алгоритмічна, аналітична, схемна.

Інваріантна форма – зображення математичної моделі безвідносно до методів, за допомогою яких може розв'язуватись поставлена задача моделювання.

Алгоритмічна форма – зображення математичної моделі у вигляді послідовності дій (Симплекс-метод).

Аналітична форма – зображення математичної моделі у вигляді формул та співвідношень між математичними виразами.

Схемна форма – зображення математичної моделі у вигляді графів, таблиць, діаграм.

Для класифікації економіко-математичних моделей використовують різні класифікаційні ознаки.

За цільовим призначенням економіко-математичні моделі поділяються на *теоретико-аналітичні*, що використовуються під час дослідження загальних властивостей і закономірностей економічних процесів, і *прикладні*, що застосовуються у розв'язанні конкретних економічних задач (моделі економічного аналізу, прогнозування, управління).

Відповідно до загальної класифікації математичних моделей вони поділяються на *функціональні* та *структурні*, а також *проміжні* форми (структурно-функціональні). Типовими структурними моделями є моделі міжгалузевих зв'язків. Прикладом функціональної моделі може слугувати модель поведінки споживачів в умовах товарно-грошових відносин.

За типом підходу до систем, які досліджуються моделі поділяють на *дескриптивні* та *нормативні*. Прикладом дескриптивних моделей є виробничі функції та функції купівельного попиту, побудовані на підставі опрацювання статистичних даних. Типовим прикладом нормативних моделей є моделі оптимального (раціонального) планування, що формалізують у той чи інший спосіб цілі економічного розвитку, можливості і засоби їх досягнення.

За характером відображення причинно-наслідкових аспектів розрізняють моделі жорстко *детерміновані* і *стохастичності* моделі, що враховують випадковість і невизначеність.

За способами відображення чинника часу економіко-математичні моделі поділяються на *статичні* й *динамічні*.

За формою математичних залежностей виділяють *лінійні* і *нелінійні* моделі. Відмінності між лінійними і нелінійними моделями є суттєвими не лише з математичного погляду, а й у теоретико-економічному плані, адже багато залежностей в економіці мають принципово нелінійний характер.

За співвідношенням екзогенних і ендогенних змінних, які включаються в модель, вони поділяються на *відкриті* і *закриті*. Повністю відкритих моделей не існує; модель повинна містити хоча б одну ендогенну змінну. Повністю закриті економіко-математичні моделі, тобто такі, що не містять екзогенних змінних, надзвичайно рідкісні. Переважна більшість економіко-математичних моделей посідає проміжну позицію і розрізняється за ступенем відкритості (закритості).

За характером вимог до результатів економіко-математичні моделі поділяються на *балансові* (у чистому вигляді майже немає) та *оптимізаційні*.

Класифікація видів математичних моделей може проводитися й за такими ознаками: аналітичне та комп'ютерне моделювання (рис. 3.3).



Рисунок 3.3 – Аналітичне та комп'ютерне моделювання

Для *аналітичного моделювання* характерним є те, що процеси функціонування елементів системи записують у вигляді деяких математичних співвідношень (алгебраїчних, інтегро-диференціальних, кінцево-різницевих тощо) чи логічних умов.

Комп'ютерне моделювання характеризується тим, що математична модель системи (використовуючи основні співвідношення аналітичного моделювання) подається у вигляді деякого алгоритму та програми, придатної для її реалізації на комп'ютері, що дає змогу проводити з нею обчислювальні експерименти. Залежно від математичного інструментарію (апарату), що використовується в побудові моделі, та способу організації обчислювальних експериментів можна виокремити три взаємопов'язані види моделювання:

чисельне, алгоритмічне (імітаційне) та статистичне.

У *чисельному моделюванні* для побудови комп'ютерної моделі використовуються методи обчислювальної математики, а обчислювальний експеримент полягає в чисельному розв'язанні деяких математичних рівнянь за заданих значень параметрів і початкових умов.

Алгоритмічне (імітаційне) моделювання (може бути детермінованим та стохастичним) — це вид комп'ютерного моделювання, для якого характерним є відтворення на комп'ютері (імітація) процесу функціонування досліджуваної складної системи.

Статистичне моделювання — це вид комп'ютерного моделювання, який дозволяє отримати статистичні дані відносно процесів у модельованій системі.

3.7 Роль прикладних економіко-математичних досліджень

Можна виокремити щонайменше чотири аспекти застосування математичних методів і моделей у вирішенні практичних проблем.

1. *Удосконалення системи економічної інформації.* Математичні методи й моделі дають змогу упорядковувати економічну інформацію, виявляти недоліки в наявній інформації та розробляти вимоги до підготовки нової інформації чи її коригування. Розроблення і застосування економіко-математичних моделей вказують шляхи вдосконалення системи економічної інформації, орієнтованої на вирішення певних завдань планування та управління.

2. *Інтенсифікація і підвищення точності економічних розрахунків.* Формалізація економічних задач і застосування комп'ютерів значно прискорюють типові, масові розрахунки, підвищують точність і скорочують трудомісткість, дають змогу проводити багатоваріантні економічні дослідження та обґрунтування складних заходів, недосяжні за панування «ручної» технології.

3. *Поглиблення кількісного аналізу економічних проблем.* Завдяки застосуванню економіко-математичного моделювання створюються нові можливості економічного аналізу; вивчення чинників, які впливають на економічні процеси; кількісного оцінювання наслідків змін умов розвитку економічних об'єктів тощо.

4. *Розв'язання принципово нових економічних задач.* За допомогою математичного моделювання вдається розв'язувати економічні задачі, які в інший спосіб розв'язати практично неможливо, наприклад, відшукування оптимального варіанта народногосподарського плану, імітація народногосподарських заходів, автоматизація контролю за функціонуванням складних економічних об'єктів.

Сфера практичного застосування економіко-математичного моделювання обмежується можливостями та ефективністю формалізації економічних проблем і ситуацій, а також станом інформаційного, математичного, технічного забезпечення використовуваних моделей.

Намагання будь-якою ціною застосувати математичну модель може не дати очікуваних результатів через відсутність необхідних умов.



Запитання для перевірки засвоєних знань

1. Що Ви розумієте під терміном “математична модель”? У якій формі зазвичай подаються математичні моделі?
2. У чому полягає сутність методології математичного моделювання?
3. Назвіть та охарактеризуйте етапи математичного моделювання.
4. У чому полягає головна особливість математичного моделювання ?
5. Назвіть та охарактеризуйте принципи математичного моделювання.
6. Яким чином можна зменшити складність моделі?
7. Що розуміють під економіко-математичною моделлю?
8. Які виділяють системотвірні елементи процесу моделювання?
9. У чому полягають особливості економічних спостережень і вимірів?
10. Які основні етапи процесу математичного моделювання економічних систем? Охарактеризуйте основні етапи побудови економіко-математичних моделей.
11. Які існують форми зображення математичної моделі? Охарактеризуйте їх.
12. Назвіть основні класифікаційні ознаки економіко-математичних моделей.
13. Як застосовується економіко-математичне моделювання для вирішенні практичних проблем?
14. Чим обмежується сфера практичного застосування економіко-математичного моделювання?



Завдання для самостійного розв’язання



Тестовий контроль для перевірки знань

1. Важливість і необхідність моделювання економічних систем викликана:
 - а) важливістю математики як розділу науки;
 - б) значним внеском стохастичного фактора у розвиток ситуацій;
 - в) необхідністю обчислення різних економічних показників;
 - г) нездатністю економістів проводити потрібні розрахунки.
2. Термін "синергетика" означає:
 - а) міждисциплінарну науку, що вивчає спільні принципи розвитку складних систем у різних галузях;
 - б) розділ енергетики, пов’язаний з економією паливних ресурсів;

- в) науковий напрямок, який розглядає економічні задачі з енергетичним змістом;
- г) спеціальну алгоритмічну мову, в основі якої лежить мова C++.

3. Яке з перелічених практичних завдань не стосується економіко-математичного моделювання?

- а) аналіз економічних об'єктів і процесів;
- б) вироблення управлінських рішень на всіх рівнях господарської ієрархії управління;
- в) пошук шляхів отримання підприємством додаткових прибутків;
- г) економічне прогнозування як передбачення розвитку економічних процесів.

4. Модель називають адекватною, якщо:

- а) вона відповідає реальному процесу, що моделюється;
- б) процес її дослідження дає очікувані результати;
- в) вона точно відображає модельований процес чи явище;
- г) у ній відсутні стохастичні фактори.

5. Модель – це:

- а) часткове подання реальності;
- б) абстракція;
- в) наближення;
- г) все перераховане вище.

6. Рішення у реальних бізнес-ситуаціях звичайно ґрунтуються на:

- а) оцінці числових даних;
- б) числових значеннях, отриманих за допомогою моделі;
- в) використанні інтуїтивних подань;
- г) всім перерахованим вище.

7. Один цикл економіко-математичного моделювання містить:

- а) один етап;
- б) чотири етапи;
- в) п'ять етапів;
- г) шість етапів.

8. Зазвичай числові розрахунки при використанні економіко-математичної моделі мають:

- а) багатоваріантний характер;
- б) одноваріантний характер;
- в) альтернативно-варіантний характер;
- г) оптимально-варіантний характер.

9. Інваріантна форма зображення моделі передбачає:

- а) багатоваріантність запису виразів моделі;
- б) незалежність від методів отримання розв'язку задачі;
- в) строго визначену форму запису виразів моделі;
- г) подібність моделі одному з вибраних варіантів.

10. Алгоритмічні моделі ще називають:

- а) схемними моделями;
- б) імітаційними моделями;
- в) комплексними моделями;
- г) оптимізаційними моделями.

11. Умовна оптимізація має на увазі:

- а) модель, що лежить в основі, є дуже точним поданням реальності;
- б) досягнення найкращого можливого результату з урахуванням обмежень;
- в) модель, яка в деяких аспектах суттєво простіша від прототипу;
- г) усе вище наведене.

12. До основних понять економіко-математичного моделювання можна віднести:

- а) мету;
- б) модель;
- в) економіко-математичну модель;
- г) а, б, в.

13. За призначенням можна виділити такі економіко-математичні моделі:

- а) сітьові;
- б) прикладні;
- в) оптимізаційні;
- г) балансові.

14. Процеси яких моделей можуть змінюватися за часом:

- а) імітаційні;
- б) аналітичні;
- в) статичні;
- г) динамічні.

15. За точністю математичного відображення явищ, що розглядаються, бувають такі моделі:

- а) ймовірні;
- б) лінійні;
- в) детерміновані;

г) імітаційні.



Завдання для розв'язання

Завдання 1.

Вірна відповідність між термінами та їх визначеннями:

	Термін		Визначення
1	Модель	<i>Д</i>	об'єкт, що заміщує оригінал і відбиває його найважливіші риси й властивості для даного дослідження, даної мети дослідження за обраної системи гіпотез
2	Математична модель	<i>А</i>	абстракція реальної дійсності, в якій відношення між реальними елементами, замінені відношеннями між математичними категоріями
3	Методологія математичного моделювання	<i>Г</i>	заміна досліджуваного об'єкта його «образом» і подальше вивчення моделі на підставі аналітичних методів та обчислювально-логічних алгоритмів, які реалізуються за допомогою комп'ютерних програм
4	Моделювання	<i>Б</i>	процес побудови, вивчення й використання моделей
5	Економіко-математична модель	<i>В</i>	концентроване вираження найсуттєвіших економічних взаємозв'язків досліджуваних процесів у вигляді математичних функцій, нерівностей і рівнянь
6	Адекватність моделі	<i>К</i>	відповідність моделі меті дослідження, прийнятій системі гіпотез за рівнем складності й організації, а також відповідність реальній системі
7	Спрощення моделі	<i>З</i>	модель ігнорує несуттєві властивості об'єкта
8	Інваріантна форма	<i>Ж</i>	зображення математичної моделі математичною мовою безвідносно до методів, за допомогою яких може розв'язуватись поставлена задача моделювання
9	Алгоритмічна форма	<i>Л</i>	зображення математичної моделі у вигляді послідовності дій

	Термін		Визначення
10	Аналітична форма	ϵ	зображення математичної моделі у вигляді формул та співвідношень між математичними виразами
11	Схемна форма	I	зображення математичної моделі у вигляді графів, таблиць, діаграм

Завдання 2.

Обведіть правильну відповідь, визначену літерою під запитанням:

1. Оптимізаційні моделі завжди забезпечують найкраще рішення у реальній ситуації:

а) так; б) ні.

2. Модель може з успіхом замінити судження й досвід менеджера:

а) так; б) ні.

3. Дані потрібні тільки після того, як модель вже побудована:

а) так; б) ні.

4. Починаючи висувати гіпотези про існування якого-небудь зв'язку між даними, ви переходите до формулювання рівнянь моделі:

а) так; б) ні.

5. Дані використовуються для побудови моделей:

а) так; б) ні.

Розділ 2. Основи лінійного моделювання

Тема 4. ПАРНА ЛІНІЙНА РЕГРЕСІЯ ТА КОРЕЛЯЦІЯ В ЕКОНОМЕТРИЧНИХ ДОСЛІДЖЕННЯХ



Основні питання:

4.1 Специфікація моделі.

4.2 Загальне поняття про вибірку лінійну регресію. Оцінювання параметрів лінійної регресії.

4.3 Поняття тісноти зв'язку, оцінка коефіцієнтів кореляції та детермінації.

4.4 Оцінка якості лінійного рівняння регресії.

4.5 Оцінка значущості параметрів лінійної регресії та кореляції.

Побудова інтервалів довіри.

4.6 Побудова інтервалів прогнозу за лінійним рівнянням регресії.



Ключові слова та поняття: регресія, функція регресії, проста регресія, множинна регресія, помилки специфікації, помилки вибірки, помилки виміру, графічний метод, аналітичний метод, експериментальний метод, метод найменших квадратів (МНК), коефіцієнт детермінації, критерій Фішера, помилка апроксимації.

4.1 Специфікація моделі

В економетриці поширено використання методів статистики. Ставлячи за мету дати кількісний опис взаємозв'язків між економічними змінними, економетрія насамперед пов'язана з методами регресії і кореляції.

Однією 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 - u та x , тобто модель виду:

$$y = f(x), \quad (4.1)$$

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:

$$y = f(x_1, x_2, \dots, x_k). \quad (4.2)$$

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 є домінуючий фактор, який і використовується в якості пояснюючої змінної.

Практично в кожному окремому випадку специфікації моделі величина y складається з двох доданків:

$$y = y_x + e, \quad (4.3)$$

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 викликана трьома джерелами: специфікацією моделі, вибіркоvim 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 **помилки 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 вибіркоvim 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.

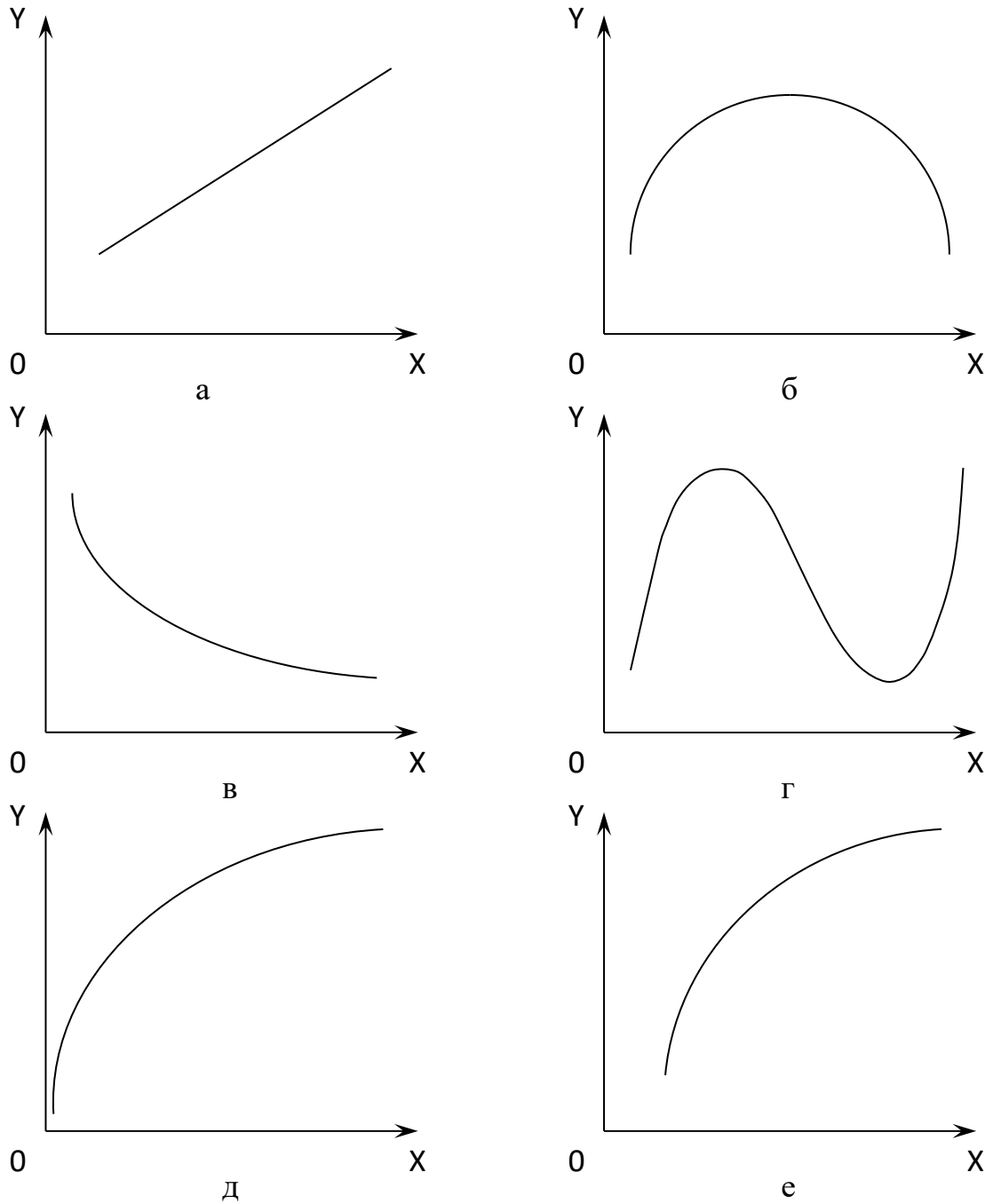
При вивченні залежності між двома ознаками **графічний метод** підбору виду рівняння регресії є досить наочний. Він заснований на побудові поля кореляції. Основні типи кривих, використовуваних при кількісній оцінці зв'язків, представлені на рис. 4.1.

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:

$$D_{зал} = \frac{1}{n} \sum (y - \hat{y}_x)^2. \quad (4.4)$$



$$a - \hat{y}_x = a + b \cdot x;$$

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

$$в - \hat{y}_x = a + \frac{b}{x};$$

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

$$д - \hat{y}_x = a \cdot x^b;$$

$$е - \hat{y}_x = a \cdot b^x.$$

Рисунок 4.1 – Основні типи кривих, використувані при кількісній оцінці зв'язків між двома змінними

Чим менше величина залишкової дисперсії, тим у меншій мірі спостерігається вплив інших факторів, що не враховуються в рівнянні регресії, тим краще рівняння регресії підходить до опису 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 більшій ступені піддаються інтерпретації і вимагають меншого обсягу спостережень.

4.2 Загальне поняття про вибірккову лінійну регресію. Оцінювання параметрів лінійної регресії

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

$$\hat{y}_x = a + bx \quad \text{або} \quad y = a + bx + e. \quad (4.5)$$

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:

$$\sum_i (y_i - \hat{y}_{x_i})^2 \rightarrow \min. \quad (4.6)$$

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:

$$b = \frac{\text{cov}(x, y)}{\sigma_x^2}, \quad (4.7)$$

де $\text{cov}(x, y)$ – коваріація ознак;

σ_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 :

$$b = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2}$$

$$a = \bar{y} - b \cdot \bar{x} . \quad (4.8)$$

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 : $Vx > Vy$.

4.3 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:

$$r_{xy} = b \frac{\sigma_x}{\sigma_y} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\sigma_x \sigma_y}. \quad (4.9)$$

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:

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i). \quad (4.10)$$

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:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2, \quad (4.11)$$

де $\sum_{i=1}^n (y_i - \bar{y})^2$ – 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;

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

$$\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} + \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{n}, \quad (4.12)$$

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

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

$\frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{n}$ – дисперсія, яку прийнято називати дисперсією, що пояснює

регресію, позначимо її через $\sigma_{регр}^2$.

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

$$\sigma_{заг}^2 = \sigma_{ном}^2 + \sigma_{регр}^2. \quad (4.13)$$

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

$$1 = \frac{\sigma_{ном}^2}{\sigma_{заг}^2} + \frac{\sigma_{регр}^2}{\sigma_{заг}^2}. \quad (4.14)$$

Як можна побачити з виразу (3.14), перша частина $(\frac{\sigma_{ном}^2}{\sigma_{заг}^2})$ є пропорцією дисперсії помилок у загальній дисперсії, тобто є частиною дисперсії, яку не можна пояснити через регресійний зв'язок. Друга частина $(\frac{\sigma_{регр}^2}{\sigma_{заг}^2})$ – 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:

$$R^2 = \frac{SSR}{SST}. \quad (4.15)$$

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

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

$$r_{xy} = b \frac{\sigma_x}{\sigma_y} \quad \text{або} \quad R^2 = (r_{xy})^2. \quad (4.16)$$

Наприклад, якщо величина $R^2 = 0,96$, то це означає, що рівнянням регресії пояснюється 96% дисперсії результативної ознаки, а на долю інших факторів приходить 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.

4.4 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 розкладається на факторну та залишкову:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \sum_{i=1}^n (\hat{y}_i - \bar{y})^2. \quad (4.17)$$

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, \dots, 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 дослідом 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 $\{(y_1 - \bar{y}), (y_2 - \bar{y}), \dots, (y_n - \bar{y})\}$ 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, \dots, 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:

$$n - 1 = 1 + (n - 2). \quad (4.18)$$

Поділивши кожен суму квадратів на відповідне їй число ступенів вільності, отримаємо середній квадрат відхилень, або дисперсію на одну ступень вільності:

$$D_{\text{заг}} = \frac{\sum (y - \bar{y})^2}{n - 1}; D_{\text{регр}} = \frac{\sum (\hat{y} - \bar{y})^2}{1}; D_{\text{ном}} = \frac{\sum (y - \hat{y})^2}{n - 2}. \quad (4.19)$$

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

$$F = \frac{D_{\text{регр}}}{D_{\text{ном}}}. \quad (4.20)$$

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: $F_{\text{факт}} > F_{\text{табл}}$.

Якщо ж величина виявиться менш табличної $F_{\text{факт}} < F_{\text{табл}}$, 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 4.1 – 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	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$\frac{\sum (\hat{y}_i - \hat{y}_i)^2}{\sum (y_i - \hat{y}_i)^2}$	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 :

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

1.5 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:

$$m_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n - 2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S^2}{\sum (x - \bar{x})^2}}, \quad (4.21)$$

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:

$$t_b = \frac{b}{m_b}, \quad (4.22)$$

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:

$$t_b = \sqrt{F}. \quad (4.23)$$

Інтервал довіри для коефіцієнта регресії визначається як

$$b \pm t_{\text{табл}} \cdot m_b \text{ або } b - t_{\text{табл}} \cdot m_b < b < b + t_{\text{табл}} \cdot m_b. \quad (4.24)$$

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:

$$m_a = \sqrt{\frac{\sum (y - \hat{y}_x)^2 \cdot \sum x^2}{(n-2) \cdot n \sum (x - \bar{x})^2}} = \sqrt{S^2 \cdot \frac{\sum x^2}{n \cdot \sum (x - \bar{x})^2}}. \quad (4.25)$$

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

$$t_a = \frac{a}{m_a}, \quad (4.26)$$

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 :

$$m_r = \sqrt{\frac{1-r^2}{n-2}}. \quad (4.27)$$

Фактичне значення t -критерія Ст'юдента визначається як

$$t_r = \frac{r}{\sqrt{1-r^2}} \cdot \sqrt{n-2}. \quad (4.28)$$

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

$$t_r^2 = t_b^2. \quad (4.29)$$

Таким чином, перевірка гіпотез про значимість коефіцієнта регресії, параметра a та коефіцієнта кореляції рівносильна перевірці 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 у і \hat{y}_x . Чим менше ця відмінність, тим ближче теоретичні значення підходять до емпіричних даних, краще якість моделі. Величина відхилень фактичних і розрахованих 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:

$$\bar{A} = \frac{1}{n} \cdot \sum \left| \frac{y - \hat{y}_x}{y} \right| \cdot 100. \quad (4.30)$$

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

4.6 Побудова інтервалів прогнозу за лінійним рівнянням регресії

У прогнозних розрахунках по рівнянню регресії визначається прогнозне (y_{np}) 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^*)

$$\hat{y}_x - m_{\hat{y}_x} \leq y^* \leq \hat{y}_x + m_{\hat{y}_x}. \quad (4.31)$$

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 :

$$m_{\bar{y}}^2 = \frac{S^2}{n}. \quad (4.32)$$

Помилка коефіцієнта регресії визначається за формулою:

$$m_b^2 = \frac{S^2}{\sum (x - \bar{x})^2}. \quad (4.33)$$

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

$$m^2_{\hat{y}_x} = \frac{S^2}{n} \cdot \frac{S^2}{\sum(x - \bar{x})^2} \cdot (x_k - \bar{x})^2 = S^2 \cdot \left(\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum(x - \bar{x})^2} \right). \quad (4.34)$$

Відповідно $m_{\hat{y}_x}$ має вираз:

$$m_{\hat{y}_x} = S \cdot \sqrt{\frac{1}{n} + \frac{(x_k - \bar{x})^2}{\sum(x - \bar{x})^2}}. \quad (4.35)$$

Розглянута формула стандартної помилки прогнозного значення у при заданому значенні x_k характеризує помилку положення лінії регресії.

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

$$\hat{y}_{xk} \pm t_{\alpha} \cdot m_{\hat{y}_x}. \quad (4.36)$$



Запитання для перевірки засвоєних 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. Поясніть 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?



Завдання для самостійного розв'язання



Тестовий контроль для перевірки знань

- Проста регресія відрізняється від множинної:
 - кількістю змінних;
 - кількістю параметрів;
 - математичною функцією.
- Якими причинами викликана наявність випадкової величини в моделі:
 - через неправильний вибір виду рівняння;
 - через наявність нелінійного зв'язку між y і x ;
 - через некомпетентність дослідника.
- Суть методу найменших квадратів складається в мінімізації:
 - $\sum (\hat{y}_x - y)^2$;
 - $\sum (y - \bar{y})^2$;
 - $\sum (y - \hat{y}_x)^2$.
- Параметр b називається:
 - коефіцієнтом кореляції;
 - коефіцієнтом регресії;
 - коефіцієнтом еластичності.
- У рівнянні $\hat{y}_x = 37 + 1.5x$ параметр b показує, що:
 - y зміниться на 1 при зміні x на 1,5;
 - y в середньому зменшиться на 1,5 при незмінності x ;
 - y в середньому збільшиться на 1,5 при збільшенні x на 1.
- Формула коефіцієнта регресії:
 - $\frac{\text{cov}(x, y)}{\sigma_x^2}$;

$$\text{б) } \frac{\sigma_x^2}{\text{cov}(x, y)};$$

$$\text{в) } \frac{\bar{y}\bar{x} - \bar{x} \cdot \bar{y}}{\sqrt{(x - \bar{x})^2}}.$$

7. Маються наступні дані: $\bar{x} = 120, \bar{y} = 100, \text{var}(x) = 700, \text{cov}(x, y) = 180$.
Рівняння лінійної регресії складе:

- а) $y = 69,16 + 0,25x$;
- б) $y = 38,7 - 0,12x$;
- в) жодне з приведених.

8. Коефіцієнт кореляції пояснює:

- а) взаємозв'язок між x та y ;
- б) кількісний вплив x на y ;
- в) тісноту зв'язку між x та y .

9. Коефіцієнт кореляції визначається за формулою:

$$\text{а) } \frac{\text{cov}(x, y)}{\sigma_x \sigma_y};$$

$$\text{б) } b \cdot \frac{\sigma_y}{\sigma_x};$$

$$\text{в) } \frac{\bar{x}\bar{y} - \bar{x} \cdot \bar{y}}{x^2 - \bar{x}^2}.$$

10. Сильний і зворотний зв'язок показує коефіцієнт кореляції, рівний:

- а) $-0,312$;
- б) $0,991$;
- в) $-0,79$.

11. Квадрат коефіцієнта кореляції – це:

- а) коефіцієнт регресії;
- б) коефіцієнт детермінації;
- в) коефіцієнт еластичності.

12. Для суми квадратів відхилень, які пояснюються за допомогою регресії, кількість ступенів вільності дорівнює:

- а) $n-2$;
- б) $n-1$;
- в) 1 .

13. $\sum (y - \bar{y})^2 = 15000, \sum (y - \hat{y})^2 = 265, n = 7$. Розрахункове значення F-критерію дорівнює:

- а) 278 ;

- б) 4,91;
- в) 0,12.

14. За допомогою якого показника оцінюється значимість рівняння регресії:

- а) F- критерію;
- б) t – критерію;
- в) МНК.

15. За допомогою якого показника оцінюється значимість параметрів рівняння регресії:

- а) F;
- б) t;
- в) МНК.

16. Стандартна помилка розраховується з метою:

- а) оцінки значимості коефіцієнту регресії;
- б) оцінки значимості параметрів рівняння регресії;
- в) оцінки коефіцієнта еластичності.

17. На ваш погляд, у якій формулі допущена помилка:

- а) $m_r = \sqrt{\frac{1-r^2}{n-2}}$;
- б) $t_r^2 = t_b^2 = \sqrt{F}$;
- в) $F = \frac{r}{1-r} \cdot (n-2)$.

18. $b = 36,84$. $m_b = 2,21$. Чому дорівнює t-тест Ст'юдента?:

- а) 16,67;
- б) 0,06;
- в) 1,66.

19. Визначити значення коефіцієнта регресії, якщо відомо, що $\sigma_x=7,1$, $\sigma_y=5,8$, $r^2=0,88$:

- а) 0,77;
- б) 1,077;
- в) 1,14.

20. Середня помилка апроксимації для визначення якості рівняння повинна знаходитися:

- а) більше 10 %;
- б) менше 1;
- в) менше 10 %.

21. Якщо в лінійному рівнянні регресії параметр $a > 0$, то:

- а) $b < 0$;
- б) $V_x > V_y$;
- в) $V_x < V_y$.

22. Середня помилка апроксимації використовується для оцінки:

- а) якості рівняння регресії;
- б) значимості параметрів рівняння регресії;
- в) значимості факторів рівняння регресії.

23. Середній квадрат залишкової суми квадратів відхилень визначається як:

- а) $MSR = \frac{SST}{n-1}$;
- б) $MSE = \frac{SSE}{n-2}$;
- в) $MSE = \frac{SSE}{n-1}$.

24. Якщо регресія має $R^2 = 0,70$, то регресійна лінія:

- а) пояснює 70 % варіації змінної x ;
- б) пояснює 70 % варіації змінної y ;
- в) матиме параметр $a = 0,70$.

25. Коваріація між x і y є:

- а) $\sum (x_i - \bar{x})^2$;
- б) $\sum (y_i - \bar{y})^2$;
- в) $(1/n) \sum (x_i - \bar{x})(y_i - \bar{y})$.

26. Коефіцієнт детермінації:

- а) точка, де лінія регресії перетинає вісь y ;
- б) вимірює придатність лінії регресії;
- в) вимірює зв'язок між незалежною та залежною змінними.

27. В рівнянні лінійної регресії параметр a - це:

- а) точка, де лінія регресії перетинає вісь ординат;
- б) вимірює придатність лінії регресії;
- в) вимірює зв'язок між незалежною та залежною змінними.



Приклади розв'язання завдань

Завдання 1

По територіях регіону наводяться дані за рік (табл. 4.2)

Таблиця 4.2

Номер регіону	Середньодушовий прожитковий мінімум у день одного працюючого, ум. грн. од. х	Середньоденна зарплата, ум. грн. од., у
1	78	133
2	82	148
3	87	134
4	79	154
5	89	162
6	106	195
7	67	139
8	88	158
9	73	152
10	87	162
11	76	159
12	115	173

Необхідно:

1. Побудувати лінійне рівняння регресії y від x ;
2. Розрахувати лінійний коефіцієнт парної кореляції і середню помилку апроксимації;
3. Оцінити статистичну значимість параметрів регресії і кореляції при $\alpha = 0,05$;
4. Виконати прогноз заробітної плати (y) при прогнозному значенні середньодушового прожиткового мінімуму в день одного працюючого (x), який складатиме 107 % від середнього рівня;
5. Оцінити точність прогнозу, розрахувавши помилку прогнозу і його інтервал довіри.

Рішення:

Для розрахунку параметрів рівняння лінійної регресії будемо розрахункову таблицю 3.

Таблиця 4.3

	x	y	yx	x^2	y^2	\hat{y}_x	$y - \hat{y}_x$	\bar{A}_i
--	-----	-----	------	-------	-------	-------------	-----------------	-------------

1	78	133	10374	6084	17689	149	-16	12,0
2	82	148	12136	6724	21904	152	-4	2,7
3	87	134	11658	7569	17956	157	-23	17,2
4	79	154	12166	6241	23716	150	4	2,6
5	89	162	14418	7921	26244	159	3	1,9
6	106	195	20670	11236	38025	174	21	10,8
7	67	139	9313	4489	19321	139	0	0,0
8	88	158	13904	7744	24964	158	0	0,0
9	73	152	11096	5329	23104	144	8	5,3
10	87	162	14094	7569	26244	157	5	3,1
11	76	159	12084	5776	25281	147	12	7,5
12	115	173	19895	13225	29929	183	-10	5,8
Разом	1027	1869	161808	89907	294377	1869	0	68,8
Середні значення	85,6	155,8	13484,0	7492,3	24531,4			5,7
σ	12,95	16,53						
σ^2	167,7	273,4						

$$b = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\sum x^2 - (\bar{x})^2} = \frac{13484 - 85,6 \cdot 155,8}{7492,3 - 85,6^2} = \frac{151,8}{164,94} = 0,92;$$

$$a = \bar{y} - b \cdot \bar{x} = 155,8 - 0,92 \cdot 85,6 = 77,0.$$

Отримано рівняння регресії: $y = 77,0 + 0,92 x$.

Зі збільшенням середньодушового прожиткового мінімуму на 1 ум. гр. од. середньоденна зарплата зростає в середньому на 0,92 ум. гр. од.

Тісноту лінійного зв'язку оцінить коефіцієнт кореляції:

$$r_{yx} = b \cdot \frac{\sigma_x}{\sigma_y} = 0,92 \cdot \frac{12,95}{16,53} = 0,721; r^2_{yx} = 0,52.$$

Це означає, що 52 % варіації заробітної плати (y) пояснюється варіацією фактора.

Якість моделі визначає середня помилка апроксимації:

$$\bar{A} = \frac{1}{n} \sum A_i = \frac{1}{n} \sum \frac{|y - \hat{y}|}{y} \cdot 100\% = \frac{68.8}{12} = 5.7\%.$$

Якість побудованої моделі оцінюється як гарна, тому що середня помилка апроксимації не перевищує 8-10 %.

1. Оцінку статистичної значимості параметрів регресії проведемо за допомогою t-статистики Ст'юдента і шляхом розрахунку довірчого інтервалу кожного з показників.

Висуваємо гіпотезу H_0 про статистично незначущість показників, тобто $a=b=r_{xy}=0$.

$t_{\text{табл}}$ для числа ступенів вільності $n-2=12-2=10$ і $\alpha = 0,05$ складе 2,23.

Визначаємо стандартні помилки m_a , m_b , m_r :

$$m_a = 12.6 \cdot \frac{\sqrt{89907}}{12 \cdot 12.95} = 24.3;$$

$$m_b = \frac{12.6}{12.95 \cdot \sqrt{12}} = 0.281;$$

$$m_{r_{yx}} = \sqrt{\frac{1-0.52}{12-2}} = 0.219.$$

Тоді

$$t_a = \frac{77}{24.3} = 3.2;$$

$$t_b = \frac{0.92}{0.281} = 3.3;$$

$$t_{r_{yx}} = \frac{0.721}{0.219} = 3.3.$$

Фактичні значення t – статистики перевершують табличні значення (3,2; 3,3; 3,3 > 2,23), тому H_0 відхиляється, тобто a, b, r_{xy} не випадково відрізняються від нуля, а статистично значимі.

Розрахуємо інтервал довіри для a і b. Для цього визначимо середню помилку для кожного показника:

$$\Delta_a = 2.23 \cdot 24.3 = 54;$$

$$\Delta_b = 2.23 \cdot 0.281 = 0.62.$$

Інтервали довіри:

$$\gamma_a = a \pm \Delta_a = 77 \pm 54; \quad \gamma_{a_{\min}} = 77 - 54 = 23; \quad \gamma_{a_{\max}} = 77 + 54 = 131;$$

$$\gamma_b = b \pm \Delta_b = 0.92 \pm 0.62; \quad \gamma_{b_{\min}} = 0.92 - 0.62 = 0.3; \quad \gamma_{b_{\max}} = 0.92 + 0.62 = 1.54.$$

Аналіз верхніх і нижніх границь довірчих інтервалів приводить до висновку про те, що з імовірністю $p = 1 - \alpha = 0,95$ параметри a і b знаходяться у зазначених границях, не приймають нульових значень, тобто не є

статистично незначущими й істотно відмінні від нуля.

2. Отримані оцінки рівняння регресії дозволяють використовувати його для прогнозу. Якщо прогнозне значення прожиткового мінімуму складе $x_p = \bar{x} \cdot 1.07 = 85.6 \cdot 1.07 = 91.6$ ум. гр. од., то тоді прогнозне значення середньоденної зарплати складе:

$$\hat{y}_p = 77 + 0.92 \cdot 91.6 = 161 \text{ ум. гр. од.}$$

3. Помилка прогнозу складе:

$$m_{\hat{y}_p} = \sigma_{\text{ост}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2}},$$

$$\text{де } \sigma_{\text{ост}} = \sqrt{\frac{\sum (y - \hat{y})^2}{n - m - 1}}.$$

$$m_{\hat{y}_p} = 12.6 \cdot \sqrt{1 + \frac{1}{12} + \frac{(91.6 - 85.6)^2}{12 \cdot 12.95^2}} = 13.2 \text{ ум. гр. од.}$$

Гранична помилка прогнозу, що у 95 % випадків не буде перевищена, складе:

$$\Delta_{\hat{y}_p} = t_{\text{табл}} \cdot m_{\hat{y}_p} = 2.23 \cdot 13.2 = 29.4$$

Довірчий інтервал прогнозу:

$$\gamma_{\hat{y}_p} = 161 \pm 29.4; \quad \gamma_{\hat{y}_{p\min}} = 161 - 29.4 = 131.6; \quad \gamma_{\hat{y}_{p\max}} = 161 + 29.4 = 190.4.$$

Виконаний прогноз середньоденної зарплати виявився надійним, але не точним, тому що діапазон верхньої і нижньої границь довірчого інтервалу D_γ складає 1,95 рази (121/62).

Завдання 2

Нехай існує наступна модель регресії, що характеризує залежність y від x : $y = 8 + 0,7x$. Відомо також, що $r_{xy} = 0.5$; $n = 20$.

Необхідно:

- Побудувати інтервал довіри для коефіцієнту регресії в цій моделі:
 - з імовірністю 90 %;
 - з імовірністю 99 %;
- Проаналізувати результати і пояснити причини їхнього розходження.

Рішення:

Знаючи залежність $t_b^2 = F$, визначимо t_b .

$$F = \frac{r^2}{1-r^2} \cdot (n-2) = \frac{0.25}{1-0.25} \cdot 18 = 6.$$

$$\text{Тоді } t_b = \sqrt{F} = \sqrt{6} = 2.449.$$

$$m_b = \frac{0.7}{2.449} = 0.286.$$

Розрахуємо довірчий інтервал для b : $b \pm t_{\text{табл}} \cdot m_b$

а) з імовірністю 90 %:

$$0,7-1,73 \cdot 0,286 < b < 0,7+1,73 \cdot 0,286$$

$$0,21 < b < 1,19$$

б) з імовірністю 99 %:

$$0,7-2,88 \cdot 0,286 < b < 0,7+2,88 \cdot 0,286$$

$$-0,12 < b < 1,52$$

Довірчий інтервал, побудований з рівнем значимості 99 % указує, що коефіцієнт регресії одночасно містить і негативні, і позитивні значення, що підтверджує нульову гіпотезу, за якою $b = 0$. У цьому випадку коефіцієнт регресії є статистично незначущим.

Завдання 3

По сукупності 30 підприємств торгівлі вивчається залежність між ознаками: x – ціна на товар, грн.; y – прибуток торговельного підприємства, тис. грн.

При оцінці лінійної регресійної моделі були отримані наступні проміжні результати:

$$\sum (y - \hat{y})^2 = 39000; \quad \sum (y - \bar{y})^2 = 120000.$$

Необхідно:

1. Поясніть, який показник кореляції можна визначити по цим даним;
2. Побудуйте таблицю дисперсійного аналізу для розрахунку значення F-критерію Фішера;
3. Порівняйте фактичне значення F-критерію з табличним. Зробіть висновки.

Рішення:

1. За даними задачі можна розрахувати коефіцієнт детермінації по формулі $r^2 = 1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2} = 1 - \frac{39000}{120000} = 0.675$. Значення коефіцієнта детермінації показує, що рівнянням регресії пояснюється 67,5 % дисперсії

результативної ознаки, а на долю інших факторів приходиться 32,5 % її дисперсії. Таким чином, величина коефіцієнта детермінації вказує на адекватність побудованої моделі. Однак, для більшої впевненості необхідно розрахувати F- критерій Фішера.

2. Побудуємо таблицю дисперсійного аналізу (табл. 4.4).

3. Оскільки $F_{\text{факт}} > F_{\text{табл}}$, то нульова гіпотеза про відсутність зв'язку між ознаками відхиляється і рівняння регресії вважається статистично значимим і надійним.

Таблиця 4.4

Джерела варіації	Число ступенів вільності	Сума квадратів відхилень	Дисперсія на одну ступень вільності	F- відношення	
				фактичне	табличне ($\alpha=0,05$)
Загальна	29	120000	-		
Факторна	1	81000	81000	58,15	4,2
Залишкова	28	39000	1392,86		

Завдання 4

Залежність середньомісячної продуктивності праці від віку робітників характеризується моделлю: $y = a + bx + cx^2$. Її використання привело до результатів, які представлені у таблиці 4.5.

Таблиця 4.5

№ з/п	Продуктивність праці робітників, грн.	
	фактична	розрахункова
1	12	10
2	8	10
3	13	13
4	15	14
5	16	15
6	11	12
7	12	13
8	9	10
9	11	10
10	9	9

Необхідно:

Оцінити якість моделі, визначивши помилку апроксимації, індекс кореляції і F-критерій Фішера.

Рішення:

Таблиця 4.6

№ з/п	y	\hat{y}	$\left \frac{y - \hat{y}}{y} \right \cdot 100$	$(y - \bar{y})^2$	$(y - \hat{y})^2$
1	12	10	16,7	0,16	4
2	8	10	25,0	12,96	4
3	13	13	0	1,96	0
4	15	14	6,7	11,56	1
5	16	15	6,3	19,36	1
6	11	12	9,1	0,36	1
7	12	13	8,3	0,16	1
8	9	10	11,1	6,76	1
9	11	10	9,1	0,36	1
10	9	9	0	6,76	0
Разом	116		92,3	60,4	14

$$\bar{y} = \frac{116}{10} = 11.6$$

$$\bar{A} = \frac{1}{n} \sum A_i = \frac{1}{n} \sum \left| \frac{y - \hat{y}}{y} \right| \cdot 100\% = \frac{92.3}{10} = 9.23\%.$$

$$\text{Індекс кореляції: } \rho_{xy} = \sqrt{1 - \frac{14.0}{60.4}} = 0.88$$

Розрахуємо F-критерій:

$$F_{\text{факт}} = \frac{\rho_{yx}^2}{1 - \rho_{yz}^2} \cdot \frac{n - m - 1}{m} = \frac{0.77}{0.23} \cdot \frac{10 - 3}{2} = 3.35 \cdot 3.5 = 11.725$$

За розрахованими показниками можна зробити однозначний висновок щодо якості побудованого нелінійного рівняння регресії. Цей результат можна пояснити порівняно високою тісністю виявленої залежності (значення індексу кореляції = 0,88)

Завдання 5

По 20 фермам області отримана інформація, представлена в таблиці 4.7.

Таблиця 4.7

Показник	Середні значення	Коефіцієнт варіації	Частка залишкової дисперсії в загальній
Врожайність, ц/га	27	20	10
Кількість добрив, кг	5	15	

Фактичне значення F- критерій Фішера = 45.

Необхідно:

1. Визначити лінійний коефіцієнт детермінації;
2. Побудувати рівняння лінійної регресії;
3. Розрахувати узагальнюючий коефіцієнт еластичності;
4. С імовірністю 0,95 % вказати довірчий інтервал очікуваного значення врожайності в припущенні росту кількості внесених добрив на 10 % від свого середнього рівня ($t_{\text{табл}}=2,1$).

Рішення:

1. Щоб визначити лінійний коефіцієнт детермінації за наявною інформацією, необхідно перетворити формулу F-критерію Фішера:

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

Тоді

$$\frac{r^2}{1-r^2} = \frac{45}{18}; r^2 = 2.5 \cdot (1-r^2);$$

$$r^2 = \frac{2,5}{3,5} = 0.714$$

$$r = \sqrt{0.714} = 0.845$$

Коефіцієнт детермінації $r^2 = 0.714$ означає, що варіація результату на 71,4 % пояснюється варіацією фактора x. Лінійний коефіцієнт кореляції, рівний 0,845, указує на наявність тісного і прямого зв'язку між ознаками.

2. Для того, щоб побудувати лінійне рівняння регресії, необхідно розрахувати його параметри – a і b. Для цього скористаємося наступними формулами:

$$b = \frac{r \cdot \sigma_x}{\sigma_y}; a = \bar{y} - b \cdot \bar{x}$$

$$b = \frac{0.845 \cdot 0.75}{5.4} = 6.08; \quad a = 27 - 6.08 \cdot 5 = -3.4$$

Таким чином, рівняння регресії прийме вид:

$$\hat{y}_x = -3.4 + 6.08x$$

Це означає, що зі збільшенням кількості внесених добрив на 1 кг врожайність підвищиться в середньому на 6,08 ц/га.

3. Коефіцієнт еластичності для лінійної функції має вигляд:

$$\mathcal{E} = \frac{b \cdot x}{a + b \cdot x}; \quad \mathcal{E} = \frac{6.08 \cdot 5}{-3.4 + 6.08 \cdot 5} = 1.13\%$$

Коефіцієнт еластичності показує, що з ростом на 1 % внесених добрив на 1 ц/га врожайність у середньому підвищиться на 1,13 %.

4. Отримані оцінки рівняння регресії дозволяють використовувати його для прогнозу. Якщо прогнозне значення внесених добрив складе $x_p = \bar{x} \cdot 1.1 = 5 \cdot 1.1 = 5.5$ кг, тоді прогнозне значення врожайності

$$\hat{y}_p = -3.4 + 6.08 \cdot 5.5 = 30.04 \text{ ц / га}$$

Гранична помилка прогнозу, що у 95 % не буде перевищена та складе:

$$\Delta \hat{y}_p = t_{\text{табл}} \cdot m_{y_p}$$

$$m_{y_p} = \sigma_{\text{ост}} \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2}}, \quad \text{де} \quad \sigma_{\text{ост}}^2 = \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}.$$

$$m_{y_p} = \sqrt{\frac{2.916}{18}} \cdot \sqrt{1 + \frac{1}{20} + \frac{(5.5 - 5)^2}{20 \cdot 0.75^2}} = 0.4 \cdot \sqrt{1.05 + \frac{0.25}{11.25}} = 0.4 \cdot 1.035 = 0.414$$

$$\Delta \hat{y}_p = t_{\text{табл}} \cdot m_{y_p} = 2.1 \cdot 0.414 = 0.87$$

Інтервал довіри прогнозу вийде: $30,04 - 0,87 < y_{\text{пр}} < 30,04 + 0,87$.

Прогноз лінії регресії в інтервалі складе $29,17 < y_{\text{пр}} < 30,91$.

Виконаний прогноз очікуваного значення врожайності виявився досить надійним.



Завдання для розв'язання

Завдання 1

По 7 районах області за рік відомі значення двох показників (табл. 4.8).

Таблиця 4.8

Район	Витрати на покупку продовольчих товарів у загальних витратах, %, (y)	Середньорічна зарплата одного працюючого, ум. гр. од., (x)
1	68,8	45,1
2	61,2	59,0
3	59,9	57,2
4	56,7	61,8
5	55,0	58,8
6	54,3	47,2
7	49,3	55,2

Необхідно:

1. Для характеристики залежності y від x розрахувати параметри наступних функцій:

- а) лінійної;
- б) рівносторонньої гіперболи;

2. Оцінити кожен модель через середню помилку апроксимації і F-критерій Фішера.

Завдання 2

По групі підприємств, що випускають однорідну продукцію, відома залежність собівартості одиниці продукції y і факторів, приведених у таблиці 4.9.

Таблиця 4.9

Ознака-фактор	Рівняння парної регресії	Середні значення фактора
Обсяг виробництва, x_1	$\hat{y}_{x_1} = 0.62 + 58.74 \cdot \frac{1}{x_1}$	$\bar{x}_1 = 2.64$
Трудомісткість одиниці продукції, x_2	$\hat{y}_{x_2} = 9.3 + 9.83 \cdot x_2$	$\bar{x}_2 = 1.38$
Оптова ціна за 1 тону енергоносія, x_3	$\hat{y}_{x_3} = 11.75 \cdot x_3^{1.6281}$	$\bar{x}_3 = 1.503$
Частка прибутку, що вилучається державою, x_4	$\hat{y}_{x_4} = 14.87 \cdot 1.016^{x_4}$	$\bar{x}_4 = 26.3$

Необхідно:

1. Визначити за допомогою коефіцієнтів еластичності силу впливу кожного фактора на результат;
2. Розташувати фактори за силою впливу.

Завдання 3

Залежність споживання продукту від середньодушового доходу за даними 20 родин характеризується наступними даними:

$$\text{рівняння регресії } \hat{y}_x = 2 \cdot x^{0.3};$$

$$\text{індекс кореляції } \rho_{xy} = 0.9;$$

$$\text{залишкова дисперсія } \sigma_{ост}^2 = 0,06.$$

Необхідно:

Провести дисперсійний аналіз отриманих результатів.

Завдання 4

Нехай існує наступна модель регресії, що характеризує залежність y від x : $y = 12 + 0,17x$. Відомо також, що $r_{xy} = 0.7$; $n = 27$.

Необхідно:

1. Побудувати інтервал довіри для коефіцієнту регресії в цій моделі:
 - а) з імовірністю 95 %;
 - б) з імовірністю 99 %.
2. Проаналізуйте результати і поясніть причини їхнього розходження.

Завдання 5

По сукупності 40 підприємств торгівлі вивчається залежність між ознаками: x – ціна на товар, грн.; y – прибуток торговельного підприємства, тис. грн.

При оцінці лінійної регресійної моделі були отримані наступні проміжні результати:

$$\sum (y - \hat{y})^2 = 550; \quad \sum (y - \bar{y})^2 = 2580.$$

Необхідно:

1. Поясніть, який показник кореляції можна визначити по цим даним;
2. Побудуйте таблицю дисперсійного аналізу для розрахунку значення F-критерію Фішера;
3. Порівняйте фактичне значення F-критерію з табличним. Зробіть висновки.

Завдання 6

Залежність середньомісячної продуктивності праці від віку робітників характеризується моделлю: $y = a + bx + cx^2$. Її використання привело до результатів, які представлені у таблиці 4.10.

Таблиця 4.10

№ з/п	Продуктивність праці робітників, грн	
	фактична	розрахункова
1	19	18
2	18	18
3	17	19
4	16	18
5	20	21
6	21	23
7	18	19
8	19	20

Необхідно:

Оцінити якість моделі, визначивши помилку апроксимації, індекс кореляції і F-критерій Фішера.

Завдання 7

Вивчалася залежність виду $y = ax^b$. Для перетворених змінних вийшли наступні дані:

$$\sum xy = 4.2; \sum x = 8.2; \sum x^2 = 9.23; \sum y = 3.9; \sum (y - \hat{y})^2 = 0.0014.$$

Необхідно:

1. Знайти параметр b ;
2. Знайти показник кореляції, припускаючи, що $\sigma_y = 0,08$. Дайте оцінку його значимості, якщо відомо, що $n = 9$.

Завдання 8

Залежність обсягу виробництва y (тис. од.) від чисельності зайнятих x (чол.) по 20 заводам концерну характеризується таким чином:

Таблиця 4. 11

Рівняння регресії	$Y = 64 - 0,5x + 0,08x^2$
Частка залишкової дисперсії в загальній	15 %

Необхідно визначити:

1. Індекс кореляції;
2. Значимість рівняння регресії;
3. Коефіцієнт еластичності, припускаючи, що чисельність зайнятих складає 35 чоловік.

Завдання 9

По групі 10 заводів, що роблять однорідну продукцію, отримано рівняння регресії собівартості одиниці продукції y (тис. грн.) від рівня технічної оснащеності (тис. грн.): $y = 20 + \frac{700}{x}$. Частка залишкової дисперсії в загальній склала 0,19.

Необхідно визначити:

1. Коефіцієнт еластичності, припускаючи, що вартість активних виробничих фондів складає 200 тис. грн.;
2. Індекс кореляції;
3. F- критерій Фішера. Зробіть висновки.

Завдання 10

По 30 фермам області отримана інформація, представлена в таблиці 4.12.

Таблиця 4.12

Показник	Середні значення	Коефіцієнт варіації	Частка залишкової дисперсії в загальній
Врожайність, ц/га	30	24	12
Кількість добрив, кг	5	15	

Фактичне значення F- критерій Фішера = 35.

Необхідно:

1. Визначити лінійний коефіцієнт детермінації;
2. Побудувати рівняння лінійної регресії;
3. Розрахувати узагальнюючий коефіцієнт еластичності;
4. З імовірністю 0,95 % вказати інтервал довіри очікуваного значення врожайності в припущенні росту кількості внесених добрив на 12 % від свого середнього рівня.

Завдання 11

Залежність обсягу продажів y (тис. грн.) від витрат на рекламу x (тис. грн.) характеризується по 12 підприємствам концерну таким чином:

Таблиця 4.13

Рівняння регресії	$Y = 10,6 + 0,6x$
Середньоквадратичне відхилення x	$\sigma_x = 4,7$
Середньоквадратичне відхилення y	$\sigma_y = 3,4$
Частка факторної дисперсії в загальній	85 %

Необхідно:

1. Визначити коефіцієнт кореляції;
2. Оцінити значимість коефіцієнта регресії через t -критерій Ст'юдента, розрахувавши стандартну помилку коефіцієнта регресії.
3. Визначити довірчий інтервал для коефіцієнта регресії з імовірністю 0,95 і зробіть економічний висновок.

Завдання 12

Досліджуючи попит на телеапаратуру, маркетинговий центр компанії заданими, зібраним по 29 торговельних точок компанії, виявив наступну залежність:

$$\ln y = 10,5 - 0,8 \ln x,$$

де y – обсяг продаж телеапаратури в окремій торговельній точці;
 x – середня ціна телеапаратури в окремій торговельній точці.

Необхідно:

До проведення цього дослідження адміністрація компанії припускала, що еластичність попиту за ціною для телеапаратури складатиме – 0,9. Чи підтвердилося припущення адміністрації результатами дослідження?

Завдання 13

Залежність попиту на товар V від його ціни характеризується по 20 спостереженням рівнянням: $\lg y = 1,75 - 0,35 \lg x$. Частка залишкової дисперсії в загальній склала 18 %.

Необхідно:

1. Записати дане рівняння у вигляді степеневі функції;
2. Оцінити еластичність попиту на товар y залежності від його ціни;
3. Визначити індекс кореляції;
4. Оцінити значимість рівняння через F -критерій Фішера. Зробіть висновки.

Завдання 14

Вивчається залежність матеріалоемності продукції від розміру підприємства по 10 однорідних заводах

Таблиця 4.14

Показник	Матеріалоємність продукції по заводах									
	1	2	3	4	5	6	7	8	9	10
Споживання матеріалів на 1 од. продукції, кг	9	6	5	4	3,7	3,6	3,5	6	7	3,5
Випуск продукції, тис. од.	100	200	300	400	500	600	700	150	120	250

Необхідно:

1. Знайти параметри рівняння $y = a + \frac{b}{x}$;
2. Оцінити тісноту зв'язку за допомогою індексу кореляції;
3. Зробити висновки про значимості рівняння регресії.

Завдання 15

Ви оцінюєте таку регресію, яка дала наступні результати:

$$y = 2300 + 10.12x$$

$$SSE = 28225;$$

$$n = 28;$$

$$\sum [x_i - \bar{x}]^2 = 3300 .$$

Необхідно:

1. Перевірити значимість b при 95%-ному рівні довіри;
2. Побудувати 90%-ний інтервал довіри для b ;
3. Зробити висновки.

Тема 5. МНОЖИННА ЛІНІЙНА РЕГРЕСІЯ ТА КОРЕЛЯЦІЯ**Основні питання:**

5.1 Поняття класичної багатофакторної регресії. Специфікація моделі

5.2 Оцінка параметрів багатофакторної регресії

5.3 Коефіцієнти множинної кореляції та детермінації

5.4 Частні рівняння множинної регресії

5.5 Оцінка надійності результатів множинної регресії та кореляції



Ключові слова та поняття: Множинна регресія, середньоквадратичне відхилення, коефіцієнт варіації, коефіцієнт «чистої» регресії, індекс множинної кореляції, скорегований індекс множинної кореляції, частні рівняння множинної регресії, частні коефіцієнти еластичності, частковий F-критерій.

5.1 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 знехтувати. 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:

$$y = a + b_1x_1 + b_2x_2 + \dots + b_px_p + e \quad (5.1)$$

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 вимірні. 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 обумовленою 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 неінтерпретируемими. 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:

$$R_{p+1}^2 \geq R_p^2; S_{p+1}^2 \leq S_p^2 \quad (5.2)$$

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:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} . \quad (5.3)$$

Коефіцієнт варіації показує відносну міру відхилення окремих значень від середньоарифметичного:

$$V = \frac{\sigma}{\bar{x}} \cdot 100 . \quad (5.4)$$

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

$$N \geq \frac{V^2 \cdot t^2}{m^2} , \quad (5.5)$$

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 і їхніх помилок ε_a , ε_e .

$$A = \frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n \cdot \sigma^3} ; \quad \varepsilon_a = \sqrt{\frac{6}{n}} ; \quad (5.6)$$

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:

$$t_y = \beta_1 \cdot t_{x_1} + \beta_2 \cdot t_{x_2} + \dots + \beta_p \cdot t_{x_p} + \varepsilon, \quad (5.12)$$

де $t_y, t_{x_1}, \dots, t_{x_p}$ - 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 $\sigma_{t_y} = \sigma_{t_{x_i}} = 1$; β - стандартизовані коефіцієнти регресії.

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

$$\begin{cases} R_{xy1} = \beta_1 + \beta_2 \cdot R_{x_2x_1} + \beta_3 \cdot R_{x_3x_1} + \dots + \beta_p \cdot R_{x_px_1}, \\ R_{xy2} = \beta_1 \cdot R_{x_1x_2} + \beta_2 + \beta_3 \cdot R_{x_3x_2} + \dots + \beta_p \cdot R_{x_px_2}, \\ \dots \\ R_{yxp} = \beta_1 \cdot R_{x_px_1} + \beta_2 \cdot R_{x_2x_p} + \dots + \beta_p. \end{cases}$$

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.

5.3 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*:

$$R_{yx_1x_2\dots x_p} = \sqrt{1 - \frac{\sigma_{\text{зал}}^2}{\sigma_y^2}}, \sqrt{1 - \frac{\sum (y - \hat{y})^2}{\sum (y - \bar{y})^2}} \quad (5.13)$$

де σ_y^2 – загальна дисперсія результативної ознаки;

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

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:

$$R_{yx_1x_2\dots x_p} \geq R_{yx_i(\max)} \quad (i = \overline{1, p}).$$

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 у 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 убик 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 $\sum (y - \hat{y}_{x_1, x_2, \dots, x_p})^2$ поділяється на число ступенів волі залишкової варіації $(n - m - 1)$, а загальна сума квадратів відхилень $\sum (y - \bar{y})^2$ – на число ступенів волі в цілому по сукупності $(n - 1)$.

Формула скорегованого індексу множинної детермінації має вид:

$$\overline{R^2} = 1 - \frac{\sum (y - \hat{y})^2 : (n - m - 1)}{\sum (y - \bar{y})^2 : (n - 1)}, \quad (5.14)$$

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

$$\overline{R^2} = 1 - (1 - R^2) \cdot \frac{(n - 1)}{(n - m - 1)}. \quad (5.15)$$

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 стати 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

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

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

5.4 Частні рівняння множинної регресії

На основі лінійного рівняння множинної регресії

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

можуть бути знайдені **частні рівняння регресії**:

$$\left\{ \begin{array}{l} y_{x_1 \cdot x_2, x_3, \dots, x_p} = f(x_1), \\ y_{x_2 \cdot x_1, x_3, \dots, x_p} = f(x_2), \\ \dots \\ y_{x_p \cdot x_1, x_2, \dots, x_{p-1}} = f(x_p), \end{array} \right. \quad (5.16)$$

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:

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

де $D_{\text{факт}}$ – факторна сума квадратів на одну ступінь вільності;
 $D_{\text{зал}}$ – 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:

$$F_{x_1} = \frac{R^2_{yx_1x_2\dots x_p} - R^2_{yx_2\dots x_p}}{1 - R^2_{yx_1x_2\dots x_p}} \times \frac{n - m - 1}{1}, \quad (5.21)$$

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

$R^2_{yx_2\dots x_p}$ – 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, \dots, x_{p-1} , то формула частного F -критерію прийме вигляд:

$$F_{x_p} = \frac{R^2_{yx_1x_2\dots x_p} - R^2_{yx_1x_2\dots x_{p-1}}}{1 - R^2_{yx_1x_2\dots x_p}} \times \frac{n - m - 1}{1}. \quad (5.22)$$

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

$$F_{x_i} = \frac{R^2_{yx_1\dots x_i\dots x_p} - R^2_{yx_1\dots x_{i-1}x_{i+1}\dots x_p}}{1 - R^2_{yx_1\dots x_i\dots x_p}} \times \frac{n - m - 1}{1}. \quad (5.23)$$

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 $F_{\text{табл}}(\alpha, df_1, df_2)$, 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:

$$t_{b_i} = \sqrt{F_{x_i}}. \quad (5.24)$$

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 двофакторного, 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

$$t_{b_i} = \frac{b_i}{m_{b_i}}, \quad (5.25)$$

де b_i – коефіцієнт чистої регресії при факторі x_i ;
 m_{b_i} – стандартна помилка коефіцієнта регресії b_i .

Для рівняння множинної регресії

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

стандартна помилка коефіцієнта регресії може бути визначена по наступній формулі:

$$m_{b_i} = \frac{\sigma_y \cdot \sqrt{1 - R^2_{yx_1 \dots x_p}}}{\sigma_{x_i} \cdot \sqrt{1 - R^2_{x_i x_1 \dots x_p}}} \cdot \frac{1}{\sqrt{n - m - 1}}, \quad (5.26)$$

де σ_y – середньоквадратичне відхилення для ознаки y ;

σ_{x_i} – середньоквадратичне відхилення для ознаки x_i ;

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

$R^2_{x_i x_1 \dots x_p}$ – 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 $R^2_{x_i x_1 \dots x_p}$. 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

$$b_i = \frac{\sigma_y}{\sigma_{x_i}} \cdot \sqrt{\frac{R^2_{yx_1 \dots x_p} - R^2_{yx_1 \dots x_{i-1} x_{i+1} \dots x_p}}{1 - R^2_{x_i x_1 \dots x_p}}}, \quad (5.27)$$

то можна переконатися, що

$$t_{b_i} = \frac{b_i}{m_{b_i}} = \sqrt{F_{x_i}} . \quad (5.28)$$



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?



Завдання для самостійного розв'язання



Тестовий контроль для перевірки знань

1. У багатофакторній регресії:
 - а) більш ніж одна залежна змінна та тільки одна незалежна;
 - б) більш ніж одна незалежна змінна та тільки одна залежна;
 - в) більш ніж одна залежна змінна та більш ніж одна незалежна.

2. При геометричній інтерпретації регресійної моделі з двома незалежними змінними ми будемо:

- а) пряму лінію, щоб показати зв'язок між залежною змінною та незалежними змінними;
- б) трикутник, щоб показати зв'язок між залежною змінною та незалежними змінними;
- в) площину, щоб показати зв'язок між залежною змінною та незалежними змінними.

3. Специфікація моделі багатofакторної регресії – це:

- а) побудова лінії регресії з декількома факторами;
- б) відбір факторів та виду математичної функції;
- в) запис рівняння багатofакторної регресії в наявному вигляді.

4. Однією з вимог відбору факторів багатofакторної регресії є:

- а) тісний зв'язок між факторами;
- б) фактори повинні підлягати якісному вимірюванню;
- в) фактори повинні кількісно вимірюватися.

5. При додатковому включенні факторів до множинної регресії повинна виконуватися тотожність:

- а) $R_{p+1}^2 \geq R_p^2$;
- б) $S_{p+1}^2 \geq S_p^2$;
- в) $R_{p+1}^2 \leq R_p^2$.

6. Інтеркореляція факторів – це умова:

- а) $R_{yx_1} \leq R_{x_1x_2}$;
- б) $R_{yx_1} \geq R_{x_1x_2}$;
- в) $R_{yx_1} = R_{x_1x_2}$.

7. Насичення моделі зайвими факторами:

- а) зменшує залишкову дисперсію;
- б) збільшує коефіцієнт детермінації;
- в) призводить до невизначеності параметрів регресії.

8. Оцінка параметрів багатofакторної регресії здійснюється за допомогою:

- а) МНК;
- б) t-тесту;
- в) F-критерію.

9. У множинній регресії кожний параметр показує:

- а) загальний вплив усіх незалежних змінних на залежну змінну;
- б) вплив незалежної змінної на залежну за умови, що всі інші незалежні змінні залишаються незмінними;
- в) де площина регресії перетинає вісь y .

10. Лінійна багатофакторна регресія має вигляд:

а) $\hat{y}_x = a + bx + cx^2$;

б) $\hat{y}_x = a + bx_1 + cx_2$;

в) $\hat{y}_x = a + \frac{b}{x} + \frac{c}{x_1}$.

11. Дана залежність $\hat{y}_x = 13.1 - 0.7x_1 + 0.8x_2$. Якщо x_1 збільшується на 1 одиницю, а x_2 не змінюється, то y :

- а) зменшується на 0,7 одиниць;
- б) збільшується на 1 одиницю;
- в) зменшується на (0,8+0,7) одиниць.

12. Якщо за даними тесту 11 y збільшиться на 1 одиницю, то:

- а) x_2 збільшиться на 0,8 одиниць;
- б) x_1 зменшиться на 0,7 одиниць;
- в) жодної вірної відповіді.

13. Якщо за даними тесту 11 x_2 збільшиться на 1 одиницю, а x_1 не змінюється, то:

- а) x_1 зменшиться на 0,7 одиниць;
- б) y в середньому збільшиться на 0,8 %;
- в) y в середньому збільшиться на 0,8 одиниць.

14. Коефіцієнти, які стоять біля параметрів багатофакторної регресії, називаються коефіцієнтами:

- а) еластичності;
- б) кореляції;
- в) регресії.

15. Частині рівняння регресії характеризують:

- а) ізольований вплив x_i на y ;
- б) сукупний вплив всіх факторів x_i на y ;
- в) щільність зв'язку між x_i та y .

16. Для рівняння $\hat{y}_x = 0.82 \cdot x_1^{-2.63} \cdot x_2^{1.11}$ чи вірним буде запис:

а) $\hat{y}_x = 0.82 + x_1^{-2.63} + x_2^{1.11}$;

б) $\hat{y}_x = 0.82 \cdot \frac{x_2^{1.11}}{x_1^{2.63}}$;

в) $\lg \hat{y}_x = 0.82 \cdot \lg x_1^{-2.63} \cdot \lg x_2^{1.11}$.

17. Для рівняння $\hat{y}_x = 0.82 \cdot x_1^{-2.63} \cdot x_2^{1.11}$ при зростанні x_1 на 1% та незмінності x_2 \hat{y}_x :

- а) зменшиться на 2,63 %;
- б) збільшиться на 1,11 %;
- в) зменшиться на (-2,63+1,11) %.

18. Показник множинної кореляції характеризує:

- а) ізолюваний вплив x_i на y ;
- б) абсолютний вплив факторів на y ;
- в) щільність зв'язку спільного впливу факторів на результат.

19. Для перевірки значущості одночасно всіх параметрів використовується:

- а) F-тест;
- б) t-тест;
- в) χ^2 -тест.

20. За інших рівних умов, якщо ми збільшуємо кількість незалежних змінних у регресії:

- а) R^2 збільшується;
- б) R^2 зменшується;
- в) R^2 може або збільшитися, або зменшитися.

21. Зв'язок між R^2 та оціненим R^2 є:

- а) оцінене $R^2 = R^2$;
- б) оцінене $R^2 = R^2(n-1)/(n-m-1)$;
- в) оцінене $R^2 = [1 - (1 - R^2)] (n-1) / (n-m-1)$.

22. Ступені вільності чисельника F-статистики в регресії, що складається з 50 спостережень та 4 незалежних змінних, такі:

- а) 50;
- б) 4;
- в) 45.

23. Ступені вільності знаменника F-статистики, що складається з 50 спостережень та 4 незалежних змінних, такі:

- а) 50;
- б) 4;
- в) 45.

24. Однією з проблем, що може виникнути у багатофакторній регресії і ніколи не буває в простій регресії, є:

- а) кореляція між величинами помилок;
- б) кореляція між помилками та незалежними змінними;
- в) кореляція між незалежними змінними.

25. Щоб перевірити значимість окремого параметра, використовують:

- а) F-тест;
- б) t-тест;
- в) χ^2 -тест.



Приклади розв'язання завдань

Завдання 1

Підприємство має деяку кількість філій. Керівництво цього підприємства бажало б знати, як y (річний товарообіг однієї філії, тис. доларів) залежить від x_1 – торговельної площі (тис. м²) та x_2 – середньоденної інтенсивності потоку покупців (тис. чол. /день).

Таблиця 15 – Дані філій підприємства

№ з/п	y	x_1	x_2
1	2,93	0,31	10,24
2	5,27	0,98	7,51
3	6,85	1,21	10,81
4	7,01	1,29	9,89
5	7,02	1,12	13,72
6	8,35	1,49	13,92
7	4,33	0,78	8,54
8	5,77	0,94	12,36
9	7,68	1,29	12,27
10	3,16	0,48	11,01
11	1,52	0,24	8,25
12	3,15	0,55	9,31

Необхідно:

1. Проінтерпретувати результати оцінки, якщо багатофакторна регресія має вигляд:

$$y = -0,832 + 4,743x_1 + 0,175x_2.$$

2. Спрогнозувати річний товарообіг філії, яка буде побудована в заселеному районі міста з середньоденною інтенсивністю потоку покупців 15

тис. чоловік на день та торговельною площею 1200 м².

Рішення:

1. Оцінки інтерпретуються таким чином: якщо значення x_2 буде постійним, а змінна x_1 збільшиться (зменшиться) на одиницю, то відповідно цій оцінці y зросте (зменшиться) на 4,743 одиниці. Для нашого випадку: збільшення торговельної площі на 1 тис. м² зумовить в середньому збільшення річного товарообігу на 4,743 тис. доларів, якщо інтенсивність потоку покупців залишиться тією же. Та якщо потік покупців зросте на 1 тис. чол. /день, то це зумовить в середньому збільшення річного товарообігу на 0,175 тис. доларів при незмінності торговельної площі.

2. Якщо багатофакторна регресійна модель має вигляд $y = -0,832 + 4,743x_1 + 17,5x_2$, то річний товарообіг філії буде дорівнювати: $y = -0,832 + 4,743 \times 1,2 + 0,175 \times 15 = -0,832 + 5,692 + 2,625 = 7,485$.

Таким чином, прогнозований за допомогою багатофакторної регресії річний товарообіг філії буде дорівнювати 7,485 тис. доларів.

Завдання 2

В таблиці наведені умовні дані, які характеризують стан промислового сектора країни, які можуть бути описані нелінійною функцією Кобба-Дугласа:

$$y = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot e^\varepsilon,$$

де y – випуск продукції, x_1 – витрати праці, x_2 – витрати капіталу, a – технологічний коефіцієнт, ε – випадкова величина (залишки або помилки), b_1 , b_2 – коефіцієнти еластичності відповідних факторів, e – основа натурального логарифма.

Таблиця 16

№ з/п	Обсяг виробництва, млн. доларів	Витрати праці, днів	Витрати капіталу, млн. доларів
1	2	3	4
1	19607,7	305,4	20803,7
2	20511,3	304,4	21096,8
3	23171,2	299,7	21271,8
4	23932,9	297,0	22167,3
5	23406,0	297,8	21647,6
1	2	3	4
6	23831,6	305,0	23803,5
7	27806,3	313,0	25076,6
8	29465,8	330,7	26445,2
9	30403,0	337,5	27939,0

10	32628,7	333,7	29713,7
11	32904,5	334,7	32957,8
12	30508,2	326,6	33585,9
13	32035,5	325,5	35474,5
14	32281,5	329,0	37821,5
15	33535,8	318,1	43794,3

На основі проведених розрахунків модель має вигляд:
 $\ln y = 1,617 \ln x_1 + 0,458 \ln x_2 - 3,775$.

Необхідно:

- Зробити аналіз розвитку промислового сектора України, тобто:
 - надайте оцінку параметрам моделі;
 - перевірте модель на адекватність.

Рішення:

1. Первісний вигляд моделі розвитку промислового сектора країни, відповідно, такий:

$$y = e^{-3,755} \cdot x_1^{1,617} \cdot x_2^{0,458}; \quad y = 0,096 \cdot x_1^{1,617} \cdot x_2^{0,458}$$

У цій моделі параметри b_1, b_2 – коефіцієнти еластичності відповідних факторів виробництва.

Коефіцієнт еластичності праці $E_1 = 1,617 > 1$, тому цей фактор є еластичним, тобто зміна його на 1% призводить до зміни обсягів виробництва на 1,617%.

У той же час чутливість капіталу є нижчою і становить $E_2 = 0,458 < 1$, тобто капітал є нееластичним фактором.

Технологічний коефіцієнт $e^{-3,755} = 0,096$ сам по собі невеликий, до того ж t-оцінка його незначна, тобто технології у виробничому секторі країни використовуються недостатньо.

Сума коефіцієнтів $b_1 + b_2 = 1,617 + 0,458 = 2,075 > 1$. Це означає, що виробництво має зростаючі масштаби, тобто його розвиток є перспективним.

2. Перевіримо побудовану модель на адекватність.

Розглянемо коефіцієнт детермінації R^2 , що є мірою пояснювальної сили незалежних змінних і розраховується за формулою

$$R^2 = SSR/SST = 1 - SSE/SST; \quad 0 \leq R^2 \leq 1.$$

У нашому випадку R^2 дорівнює 0,864. Це свідчить про те, що отримане рівняння регресії на 86,4% пояснює загальний розкид даних відносно середнього загального обсягу виробництва.

Для підтвердження адекватності моделі використаємо F-критерій Фішера.

Задамо рівень значущості $\alpha = 5\%$. За спеціальною таблицею F-розподілу Фішера при 2 та 12 ступенях вільності та рівні значущості 5% знайдемо

критичне значення $F_{кр}$ і порівняємо його з розрахованим F-відношенням.

$F_{кр(2,12,95\%)} = 3,88$, тобто $F > F_{кр}$, що свідчить про адекватність побудованої моделі.

Завдання 3

Відомо, що у деякої статистиці з 17-ти спостережень

$$SSE = \sum (y_i - \hat{y}_i)^2 = 679; SST = \sum (y - \bar{y})^2 = 3719.$$

Необхідно:

1. Використати дані для перевірки якості двофакторної моделі при 5%-му рівні значущості.

Рішення:

Якість побудованої моделі використання F-тесту.

$$F_{p,n-p-1} = (SSR/p) / (SSE/n-p-1);$$

$$SST = SSE + SSR;$$

$$SSR = 3719 - 679 = 3040;$$

$$F = [3040/2] / [679/14] = 31,3.$$

Критичне значення для $F_{2,14,0,05\%} = 3,74$, отже, ми можемо відкинути нульову гіпотезу, тому що $31,3 > 3,74$ ($F > F_{кр}$).

Завдання 4

Ви оцінюєте таку регресію, використовуючи 30 спостережень:

$$y_i = 122 + 32,4x_1 - 0,78x_2;$$

$$r_{x_1,x_2} = 0,59;$$

$$\sum (x_{1i} - \bar{x}_1)^2 = 23,6;$$

$$\sum (x_{2i} - \bar{x}_2)^2 = 12,3; \hat{\sigma}_e^2 = 41,7.$$

Необхідно:

Перевірити значущість параметрів при рівні значущості 5%.

Рішення:

$$\hat{\sigma}_{b1}^2 = \frac{\hat{\sigma}_e^2}{(1 - r_{x_1,x_2}^2) \sum (x_{1i} - \bar{x}_1)^2} = \frac{41,7}{(1 - 0,59) \times (23,6)} = 4,31;$$

$$\hat{\sigma}_{b1} = \sqrt{4,31} = 2,08;$$

$$\hat{\sigma}_{b2}^2 = \frac{\hat{\sigma}_e^2}{(1 - r_{x_1,x_2}^2) \sum (x_{2i} - \bar{x}_2)^2} = \frac{41,7}{(1 - 0,59) \times (12,3)} = 8,27;$$

$$\hat{\sigma}_{b2} = \sqrt{8,27} = 2,88;$$

$$t_{b1} = 32,4 / 2,08 = 15,58;$$

$$t_{b_2} = -0,78/2,88 = -2,71.$$

$t_{0,025;27} = \pm 2,052$, отже, ми відкидаємо нульову гіпотезу, що $b_1=0$, але приймаємо нульову гіпотезу, що $b_2=0$.

Завдання 5

Припустимо, що по ряду регіонів множинна регресія величини імпорту на визначений товар у відносно вітчизняного його виробництва x_1 , зміни запасів x_2 і споживання на внутрішньому ринку x_3 виявилася наступною:

$$\hat{y} = -66,028 + 0,135x_1 + 0,476x_2 + 0,343x_3.$$

При цьому середні значення для розглянутих ознак склали:

$$\bar{y} = 31,5; \quad \bar{x}_1 = 245; \quad \bar{x}_2 = 3,7; \quad \bar{x}_3 = 182,5.$$

Необхідно визначити:

1. Значення середніх по сукупності показників еластичності;
2. Часткові рівняння регресії.

Рішення

1. Середні коефіцієнти еластичності визначаються по формулі:

$$\bar{E} = b_i \frac{\bar{x}_i}{\bar{y}}.$$

Для даної умови $\bar{E}_1 = 0,135 \frac{245,7}{31,5} = 1,053$ %, тобто з ростом величини вітчизняного виробництва на 1 % розмір імпорту в середньому по сукупності регіонів зросте на 1,053 % при незмінних запасах і споживанні родин.

$\bar{E}_2 = 0,476 \frac{3,7}{31,5} = 0,056$ %. Це означає, що з ростом зміни запасів при незмінному виробництві і внутрішнім споживанні величина імпорту збільшується в середньому на 0,056 %.

$\bar{E}_3 = 0,343 \frac{182,5}{31,5} = 1,987$ %. Це значить, що при незмінному обсязі виробництва і величини запасів зі збільшенням внутрішнього споживання на 1 % імпорт товару зростає в середньому по сукупності регіонів на 1,987 %.

2. Часткові рівняння регресії складуть:

$$\hat{y}_{x_1 x_2 x_3} = a + b_1 x_1 + b_2 \bar{x}_2 + b_3 \bar{x}_3 = -1,669 + 0,135 x_1;$$

$$\hat{y}_{x_2 x_1 x_3} = a + b_1 \bar{x}_1 + b_2 x_2 + b_3 \bar{x}_3 = 29,739 + 0,476 x_2;$$

$$\hat{y}_{x_3 x_1 x_2} = a + b_1 \bar{x}_1 + b_2 \bar{x}_2 + b_3 x_3 = -31,097 + 0,343 x_3.$$



Завдання для розв'язання

Завдання 1

Дано рівняння $y=1,215+0,469x_1+1,026x_2+0,65x_3$.

Необхідно визначити:

1. До якого типу моделей відноситься наведене рівняння?
2. На скільки зміна одного з факторів зумовить зміну показника?

Завдання 2

Припустимо, що нас цікавить зв'язок між досвідом, навчання та доходом. На підставі певної статистики ми одержали таку модель:

$$y=202,1+14x_1+35,1x_2,$$

де y – доход людини (грн);

x_1 – досвід (роки);

x_2 – роки навчання.

Необхідно:

1. Проаналізувати коефіцієнти регресії для x_1 та x_2 .

Завдання 3

Маємо таку статистику по філіях підприємства, де y (річний товарообіг однієї філії, тис. доларів) функціонально залежить від x_1 – торговельної площі (m^2) та x_2 – середньоденної інтенсивності потоку покупців (тис. чол. /день).

Таблиця 17

№ з/п	y	x_1	x_2
1	2	3	4
1	2,93	0,31	10,24
2	5,27	0,98	7,51
3	6,85	1,21	10,81
4	7,01	1,29	9,89
5	7,02	1,12	13,72
6	8,35	1,49	13,92
7	4,33	0,78	8,54
1	2	3	4
8	5,77	0,94	12,36
9	7,68	1,29	12,27
10	3,16	0,48	11,01
11	1,52	0,24	8,25
12	3,15	0,55	9,31

Використовуючи отриману регресійну модель

$$y = -0,832 + 4,743x_1 + 0,175x_2,$$

складіть прогноз товарообігу за рік філії, яка буде побудова у заселеному районі міста зі середньоденною інтенсивністю потоку покупців 7500 чоловік на день та торговельною площею 900 м².

Завдання 4

Припустимо, що у нашому прикладі (завдання 3) керівництво підприємства ставить собі за мету побудувати філію на вулиці з середньоденною інтенсивністю потоку покупців 10000 чол./день. При цьому воно бажає отримати товарообіг за рік у обсягу 7 тис. доларів. Інструментом досягнення цієї мети слугуватиме торговельна площа філії.

Завдання 5

Продуктивність праці, фондівіддача та рівень рентабельності характеризуються такими даними (табл. 18) по плодоконсервних заводах області за рік.

Таблиця 18

№ завод у	Чинник		Рівень рентабельності, %
	Продуктивність праці у розрахунку на 1 робітника, грн.	Фондовіддача, грн.	
1	2	3	4
1	11540	1,24	39,4
2	2911	0,63	23,2
3	6630	1,18	37,2
4	8492	1,12	35,1
5	2901	0,44	20,0
6	9410	1,19	37,9
7	1920	0,48	20,1
8	2569	0,65	23,4
9	3520	0,26	13,4
10	2340	0,75	24,8
1	2	3	4
11	6921	1,03	32,2
12	7671	0,89	30,2
13	1586	0,16	10,3
14	3223	0,67	23,7
15	7224	0,9	31,3

Необхідно:

1. Побудувати регресійну модель, оцінивши її параметри.

Завдання 6

Ви оцінюєте таку модель:

$$y = a + b_1x_1 + b_2x_2,$$

що базується на 30 спостереженнях, і отримали

$$\sum (y_i - \hat{y}_i)^2 = 1203,7, \quad \sum (y - \bar{y})^2 = 1500,36.$$

Необхідно:

1. Підрахувати коефіцієнт детермінації та дайте якісну оцінку моделі;
2. Чи добре модель пояснює виявлену закономірність?

Завдання 7

Ми оцінюємо регресію, використовуючи 50 спостережень та 4 незалежних змінні. R^2 для регресії становить 0,64.

Необхідно:

1. Визначити оцінений коефіцієнт детермінації R^2 для регресії.

Завдання 8

Регресія залежної змінної Y на три незалежні змінні на основі $n = 30$ спостережень дала такі результати:

$$Y = 25,1 + 1,2 X_1 + 1,0 X_2 - 0,50 X_3$$

Стандартні помилки	(2,1)	(1,5)	(1,3)	(0,06)
t - значення	(11,9)	()	()	()
95 %-ні інтервали довіри	(±4,3)	()	()	()

Необхідно:

1. Заповнити пропуски в дужках, використовуючи вихідну інформацію.

Завдання 9

Маємо таку статистику:

Таблиця 19

y	X ₁	X ₂
925	210	4,8
850	185	4,7
1622	225	4,7
1121	215	4,6
658	180	4,9

977	212	4,6
574	195	5,0

Необхідно:

1. Побудувати модель багатофакторної регресії, оцінивши її параметри.

Завдання 10

По 20 регіонах вивчаються наступні дані (табл. 20): залежність доходу y (ум. од.) від частки зайнятих важкою фізичною працею в загальній чисельності зайнятих x_1 (%) і від частки економічно активного населення в чисельності всього населення x_2 (%).

Таблиця 20

Ознака	Середнє значення	Середнє-квадратичне відхилення	Характеристика щільності зв'язку	Рівняння зв'язку
y	112,76	31,58	$R_{yx_1x_2} = 0,733$	$\hat{y}_{x_1x_2} = -130,49 + 6,14x_1 + 4,13x_2$
x_1	5,4	3,34	$r_{yx_1} = 0,746$	$\hat{y}_{x_1} = 74,4 + 7,1x_1$
x_2	50,88	1,74	$r_{yx_2} = 0,507$ $r_{x_1x_2} = 0,432$	$\hat{y}_{x_2} = -355,3 + 9,2x_2$

Необхідно:

1. Скласти таблицю дисперсійного аналізу для перевірки при рівні значимості $\alpha = 0,05$ статистичної значимості рівняння багатофакторної регресії та її показника щільності зв'язку.

Розділ 3. Нелінійне моделювання та випадки множинної регресії

Тема 6. НЕЛІНІЙНА РЕГРЕСІЯ



Основні питання

6.1 Специфікація нелінійної моделі

6.2 Коефіцієнти еластичності для математичних функцій



Ключові слова та поняття: нелінійна регресія, класи нелінійних регресій, ступенева функція, показова функція, експонентна функція, коефіцієнти еластичності для нелінійних функцій, кореляція для нелінійної регресії

6.1 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 .

6.2 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:

$$E = f(x) \frac{x}{y}, \quad (6.1)$$

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:

$$\bar{E} = b \times \frac{\bar{x}}{\bar{y}} \quad (6.2)$$

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 6.1 - 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 $y = a + b \ln x$	$E = \frac{c \cdot x}{\frac{1}{b} \cdot e^{cx} + 1}$

6.3 Кореляція для нелінійної регресії

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

$$R = \left(1 - \frac{\sigma_{\text{зал}}^2}{\sigma_y^2} \right)^{\frac{1}{2}}, \quad (6.3)$$

де σ_y^2 - загальна дисперсія результативної ознаки y ;
 $\sigma_{\text{зал}}^2$ - залишкова дисперсія.

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

виразити як

$$R = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}} \quad (6.4)$$

Величина даного показника знаходиться в межах $0 \leq R \leq 1$: 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:

$$F_{\text{факт}} = \frac{R^2}{1-R^2} \times \frac{n-m-1}{m}, \quad (6.5)$$

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:

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

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

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

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*:

$$t = \frac{R_{yx}^2 - r_{yx}^2}{m_{|R-r|}}, \quad (6.6)$$

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

$$m_{|R-r|} = 2 \cdot \sqrt{\frac{(R^2 - r^2) - (R^2 - r^2)^2 \cdot (2 - (R^2 + r^2))}{n}}. \quad (6.7)$$

Якщо $t_{\text{факт}} > t_{\text{табл}}$, 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 перевірки засвоєних 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 7. МУЛЬТИКОЛІНЕАРНІСТЬ ЯК ОСОБЛИВИЙ ВИПАДОК МНОЖИННОЇ РЕГРЕСІЇ



Основні питання

7.1 Суть мультиколінеарності

7.2 Наслідки мультиколінеарності

7.3 Визначення мультиколінеарності та методи її усунення



Ключові слова та поняття: мультиколінеарність, строга лінійна залежність, наслідки мультиколінеарності, частний коефіцієнтом кореляції, перетворення змінних

7.1 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

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon. \quad (7.1)$$

Нехай також між пояснюючими змінними існує строга лінійна залежність

$$X_2 = \gamma_0 + \gamma_1 X_1. \quad (7.2)$$

Підставивши (7.2) в (7.1), 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:

$$Y = a + b X_1 + \varepsilon. \quad (7.3)$$

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

$$\begin{cases} \beta_0 + \beta_2 \gamma_0 = a, \\ \beta_1 + \beta_2 \gamma_1 = b. \end{cases} \quad (7.4)$$

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 недосконалою 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 поведінку залежної змінної й робить оцінки коефіцієнтів регресії ненадійними. Даний висновок наочно підтверджується за допомогою діаграми Венна (рис. 7.1)

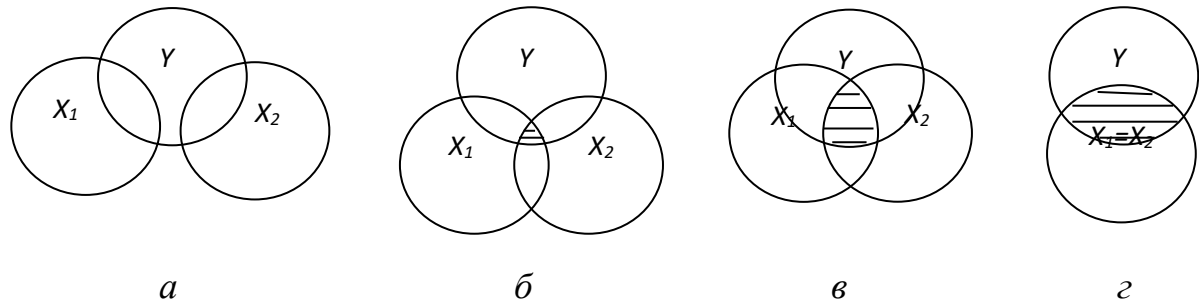


Рис. 7.1

На рис. 7.1, 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. 7.1, г при досконалої мультиколінеарності неможливо розмежувати ступені індивідуального впливу пояснюючих змінних X_1 і X_2 на залежну змінну Y .

7.2 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 не виправданого 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.

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

2. Можливе одержання невірної знака у коефіцієнта регресії.

7.3 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 *частковий 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

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} \quad (7.5)$$

У загальному випадку вибіркового часткового коефіцієнта кореляції між змінними X_i і X_j ($1 \leq i < j \leq m$), очищений від впливу інших $(m - 2)$ пояснюючих змінних, символічно позначається

$$r_{yj.12\dots(j-1)(j+1)\dots(j-1)(j+1)\dots m} = \frac{-c_{ij}^*}{\sqrt{c_{ii}^* c_{jj}^*}}, \quad (7.6)$$

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

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

$$r_{ij.kl} = \frac{r_{ij/k} - r_{il.k} \cdot r_{jl.k}}{\sqrt{(1 - r_{ij.k}^2)(1 - r_{jl.k}^2)}}. \quad (7.7)$$

Нехай $r_j = r_{yj.12\dots(j-1)(j+1)\dots m}$ – 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, $r_j^2, j = 1, 2, \dots, m$, 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 $X_j, j = 1, 2, \dots, m$, на пояснючі змінні, які залишилися. Обчислюються відповідні коефіцієнти детермінації R_j^2 й розраховується їх статистична значимість на основі F -статистики:

$$F_j = \frac{R_j^2}{1 - R_j^2} \cdot \frac{n - m}{m - 1}, \quad (7.8)$$

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 протилежному 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:

$$VIF_i = \frac{1}{1 - R_i^2} \quad (7.9)$$

Якщо $VIF_i \leq 10$, то можна стверджувати, що зв'язок між i -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 (7.1). 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 (7.1) прийме вид:

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

$$Y - 0,8X_1 = \beta_0 + \beta_2 X_2 + \varepsilon. \quad (7.10)$$

Рівняння (7.10) 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

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2, \quad (7.11)$$

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

$$\frac{\hat{Y}}{X_1} = b_0 + b_1 \frac{X_2}{X_1},$$

$$\frac{\hat{Y}}{X_2} = b_0 + b_1 \frac{X_1}{X_2}.$$
(7.12)

Цілком імовірно, що в моделях, аналогічних (7.12), 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 перевірки засвоєних знань

1. Поясніть значення термінів «колінеарність» і «мультиколінеарність».
2. У чому розходження між досконалою й недосконалою мультиколінеарністю?
3. Які основні наслідки мультиколінеарності?
4. Як можна виявити мультиколінеарність?
5. Як оцінюється корельованість між двома пояснюючими змінними?
6. Перелічіть 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 v . Будується нове рівняння регресії с залежною змінною Y і двома пояснюючими змінними – X_2 и v . Чи буде вирішена таким чином проблема мультиколінеарності?

Тема 8. СУТНІСТЬ ГОМО- ТА ГЕТЕРОСКЕДАСТИЧНОСТІ



Основні питання

8.1 Сутність гетероскедастичності та її наслідки

8.2 Виявлення гетероскедастичності. Методи пом'якшення проблеми гетероскедастичності



Ключові слова та поняття: передумови МНК, гомоскедастичність, гетероскедастичність, проблема гетероскедастичності, виявлення гетероскедастичності, тест Парка, тест Глейзера, тест Голдфелда-Квандта

8.1 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 $\text{var}(\varepsilon_i) = \sigma_\varepsilon^2 = \text{const}$.

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

Сутність припущення гомоскедастичності полягає в тому, що σ_ε^2 не є функцією x_{ij} , тобто $\sigma_\varepsilon^2 \neq f(x_{1i}, x_{2i}, \dots, x_{pi})$. Графічно випадок гомоскедастичності зображений на рис. 8.1.

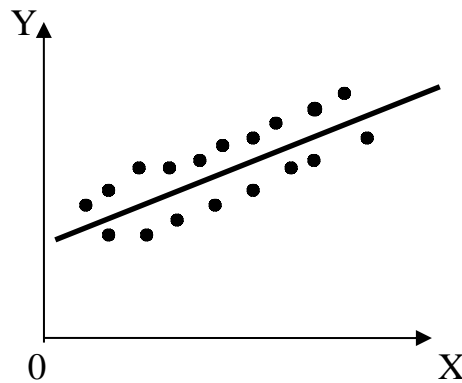


Рис 8.1. Гомоскедастичність

Якщо σ_ε^2 не є постійною, її значення залежать від значень x_{ij} , то можна записати $\sigma_\varepsilon^2 = f(x_{1i}, x_{2i}, \dots, x_{pi})$. 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).

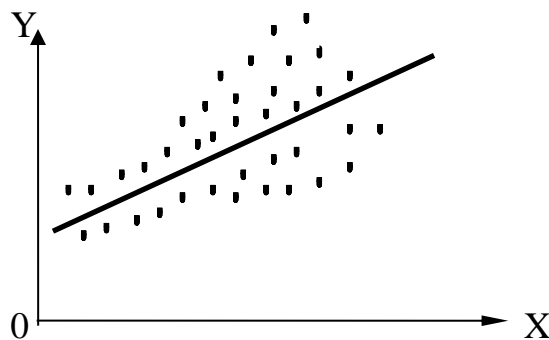


Рис 8.2. Гетероскедастичність (зростання дисперсії ε_i).

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, ВВП, 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, однією з яких є сталість дисперсії відхилень (гомоскедастичність): $\text{var}(\varepsilon_i) = \sigma_\varepsilon^2 = \text{const}$ для всіх спостережень i , $i = 1, 2, \dots, n$.

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 не пояснена рівнянням регресії дисперсія

$$\sigma^2 = \frac{\sum e_i^2}{n - m - 1}$$
 (m – число пояснюючих змінних), що використовується 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.

8.2 Виявлення гетероскедастичності. Методи пом'якшення проблеми гетероскедастичності

У ряді випадків, знаючи характер даних, поява проблеми гетероскедастичності 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:

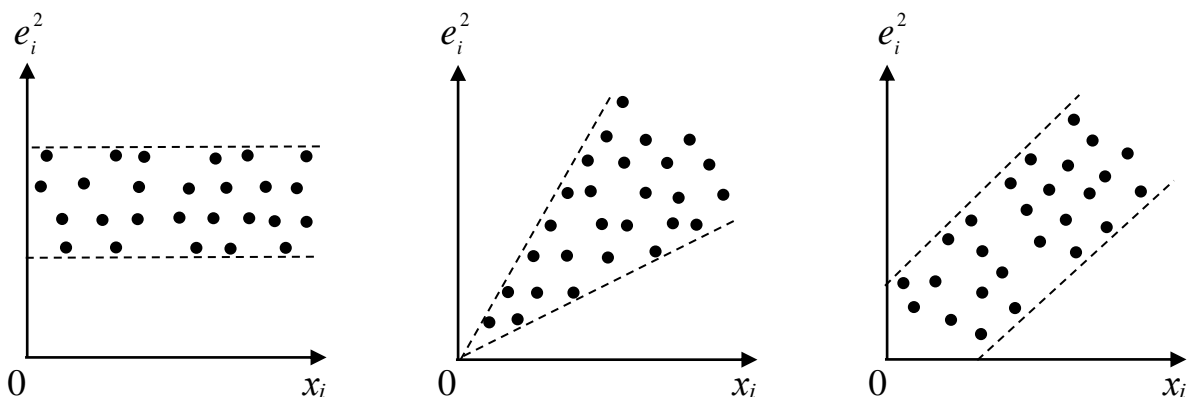
1. Графічний аналіз відхилень,
2. Тест рангової кореляції Спірмана,
3. Тест Парка,
4. Тест Глейзера,
5. Тест Голдфелда-Квандта.

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 $Y = b_0 + b_1X_1 + \dots + b_mX_m$), а по осі ординат або відхилення e_i , або їхні квадрати e_i^2 , $i = 1, 2, \dots, n$. 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 їхньої сталості, тобто в цьому випадку виконуються умови гомоскедастичності.

На рис. 8.3, *б-д* спостерігаються деякі систематичні зміни в співвідношеннях між значеннями x_i змінної X і квадратами відхилень e_i^2 . На рис. 8.3, *б* и *в* відбита лінійна, *г* – квадратична, *д* – гіперболічна залежності між квадратами відхилень і значеннями пояснюючої змінної X . Інакше кажучи, ситуації, представлені на рис. 8.3, *б* — *д*, відбивають більшу ймовірність наявності гетероскедастичності для розглянутих статистичних даних.



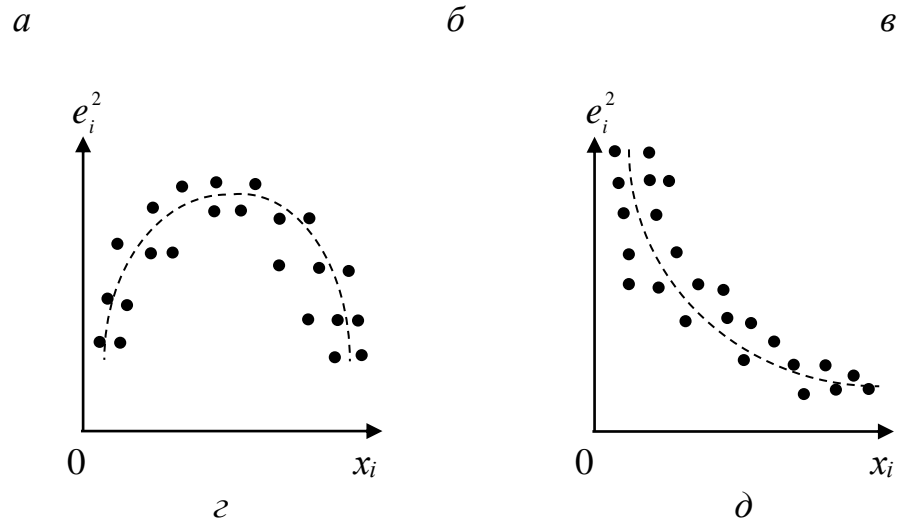


Рис. 8.3

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 , $j = 1, 2, \dots, m$, окремо. Частіше ж замість пояснюючих змінних X_j по осі абсцис відкладають значення \hat{y}_i , $i = 1, 2, \dots, n$, одержані з емпіричного рівняння регресії. Оскільки за рівнянням множинної лінійної регресії \hat{y}_i є лінійною комбінацією x_{ij} , $j = 1, 2, \dots, m$, $i = 1, 2, \dots, n$, 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 і 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:

$$r_s = 1 - 6 \left[\frac{\sum_{i=1}^n d_i}{n(n^2 - 1)} \right], \quad (8.1)$$

де d_i – різниця між рангами e_i і x_i , $i = 1, 2, \dots, n$, 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 $\nu = n - 2$:

$$t = \frac{r_{x,e} \sqrt{n-2}}{\sqrt{1-r_{x,e}^2}}, \quad (8.2)$$

де n – кількість спостережень.

Якщо спостережуване значення t -статистики, обчислене за формулою, перевищує $t_{кр} = t_{\alpha/2, n-2}$ (обумовлене за таблицею критичних крапок розподілу Ст'юдента), це підтверджує гіпотезу про гетероскедастичність. У протилежному 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:

$$\sigma_i^2 = \sigma^2 x_i^\beta e^{\nu_i}. \quad (8.3)$$

Прологарифмував (7.3), одержимо:

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln x_i + \nu_i. \quad (8.4)$$

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

$$\ln \sigma_i^2 = \alpha + \beta \ln x_i + \nu_i. \quad (8.5)$$

де $\alpha = \ln \sigma^2$.

У випадку множинної регресії залежність (8.5) будується для кожної пояснюючої змінної.

4. Перевіряється статистична значимість коефіцієнта β рівняння (8.5) 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 (8.5) може привести до необґрунтованих висновків (наприклад, коефіцієнт β статистично незначимий, а гетероскедастичність має місце). Можлива ще одна проблема. Для випадкового відхилення v_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^2 і значеннями змінної x_i . За даним методом оцінюється регресійна залежність модулів відхилень $|e_i|$ (тісно зв'язаних з σ_i^2) від x_i . При цьому розглянута залежність моделюється наступним рівнянням регресії:

$$|e_i| = \alpha + \beta x_i^k + v_i. \quad (8.6)$$

Змінюючи значення k , можна побудувати різні регресії. Звичайно $k = \dots, -1, -0,5, 0,5, 1, \dots$. 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 v_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 $\sigma_i^2 = \sigma^2 x_i^2$, $i = 1, 2, \dots, n$. 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 убутання.

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:

$$F = \frac{\sum e_2^2 / [\{(n - c) / 2\} - k]}{\sum e_1^2 / [\{(n - c) / 2\} - k]} = \frac{\sum e_2^2}{\sum e_1^2}, \quad (8.7)$$

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

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

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

Природним є питання: якими повинні бути розміри підвбірок 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: $F = \frac{\sum e_1^2}{\sum e_2^2}$.

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 $\varepsilon_i, i = 1, 2, \dots, n$.

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:

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

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

$$\frac{y_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{x_i}{\sigma_i} + \frac{\varepsilon_i}{\sigma_i}. \quad (8.9)$$

Поклавши $\frac{y_i}{\sigma_i} = y_i^*$, $\frac{x_i}{\sigma_i} = x_i^*$, $\frac{\varepsilon_i}{\sigma_i} = v_i$, $\frac{1}{\sigma_i} = z_i$, одержимо рівняння регресії без вільного члена, але з додатковою пояснюючою змінною Z і з «перетвореним» відхиленням v :

$$y_i^* = \beta_0 z_i + \beta_1 x_i^* + v_i \quad (8.10)$$

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. 8.4, а) або значенням x_i^2 (рис. 8.4, б).

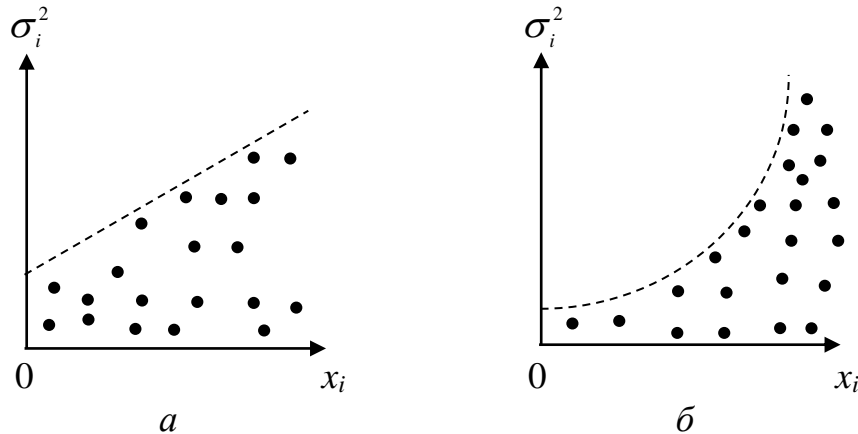


Рис. 8.4.

1. Дисперсії σ_i^2 пропорційні x_i (рис. 8.4, а):

$$\sigma_i^2 = \sigma^2 x_i \quad (\sigma^2 - \text{коефіцієнт пропорційності}).$$

Тоді рівняння (8.8) перетвориться діленням його лівої й правої частин на $\sqrt{x_i}$:

$$\frac{y_i}{\sqrt{x_i}} = \frac{\beta_0}{\sqrt{x_i}} + \beta_1 \frac{x_i}{\sqrt{x_i}} + \frac{\varepsilon_i}{\sqrt{x_i}} \Rightarrow \frac{y_i}{\sqrt{x_i}} = \beta_0 \frac{1}{\sqrt{x_i}} + \beta_1 \sqrt{x_i} + v_i. \quad (8.11)$$

Для випадкових відхилень $v_i = \frac{\varepsilon_i}{\sqrt{x_i}}$ виконується умова гомоскедастичності. Отже, для регресії (8.11) застосуємо звичайний МНК.

Таким чином, оцінивши для (8.11) за МНК коефіцієнти β_0 й β_1 , потім повертаються до вихідного рівняння регресії (8.8).

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 $\hat{Y} = b_0 + b_1 X_1 + \dots + b_m X_m$, тобто фактично лінійна комбінація пояснюючих змінних. У цьому випадку отримують наступну регресію:

$$\frac{y_i}{\sqrt{\hat{y}_i}} = \beta_0 \frac{1}{\sqrt{\hat{y}_i}} + \beta_1 \frac{x_{i1}}{\sqrt{\hat{y}_i}} + \dots + \beta_m \frac{x_{im}}{\sqrt{\hat{y}_i}} + \frac{\varepsilon_i}{\sqrt{\hat{y}_i}}. \quad (8.12)$$

Іноді із всіх пояснюючих змінних вибирається найбільш підходяща виходячи із графічного подання (рис. 8.4).

2. Дисперсії σ_i^2 пропорційні x_i^2 (рис. 8.4, б).

У випадку, якщо залежність σ_i^2 від x_i доцільніше виразити не лінійною функцією, а квадратичною, то відповідним перетворенням буде ділення рівняння регресії (8.8) на x_i :

$$\frac{y_i}{x_i} = \beta_0 \frac{1}{x_i} + \beta_1 + \frac{\varepsilon_i}{x_i} \Rightarrow \frac{y_i}{x_i} = \beta_0 \frac{1}{x_i} + \beta_1 + v_i, \quad (8.13)$$

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 (8.13) повертаються до вихідного рівняння (8.8).

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 перевірки засвоєних 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 9. АВТОКОРЕЛЯЦІЯ В ЕКОНОМЕТРИЧНИХ МОДЕЛЯХ



Основні питання

9.1 Сутність і причини автокореляції

9.2 Наслідки автокореляції

9.3 Виявлення автокореляції

9.4 Методи усунення автокореляції

9.5 Моделі розподіленого лагу



Ключові слова та поняття: послідовна кореляція, автокореляція, негативна автокореляція, позитивна автокореляція, помилки специфікації, інерція, ефект павутини, критерій Дарбіна–Уотсона

9.1 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 ($\sigma(\varepsilon_{i-1}, \varepsilon_i) = 0$), $i = 2, 3, \dots, n$.

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) у край 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* ($\sigma(\varepsilon_{t-1}, \varepsilon_t) < 0$).

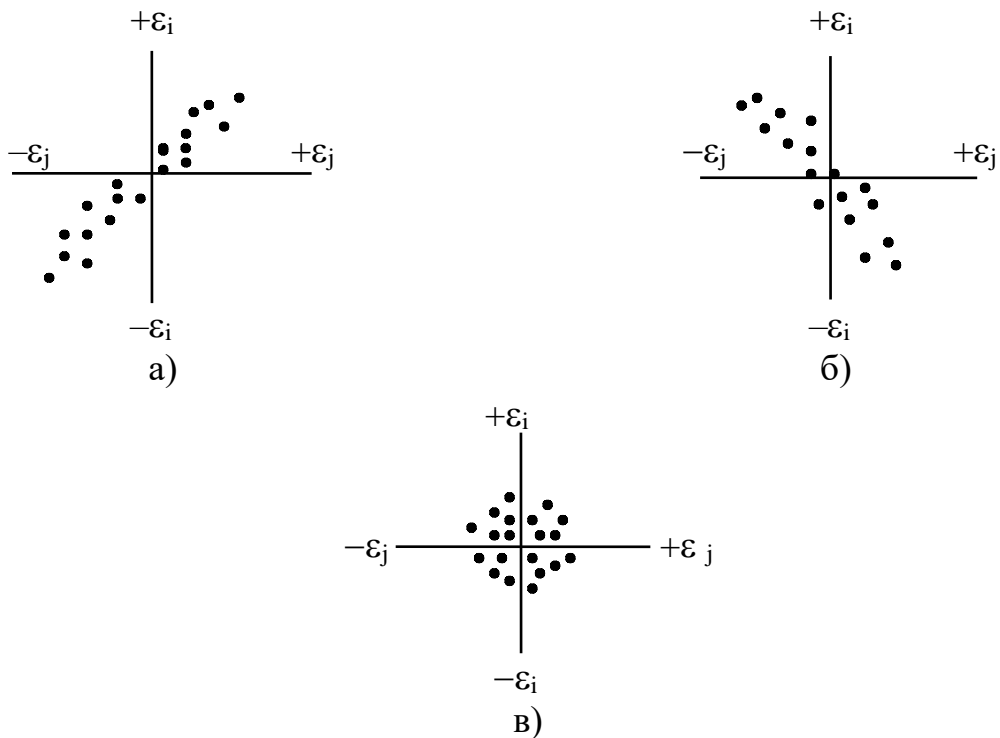


Рис. 9.1.

На рис. 9.1 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 настає негативне й навпаки. Можлива схема розсіювання крапок у цьому випадку представлена на рис. 9.1 б) негативне значення ε_i супроводжується позитивним $-\varepsilon_j$ і навпаки. На рис. 9.1 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, ВВП і т. ін.) мають певну циклічність, пов'язаної з хвилеподібністю ділової активності. Дійсно, економічний підйом приводить до росту зайнятості, скороченню інфляції, збільшенню ВВП 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.

9.2 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 $\sigma^2 = \sum_{i=1}^T \frac{e_i^2}{T - m - 1}$ 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.

9.3 Виявлення автокореляції

В силу невідомості значень параметрів рівняння регресії невідомими будуть також і дійсні значення відхилень ε_t , $t = 1, 2, \dots, T$. Тому висновки о їх незалежності здійснюються на основі оцінок e_t , $t = 1, 2, \dots, T$, 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. 9.2. 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).

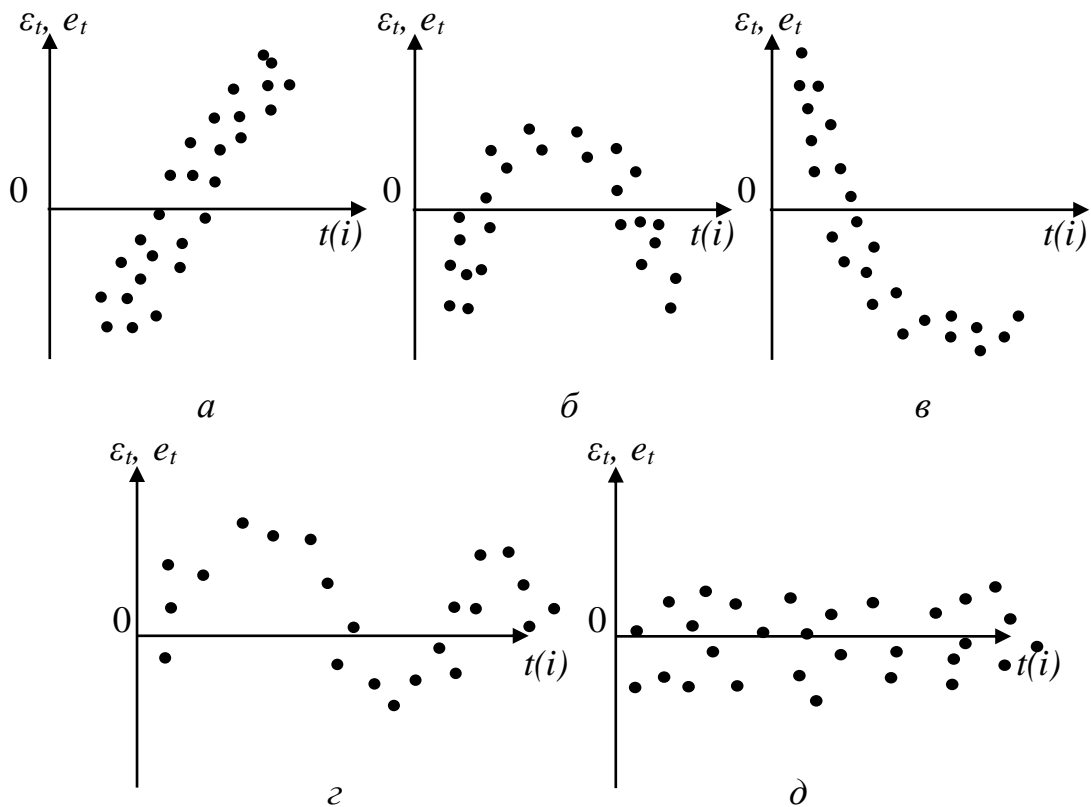


Рис. 9.2

Природно припустити, що на рис. 9.2, *а-г* є певні зв'язки між відхиленнями, тобто автокореляція має місце. Відсутність залежності на рис. 9.2, *д* швидше за все свідчить про відсутність автокореляції.

На рис. 9.2, *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 наочно, якщо графік 9.2, б доповнити графіком залежності e_t від e_{t-1} (рис. 9.3).

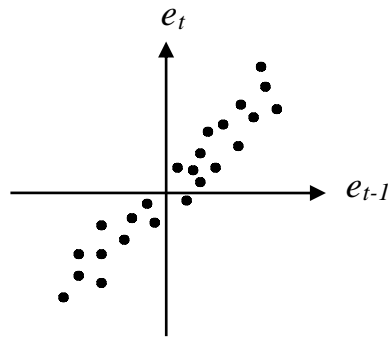


Рис. 9.3

Переважає 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 , $t = 1, 2, \dots, T$. 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 не випадковий 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.

$$DW = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}. \quad (9.1)$$

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

$$DW \approx \frac{2(\sum e_t^2 - \sum e_t e_{t-1})}{\sum e_t^2} = 2(1 - r_{e_t e_{t-1}}). \quad (9.2)$$

Якщо $e_t = e_{t-1}$, то $r_{e_t e_{t-1}} = 1$ й $DW = 0$. Якщо $e_t = -e_{t-1}$, то $r_{e_t e_{t-1}} = -1$ й $DW = 4$.

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

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

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

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:

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

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

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

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

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

Всі випадки зображені на рис (9.4).

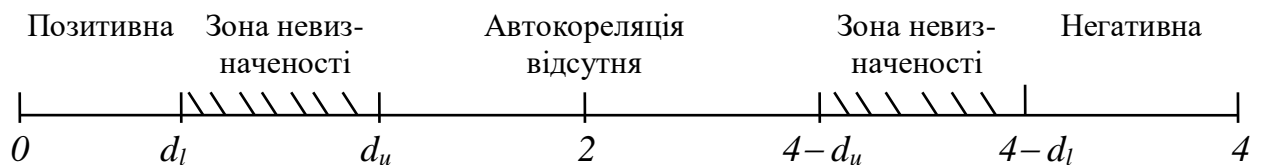


Рис. 9.4.

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: $\varepsilon_t = \rho \varepsilon_{t-1} + v_t$, називаної авторегресійною схемою першого порядку AR(1). Тут v_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:

$$y = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \gamma_{t-1} + \varepsilon_t. \quad (9.3)$$

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

$$y_t = \beta_0 + \beta_1 x_t + \gamma_{t-1} + \varepsilon_t. \quad (9.4)$$

Нехай випадкове відхилення ε_t піддано впливу авторегресії першого порядку:

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t. \quad (9.5)$$

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

$$y_t = \beta_0 + \beta_1 x_t + \gamma_{t-1} + \rho \varepsilon_{t-1} + v_t. \quad (9.6)$$

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 4⁰) — пояснюючі змінні не повинні бути випадковими (не мати випадкової складової). Значення будь-якої пояснюючої змінної повинне бути екзогенним, повністю визначеним. У протилежному 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

$$h = \hat{\rho} \sqrt{\frac{n}{1 - n \text{var}(g)}}, \quad (9.7)$$

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

$\text{var}(g)$ – 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 протилежному випадку вона не відхиляється.

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

У такий спосіб можна записати

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

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 $nvar(g) > 1$.

9.4 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

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i. \quad (9.8)$$

Тоді спостереженням t й $(t-1)$ відповідають формули:

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t. \quad (9.9)$$

$$y_{t-1} = \beta_0 + \beta_1 x_{t-1} + \varepsilon_{t-1}. \quad (9.10)$$

Нехай випадкові відхилення піддаються впливу авторегресії першого порядку (9.5):

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

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

Віднімемо з (9.9) співвідношення (9.10), помножене на ρ :

$$y_t - \rho y_{t-1} = \beta_0 (1 - \rho) + \beta_1 (x_t - \rho x_{t-1}) + (\varepsilon_t - \rho \varepsilon_{t-1}). \quad (9.11)$$

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:

$$y_t^* = \beta_0^* + \beta_1 x_t^* + v_t. \quad (9.12)$$

Тому що по припущенню коефіцієнт ρ відомий, то очевидно, y_t^* , x_t^* , v_t обчислюються досить просто. У силу того що випадкові відхилення v_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*:

$$x_1^* = \sqrt{1 - \rho^2} \cdot x_1, \quad (9.13)$$

$$y_1^* = \sqrt{1 - \rho^2} \cdot y_1.$$

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.:

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + v_t,$$

$$\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \rho_3 \varepsilon_{t-3} + v_t. \quad (9.14)$$

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 (9.2):

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

Тоді як оцінку коефіцієнта ρ може бути взятий коефіцієнт $r = r_{e_t e_{t-1}}$. З (9.2) маємо:

$$r \approx 1 - \frac{DW}{2}. \quad (9.15)$$

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 (9.8):

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

і авторегресійної схеми (9.5) першого порядку

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

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

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

$$e_t = \hat{\rho} e_{t-1} + v_t, \quad (9.16)$$

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

в) На основі даної оцінки будується рівняння:

$$(y_t - \hat{\rho} y_{t-1}) = \alpha(1 - \hat{\rho}) + \beta(x_t - \hat{\rho} x_{t-1}) + (\varepsilon_t - \hat{\rho} \varepsilon_{t-1}). \quad (9.17)$$

за допомогою якого оцінюються коефіцієнти α і ρ (у цьому випадку значення ρ відомо).

г) Значення $\beta_0 = \alpha (1 - \rho)$ і $\beta_1 = \beta$ підставляються в (9.8). 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 (9.11) 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 (9.11) саме з даним значенням $\hat{\rho}$.

Цей ітераційний метод широко використовується в пакетах прикладних програм.

4. Метод перших різностей.

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

Для часових рядів характерна позитивна автокореляція залишків. Тому при високій автокореляції думають $\rho = 1$, і, отже, рівняння (9.11) приймає вид

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

або (9.18)

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

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

$$\Delta y_t = \beta_1 \Delta x_t + v_t. \quad (9.19)$$

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 (9.9) і (9.10) з урахуванням (9.5), можна одержати наступне рівняння регресії:

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

або (9.20)

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

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

9.5 Моделі розподіленого лагу

При аналізі багатьох економічних показників (особливо в макроекономіці) часто використовують щорічні, щоквартальні, щомісячні, щоденні дані. Наприклад, це можуть бути річні дані по ВВП, 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 $y_{t+1}, y_{t+2}, \dots, y_{t+k}, \dots$; значення Y у попередні моменти позначаються $y_{t-1}, y_{t-2}, \dots, y_{t-k}, \dots$.

Неважко зрозуміти, що при вивченні залежностей між такими показниками або при аналізі їхнього розвитку в часі в якості пояснюючих змінних використовуються 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

$$y = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_k x_{t-k} + \varepsilon_t \quad (9.21)$$

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

$$y_t = \alpha + \beta x_t + \gamma y_{t-1} + \varepsilon_t \quad (9.22)$$

В економетричному аналізі динамічні моделі використовуються 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 поведженні 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:

$$\beta_k = \beta_0 \lambda^k, \quad k = 0, 1, \dots \quad (9.23)$$

де $0 < \lambda < 1$ характеризує швидкість убуття 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 9.21 перетворюється у

$$y = \alpha + \beta_0 x_t + \beta_0 \lambda x_{t-1} + \beta_0 \lambda^2 x_{t-2} + \dots + \varepsilon_t \quad (9.24)$$

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

$$\lambda y_{t-1} = \lambda \alpha + \beta_0 \lambda x_t + \beta_0 \lambda x_{t-1} + \beta_0 \lambda^2 x_{t-2} + \dots + \lambda \varepsilon_{t-1}, \quad (9.25)$$

одержимо наступне рівняння

$$\begin{aligned} y_t - \lambda y_{t-1} &= (1 - \lambda) \alpha + \beta_0 x_t + (\varepsilon_t - \lambda \varepsilon_{t-1}) \Rightarrow \\ y_t &= \alpha(1 - \lambda) + \beta_0 x_t + \lambda y_{t-1} + v_t, \end{aligned} \quad (9.26)$$

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

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

Порівнюючи первісну модель

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

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 v_t .

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

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

Модель Койка має дві модифікації:

а) модель адаптивних очікувань

$$y_t = \gamma \alpha + \gamma \beta x_t + (1 - \gamma) y_{t-1} + v_t, \quad (9.27)$$

де γ – коефіцієнт очікування;

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

б) модель часткового коригування

$$y_t = \lambda\alpha + \lambda\beta x_t + (1-\lambda)y_{t-1} + \lambda\varepsilon_t, \quad (9.28)$$

λ – коефіцієнт коригування.

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

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

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 v_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 v_t . Такий замінник називається допоміжною змінною. 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 перевірки засвоєних 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. Перелічіть основні методи виявлення автокореляції.
7. Опишіть схему використання статистики DW Дарбіна–Уотсона.
8. Перелічіть обмеження використання статистики DW Дарбіна–Уотсона.
9. Яка статистика використовується 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 використовується 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 10. ПОБУДОВА ЕКОНОМЕТРИЧНОЇ МОДЕЛІ НА ОСНОВІ ОДНОЧАСНИХ СТРУКТУРНИХ РІВНЯНЬ



10.1 Поняття економетричних систем рівнянь. Структурна та зведена форма моделі

10.2 Проблема ідентифікації. Оцінювання параметрів структурної моделі



Ключові слова та поняття: системами одночасних структурних рівнянь, ендогенні змінні, екзогенні змінні, структурна форма моделі, структурні коефіцієнти моделі, зведена форма моделі, проблема ідентифікації, надідентифікацію, двокроковий метод найменших квадратів (2 МНК)

10.1 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:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + e_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + e_2, \end{cases} \quad (10.1)$$

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 вибрати 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 зміщені 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:

$$\begin{cases} \hat{y}_1 = c_{11} \cdot x_1 + c_{12} \cdot x_2 + \dots + c_{1m} \cdot x_m, \\ \hat{y}_2 = c_{21} \cdot x_1 + c_{22} \cdot x_2 + \dots + c_{2m} \cdot x_m, \\ \dots \\ \hat{y}_n = c_{n1} \cdot x_1 + c_{n2} \cdot x_2 + \dots + c_{nm} \cdot x_m, \end{cases} \quad (10.2)$$

де c_i – коефіцієнти зведеної форми моделі.

10.2 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.

Моделі 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

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2, \\ y_2 = b_{21}y_1 + a_{22}x_2 + a_{23}x_3 \end{cases} \quad (10.3)$$

з двома ендогенними та трьома екзогенними змінними є точно ідентифікована.

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

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1, \end{cases}$$

$$y_2 = b_{21}y_1 + a_{22}x_2, \quad (10.4)$$

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

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1, \\ y_2 = b_{21}y_1 + a_{22}x_2, \end{cases} \quad (10.5)$$

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

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2 + a_{13}x_3, \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + a_{23}x_3, \end{cases} \quad (10.6)$$

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

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2, \\ y_2 = b_{21}y_1 + a_{23}x_3. \end{cases} \quad (10.7)$$

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

$$AY_t + BX_t = u_t, \quad (10.8)$$

де A – невироджена матриця невідомих параметрів при ендогенних змінних розмірності $(k \times k)$;

B – матриця ендогенних змінних розмірності $(k \times l)$;

Y_t – вектор екзогенних змінних розмірності $(m \times l)$;

X_t – вектор випадкових залишків розмірності $(k \times l)$.

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

$$Y_{1t} = r_{11}X_{1t} + r_{12}X_{2t} + \dots + r_{1k}X_{kt} + u_{1t};$$

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:

$$\begin{bmatrix} \hat{b} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} Y_1'X(X'X)^{-1}X'Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X \end{bmatrix}^{-1} \begin{bmatrix} Y_1'X(X'X)^{-1}X'Y \\ X_1'Y \end{bmatrix}. \quad (10.12)$$

2. Знаходження добутку матриць поточних ендогенних змінних, які містяться у правій частині моделі, то матриці всіх екзогенних змінних моделі, тобто $Y_1'X$.

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 добутку $Y_1'X (X'X)^{-1} X'Y_1$ матриць, які здобуто на кроках 2, 3, 4.

6. Визначення добутку $Y_1'X$ матриць ендогенних змінних у правій частині моделі і екзогенних змінних, які внесені до даного рівняння.

7. Знаходження добутку $X'Y_1$ матриць екзогенних змінних, які входять у дане рівняння, і ендогенних змінних правої частини системи рівнянь.

8. Визначення добутку $Y_1'X_1$ матриць екзогенних змінних даного рівняння.

9. Знаходження матриці, оберненої до блочної:

$$Q_S^{-1} = \begin{bmatrix} Y_1'X(X'X)^{-1}X'Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X \end{bmatrix}^{-1} \quad (10.13)$$

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:

$$q_s = Y_1'X_1X_1(X'X)^{-1}X'Y_1. \quad (10.14)$$

12. Визначення параметрів моделі:

$$\begin{bmatrix} \hat{b}_s \\ \hat{a}_s \end{bmatrix} = Q_s^{-1} q_s. \quad (10.15)$$

13. Обчислення s -ї залежної ендогенної змінної на основі знайдених параметрів a_s і b_s :

$$\hat{Y}_s = a_s Y_1 + b_s X_1. \quad (10.16)$$

14. Обчислення вектора залишків в s -му рівнянні системи:

$$u_s = Y_s - \hat{Y}_s. \quad (10.17)$$

15. Визначення дисперсії залишків для кожного рівняння:

$$\sigma_{u_s}^2 = \frac{1}{n - k - m + 1} u_s' u_s. \quad (10.18)$$

16. Знаходження матриці коваріацій для параметрів кожного рівняння:

$$\text{asy var} \begin{bmatrix} \hat{b}_s \\ \hat{a}_s \end{bmatrix} = \sigma_{u_s}^2 Q_s^{-1}. \quad (10.19)$$

17. Знаходження стандартної похибки параметрів і визначення довірчих інтервалів:

$$S \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \sqrt{\sigma_{u_s}^2 Q_s^{-1}}; \quad (10.20)$$

$$\hat{a} - t_{(\alpha)} S \leq a \leq \hat{a} + t_{(\alpha)} S; \quad (10.21)$$

$$\hat{b} - t_{(\alpha)} S \leq b \leq \hat{b} + t_{(\alpha)} S. \quad (10.22)$$



Запитання для перевірки засвоєних 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 понадіентифіковане?
9. Яка система рівнянь називається точно ідентифікованою?
10. Яка система рівнянь називається надіентифікованою?
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 надіентифіковані рівняння?
17. Чи можна виконувати оцінку параметрів моделі окремо для групи точно ідентифікованих і над ідентифікованих рівнянь?



Перелік рекомендованої літератури

1. Григорків, В.С. Моделювання економіки: підручник. Чернівці : ЧНУ ім. Ю. Федьковича, 2019. 360 с.
2. Диха М. В., Мороз В. С. Економетрія: Навчальний посібник. К.: Центр навчальної літератури (ЦУЛ), 2019. 206 с.
3. Економіко-математичне моделювання: навчальний посібник / за ред. О. Т. Івашука. – Тернопіль : ТНЕУ Економічна думка, 2008. – 704 с.
4. Єлейко В.І., Копич І.М., Боднар Р.Д., Демчишин М.Я. Економетрія: Навч. посібник / Укоопспілка; Львівська комерційна академія. – Л. : Видавництво Львівської комерційної академії, 2007. – 352с.
5. Клебанова Т.С., Дубровина Н.А., Раєвнева Е.В. Эконометрия: Учеб. пособие [для студентов вузов] / Т.С. Клебанова, Н.А. Дубровина, Е.В. Раєвнева; Харьк. нац. экон. ун-т. – Х.: ИНЖЭК, 2005. – 156 с.
6. Княженко І.І., Павленко Л.В. Економетрія: Навч.-метод. посібник для самостійного вивчення дисципліни / Бердянський ін-т підприємництва. – Донецьк : ТОВ "Юго-Восток, Лтд", 2006. – 187с.
7. Козьменко О. В., Кузьменко О. В. Економіко-математичні методи та моделі (економетрика): Навч. посібник. Суми: Університетська книга, 2018. 406 с.
8. Корольов О. А., Рязанцева В.В. Економетрія: Практикум: Навч. посіб. / О.А. Корольов, В.В. Рязанцева; Київ. нац. торг.-екон. ун-т. – К. :КНТЕУ, 2005 – 277 с.
9. Лугінін О.Є., Білоусова С.В., Білоусов О.М. Економетрія: Навч. посібник. – К.: Центр навч. літератури, 2005. – 252 с.
10. Лукьяненко І. Г. Економетрика : підручник / І. Г. Лукьяненко, Л. І. Краснікова – К. : Товариство «Знання», КОО, 1998. – 494 с.
11. Назаренко О.М. Основи економетрики: Підручник. – Київ: “Центр навчальної літератури”, 2004. – 392 с.
12. Пехота М.А., Грищенко О.Ю. Основи економетрії / М.А. Пехота, О.Ю. Грищенко. – К.: Ін-т аграр. економіки, 2007. – 178 с. – Бібліогр.: с. 170.
13. Тімченко О.Д, Филипенко О.М. Економетрія: конспект лекцій / Харк. держ. університет харчування та торгівлі. – Харків, 2006. – 114 с.
14. Тімченко О.Д, Филипенко О.М. Економетрія: практикум / Харк. держ. університет харчування та торгівлі. – Харків, 2005. – 76 с.
15. Тімченко, О.Д. Економетрика : методичні вказівки та завдання до виконання самостійної та індивідуальної роботи / О.Д. Тімченко, С.О. Зубков, Є.Ю. Стоян. – Харків : ХДУХТ, 2019. – 98 с.
16. Тімченко, О.Д. Економетрика : методичні вказівки та завдання до виконання лабораторних робіт / О.Д. Тімченко, О.Д. Рачкован, В.О. Олім. – Харків : ХДУХТ, 2019. – 70 с.

Інформаційні ресурси

17. Економетрика [Електронний ресурс] – Режим доступу : <<http://znaimo.com.ua/>>.
18. Економетрика – бібліотека ресурсів [Електронний ресурс]. –Режим доступу : <http://efaculty.kiev.ua/ekon.htm>.
19. Електронна бібліотека підручників [Електронний ресурс] – Режим доступу : <<http://studentam.kiev.ua/>>.
20. Офіційний сайт Державного комітету статистики України [Електронний ресурс] – Режим доступу : <<http://www.ukrstat.gov.ua/>>.
21. Офіційний сайт Міністерства економічного розвитку і торгівлі України [Електронний ресурс] – Режим доступу : <<http://www.me.gov.ua/>>.
22. Сайт Национального банка Украины. – Режим доступа : www.bank.gov.ua.
23. Сайт ПФТС. – Режим доступа : <http://pfts.com/>.
24. Сайт Агентства по развитию инфраструктуры фондового рынка Украины.–Режим доступа : <http://www.smida.gov.ua/db>

НАВЧАЛЬНЕ ВИДАННЯ

Укладачі:
ФИЛИПЕНКО Олена Михайлівна
ТІМЧЕНКО Ольга Дмитрівна

ЕКОНОМЕТРИКА

Навчальний посібник

**для здобувачів вищої освіти першого рівня (бакалавр)
спеціальностей 051 Економіка, 071 Облік і оподаткування,
072 Фінанси, банківська справа і страхування**

Підписано до друку 2021 р. Папір офсет.
Друк офсет. Умов. друк. арк. 6.3 Тираж 100 прим.

Видавництво та друк
ФОП Іванченко І. С.

пр. Тракторобудівників, 89-а/62, м. Харків, 61135.
Тел.: +38-057-756-09-25, +38-050-40-243-50.

Свідоцтво про внесення суб'єкта видавничої справи
до державного реєстру видавців, виготівників та розповсюджувачів
видавничої продукції серія ДК №4388 від 15.08.2012 р.

www.monograf.com.ua