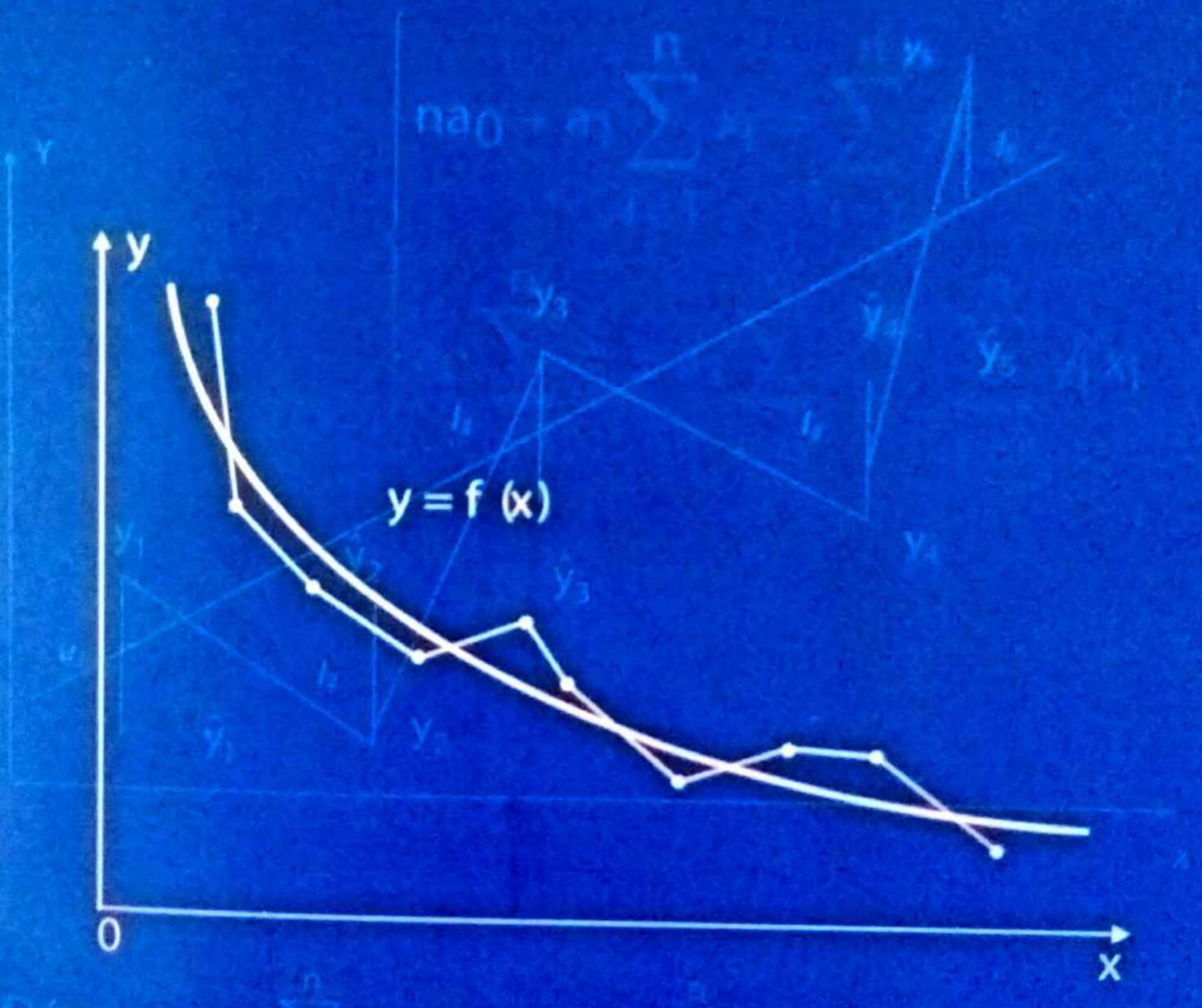
М. В. Диха, В. С. Мороз

ЕКОНОМЕТРІЯ



МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ НІВЕРСИТЕТ

М. В. Диха, В. С. Мороз

ЕКОНОМЕТРІЯ

НАВЧАЛЬНИЙ ПОСІБНИК

Київ
«Центр учбової
літератури»
2022

УДК 330.43(075.8)
ББК 65.01я73
Д 49

Рекомендований до друку Вченою радою
Хмельницького національного університету як навчальний посібник
для студентів вищих навчальних закладів
(Протокол № 7 від 29 червня 2016 року)

Рецензенти:

О. Б. Альохін – професор кафедри економічної кібернетики та інформаційних технологій Одеського національного політехнічного університету, доктор економічних наук, професор;

І. С. Благун – завідувач кафедри економічної кібернетики Прикарпатського національного університету ім. В. Стефаника, доктор економічних наук, професор, Заслужений діяч науки і техніки України;

П. М. Григорук – професор кафедри автоматизованих систем та моделювання в економіці Хмельницького національного університету, доктор економічних наук, професор.

Диха М. В. Економетрія: навчальний посібник [текст] / М. В. Диха, В. С. Мороз
Д 49 – К. : «Центр учбової літератури», 2022. – 206 с.

ISBN 978-617-673-486-4

Навчальний посібник «Економетрія» надає можливість оволодіти основами побудови та аналізу економетричних моделей економічних явищ та процесів. Економетричні моделі взаємозв'язків економічних процесів, досліджуваних показників дозволяють проводити економічні експерименти для виявлення можливостей зростання економіки підприємства, для пошуку менеджерами найкращих управлінських рішень, дозволяють перевіряти економічні гіпотези на предмет їх практичної реалізації.

Навчальний посібник розрахований для студентів вищих навчальних закладів, аспірантів, викладачів, економістів та аналітиків.

УДК 330.43(075.8)
ББК 65.01я73

ISBN 978-617-673-486-4

ЗМІСТ

ВСТУП

«Економетрія» або «Економетрика» відноситься до прикладних економічних дисциплін з широким використанням методів прикладної математики, теорії ймовірностей, математичної статистики та економічної теорії.

«Економетрія» як наука сформувалася у наукову дисципліну в середині XX ст. Вагомий внесок у становлення дисципліни зробили англійські математики та статистики Ф. Гальтоні, К. Пірсон, а також відомі економісти Г. Мур і Т. Шульц. У перших працях розроблялись аналітико-статистичні моделі у вигляді лінійної регресії попиту та пропозиції, їх залежність від обсягів випуску продукції, рівня цін, прибутків. В 1928 р. Х. Коббом та П. Дугласом була побудована перша виробнича функція для опису залежності обсягу виробництва від кількості спожитої праці та виробничого капіталу. Починаючи з 30-х рр. XX ст. відомі економісти Я. Танберген, Л. Клейн, Р. Стоун та ін. розробили моделі економіки, які описували статистичні зв'язки виробництва, кінцевого індивідуального і державного попиту, цін, податків, пропозиції робочої сили, накопичення та зношування капіталу.

У другій половині XX ст. були видані праці Г. Тінтнера, Е. Маленво, Т. Андерсона, Дж. Джонстона, М. Дж. Кендалла, А. Стьюарта, Е. Кейна, які мали важливу роль у становленні науки «Економетрія».

Наука «Економетрія» є триєдиною. Вона базується у своїх дослідженнях на використанні економічної теорії, математичних методів та статистики. Тому економетричні моделі часто ототожнюють з економіко-статистичними.

«Економетрія» відноситься до базових курсів у підготовці фахівців економічних спеціальностей у вищих навчальних закладах. Оволодіння економетричними методами моделювання економічних явищ та процесів вимагає від студентів певної економічної, математичної та комп'ютерної підготовки.

Головним завданням економетрії є розробка економетричних моделей, оцінювання їх невідомих параметрів, перевірка моделей та всебічний аналіз отриманих результатів.

Для потреб ринкової економіки економетрія розробляє і вивчає виробничі функції, функції попиту та пропозиції, моделі рівноваги попиту та пропозиції, статистичні та динамічні моделі міжгалузевого балансу, моделі загальної економічної рівноваги.

Економетричні моделі взаємозв'язків окремих економічних показників та процесів дозволяють перевіряти, уточнювати економічні гіпотези та робити висновки, виявляти можливі резерви зростання економіки, дозволяють проводити економічні експерименти для пошуку менеджерами оптимальних управлінських рішень.

Економетричні моделі знаходять широке застосування при проведенні економічного та фінансового аналізу, в соціально-економічному прогнозуванні, прийнятті управлінських рішень.

Економетричні моделі знаходять застосування не тільки в економічних дослідженнях. Вони також використовуються у біології, історії, соціології, медицині та інших суспільних і природничих науках, коли виникає необхідність виявляти та досліджувати зв'язки між великою кількістю змінних.

Тема 1

ПРЕДМЕТ, МЕТОД І ЗАВДАННЯ ДИСЦИПЛІНИ «ЕКОНОМЕТРІЯ»

1.1 Математичне моделювання як метод наукового пізнання економічних явищ і процесів

Весь навколишній світ, що оточує нас, являє собою деяку систему. Ідеї системності у природі та системної методології виникли ще в античні часи. Носіями цих ідей були античні філософи Платон та Аристотель. Загальна теорія систем як наука сформувалась у 30-40-х рр. 20 ст. Цей науковий напрямок розробив та запровадив австрійський вчений-біолог К. Берталанфі. Згідно його концепції загальна теорія систем є універсальною наукою про системи будь-якого типу.

Поняття системи трактується з філософських та наукових позицій. Слово система походить від старогрецького слова «systema», що означає ціле, складене із частин. Таким чином, **система** – це сукупність взаємно пов'язаних елементів, що становлять ціле або єдність.

Кількість елементів, що входять у склад системи не визначається, бо це може привести до деякого парадоксу. Наприклад, старогрецькі філософи дискутували: скільки каменів можуть утворити купу?

Приклади систем – людина, інші живі організми, підприємство, окремі технічні комплекси. Усе в природі утворює систему.

Елемент системи – це частина системи, яка має певне функціональне призначення. Елементи системи можуть бути складними або простими. Складні в свою чергу складаються з підсистем, простіших взаємопов'язаних елементів. Всяка система має організацію та структуру.

Організація системи – це внутрішня впорядкованість та узгодженість взаємодії елементів системи. Приклад: взаємодія окремих виробничих підрозділів підприємства впорядкована та узгоджена між собою в рамках підприємства. Організація системи – це одна з основних властивостей системи.

Структура системи – це сукупність внутрішніх сталих зв'язків між елементами системи. Ця сукупність визначає основні властивості системи. Структура системи може бути ієрархічною. Ієрархічною вважають структуру, в якій мають місце зв'язки між елементами вищого та нижчого рівнів.

Усі системи поділяються на **абстрактні та матеріальні** системи.

Абстрактні системи – це продукт людського мислення. До цих систем відносять гіпотези, знання, теореми і т.д.

Матеріальні системи – це сукупність матеріальних об'єктів. До матеріальних відносять соціальні системи, тобто суспільство або інші соціально-економічні системи. Відносини між елементами у соціальних системах є суспільними відносинами між людьми в процесі суспільної діяльності.

В залежності від сталості систем у часі розрізняють **статичні та динамічні системи**. Статичні з плином часу не змінюються. Динамічні з плином часу змінюють свій стан.

Динамічні системи у свою чергу поділяються на **детерміновані та стохастичні (імовірнісні)**. Динамічні системи, у яких стан її елементів у даний момент часу повністю визначає її стан у будь-який попередній або наступний моменти часу називається детермінованою системою. В детермінованій системі складові частини взаємодіють точно за передбаченим алгоритмом. При її дослідженні не виникає ніякої невизначеності. Якщо відомий попередній стан і програма (алгоритм) переходу в інший стан, то завжди можна безпомилково передбачити наступний стан. Якщо передбачити стан системи неможливо, то така система відноситься до класу стохастичних систем (економічні, соціальні системи). Стохастичність системи можна розглядати як ознаку її складності, яка може проявлятись як об'єктивна властивість системи (або через складність системи людина не в змозі її вивчити достатньо повно).

За характером взаємодії з навколишнім середовищем системи поділяються на **закриті та відкриті**. Закриті системи ізольовані від зовнішнього середовища. Усі процеси, крім енергетичних, відбуваються лише в середині самої системи. Відкриті системи активно взаємодіють з навколишнім середовищем (підприємство).

За складністю всі системи поділяються на **прості, складні та дуже складні (великі)**.

Системи, які мають зазначені вище властивості називаються **кібернетичними**, тобто керованими. Для вивчення кібернетичних систем розробляють та досліджують відповідні поняття і методи. До них належить загальна теорія систем, яка є універсальним інструментом дослідження; теорія економічних систем для вивчення економічних явищ і процесів, об'єктів економіки.

Економічні системи – це частковий випадок складних динамічних систем. Ці системи мають усі властивості кібернетичних систем, але на відміну від абстрактних кібернетичних економічні системи завжди конкретні. **Економічна система** – це деяка сукупність (комплекс) відомостей про економічний об'єкт або процес, необхідних для вирішення певного завдання управління цим об'єктом (процесом). Під економічною системою розуміють сукупність (перелік) взаємозв'язаних змінних (показників), які характеризують стан об'єкта або процесу з точки зору вирішення певної економічної задачі. Економічні системи складаються з підсистем та елементів.

Універсальним методом вивчення та дослідження економічних систем є **побудова їх математичних моделей**.

Термін «модель» (лат. «modulus» – зразок). Поняття моделі дуже розпливчасте. **Моделлю** системи, об'єкта чи явища називають штучну систему або об'єкт, що в певних умовах може замінити систему-оригінал або об'єкт, шляхом відтворення найбільш характерних параметрів та властивостей об'єкта, які нас цікавлять. Заміна системи або реального об'єкта його моделлю дозволяє отримати додаткову інформацію про цей об'єкт та створити зручності досліднику при його дослідженні. В залежності від мети дослідження можуть бути побудовані різні моделі одного і того ж об'єкта. Характерною рисою моделей є їх спрощеність по відношенню до оригіналу. Таким чином, модель – це спрощене умовне відображення оригіналу. Моделлю також є алгоритм поведінки системи.

Усі моделі класифікуються на **матеріальні та ідеальні (абстрактні)**. До матеріальних відносяться моделі взуття, одягу, літака, штучного супутника Землі і т.п..

У наукових дослідженнях найбільший інтерес представляють абстрактні моделі. До абстрактних моделей належать **математичні**

моделі, які є системою математичних залежностей між величинами, що характеризують модельований об'єкт. Абстрактні моделі – це ідеальні (уявні) моделі, які є продуктом людського розуму, мислення; їх ще називають логіко-математичними. Ці моделі використовують мову математики та логіки і є певною системою математичних співвідношень та логічних виразів (рівнянь, нерівностей, функцій, алгоритмів тощо). Прикладом абстрактної моделі є система знань про навколишній світ. Серед абстрактних моделей найбільший інтерес для дослідників економічних систем викликають економіко-математичні моделі (математичні моделі економічних процесів і явищ). Розрізняють **теоретико-аналітичні та прикладні економіко-математичні** моделі. Перші використовуються для дослідження загальних властивостей і закономірностей економічних явищ і процесів; другі – для розв'язку конкретних задач економічного аналізу, прогнозування, управління економічними процесами та економічними об'єктами.

Економіко-математичні моделі поділяються на **структурні і функціональні**. У структурних моделях переважає відображення внутрішньої структури об'єкта або системи внутрішньої організації і взаємозв'язки елементів. Функціональні моделі відображають залежності між вхідними і вихідними параметрами певних об'єктів чи систем. За допомогою функціональних моделей пізнають сутність об'єкта через його діяльність, поведінку, функціонування.

Моделі, призначені для пошуку найкращих станів об'єкта щодо обраного критерію, називаються **нормативними (оптимізаційними)**, а моделі, призначені для пояснення факторів, станів, прогнозів, поведінки об'єкта – **дескриптивними (описовими)**.

Залежно від визначеності використовуваної інформації розрізняють **детерміновані та ймовірнісні** моделі. В детермінованих моделях результати на виході однозначно визначаються управляючими впливами. Ймовірнісні моделі ґрунтуються на теорії ймовірності. Невизначеність, властива таким моделям, описується відповідними законами розподілу випадкових величин.

Якщо залежність в моделі належить до одного й того ж моменту або періоду часу, то така модель називається **статичною**. Модель, яка відображає зміни економічних процесів в часі, є **динамічною** моделлю.

За формою математичних залежностей економіко-математичні моделі поділяються на **лінійні та нелінійні**, а за рівнем агрегування об'єктів моделювання – на **макроекономічні** (описують функціонування економіки країни) та **мікроекономічні** (описують функціонування суб'єктів господарювання економіки (підприємства)) моделі.

Процес побудови та дослідження моделі називається **моделюванням**. Моделювання як пізнавальний прийом, як форма відображення дійсності, зародилось ще в античну епоху одночасно з виникненням наукового пізнання.

Необхідність моделювання економічних явищ і процесів та перенесення результатів дослідження моделей на реальний об'єкт замість того, щоб досліджувати цей об'єкт безпосередньо обумовлена тим, що моделі, по-перше, є простішими для дослідника, ніж реальні об'єкти, що досліджуються. По-друге, деякі об'єкти взагалі неможливо досліджувати активно, тобто проводити з ними експеримент, який має виключно пізнавальне значення. Крім того, моделювання має інше, більш важливе для науки значення. Оскільки в моделі розглядаються лише деякі найбільш суттєві в даному дослідженні сторони досліджуваного об'єкту, то моделювання дозволяє виявити суттєві фактори, які формують ті або інші властивості досліджуваних об'єктів.

Сьогодні важко назвати галузь науки, де моделювання б не використовувалось. Економіко-математичне моделювання є науковим методом пізнання економічних явищ та процесів.

Метод (від грецького «methodos» – шлях дослідження) – це спосіб досягнення мети, вирішення конкретного завдання, сукупність прийомів чи операцій практичного або теоретичного пізнання дійсності. Складові наукового методу одержання результатів такі:

1) збір і накопичення емпіричних даних (безперечних фактів), що отримують шляхом спостереження та експерименту;

2) формування гіпотез на підставі зібраних даних шляхом пошуку моделей взаємовідносин між даними і наступне індуктивне узагальнення;

3) перевірка гіпотез шляхом виведення передбачень, які з них випливають, і подальше планування та проведення експериментів для перевірки правдивості гіпотез;

4) відкидання гіпотез, які не підтверджені експериментами, побудова теорії на основі підтверджених гіпотез.

Звернемося до деяких логічних понять, покладених в основу наукової аргументації і доказу.

Індукція (лат. «inductio» – наведення, спонукання) – метод пізнання, що грунтується на отриманні формально-логічного висновку, який робиться на підставі знання про окремі факти. Індукція – спосіб пізнання, що грунтується на отримані загальних висновків і положень на основі вивчення окремих фактів, явищ і процесів, отримання більш загального знання на основі менш загального, на основі експериментів і спостережень фактів.

В індукції думка йде від знання одиничного чи менш загального до загального або більш загального. Об'єктивною основою індукції є діалектичний зв'язок між одиничним і загальним в явищах природи і суспільства, який дає можливість через знання окремого досягти знання загального.

Як один з логічних методів пізнання – індукція розроблялась Ф. Беконом, пізніше Дж. Міллем. Індукція широко застосовується в усіх процесах наукового пізнання поряд з дедукцією.

Дедукція (лат. «deduction» – виведення) – логічний метод пізнання, який грунтується на отриманні окремих конкретних висновків, положень на основі знання загальних положень, закономірностей розвитку цілісної системи. Логічною підставою дедуктивного висновку є аксіома: «Все, що стверджується або заперечується відносно всього класу, стверджується або заперечується відносно окремих елементів цього класу». Теорію дедукції вперше створив Аристотель. Подальший розвиток вона отримала в працях Р. Декарта та інших філософів 17-18 ст.

Розглянемо етапи економіко-математичного моделювання.

1. Постановка економічної проблеми та її якісний аналіз. Головним на цьому етапі є чітке формулювання суті проблеми та допущень; виділення найважливіших характеристик і властивостей об'єкта, який

моделюється, основних залежностей між елементами; формування гіпотез, що пояснюють поведінку і розвиток об'єкта.

2. Побудова математичної моделі. Це етап формалізації економічної проблеми, тобто її запису у вигляді конкретних математичних залежностей та відношень (функцій, рівнянь, тотожностей, нерівностей і т.п.). Модель повинна включати тільки основні фактори та умови, які характеризують об'єкт.

3. Математичний аналіз моделі. На цьому етапі з'ясовують загальні властивості моделі та методи її рішень.

4. Підготовка вхідної інформації. Цей етап є найбільш трудомісткий. В процесі підготовки інформації широко використовують методи теорії ймовірностей, математичної статистики.

5. Чисельне рішення моделі. Цей етап включає розробку алгоритмів для чисельного рішення задачі, розробку програм та безпосереднє проведення розрахунків.

6. Аналіз чисельних результатів та їх застосування. На цьому заключному етапі вирішують питання правильності та повноти результатів моделювання.

Важливим при економіко-математичному моделюванні є адекватність моделі, тобто відповідність моделі об'єкту або процесу, який моделюється. Без такої перевірки застосування результатів моделювання може не тільки не принести очікуваних позитивних результатів, а й бути шкідливим через хибність результатів.

1.2 Економетричні моделі в системі економіко-математичного моделювання

Класифікація економіко-математичних моделей чітко не встановлена. Умовно всі економіко-математичні моделі класифікують **на економетричні та оптимізаційні**.

В свою чергу економетричні моделі складають групу **детермінованих та стохастичних моделей.** Для рішення детермінованих моделей використовують методи лінійної алгебри, математичного аналізу та диференціальні рівняння. Рішення стохастичних моделей базується на використанні методів теорії ймовірностей та математичної статистики, теорії випадкових процесів, теорії масового обслуговування.

Економетричні моделі належать до функціональних моделей, які кількісно описують зв'язок між вхідними показниками економічної системи (X) та результативним показником (Y). В загальному вигляді економетричну модель можна записати так:

$$Y = f(X, u), \tag{1.1}$$

де X – вхідні економічні показники , u – випадкова або стохастична складова.

Показники X найчастіше є детермінованими. Оскільки залежна (результативна) змінна Y залежить від u, то вона також є стохастичною.

Економетричні моделі кількісно описують кореляційний зв'язок між економічними величинами. Для побудови економетричної моделі необхідно:

1) мати достатню сукупність спостережень вхідних даних;
2) забезпечити однорідність сукупності спостережень;
3) забезпечити точність вхідних даних.

В залежності від джерела виникнення розрізняють апріорну та апостеріорну інформацію. **Апріорі** (від лат. «a priori» – буквально із попереднього) – знання, що передує досліду і незалежне від нього. Термін «апріорі» впроваджений середньовічними схоластами, які акцентували на тому, що деякі знання передують досліду, експерименту і таким чином є незалежними від них. Згідно Декарту та Лейбніцу, найбільш глибоке знання досягається поза дослідом шляхом безпосереднього пізнання істини та інтелектуальної інтуїції.

Апостеріорі (лат. «a posteriori» – буквально із наступного) – знання, отримані з досліду. Такий спосіб отримання знання розглядався у роботах Платона, Аристотеля, Боеція, Фоми Аквінського, а пізніше – в роботах І. Канта, який вважав, що спеціальні закони науки можуть пізнаватися лише апостеріорно, тоді як загальні принципи пізнання незалежні від всякого досліду, тобто апріорні.

При побудові економетричних моделей використовують як апріорні, так і апостеріорні знання (інформацію). На основі інтуїції, тобто апріорної інформації дослідник робить загальні висновки про можливість та вид кореляційної залежності між економічними показниками, а на основі апостеріорних знань (інформації) – будує економетричну модель.

Важливою передумовою застосування кореляційного аналізу в економічних дослідженнях є дотримання певних вимог до вхідної інформації: це достовірність і однорідність вхідної інформації та відповідність її закону нормального розподілу.

Достовірність інформації зумовлюється джерелом її отримання. Однорідність вхідної інформації оцінюється за критеріями середньоквадратичного відхилення та коефіцієнта варіації, які розраховуються по кожному об'єкту інформаційної сукупності.

Середньоквадратичне відхилення σ показує абсолютне відхилення індивідуальних значень від середньоарифметичного і розраховується за формулою:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}}, \tag{1.2}$$

де $\bar{x}$ – середнє арифметичне значення об'єкта інформаційної сукупності, n – кількість дослідів.

Коефіцієнт варіації характеризує відносну міру відхилень окремих значень об'єкта інформаційної сукупності від середньоарифметичної і розраховується за формулою:

$$V = \frac{\sigma}{\bar{x}} \cdot 100\%, \tag{1.3}$$

Чим більша величина коефіцієнта варіації, тим відносно більша розкиданість та менша однорідність значень об'єкта інформаційної сукупності. Варіація значень об'єкта інформаційної сукупності (показника) вважається незначною, якщо коефіцієнт варіації не перевищує 10%; середньою – якщо складає 10-20%, значною – якщо вона більше 20% і менше 33%; якщо варіація перевищує 33% – це свідчить про неоднорідність інформації та необхідність виключення нетипових спостережень.

До вхідної інформації пред'являється умова відповідності її закону нормального розподілу. Згідно з цим законом основна кількість досліджуваних даних по кожному показнику повинна бути згрупована навколо їх середнього значення: дані з дуже малими та дуже великими значеннями повинні зустрічатися по можливості рідше. Для кількісної оцінки ступеня відхилення інформації від нормального розподілу використовують відношення показника асиметрії до її похибки та відношення показника ексцесу до його похибки.

Показник асиметрії A та її похибку m_a розраховують за формулами:

$$A = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{n \cdot \sigma^3}; \qquad m_a = \sqrt{\frac{6}{n}}. \tag{1.4}$$

Показник ексцесу E та його похибку m_e розраховують за формулами:

$$E = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n \cdot \sigma^4}; \qquad m_e = \sqrt{\frac{24}{n}}, \quad \text{або} \quad m_e = 2m_a. \tag{1.5}$$

У симетричному розподілі показник асиметрії A=0. Якщо $A \neq 0$, то це свідчить про наявність асиметрії у розподілі даних навколо середньої величини $\bar{x}$. Від'ємна асиметрія свідчить, що у масиві даних показника переважають дані з великими значеннями, а дані з малими значеннями зустрічаються рідше. У додатній асиметрії тенденція зворотня, тобто у масиві значень показника дані з малими значеннями зустрічаються частіше, ніж з великими.

У нормальному розподілі показник ексцесу E=0. Якщо $E > 0$, то дані густо згруповані навколо середньої і крива розподілу буде гостро вершинною; і навпаки, якщо $E < 0$, то крива розподілу буде плоско вершинною. Однак, якщо відношення $\frac{A}{m_a}$ та $\frac{E}{m_e}$ менші 3, то асиметрія і ексцес не мають суттєвого значення і досліджувана інформація підпадає під нормальний закон розподілу.

Економетричний аналіз складається з таких етапів:

1. Формування теорії чи гіпотези та постановка конкретної економічної задачі.

На основі апріорної або інтуїтивної інформації про об'єкт дослідження висуваємо попередню гіпотезу про можливі причинні зв'язки та її напрямки; формуємо постановку задачі, на основі вивчення літературних джерел, логічних висновків економічної суті приймаємо гіпотезу про форму аналітичної залежності між досліджуваним чинниками.

2. Побудова та оцінка параметрів економетричної моделі.

Обґрунтувавши вид аналітичної функції, формуємо масив вхідної інформації та проводимо її попередній аналіз на кількісну і якісну однорідність та відповідність її закону нормального розподілу. При побудові моделі для розрахунку її параметрів використовують типові алгоритми та пакети прикладних програм.

3. Перевірка та коригування моделі, статистичні висновки.

Якість економетричної моделі оцінюємо за допомогою коефіцієнта детермінації, t – критерія Стьюдента та адекватності моделі за F – критерієм Фішера. Якщо виявиться, що кореляційний зв'язок помірний, а фактичні значення t – критерії Стьюдента та F – критерії Фішера менші за їх критичні значення, то необхідно повернутись на попередній крок алгоритму.

4. Застосування моделі для прогнозування, аналізу, вивчення причинних зв'язків.

Після оцінки якості економетричної моделі та її адаптації проводять застосування економетричної для цілей економічного прогнозування, економічного та фінансового аналізу, вивчення взаємозв'язків між економічними показниками.

Важко вказати такі теоретичні і практичні проблеми ринкової економіки, у вирішенні яких в даний час не застосовувались би економіко-математичні методи. Математичне моделювання стало найбільш затребуваним напрямком в економічній науці Заходу. Не випадково, що з моменту заснування Нобелівських премій по економіці (1969 р.) вони присуджуються, як правило, за економіко-математичні дослідження. Серед Нобелівських лауреатів – найвизначніші економетрики: Р. Фріш, Я. Тинберген, П. Самуельсон, Д. Хікс, К.Ерроу, В. Леонтьєв, Т. Купманс.

Ще на початку ХХ ст. були зроблені спроби розробки так званих «барометрів розвитку» – різновидності індексу господарського розвитку, які мали «передбачувати» майбутню кон'юнктуру товарного ринку. Найбільшу популярність одержала Гарвардська школа економічних барометрів. На цей «барометр» покладалась надія передбачування поведінки товарного і грошового ринків. Однак «барометр» не зміг спрогнозувати світової кризи, що уразила капіталістичні країни у 1929-1933 рр.. Криза обумовила критичний перегляд методів економетрики. Наслідком такого перегляду було впровадження методів, які враховують випадкові процеси в економіці. Ще в 1928 р. були опубліковані відомі досліди Х. Кобба і П. Дугласа про виробничу функцію з можливістю заміни факторів. Ця функція є важливим інструментом економетричного аналізу:

$$y = a \cdot L^{\alpha_L} \cdot C^{\alpha_C}, \tag{1.6}$$

де L – витрати праці (labor),

C – витрати виробничих фондів або «капіталу» (capital); при цьому $\alpha_L + \alpha_c = 1$,

y – обсяг випуску продукції,

α_L, α_c – коефіцієнти еластичності праці та капіталу.

За межами моделей завжди залишається ряд неврахованих факторів, а тому одержані результати є наближеними; ні одна модель не може замінити логічні висновки дослідника і описати безліч неформальних умов.

Економетричні моделі являють собою системи регресійних рівнянь і тотожностей, кожне із яких використовується для визначення одного показника, що досліджується. При цьому показники, які виступають в одних рівняннях в якості змінних, в других використовуються в якості аргументу, який впливає на значення решти змінних. В більш вузькому значенні економетричними моделями вважаються системи рівнянь, які враховують імовірний характер економічних процесів. Звідси випливає, що рівняння економетричної моделі містять також і випадкові змінні, а її параметри встановлюються статистично на основі часових рядів або, наприклад, вибіркових даних.

Економетричні моделі в їх класичному вигляді відносяться до безкритеріальних моделей. Вони виступають в якості гнучкого і ефективного інструменту прогнозування економічних тенденцій.

На основі економетричних моделей можна прогнозувати як характеристики еволюційного розвитку, так і параметри стрибків. Основним методом прогнозування еволюційного розвитку є метод прогнозної екстраполяції, який базується на використанні принципу (закону) інерції, тобто перенесенні тенденції (тренду) розвитку економічного явища, що мала місце у минулому і теперішньому періодах на майбутній період.

Розрізняють базу екстраполяції та строки прогнозування. База екстраполяції – кількість інтервалів часу (10-15), за які вивчається тенденція, а строки прогнозування – кількість періодів часу, на які розраховують прогноз. В залежності від строків прогнозування прогнози класифікують на короткострокові (на 1-2 періоди), середньострокові (3-5 періодів) та довгострокові (більше 5 періодів). Прогноз, що характеризується одним значенням, називається точковим, трьома значеннями – інтервальним (нижня та верхня границі прогнозу та точкове значення прогнозу). Чим довша база прогнозу та коротші строки прогнозу, тим точніший прогноз.

Економетричні моделі будуються на основі статистичної інформації часових рядів і не вимагають затрат на розрахунок на перспективу різних нормативних показників (наприклад, коефіцієнтів прямих матеріальних витрат) на відміну від моделей структурних (моделі міжгалузевого балансу).

На основі економетричних моделей, побудованих на основі часових рядів, проводиться короткострокове та середньострокове прогнозування розвитку економіки країни.

При вивченні складних соціально-економічних явищ все більшого значення набувають багатомірні статистичні моделі, в основі побудови яких лежить факторний аналіз, багатомірний шкалограмний аналіз, методи розпізнавання образів та інші методи прикладної математики.

Недоліком економетричних моделей є те, що статистичні моделі можуть бути перевіреними лише після реалізації/настання прогнозованих явищ і процесів.

1.3 Економетрія – наука про економіко-статистичне моделювання

Усі соціально-економічні явища та процеси навколишнього світу знаходяться у причинній залежності та взаємозв'язку. В процесі прийняття управлінських рішень виникає необхідність кількісної оцінки співвідношень між цими явищами та процесами, які можна виразити системою економічних показників. Кількісні зв'язки між показниками економічних явищ та процесів вивчаються прикладною економічною дисципліною економетрією.

Термін «економетрія» вперше запропонував львівський вчений П. Чомпа у своїй книзі: «Нариси економетрії і природної теорії бухгалтерії, яка ґрунтується на політичній економії», що була видана в 1910 р. у Львові.

Як самостійна наука «економетрія» сформувалась у 30-х рр. XX ст. Термін «економетрія» (або «економетрика») для позначення нового напрямку в економічній науці широко використовував норвезький вчений Р. Фріш, який у 1926 р. дав таке визначення економетрики: «Економетрика є синтез економічної теорії, математики і статистики». Офіційною датою започаткування економетрії вважається 1931 р., коли було створено Міжнародне економетричне товариство (повна назва «Міжнародне товариство розвитку економічної теорії в його зв'язку з статистикою і математикою»). З 1933 р. це товариство видає журнал «Економетрика».

Термін «економетрія» в перекладі з давньогрецької означає вимірювання в економіці. Класичне визначення науки економетрія таке: **економетрія** – це наука, що вивчає кількісні закономірності та взаємозв'язки економічних явищ та процесів за допомогою математико-статистичних методів та моделей.

Засновниками економетрики вважаються Р. Фріш, Я. Тинберген, Е. Штумпеттер. У своїх роботах вони пробували узгодити економічну теорію з математикою і статистикою.

На перших порах вчені старались слідувати триєдиній формулі Р. Фріша, дуже високо цінуючи ідею поєднання в економічному дослідженні теоретичного аналізу, використання емпіричних даних і застосування методів математики.

Слід підкреслити, що економетрія ніколи не представляла собою єдиної течії з точки зору теоретичних поглядів її прихильників. В теоретичному відношенні економетрика еклектична.

Початкове поняття економетрики було загальноприйнятим тільки в період її створення. Надалі швидко розвиватися тенденції розшарування економетрики. Цей процес здійснювався в трьох напрямках: поряд з теоретичними дослідженнями на основі застосування математики і статистики все більшого значення набували прикладні розробки, які не мали безпосереднього відношення до політекономічних проблем; відділялись абстрактно-теоретичні дослідження математичних моделей економіки, які не використовували емпіричних даних; одержали розвиток роботи емпірико-статистичного характеру.

Наслідком цих відцентрових тенденцій стала розпливчастість самого поняття «економетрика».

Під економетрією в широкому змісті розуміється сукупність різного роду економічних досліджень, що проводяться з використанням математичних методів, тобто економетрія являє собою досить широкий методологічний напрямок з великою різноманітністю об'єктів дослідження, що не мають своєї власної економічної теорії.

Економетрія у вузькому розумінні – застосування математико-статистичних методів та моделей в економічних дослідженнях: побудова математико-статистичних моделей економічних явищ, оцінка параметрів у моделях будь-якого типу і т.п.

Економетрія як наука швидко розвивається.

На сьогодні сформувались дві ведучі школи економетрики: американська та голландська. В економетричних методах обох шкіл широко використовуються відносні показники, множини, рівняння і системи рівнянь регресії, які містять часові змінні. Так, Бруклінська економерична модель американської школи містить 359 рівнянь і 56 тотожностей. Найбільш відома голландська економетрична модель –

модель розвитку Голландії – побудована на системі міжгалузевих балансів і містить 53 рівняння.

В останні роки в економетрику проникають формально-математичні методи. Багато досліджень дуже математизовані і по своїй суті абстрактні. В таких дослідженнях під нашаруванням теорем, лем, формул часто на другий план відходить емпіричний матеріал, який власне підлягає дослідженню. Президент американської асоціації економетристів В. Леонтьєв писав, що збільшення кількості математичних формул інколи зменшує цінність економічного дослідження, тому що за формулами ми не завжди помічаємо суті економічної проблеми.

Отже, економетрія є інструментом, який дозволяє перейти від якісного економічного аналізу до кількісного, використовуючи математико-статистичні методи та моделі. Економетрія намагається поєднати можливості математичної школи, статистики та економічної теорії. **Економетрія** – наука, що займається побудовою, аналізом і вдосконаленням математичних моделей і методів на основі реальних емпіричних (статистичних) даних про взаємозв'язок економічних об'єктів і процесів.

Метою економетрики є кількісний аналіз, пояснення і прогнозування взаємозв'язків економічних процесів. Економетрія дозволяє знайти кількісне підтвердження або спростування того або іншого економічного закону або гіпотези.

Об'єктом економетрії є економічні процеси і явища, а предметом – кількісне моделювання.

В рамках дисципліни «Економетрія» вивчаються основи математичного моделювання економічних процесів; однофакторні лінійні та нелінійні економетричні моделі; багатофакторні економетричні моделі; економетричні моделі на основі системи структурних рівнянь; прикладні економетричні моделі.

Економетрія базується на вивченні взаємодії різних економічних процесів і показників та відображення цієї взаємодії у формалізованому вигляді та побудові моделей. Застосування економетрії має важливе значення при вивченні динаміки і тенденцій розвитку економіки, виявленні впливу найважливіших факторів на досліджувані процеси/результати, прийнятті рішень.

Питання для самоконтролю:

1. Сутність та класифікація систем.
2. Сутність та класифікація моделей.
3. Етапи економіко-математичного моделювання.
4. Охарактеризуйте сутність апріорного підходу до відбору чинників моделі.
5. Охарактеризуйте сутність апостеріорного підходу до відбору чинників моделі.
6. Сфера застосування економіко-математичних методів та моделей.
7. Дайте визначення науки «econометрія».
8. Взаємозвязок економетрії з іншими дисциплінами.
9. Охарактеризуйте об'єкт, предмет і основні цілі економетрики.

Тема 2

ОДНОФАКТОРНІ ЛІНІЙНІ ЕКОНОМЕТРИЧНІ МОДЕЛІ

2.1 Поняття кореляційних зв'язків та регресійної залежності

В природі, суспільстві, економіці явища, процеси, об'єкти знаходяться між собою в причинній залежності.

Зв'язок між двома величинами називається функціональним, якщо будь-якому визначеному значенню величини x (із множини її можливих значень) відповідає одне і тільки одне значення y, тобто y можемо представити функцією від x:

$$y = f(x). \tag{2.1}$$

Зв'язок між двома величинами називається стохастичним, якщо після визначення величини x величина y залишається випадковою і може приймати різні значення з обумовленими імовірностями.

При вивченні зв'язку між явищами стохастична залежність частково вказує на відповідну причинну залежність (наприклад, залежність продуктивності праці робітника від стажу його роботи за фахом). Але при наявності стохастичного зв'язку між явищами може і не бути причинної залежності між ними. Це виникає тому, що обидва явища окремо залежать від загальних факторів. Так, зв'язок між фондовіддачею і собівартістю є стохастичним і непричинним, так як обидва ці показники залежать від фондоозброєності, електроозброєності і т.д.

Окремими випадками стохастичної форми зв'язку можуть бути кореляційні зв'язки. Дві випадкові величини є кореляційно залежними, якщо математичне очікування однієї з них змінюється в залежності від зміни другої.

«Кореляція» походить від англійського «corelation» і означає співвідношення або відповідність між факторами й ознаками. Основоположниками теорії кореляції вважаються англійські статистики Ф. Гальтон і К. Пірсон. Термін «кореляція» застосовується в різних галузях науки і техніки для позначення взаємозалежності, взаємної відповідності.

При виконанні кореляційних розрахунків необхідно розрізняти факторну та результативну ознаку. Факторною називається така ознака, від якої залежить інша ознака, а сама вона є незалежною. Залежна ознака називається результативною.

У процесі формалізації економіко-статистичної моделі факторна ознака позначається через X_i, а результативна через $У_i$, тобто умовно можна сказати, що факторна ознака виражає аргумент, а результативна – функцію. Факторна ознака X_i є незалежною від змінної $У_i$ так як відсутній зворотний вплив $У_i$ на X_i. У зв'язку з цим чинники X_i часто називають екзогенними (зовнішніми), а змінну $У_i$ – ендогенною (внутрішньою) змінною моделі. Значення змінних X_i визначаються поза моделлю, тобто задаються як початкові дані.

Факторна ознака або фактор – це технічні, технологічні, природні, кліматичні, економічні, організаційні, соціально-демографічні та інші показники, що проявляють вплив на який-небудь результативний економічний показник: прибуток, собівартість, продуктивність праці та ін. Задача математичного моделювання полягає у виявленні кількісного зв'язку між факторами та результативним економічним показником.

Фактор, що включається в економетричну модель, повинен відповідати таким вимогам: 1) мати кількісне вираження; 2) між фактором і результуючим показником повинен бути причинний зв'язок і статистичний зв'язок; 3) між факторами у багатофакторній моделі не повинно бути мультиколінеарності (тісного зв'язку між факторами).

Кореляційний зв'язок між факторами в економіці класифікують за ознаками:

– за типом: прямий і обернений;

– за формою: лінійний і нелінійний;

– за тіснотою зв'язку: слабкий, помірний, помітний, сильний, дуже сильний;

– за участю факторних ознак: парний, множинний.

Розглянемо приклади факторних та результативних ознак.

Приклад 2.1. Досліджується кореляційна залежність між рівнем продуктивності праці та рівнем механізації праці. Отже, рівень механізації праці – факторна ознака (x), а рівень продуктивності праці – результативна ознака (y). Зв'язок між ознаками прямий та лінійний.

Приклад 2.2. Досліджується залежність між рівнем продуктивності праці робітника та його віком. У цьому випадку вік робітника – факторна ознака (x), а рівень продуктивності праці – результативна ознака (y). Зв'язок між ознаками нелінійний: з початку прямий, а потім обернений, а залежність описується квадратичною параболою.

Приклад 2.3. Досліджується залежність між собівартістю одиниці продукції та врожайністю. Врожайність сільськогосподарської культури є факторною ознакою (x), а собівартість одиниці продукції – результативною (y). Зв'язок між ознаками: нелінійний та обернений.

Кількісний вплив факторів на результативний показник вивчається за допомогою регресійного аналізу, який дозволяє встановити вид аналітичної залежності між змінними x, y та оцінити параметри економетричної моделі.

Термін **«регресія»** (рух назад, повернення в колишній стан) був введений Ф. Галтоном в кінці XIX ст. при аналізі залежності між старінням батьків і зростанням дітей. Під **регресією** розуміється функціональна залежність між пояснюючими змінними і умовним математичним очікуванням (середнім значенням) залежної змінної, яка будується з метою прогнозу (прогнозування) цього середнього значення при фіксованих значеннях перших.

Прикладом можливого застосування регресійного аналізу в економіці може бути дослідження продуктивності праці, собівартості та інших якісних економічних показників від таких факторів як вартість основних засобів, питома вага заробітної плати у витратах на виробництво, рівень спеціалізації, кооперування, плинність та рівень кваліфікації кадрів. Регресійні моделі також широко застосовуються в прогнозуванні.

При виборі форми кореляційної залежності $y = f(x)$ виходять перш за все із економічної природи явищ, простоти функції і вимоги на обмеження числа параметрів. Форму кореляційного зв'язку можна визначити як графічним, так і аналітичним методами.

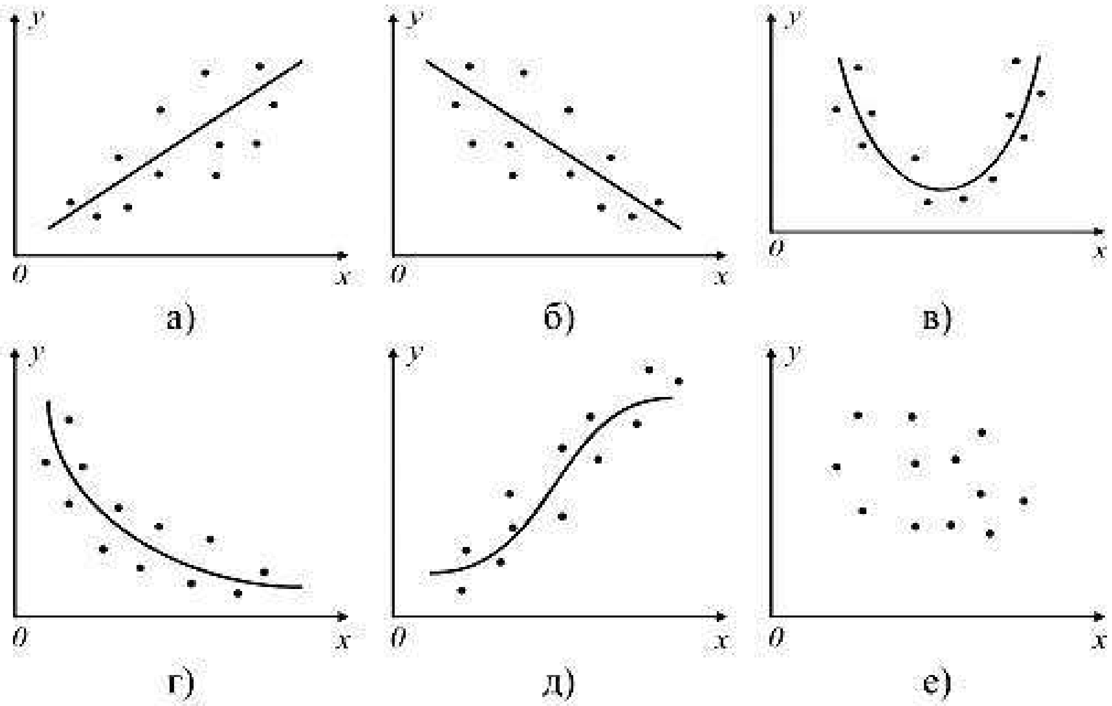


Рисунок 2.1 – Поле кореляції

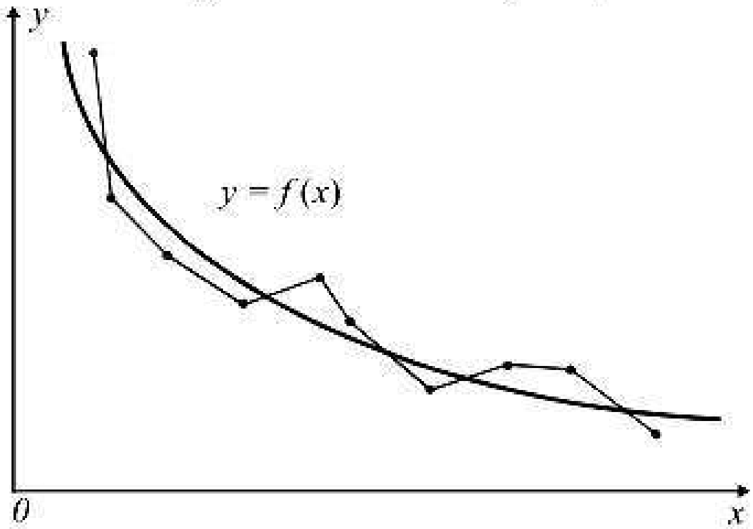


Рисунок 2.2 – Емпірична та теоретична лінії регресії

У випадку парної кореляції вихідні дані n пар $x_i, y_i \left(i = \overline{1,n}\right)$ в прямокутній системі координат утворюють кореляційне поле (рисунок 2.1). Розміщення точок на кореляційному полі дозволяє визначати характер залежності: $y = f(x)$ (а, б – лінійна; в – параболічна; г – гіперболічна; д – логістична; е – відсутня). Якщо x_i – різні $\left(i = \overline{1,n}\right)$, то точки кореляційного поля з'єднують в послідовності зростання абсциси і одержують так звану емпіричну лінію регресії (рисунок 2.2).

Графік функції $y = f(x)$ називають теоретичною лінією регресії. Щоб вибрати ту чи іншу форму кореляційної залежності, слід зіставити кореляційне поле або емпіричну лінію регресії з графіками відомих функцій. Для більш точного встановлення форми зв'язку вихідні дані обробляють на ЕОМ по програмах кореляційного аналізу. При цьому аналізують кілька функцій $y = f(x)$ і беруть ту, для якої кореляційне відношення η або коефіцієнт парної кореляції r найбільші (або середня похибка апроксимації ε найменша).

2.2 Лінійний кореляційний та регресійний аналіз двох змінних

Економетричні моделі в залежності від обсягу вибірки статистичних даних поділяються на узагальнені та вибіркові.

Узагальненою вважається регресійна модель, побудована по статистичних даних генеральної сукупності і має вигляд: $y = \beta_0 + \beta_1 x + u$, де β_0, β_1 – параметри моделі, u – випадкова величина (відхилення).

Вибіркова модель будується по статистичних даних вибіркової сукупності. У загальному вигляді вибіркова регресійна модель між факторною ознакою $X = \{x_1, x_2, \cdots, x_n\}$ та результативною ознакою $Y = \{y_1, y_2 \cdots, y_n\}$ з урахуванням фактора випадкових величин (помилок) $U = \{u_1, u_2, \cdots, u_n\}$ записується у вигляді:

$$y = a_0 + a_1 x + u, \qquad (2.2)$$

a_0, a_1 – невідомі параметри економетричної моделі;

u – випадкова величина (відхилення).

Тут і надалі з метою уникнення неоднозначності великими літерами X, Y, U ми позначаємо дискретні (векторні) величини, а малими x, y, u – неперервні.

Причини обов'язкової присутності в регресійних моделях випадкової змінної (відхилення) u такі:

1. Невключення до моделі всіх пояснюючих змінних. Будь-яка регресійна модель є спрощенням реальної ситуації. Остання завжди

являє собою взаємодію різних чинників, багато з яких в моделі не враховуються, що обумовлює відхилення реальних значень залежної змінної від її модельних значень. Проблема полягає ще й в тому, що наперед не відомо, які фактори при певних умовах дійсно є визначальними, а якими можна нехтувати.

2. Неправильний вибір функціональної форми моделі. Через недостатню вивченість процесу чи явища, що моделюється, може бути невірно підібрана аналітична функція, якою проводиться моделювання.

3. Агрегація змінних. У багатьох моделях розглядаються залежності між чинниками, які представляють складну комбінацію інших, простіших змінних. Наприклад, чинник сукупний попит є складною композицією індивідуальних попитів, які впливають на результативний показник. Це може виявитись причиною відхилення реальних значень від модельних.

4. Помилка вимірювань. Якою б якісною не була модель, помилки вимірювань змінних вплинуть на невідповідність модельних значень емпіричним даним, що також відобразиться на величині випадкового члена (відхилення).

5. Обмеженість статистичних даних. Часто будуються моделі, що виражаються безперервними функціями. Але для цього використовується набір даних, що мають дискретну структуру. Ця невідповідність знаходить свій вираз у випадковому відхилені.

6. Непередбачуваність людського чинника. Ця причина може «зіпсувати» найякіснішу модель. Дійсно, при правильному виборі форми моделі, скрупульозному підборі пояснюючих змінних все одно неможливо спрогнозувати поведінку кожного індивідуума.

Таким чином, відхилення (випадкова величина) є віддзеркаленням впливу всіх описаних вище причин. До того ж, цей перелік може бути доповненим.

Метод математичної статистики, який вивчає кореляційні зв'язки між явищами, називається кореляційним аналізом. Кореляційний аналіз представляє собою інструмент, який дозволяє кількісно оцінити зв`язки між великою кількістю взаємодіючих економічних явищ, при цьому деякі з них невідомі. Застосування кореляційного аналізу дає можливість перевірити різні економічні гіпотези про наявність і силу зв'язку між двома явищами або явищем та групою явищ, а також гіпотезу про форму зв'язку.

Задачею регресійного аналізу є обчислення невідомих параметрів a_0, a_1 рівняння регресії $y = a_0 + a_1 x$. При цьому необхідно досягти «найкращої» апроксимації. Найчастіше при цьому використовують метод найменших квадратів, що передбачає мінімізацію виразу:

$$Q(a_0, a_1) = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} u_i^2 \to \min, \qquad (2.3)$$

де $y_i, \hat{y}_i$ – фактичні (емпіричні) та розрахункові (теоретичні) значення результативної ознаки.

На рисунку 2.3 пряма є теоретичною лінією регресії. Із множини прямих необхідно вибрати «найкращу» з точки зору мінімізації суми квадратів відхилень u_i: $u_i = y_i - \hat{y}_i = y_i - a_0 - a_1 x_i$; $i = \overline{1,n}$

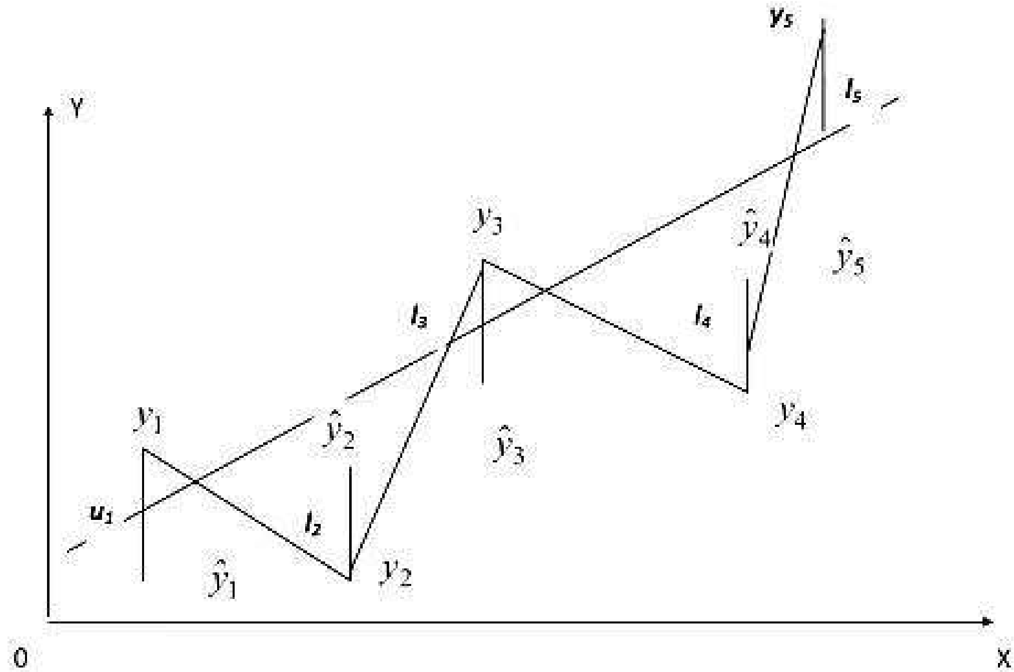


Рисунок 2.3 – Графічна інтерпретація методу найменших квадратів

Відхилення або помилки u_i ще іноді називають залишками. Теоретичну лінію регресії необхідно проводити таким чином, щоб сума квадратів відхилень була мінімальною. У цьому і полягає метод найменших квадратів: невідомі параметри a_0 та a_1 визначаються таким чином, щоб мінімізувати $\sum_{i=1}^{n} u_i^2$. Мінімум функції (2.2) досягається за умови, коли перші похідні дорівнюють нулю.

Тому підставивши в вираз (2.2), взявши частинні похідні $\frac{\partial Q}{\partial a_0}$ і $\frac{\partial Q}{\partial a_1}$, після елементарних перетворень одержимо систему нормальних рівнянь:

$$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \end{cases}, \tag{2.4}$$

де n – кількість спостережень або довжина вибірки.

Шляхом розв'язання системи нормальних рівнянь на основі метода найменших квадратів оцінюються параметри лінійної економетричної моделі a_0 та a_1:

$$a_0 = \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i x_i \cdot \sum_{i=1}^{n} x_i}{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}; \quad a_1 = \frac{n \cdot \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}.$$

Параметри a_0, a_1 мають таку економічну інтерпретацію або зміст: параметр a_0 характеризує деяке середнє значення результативного показника y, а параметр a_1 показує, як в середньому зміниться y при зміні x на одну одиницю.

Приклад 2.4. Нехай залежність денного виробітку робітника від рівня механізації праці описується рівнянням регресії:

$$y = 2{,}142 + 0{,}051x$$

У цьому рівнянні параметр a_0 є середнім денним виробітком при виконанні операції вручну, а a_1 – перевищення середнього виробітку при механізованому виконанні операції. А тому параметр a_1 (коефіцієнт нахилу) показує, що при підвищенні рівня механізації на 1% денний виробіток зростає в середньому на 0,051 одиниць.

Отже, при моделюванні та аналізі багатьох соціально-економічних явищ та процесів виникає задача виявлення та оцінки зв'язку між ними, одне з яких є незалежною змінною (x), чи фактором, а інше (y) – залежною або результативною ознакою. Форма зв'язку між змінними x та y встановлюється шляхом логічного аналізу їх природи та зовнішнього вигляду кореляційного поля та емпіричної лінії регресії, а тіснота зв'язку – величиною коефіцієнта кореляції.

Тіснота (щільність) зв'язку між змінними x та y оцінюється коефіцієнтом парної кореляції r_{xy} (якщо зв'язок лінійний) і кореляційним відношенням η_{xy} (якщо зв'язок нелінійний).

Для обчислення коефіцієнта кореляції пропонуються різні формули. Розглянемо деякі з них:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}, \tag{2.5}$$

де $\overline{xy}$ – середнє значення добутку змінної x та змінної y; $\bar{x}$, $\bar{y}$ – середнє значення змінних x та y.

$$r_{xy} = \frac{Cov(x, y)}{\sqrt{Var(x) \cdot Var(y)}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \cdot \sigma_x \cdot \sigma_y}, \tag{2.6}$$

$$r_{xy} = \frac{n \cdot \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{\sqrt{\left[n \cdot \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2\right] \cdot \left[n \cdot \sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right]}}, \tag{2.7}$$

$$\overline{xy} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i y_i; \ \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i; \ \bar{y} = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i;$$

n – довжина вибірки або кількість спостережень;

$Cov(x, y)$ – коефіцієнт коваріації між змінними x та y;

$$Cov(x,y)=\frac{\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})}{n-1}$$

$Var(x)$ – дисперсія змінної x,

$$Var(x)=\sigma_x^2=\frac{1}{n}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$Var(y)$ – дисперсія змінної y;

$$Var(y)=\sigma_y^2=\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i-\bar{y})^2$$

Коефіцієнт кореляції змінюється в інтервалі:

$$-1\le r_{xy}\le 1$$

При коефіцієнті кореляції рівному 0, між y та x не існує кореляційного зв'язку. Якщо коефіцієнт кореляції знаходиться в інтервалі $-1\le r_{xy}\le 1$ або $0\le r_{xy}\le 1$, між y та x існує обернена або пряма кореляційна залежність.

За щільністю зв'язку можна виділити:

а) слабкий зв'язок, якщо $r_{xy}\le 0{,}3$;

б) середній зв'язок, якщо $r_{xy}=0{,}31-0{,}65$;

в) сильний зв'язок, якщо $r_{xy}=0{,}66-0{,}95$.

Для визначення варіації результативного показника під впливом факторів обчислюють коефіцієнт детермінації r_{yx}^2. Припустимо, що $r_{yx}^2=0{,}8$; тоді можна сказати, що 80% варіації результативного показника відбувається під впливом фактора x, а решта 20% приходиться на інші фактори та випадкові величини.

При виявленні зв'язку між варіацією факторної ознаки *(x)* і варіацією результативної ознаки *(y)* використовують такі дисперсії:

1) дисперсія, яка вимірює загальну варіацію за рахунок дії всіх факторів, або загальна дисперсія:

$$\sigma_y^2=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}, \qquad (2.8)$$

2) дисперсія, яка вимірює варіацію результативної ознаки y за рахунок дії факторної ознаки x або дисперсія, що пояснює регресію:

$$\delta_y^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{n}, \qquad (2.9)$$

3) залишкова дисперсія, яка характеризує варіацію ознаки y за рахунок всіх факторів, крім x (тобто при виключенні x) або дисперсія помилок:

$$\sigma_u^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}, \qquad (2.10)$$

тоді по правилу додавання дисперсій:

$$\sigma_y^2 = \delta_y^2 + \sigma_u^2, \qquad (2.11)$$

або

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \hat{y})^2 + \sum(\hat{y}_i - \bar{y})^2, \qquad (2.12)$$

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

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

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

Вираз (2.11) запишемо у скороченому вигляді:

$$SST=SSE+SSR. \qquad (2.13)$$

Таким чином, ми розклали загальну дисперсію на дві частини: дисперсію, яка пояснює регресію, та дисперсію помилок (або дисперсію випадкової величини).

Поділивши обидві частини виразу (2.13) на σ_y^2, отримаємо:

$$1 = \frac{\sigma_u^{\ 2}}{\sigma_y^2} + \frac{\delta_y^2}{\sigma_y^2}. \quad (2.14)$$

Із виразу (2.14) виплаває, що перша частина $\frac{\sigma_u^2}{\sigma_y^2}$ є питомою вагою помилок у загальній дисперсії, тобто часткою дисперсії, яку не можна пояснити через регресійний зв'язок. Друга частина $\frac{\delta_y^2}{\sigma_y^2}$ є складовою дисперсії, яку можна пояснити через лінію регресії.

Частина дисперсії, що пояснюється регресією, називається коефіцієнтом детермінації і позначається R^2. Коефіцієнт детермінації використовується як критерій адекватності моделі, бо є мірою пояснювальної сили незалежності змінної x.

Таким чином, коефіцієнт детермінації:

$$R^2 = \frac{\delta_y^2}{\sigma_y^2} \quad (2.15)$$

або

$$R^2 = \frac{SSR}{SST}. \quad (2.16)$$

Із виразу (2.15) випливає, що коефіцієнт детермінації завжди додатній і знаходиться у межах $0 \le R^2 \le 1$.

Між коефіцієнтом кореляції і нахилом a_1 та середнім квадратичним відхиленням σ_x, σ_y існує певний зв'язок. Це дає можливість розрахувати параметри вибіркового рівняння регресії $y = a_0 + a_1 x + u$ через ці величини.

Оскільки
$$r_{xy} = \frac{\operatorname{cov}(x,y)}{\sqrt{\operatorname{var}(x)\cdot\operatorname{var}(y)}}, \quad (2.17)$$

$$a_1 = \frac{\operatorname{cov}(x,y)}{\sqrt{\operatorname{var}(x)}} = \frac{\operatorname{cov}(x,y)}{\sigma_x^2}, \quad (2.18)$$

можна записати вираз для коефіцієнта кореляції через параметр a_1:

$$r = \left[\frac{\operatorname{cov}(x,y)}{\sigma_x^2}\right]\cdot\left(\frac{\sigma_x}{\sigma_y}\right) = a\cdot\frac{\sigma_x}{\sigma_y}. \quad (2.19)$$

Запишемо формули для розрахунку параметрів економетричної моделі:

$$a_1 = \frac{\operatorname{cov}(x,y)}{\operatorname{var}(x)} = \frac{\operatorname{cov}(x,y)}{\sigma_x^2} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})}{\sigma_x^2} = \frac{\overline{xy}-\bar{x}\cdot\bar{y}}{\sigma_x^2} = r_{xy}\cdot\frac{\sigma_y}{\sigma_x}. \quad (2.20)$$

$$a_0 = \bar{y} - r_{xy}\cdot\frac{\sigma_y}{\sigma_x}\cdot\bar{x} = \bar{y} - a_1\cdot\bar{x}. \quad (2.21)$$

Необхідно відмітити, що при лінійній формі зв'язку коефіцієнт кореляції r_{xy} є оцінкою точності апроксимації, тобто адекватності моделі і дорівнює кореляційному відношенню η_{xy}. Після побудови моделі обчислюється також середня відносна похибка апроксимації, %:

$$\varepsilon = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{y_i-\hat{y}_i}{y_i}\right|. \quad (2.22)$$

Середня похибка апроксимації показує в процентах середнє для всіх значення результативного показника відхилення розрахункових значень. Модель можна вважати адекватною, якщо середня похибка апроксимації буде знаходитись у межах 12-15%.

2.3 Умови застосування метода найменших квадратів (1МНК)

Запишемо вибіркову економетричну модель у матричній формі:

$$Y = XA + U, \quad (2.23)$$

де Y – вектор значень залежної змінної;

X – матриця незалежних змінних розміром $n\times m$ (n – кількість спостережень, m – кількість незалежних змінних);

A – вектор оцінок параметрів моделі;

U – вектор залишків (похибок).

Застосування однокрокового метода найменших квадратів (1МНК) для оцінки параметрів моделі передбачає такі умови:

1) математичне сподівання залишків дорівнює нулю, тобто:

$$M(U) = 0; \qquad (2.24)$$

а залишки мають нормальний розподіл.

2) значення U_i вектора залишків U незалежні між собою і мають постійну дисперсію:

$$M(UU') = \delta^2 I, \qquad (2.25)$$

де I – одинична матриця,

3) незалежні змінні моделі не пов'язані із залишками;

$$M(x'U) = 0, \qquad (2.26)$$

4) незалежні змінні моделі утворюють лінійно незалежну систему векторів або, іншими словами, незалежні змінні не повинні бути мультиколінеарними, тобто $|x' \cdot x| \neq 0$;

$$VAR(x'_\kappa \cdot x_j) = 0, \quad \kappa \neq j; \qquad (2.27)$$

$$VAR(x'_\kappa \cdot x_j) = 1, \quad \kappa = j;$$

де x_κ – *к-й* вектор матриці пояснювальних змінних; x_j – *j-й* вектор цієї матриці пояснювальних змінних X, $\kappa = \overline{1,m}$, $j = \overline{1,m}$.

Перша умова є очевидною. Адже коли математичне сподівання залишків не дорівнює нулю, то це означає, що існує систематичний вплив на залежну змінну, а до модельної специфікації не введено всіх основних незалежних змінних. Якщо ця умова не виконується, то має місце помилка специфікації.

В економетричних моделях з вільним членом за рахунок його значення можна скоригувати рівняння так, щоб математичне сподівання залишків дорівнювало нулю. Отже, для таких моделей перша умова практично виконується завжди.

Друга умова передбачає наявність сталої дисперсії залишків. Цю властивість називають **гомоскедастичністю.** Проте вона може використовуватись лише тоді, коли залишки є помилками вимірювання. Якщо залишки акумулюють загальний вплив змінних, які не враховані в моделі, то дисперсія залишків не може бути сталою величиною. Вона змінюється для окремих груп спостережень. У цьому випадку має місце явище **гетероскедастичності**, яке впливає на методи оцінювання параметрів.

Третя умова передбачає незалежність між залишками U_i та пояснювальними змінними x_i, яка порушується насамперед тоді, коли економетрична модель будується на базі одночасових структурних рівнянь або має лагові змінні. Тоді для оцінювання параметрів моделі використовується, як правило, дво- або трикровий метод найменших квадратів.

Четверта умова означає, що всі пояснювальні змінні (x_j), які входять до економетричної моделі, мають бути незалежними між собою. Проте очевидно, що в економіці дуже важко сформувати такий масив незалежних пояснювальних змінних, які були б зовсім не пов'язані між собою. Тоді кожного разу необхідно з`ясувати, чи не впливатиме залежність пояснювальних змінних на оцінку параметрів моделі.

Це явище називаються мультиколінеарністю змінних. Воно призводить до ненадійності оцінки параметрів моделі, робить їх чутливими до вибраної специфікації моделі та до конкретного вибору даних. Знижується рівень довіри до результатів верифікації моделей з допомогою однокрокового метода найменших квадратів (1МНК). Отже, це явище з усіх точок зору є дуже небажаним. Але воно досить поширене. Існують методи для виявлення мультиколінеарності і способи її врахування з допомогою специфікації моделі чи спеціальних методів оцінювання параметрів (методу Ейткена).

2.4 Специфікація моделі

Економетрична модель – це функція чи система функцій, що описує кореляційно-регресійний зв`язок між економетричними показниками, один чи кілька з яких є залежною зміною, інші – незалежними.

Незалежні змінні моделі називаються пояснюючими, наперед заданими змінними. Залежні змінні називаються пояснювальними змінними.

Специфікація моделі – це аналітична форма економетричної моделі. Специфікація моделі може здійснюватись за допомогою різних видів функцій, які можуть бути застосовані для вивчення взаємозв'язків, а саме:

1) лінійної функції: $y = a_0 + \sum_{j=1}^{m} a_j x_j + u$, (2.28)

2) степеневої функції: $y = a_0 \prod_{j=1}^{m} a_j x_j + u$ (2.29)

та ін.

При цьому в процесі такого дослідження можна кілька разів повертатись до етапу специфікації моделі, уточнюючи перелік незалежних змінних та вид функції, що застосовується.

Специфікація моделі передбачає також добір факторів (чинників) для економічного дослідження. Адже коли вид функції та її складові не відповідають реальним залежностям, то йдеться про помилки специфікації.

Помилки специфікації моделі можуть бути трьох видів:

1) ігнорування істотної пояснюючої змінної при побудові економетричної моделі;

2) введення до моделі незалежної змінної, яка не стосується вимірюваного зв'язку;

3) використання невідповідних математичних форм залежності.

Перша з цих помилок призводить до зміщення оцінок, причому зміщення буде тим більшим, чим більша кореляція між введеними та не введеними до моделі змінними, а напрям залежить від знака оцінок параметрів при введених змінних і від характеру кореляції між введеними та не введеними змінними. Оцінки параметрів також будуть зміщеними (у такому разі вони вищі), тому можуть бути хибні висновки щодо значень параметрів генеральної сукупності.

Друга помилка специфікації. Якщо до моделі вводиться змінна, яка неістотно впливає на залежну змінну, то на відміну від першої помилки специфікації оцінки параметрів моделі не будуть зміщеними. Причому, за допомогою звичайних процедур можна отримати також не зміщені оцінки дисперсій цих параметрів. Але це не означає, що економетричну модель можна беззастережно розширювати за рахунок «неістотних» змінних. По-перше, існує ненульова ймовірність того, що в результаті використання вибіркових даних змінна, яка зовсім не стосується моделі, покаже істотний зв'язок із залежною змінною. А це означає, що кількісний зв'язок між змінними буде виміряний неправильно.

Третя помилка специфікації. Припускається, що залежна змінна є лінійною функцією від деякої пояснювальної змінної, тоді як насправді тут краще підійшла б квадратична, кубічна чи якась поліноміальна залежність вищого порядку. У цьому разі наслідки такі самі, як і при першій помилці, тобто оцінки параметрів моделі матимуть зміщення. Питання про вибір найкращої форми залежності має базуватись на перевірці ступеня узгодженості виду функції з вихідними даними спостережень.

Адекватність побудованої моделі можна встановити, аналізуючи залишки моделі. Вони обчислюються як різниці між фактичними значеннями залежної змінної і обчисленими за моделлю. Перевірити розподіл залишків на невипадковість їх характеру можна за критерієм Дарбіна-Уотсона. Тоді перевірка моделі на існування автокореляції першого порядку аналогічна перевірці того, наскільки вдало вибрано форму економетричної моделі.

Питання для самоконтролю:

1. Які існують види причинної залежності?

2. Кореляційний зв'язок, класифікація за ознаками.

3 Поняття регресії, регресійна залежність.

4. Що являє собою кореляційне поле? Для чого воно будується?

5. Узагальнена та вибіркова економетричні моделі.

6. Поясніть причини присутності в регресійних моделях випадкової змінної (відхилення).

7. Інтерпретація метода найменших квадратів та оцінка параметрів однофакторної лінійної моделі.

8. Оцінка щільності кореляційного зв'язку. Коефіцієнт кореляції.

9. Варіація та коваріація, їх зміст та застосування. Коефіцієнт детермінації.

10. Види та правило додавання дисперсій.

11. Умови застосування метода найменших квадратів.

12. Що таке специфікація моделі?

Тема 3

ОДНОФАКТОРНІ НЕЛІНІЙНІ ЕКОНОМЕТРИЧНІ МОДЕЛІ

3.1 Побудова однофакторної економетричної моделі

Зв'язки між економічними явищами часто бувають нелінійними. Наприклад, зв'язок між собівартістю зерна і врожайністю має вид гіперболи, зв'язок між фондовіддачею основних засобів і коефіцієнтом використання виробничих потужностей являє собою параболу, зв'язок між витратами на харчування (в перерахунку на одну споживчу одиницю) і кількістю дітей у сім'ї виражається степеневим рівнянням.

В економетричних дослідженнях досить часто використовуються такі нелінійні моделі:

гіпербола: $y = a_0 + \frac{a_1}{x} + u$;

показникова функція: $y = a_0 \cdot a_1^{x} + u$;

напівлогарифмічна функція: $y = a_0 + a_1 \cdot \ln x + u$;

логарифмічна функція: $\ln y = a_0 + a_0 \cdot x + u$;

парабола: $y = a_0 + a_1 \cdot \sqrt{x} + u$, $y = a_0 + a_1 \cdot x^2 + u$, $y = a_0 + a_1 \cdot x + a_2 \cdot x^2 + u$;

степенева функція: $y = a_0 \cdot x^{a_1} + u$.

Для вивчення нелінійних зв'язків необхідно перетворити криву регресії в функцію, яка лінійна відносно параметрів a_0 і a_1, тобто провести лінеаризацію рівняння регресії. Тобто, у випадках, коли функція є нелінійною (але не поліномом), її зводять до лінійної, наприклад шляхом логарифмування та наступної заміни нелінійних змінних на лінійні (заміни в залежності від виду функції).

Метод приведення нелінійної функції до лінійного виду називають **лінеаризацією** рівняння регресії.

Розглянемо методику лінеаризації наведених вище функцій:

1. Якщо $y = a_0 + \frac{a_1}{x}$, то заміною аргументу $x' = \frac{1}{x}$ зводимо рівняння до виду $y = a_0 + a_1 \cdot x'$.

2. Якщо $y = a_0 \cdot a_1^x$, то логарифмуючи це рівняння, отримуємо $\ln y = \ln a_0 + x \cdot \ln a_1$. Заміною змінних $y' = \ln y, a_0' = \ln a_0$ та $a_1' = \ln a_1$ отримуємо $y' = a_0' + a_1' \cdot x$.

3. У випадку $y = a_0 + a_1 \cdot \ln x$, приймаючи $x' = \ln x$, отримуємо лінеаризовану функцію $y = a_0 + a_1 \cdot x'$.

4. Якщо $y = a_0 + a_1 \cdot x^2$ чи $y = a_0 + a_1 \cdot \sqrt{x}$, то заміною $x' = \sqrt{x}$ або $x' = x^2$ зведемо їх до лінійної функції.

5. Для функції $\ln y = a_0 + a_1 \cdot x$ робимо заміну $y' = \ln y$ і отримуємо лінійну економетричну модель $y' = a_0 + a_1 \cdot x$.

6. Степеневу функцію виду $y = a_0 x^{a_1}$ можна привести до лінійного виду шляхом логарифмування лівої і правої частини рівняння:

$$\ln y = \ln a_0 + a_1 \ln x.$$

Зробивши заміну нелінійних величин на лінійні $y' = \ln y$, $a_0' = \ln a_0$, $x' = \ln x$, отримаємо лінійну модель: $y' = a_0' + a_1 \cdot x'$.

Параметри цієї моделі можна знайти, розв'язавши лінеаризовану систему нормальних рівнянь:

$$\begin{cases} n \cdot \ln a_0 + a_1 \cdot \sum_{i=1}^{n} \ln x_i = \sum_{i=1}^{n} \ln y_i \\ \ln a_0 \cdot \sum_{i=1}^{n} \ln x_i + a_1 \cdot \sum_{i=1}^{n} (\ln x_i)^2 = \sum_{i=1}^{n} \ln y_i \cdot \ln x_i \end{cases} \quad (3.1)$$

$$a_0' = \exp(\ln a_0).$$

Такими ж методами можна лінеаризувати інші нелінійні функції. Графіки та системи рівнянь найбільш уживаних функцій зростання наведені у додатку.

Алгоритм побудови однофакторної економетричної моделі

1. Побудова поля кореляції та емпіричної лінії регресії. Вибір форми зв'язку. Лінеаризація моделі у випадку її нелінійності.

2. Розрахунок параметрів лінійної або лінеаризованої моделі a_0 та a_1 шляхом розв'язання системи нормальних рівнянь.

3. Розрахунок середніх дисперсій та середньоквадратичних відхилень для змінних x та y і визначення тісноти зв'язку між ними.

Для оцінки тісноти лінійного зв'язку обчислюють коефіцієнт парної кореляції (формули 2.5, 2.6, 2.7), а для оцінки нелінійного кореляційного зв'язку використовують так зване кореляційне відношення, яке обчислюється за формулою:

$$\eta_{xy} = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad , \tag{3.2}$$

де y_i – фактичні (емпіричні, поточні) значення результативного показника; $\hat{y}_i$ – теоретичні (розрахункові) значення результативного показника; $\bar{y}$ – середнє значення y.

Величина кореляційного відношення знаходиться у межах $0 \le \eta_{xy} \le 1$. У випадку функціонального зв'язку $\eta_{xy} = 1$. Якщо кореляційний зв'язок відсутній, то $\eta_{xy} = 0$. Однак, значення $\eta_{xy} = 0$ не свідчить про відсутність причинної залежності між економічними явищами.

Теоретично показники η_{yx} і η_{xy} рівнозначні, але $\eta_{yx} = \eta_{xy}$ лише у випадку лінійної залежності. У цьому випадку $\eta_{yx} = \eta_{xy} = r_{yx} = r_{xy}$

В усіх інших випадках, тобто при нелінійній залежності $\eta_{yx} \ne \eta_{xy}$, коефіцієнт кореляції по абсолютній величині менший за найменше із цих двох кореляційних відношень.

У практичній діяльності нас зазвичай цікавить тільки одне із кореляційних відношень, а саме: кореляційне відношення між фактором, прийнятим за функцію, і фактором, прийнятим за аргумент.

На співвідношеннях між η_{xy} і r_{xy} можемо побудувати критерій криволінійності зв'язку між змінними y і x. Спрощена формула такого ряду критерію має такий вигляд:

$$K = \frac{\sqrt{n}}{0.67449} \cdot \frac{1}{2}\sqrt{\eta^2 - r^2} \ , \tag{3.3}$$

Якщо значення обчисленого K>2.5, то кореляційний зв'язок можна вважати криволінійним.

4. Оцінка значущості коефіцієнта парної кореляції (кореляційного відношення) та розрахунок його довірчих меж. Довірчі межі коефіцієнта парної кореляції (кореляційного відношення) визначаються через оцінку дисперсії:

$$\sigma_r = \frac{1-r^2}{\sqrt{n-m-1}} = \sigma_r = \frac{1-r^2}{\sqrt{k}}; \qquad \Delta r = \pm t_{\alpha,k} \cdot \sigma_r, \tag{3.4}$$

де σ_r – дисперсія коефіцієнта парної кореляції,

$t_{\alpha,k}$ – нормовані (табличні) значення функції Стьюдента, визначені по заданому рівню значущості α та числу ступенів свободи k.

Таким чином, з довірчою ймовірністю $p=1-\alpha$ оцінка показника тісноти зв'язку для генеральної сукупності буде знаходитись у межах:

$$\rho = r \pm t_{\alpha,k} \cdot \sigma_r$$

або

$$r - t_{\alpha,k} \cdot \sigma_r \le \rho \le r + t_{\alpha,k} \cdot \sigma_r.$$

Оцінку значущості коефіцієнта кореляції (кореляційного відношення) виконують за допомогою t – критерія Стьюдента, для чого розраховують фактичне значення t – критерія за формулою:

$$t_r = \frac{|r|}{\sigma_r}, \tag{3.5}$$

Розраховане за формулою 3.5 значення t_r порівнюють з його критичним значенням при прийнятому рівні значущості α та числі ступенів свободи $k=n-m-1$. В соціально-економічних дослідженнях рівень значущості α звичайно приймають рівним 0,05; тоді довірча ймовірність $p=1-\alpha=0{,}95$. Якщо розрахункове значення t_r більше критичного $t_{\alpha,k}$, то параметр вважається значущим, тобто відхиляється гіпотеза про те, що параметр в дійсності дорівнює нулю (для генеральної сукупності) і лише в силу деяких випадкових обставин він дорівнює величині, що перевіряється.

5. Оцінка значущості довірчих границь параметрів a_0 та a_1 лінійної (лінеаризованої) функції $y=a_0+a_1 \cdot x$ здійснюється за формулами:

$$\sigma_{a_0}=\sigma_u\cdot\sqrt{\frac{\sum_{i=1}^{n}x_i^2}{n\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2}},\ \sigma_{a_1}=\frac{\sigma_u}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}},\tag{3.6}$$

$$\sigma_u=\sqrt{\frac{\sum_{i=1}^{n}u_i^2}{n-2}}=\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-2}}.\tag{3.7}$$

Тоді

$$\tilde{a}_0=a_0\pm t_{\alpha.k}\cdot\sigma_{a_0};\quad \tilde{a}_1=a_1\pm t_{\alpha.k}\cdot\sigma_{a_1},\tag{3.8}$$

де $t_{\alpha.k}$ – табличне значення t – критерія Стьюдента при рівні значущості α та $k=n-m-1$ степенях свободи.

Рівняння регресії для генеральної сукупності можна записати у вигляді:

$$\tilde{y}=\left(a_0\pm t_{\alpha.k}\cdot\sigma_{a_0}\right)+\left(a_1\pm t_{\alpha.k}\cdot\sigma_{a_1}\right)\cdot x.\tag{3.9}$$

Значущість параметрів a_0 та a_1 визначають за t – критерієм Стьюдента:

$$t_{a_0}=\frac{|a_0|}{\sigma_{a_0}};\qquad t_{a_1}=\frac{|a_1|}{\sigma_{a_1}}.\tag{3.10}$$

Якщо $t_{a_0}, t_{a_1}\geq t_{\alpha.k}$, то параметри a_0 та a_1 генеральної сукупності значущі.

Адекватність економетричної моделі статистичним даним оцінюється за F – критерієм Фішера. Під адекватністю економетричної моделі розуміють відповідність моделі статистичним даним, по яких вона побудована.

Для цього знаходять розрахункове (фактичне) значення F – критерія Фішера за формулою:

$$F_p=\frac{\delta_y^2}{\sigma_u^2},\tag{3.11}$$

де $\delta_y^2=\frac{\sum_{i=1}^{n}\left(\hat{y}_i-\bar{y}\right)^2}{m}$, $\sigma_u^2=\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-m-1}$

$$\sigma_u^2=\frac{1}{n-m-1}\cdot\sum_{i=1}^{n}u_i^2=\frac{1}{n-m-1}\cdot\sum_{i=1}^{n}(y_i-\hat{y}_i)^2,\ \delta_y^2=\frac{1}{m-1}\cdot\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$$

n – кількість дослідів(спостережень); m – кількість факторів, включених у модель.

На практиці при розрахунку F – статистики часто суму $\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$ замінюють різницею (див. 2.12).

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

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

Тоді формула розрахунку фактичного значення F – критерія Фішера буде мати вигляд:

$$F_p = \frac{\sigma_y^2 - \sigma_u^2}{\sigma_u^2}. \tag{3.12}$$

Знаходимо табличне значення F_{α,k_1,k_2} Отримане розрахункове значення порівнюється з табличним. Якщо $F_p > F_{\alpha,k_1,k_2}$, то з ймовірністю $p=1-\alpha$ можна вважати, що розглянута економетрична модель адекватна експериментальним даним, у протилежному випадку з ймовірністю p розглянуту парну регресію не можна вважати адекватною.

6. Оцінка довірчих границь базисних середніх значень $\hat{y}_i$ та довірчих зон рівняння регресії. Для цього визначаємо довірчі границі для кожного базисного значення при заданому рівні значимості за формулою:

$$\Delta y_i = \frac{t_{\alpha,k} \cdot \sigma_u}{\sqrt{n}} \cdot \sqrt{1 + \frac{(x_i - \bar{x})^2}{\sigma_x^2}} . \tag{3.13}$$

Тоді оцінка кожного базисного середнього складе:

$$\widetilde{y}_i = \hat{y}_i \pm \Delta y_i . \tag{3.14}$$

З'єднавши на графіку плавною лінією усі значення $\max\widetilde{y}_i = \hat{y}_i + \Delta y_i$ та відповідно $\min\widetilde{y}_i = \hat{y}_i - \Delta y_i$, отримаємо так звану довірчу зону рівняння регресії.

7. Визначення середнього значення прогнозу, його довірчої границі та зони.

Середнє значення прогнозу показника y_p визначаємо методом математичної екстраполяції шляхом підстановки в економетричну модель замість x його відносного прогнозного значення x_p. Такий метод прогнозування називається *методом математичної екстраполяції*, а прогнозне значення результативного показника – *точковим прогнозом*.

Для знаходження інтервального прогнозу розраховують його довірчі межі за формулою:

$$\Delta y_p = t_{a,k} \cdot \frac{\sigma_u}{\sqrt{n}} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sigma_x^2}}, \tag{3.15}$$

Похибку прогнозу оцінюють за формулою:

$$\Delta y_p = \pm t_{\alpha,k} \cdot \frac{\sigma_u}{\sqrt{n}} \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sigma_x^2}}. \tag{3.16}$$

Прогноз («істинне» значення прогнозу) буде знаходитись у межах:

$$\tilde{\Delta}_p = \hat{y}_p \pm \Delta y_p, \tag{3.17}$$

або

$$\hat{y}_i - \Delta\hat{y}_i \le \tilde{y}_i \le \hat{y}_i + \Delta\hat{y}_i.$$

При щільній кореляції (більше 0,75) прогнозні значення можуть бути прийняті у бізнес-плануванні результативного показника на найближчу перспективу. При цьому слід пам'ятати, що чим дальше від базисних показників взято прогнозне значення фактора x_p, тим менш надійний прогноз, тим більша імовірність значного відхилення від середнього розрахункового значення, тобто довірча зона прогнозування розширюється.

3.2 Оцінка значимості параметрів однофакторної економетричної моделі

Кореляційний аналіз базується на досить великій сукупності вхідних даних (одиниць спостереження). Однак на практиці не завжди можна охопити всі без виключення аналогічні, однорідні в якісному відношенні одиниці. При дослідженні залежності між економічними показниками обмежуються лише частиною одиниць спостережень, відбираючи найбільш типові. Часто для цього використовують спосіб простого випадкового відбору.

Множина всіх якісно однорідних одиниць, які володіють подібними властивостями, називається *генеральною сукупністю*, а об'єкти спостереження, які представляють частину генеральної сукупності й відібрані для аналізу, утворюють *вибіркову сукупність*.

Припустимо, що в ході дослідження зв'язку між показниками на основі вибіркових даних розраховані коефіцієнти рівняння регресії та кореляції. Оскільки в умовах експерименту відсутній функціональний зв'язок між показниками x та y, тобто $y = f(x) + u$, де u – деяка випадкова величина, то параметри рівняння регресії і коефіцієнт кореляції, розраховані по даних вибіркової і генеральної сукупностей, не будуть співпадати. Параметри моделі, розраховані по даних вибіркової сукупності, мають деяку «помилку» в порівнянні з відповідними параметрами моделі, розрахованими по даних генеральної сукупності. Тому обчислення ймовірнісних помилок і значимості параметрів моделі генеральної сукупності по розрахунку даних вибірки є необхідною складовою частиною оцінки економіко-статистичної моделі.

Приклад 3.1. Знайти оцінки параметрів лінії регресії на основі даних показника y та фактора x (таблиця 3.1), якщо статистична залежність між фактором x і показником y має вигляд $y = a_0 + \frac{a_1}{x}$.

Використовуючи критерій Фішера, з надійністю $p = 0{,}95$ оцінити адекватність прийнятої моделі статистичним даним. Якщо із заданою надійністю прийнята математична модель адекватна експериментальним даним, то знайти:

- з надійністю $p = 0{,}95$ довірчу зону базисних середніх;
- точкову оцінку прогнозу для $x_p = 110$;
- з надійністю $p = 0{,}95$ інтервальну оцінку прогнозу;
- оцінку індексу кореляції.

Побудувати графіки:

- фактичних даних;
- лінії регресії та довірчі зони.

Розв'язок. Будуємо поле кореляції, емпіричну лінію регресії та визначаємо форму зв'язку між змінними x та y (рисунок 3.1).

Зовнішній вигляд емпіричної лінії регресії нагадує графік гіперболи

$$y = a_0 + \frac{a_1}{x} + u$$

Невідомі параметри економетричної моделі a_0 і a_1 знаходимо із системи нормальних рівнянь:

$$\begin{cases} na_0 + a_1 \cdot \sum_{i=1}^{n} \frac{1}{x_i} = \sum_{i=1}^{n} y_i \\ a_0 \sum_{i=1}^{n} \frac{1}{x_i} + a_1 \sum_{i=1}^{n} \frac{1}{x_i^2} = \sum_{i=1}^{n} \frac{y_i}{x_i} \end{cases}$$

Для цього в таблиці 3.2 накопичуємо суми:

$$S_x = \sum_{i=1}^{n} \frac{1}{x_i} = 0{,}221,\ S_{xx} = \sum_{i=1}^{n} \frac{1}{x_i^2} = 0{,}007,\ S_{xy} = \sum_{i=1}^{n} \frac{y_i}{x_i} = 1{,}415,\ S_y = \sum_{i=1}^{n} y_i = 51{,}6$$

і підставляємо їх у систему нормальних рівнянь.

Маємо:

$$\begin{cases} 10a_0 + 0{,}221a_1 = 51{,}6 \\ 0{,}221a_0 + 0{,}007a_1 = 1{,}415 \end{cases}$$

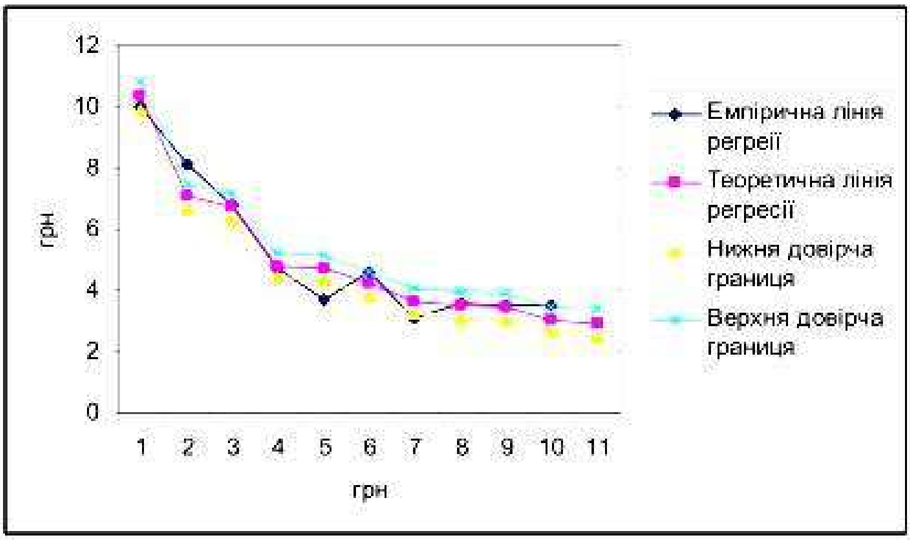


Рисунок 3.1 – Емпірична, теоретична лінії регресії та графіки довірчих границь прогнозу

Таблиця 3.1 – Вхідні дані та результати розрахунків

сп	x_i	y_i	$\frac{1}{x_i}$	$\left(\frac{1}{x_i}\right)^2$	$\frac{y_i}{x_i}$	$\hat{y}_i$	$(y_i-\hat{y}_i)^2$	$(y_i-\bar{y})^2$	$(\hat{y}_i-\bar{y})^2$	$\left(\frac{1}{x_i}-\bar{x}'\right)^2$	$\Delta\hat{y}_i$	Довірчі границі	
												$\hat{y}_i-\Delta\hat{y}_i$	$\hat{y}_i+\Delta\hat{y}_i$
1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	80	3,6	0.013	0,0002	0.045	3,515	0,007	2.434	2.706	0.00009	0,406	3,109	3.921
2	75	3,1	0,013	0,0002	0,041	3,657	0,311	4,244	2,258	0,00008	0,404	3,254	4,061
3	102	3,5	0,010	0,0001	0,034	3,055	0,198	2,756	4,433	0,00015	0,413	2,642	3,467
4	82	3,5	0,012	0,0001	0,043	3,463	0,001	2,756	2,880	0,00010	0,406	3,057	3,869
5	50	4,7	0,020	0,0004	0,094	4,796	0,009	0,212	0,133	0,00000	0,395	4,401	5,191
6	30	8,1	0.033	0,0011	0.270	7,073	1,055	8,644	3,659	0.00013	0,410	6,663	7.483
7	60	4,6	0.017	0,0003	0.077	4,227	0,139	0.314	0.871	0.00003	0,398	3,829	4.625
8	32	6,8	0.031	0,0010	0.213	6,717	0,007	2,690	2,425	0.00008	0,405	6,313	7.122
9	19	10	0,053	0,0028	0,526	10,37	0,136	23,426	27.131	0,00093	0,496	9,873	10,865
10	51	3,7	0,020	0,0004	0,073	4,729	1,059	2,132	0,186	0,00001	0,395	4,334	5,124
Σ	581	51,6	0,221	0,0065	1,415		2,923	49,604	46,681	0,0016			
П			0,009			2,933				0,00017	0,433	2,500	3,366

З останньої системи знаходимо: $a_0 = 1,38$, $a_1 = 170,78$.

Таким чином, економетрична модель має вигляд:

$$y = 1,38 + \frac{170,78}{x} + u.$$

Розраховуємо теоретичні значення $\hat{y}_i$ та прогнозне y_p значення результативного показника і будуємо теоретичну лінію регресії (рисунок 3.1).

$$\hat{y}_1 = 1,38 + \frac{170,78}{80} = 3,515,$$

$$\hat{y}_2 = 1,38 + \frac{170,78}{75} = 3,657, \text{ і т.д.}$$

$$y_p = 1,38 + \frac{170,78}{110} = 2,933.$$

Тісноту зв'язку між ознаками x та y оцінюємо кореляційним відношенням:

$$\eta_{xy} = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} = \sqrt{1 - \frac{\sum_{i=1}^{n}u_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}}.$$

Для розрахунку η_{xy} у таблиці 3.1 знайдемо суми $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ та $\sum_{i=1}^{n}(y_i - \bar{y})^2$.

Тоді $\eta_{xy}=\sqrt{1-\frac{2,923}{49,604}}=0,97$.

Оцінимо значущість кореляційного відношення за t – критерієм Стьюдента. Маємо:

$$\sigma_{\eta_{xy}}=\frac{1-\eta_{xy}^2}{\sqrt{n-2}}=\frac{1-0,97^2}{\sqrt{8}}=0,021;\quad t_{\eta_{xy}}=\frac{\eta_{xy}}{\sigma_{\eta_{xy}}}=\frac{0,97}{0,021}=46,19.$$

Табличне значення $t_{\alpha,k}$ критерія Стьюдента для рівня значущості $\alpha=0,05$ і степенів свободи $k=8$ становить $t_{\alpha,k}=2,31$. Оскільки $t_{\eta_{xy}}>t_{\alpha,k}$, то з ймовірністю $p=1-0,05=0,95$ можна стверджувати, що кореляційне відношення значуще.

Оцінимо значущість параметрів економетричної моделі a_0 та a_1. Спочатку обчислимо σ_u^2, σ_{a_0}, σ_{a_1}. Маємо:

$$\sigma_u^2=\frac{1}{n}\cdot\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=0,292;\qquad \sigma_{a_1}=\sqrt{\frac{\sigma_u^2}{\sum_{i=1}^{n}\left(\frac{1}{x_i}-\bar{x}'\right)^2}}=\sqrt{\frac{0,292}{0,0016}}=13,51;$$

$$\sigma_{a_0}=\sqrt{\sigma_u^2\cdot\frac{\sum_{i=1}^{n}\frac{1}{x_i^2}}{n\cdot\sum_{i=1}^{n}\left(\frac{1}{x_i}-\bar{x}'\right)^2}}=\sqrt{0,292\cdot\frac{0,007}{0,016}}=0,344.$$

Тоді:

$$t_{a_0}=\frac{|a_0|}{\sigma_{a_0}}=\frac{1,38}{0,344}=4,012,\qquad t_{a_1}=\frac{|a_1|}{\sigma_{a_1}}=\frac{170,78}{13,51}=12,64.$$

Оскільки $t_{a_0}>t_{\alpha,k}$, $t_{a_1}>t_{\alpha,k}$, то параметри a_0 і a_1 економетричної моделі значущі.

Проведемо оцінку адекватності моделі за F – критерієм Фішера. Для цього обчислимо:

$$F_p=\frac{\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2/(m-1)}{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2/(n-m-1)}=\frac{46,68/(2-1)}{2,923/(10-2-1)}=111,789.$$

Табличне значення критерія Фішера для рівня значущості $\alpha=0,05$ і степенів свободи $k_1=1$ і $k_2=7$ становить $F_{\alpha,k_1,k_2}=5,59$. Оскільки

$F_p > F_{\alpha,k_1,k_2}$, то економетрична модель адекватно описує економічне явище.

Проведемо оцінку довірчих меж базисних середніх та прогнозу за формулою:

$$\Delta \hat{y}_i = \frac{t_{\alpha,k} \cdot \sigma_u}{\sqrt{n}} \sqrt{1 + \frac{\left(\frac{1}{x} + \bar{x}'\right)}{\sigma_x^2}}.$$

Маємо:

$$\Delta \hat{y}_1 = \frac{2.31 \cdot 0.54}{\sqrt{10}} \sqrt{1 + \frac{0.00009}{0.0016}} = 0{,}406;$$

$$\Delta \hat{y}_2 = \frac{2.31 \cdot 0.54}{\sqrt{10}} \sqrt{1 + \frac{0.00008}{0.0016}} = 0.404; \text{ і т.д.}$$

$$\Delta y_p = 2.31 \cdot \frac{0.54}{\sqrt{10}} \sqrt{1 + \frac{1}{10} + \frac{(1{,}110 - 0{,}022)^2}{0{,}0016}} = 0{,}433.$$

Тоді $\hat{y}_i - \Delta\hat{y}_i \le \tilde{y}_i \le \hat{y}_i + \Delta\hat{y}_i$.

Будуємо довірчі межі базисних середніх та прогнозу (рисунок 3.1).

Приклад 3.2. По статистичних даних таблиці 3.2 побудувати та дослідити економетричну модель: $y = a_0 + a_1 \cdot \ln x + u$.

Розв'язок. Економетрична модель нагадує графік напівлогарифмічної кривої $y = a_0 + a_1 \cdot \ln x + u$.

Розраховуємо параметри моделі a_0 та a_1. Записуємо систему нормальних рівнянь:

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum_{i=1}^{n} \ln x_i = \sum_{i=1}^{n} y_i \\ a_0 \cdot \sum_{i=1}^{n} \ln x_i + a_1 \cdot \sum_{i=1}^{n} \ln^2 x_i = \sum_{i=1}^{n} y_i \cdot \ln x_i \end{cases}$$

Із таблиці 3.2 у систему нормальних рівнянь підставляємо відповідні суми:

$$\begin{cases} 10 \cdot a_0 + 17.11 \cdot a_1 = 158.5 \\ 17.11 \cdot a_0 + 30.81 \cdot a_1 = 293.6 \end{cases}$$

звідки $a_0 = -9{,}094$, $a_1 = 14{,}579$.

Запишемо економетричну модель: $y = -9{,}094 + 14{,}579 \ln x + u$.

Таблиця 3.2 – Вхідні дані та результати розрахунків

x_i	$\ln x_i$	$\ln^2 x_i$	$y_i \ln x_i$	$\hat{y}_i$	$(y_i - \hat{y}_i)^2$	$(y_i - \bar{y})^2$	$\exp(\ln x_i - \bar{x}')^2$	Δy_i	Довірчі межі $\hat{y}_i - \Delta y_i$	$\hat{y}_i + \Delta y_i$
2,7	0,99	0,99	6,5	5,387	1,23871	87,423	0,5150	0,798	4,59	6,19
3,4	1,22	1,50	9,9	8,748	0,41980	60,063	0,2373	0,610	8,14	9,36
4,1	1,41	1,99	15,4	11,477	0,33333	24,503	0,0900	0,482	11,00	11,96
4,7	1,55	2,39	20,4	13,469	0,07211	7,023	0,0267	0,415	13,05	13,88
5,5	1,70	2,91	27,1	15,760	0,01955	0,002	0,00001	0,383	15,38	16,14
6,3	1,84	3,39	32,0	17,740	0,11567	2,402	0,0168	0,403	17,34	18,14
6,9	1,93	3,73	36,3	19,066	0,07097	8,702	0,0487	0,439	18,63	19,51
7,7	2,04	4,17	43,1	20,666	0,18857	27,562	0,1091	0,501	20,17	21,17
8,8	2,17	4,73	48,9	22,613	0,01267	44,222	0,2152	0,593	22,02	23,21
9,4	2,24	5,02	54,0	23,574	0,27649	68,062	0,2807	0,643	22,93	24,22
59,5	17,11	30,81	293,6		2,74785	329,96	1,5394			
10,2	2,32			24,765			0,3739	0,807	23,96	25,57

Розраховуємо теоретичні та прогнозні значення результативного показника $\hat{y}_i$.

Тісноту зв'язку між змінними y та x оцінюємо кореляційним відношенням:

$$\eta_{yx} = \sqrt{1 - \frac{2{,}74785}{329{,}96}} = 0{,}996,$$

$$\sigma_\eta = \frac{1 - 0{,}996^2}{\sqrt{10}} = 0{,}0025248, \qquad t_\eta = \frac{\eta}{\sigma_\eta} = \frac{0{,}996}{0{,}0025248} = 394{,}49.$$

Оскільки $t_\eta > t_{табл.}$, то кореляційне відношення значуще.

Оцінка значущості довірчих меж параметрів економетричної моделі:

$$t_{a_0} = \frac{|a_0|}{\sigma_{a_0}} = \frac{9{,}094}{0{,}7416} = 12{,}262;$$

$$\sigma_{a_0} = \sigma_u \cdot \sqrt{\frac{\sum_{i=1}^{n} \ln^2 x_i}{n \cdot \sum_{i=1}^{n} (\ln x_i - x')^2}} = \sqrt{\frac{2{,}74785}{10}} \cdot \sqrt{\frac{30{,}81}{10 \cdot 1{,}5394}} = 0{,}7416;$$

$$\bar{x}' = \frac{1}{n} \cdot \sum_{i=1}^{n} \ln x_i = \frac{17{,}11}{10} = 1{,}711; \quad t_{a_1} = \frac{|a_1|}{\sigma_{a_1}} = \frac{14{,}579}{0{,}4225} = 18{,}09;$$

$$\sigma_{a_1} = \frac{\sigma_u}{\sqrt{\sum_{i=1}^{n}(\ln x_i - \bar{x}')^2}} = \frac{0,5242}{\sqrt{1,5394}} = 0,4225 \,.$$

Оскільки $t_{a_0}, t_{a_1} > t_{табл}$, то параметри моделі значущі.

Оцінюємо адекватність моделі за F – критерієм Фішера:

$$F_p = \frac{\delta_y^2}{\sigma_u^2} = \frac{327,2125}{0,34350} = 952,63 \,, \quad \sigma_u^2 = \frac{2,74785}{10-1-1} = 0,34350 \,,$$

$$\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 329,96 - 2,74785 = 327,21215 \,,$$

$$\delta_y^2 = \frac{1}{m-1}\cdot\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = 327,21215 \,;$$

$F_{табл.} = 5,32$. Оскільки $F_p > F_{табл.}$, то з ймовірністю $p = 0,95$ можна стверджувати, що модель адекватна статистичним даним.

Оцінка довірчих меж базисних середніх та прогнозу:

$$\Delta y_1 = \frac{t_{\alpha,k}\sqrt{\sigma_u^2}}{\sqrt{n}}\cdot\sqrt{1+\frac{(\ln x_1 - \bar{x}')^2}{\sigma_x^2}} = \frac{2,31\cdot\sqrt{0,3435}}{\sqrt{10}}\cdot\sqrt{1+\frac{0,5150}{1,5394}} \,.$$

Приклад 3.3. По статистичних даних таблиці 3.3 побудувати та дослідити економетричну модель виду:

$$y = a_0 \cdot x^{a_1} + u \,.$$

Розв'язок. Прологарифмувавши економетричну модель, отримаємо: $\ln y = \ln a_0 + a_1 \cdot \ln x$.

Проміжні розрахунки ($\sum \ln y_i$, $\sum \ln x_i$, $\sum (\ln x_i)^2$, $\sum \ln x_i \cdot \ln y_i$.) представлені в таблиці 3.3.

Запишемо систему нормальних рівнянь:

$$\begin{cases} n\cdot\ln a_0 + a_1\cdot\sum_{i=1}^{n}\ln x_i = \sum_{i=1}^{n}\ln y_i \\ \ln a_0\cdot\sum_{i=1}^{n}\ln x_i + a_1\cdot\sum_{i=1}^{n}(\ln x_i)^2 = \sum_{i=1}^{n}\ln y_i\cdot\ln x_i \end{cases}$$

Після підстановки обчислених значень сум система матиме вигляд:

$$\begin{cases} 10\cdot\ln a_0 + 17,11\cdot a_1 = 26,83 \\ 17,11\cdot\ln a_0 + 30,81\cdot a_1 = 47,49 \end{cases}$$

звідки

$$\ln a_0 = 0.9173\,,\ a_1 = 1{,}03198\,,\ a_0 = \exp(\ln a_0) = 2.50248\,.$$

Таблиця 3.3 – Вхідні дані та результати розрахунків

i	y_i	x_i	$\ln x_i$	$\ln y_i$	$\ln y_i \cdot \ln x_i$	$\ln^2 x_i$	$\hat{y}_i$	$(y_i - \hat{y}_i)^2$	$(y_i - \bar{y})^2$	$\exp(\ln y_i - \bar{y})^2$	Δy_i	Довірчі межі $\hat{y}_i - \Delta y_i$	$\hat{y}_i + \Delta y_i$
1	6,49	2,7	0,99	1,87	1,85	0,99	7,02	0,2704	87,423	0,5150	1,26	5,76	8,28
2	8,08	3,4	1,22	2,09	2,55	1,50	8,81	0,5041	60,063	0,2373	0,96	7,85	9,77
3	10,911	4,1	1,41	2,39	3,37	1,99	10,72	0,0324	24,503	0,0900	0,76	9,96	11,48
4	13,2	4,7	1,55	2,58	4,00	2,39	12,39	0,6561	7,023	0,0267	0,65	11,74	13,04
5	16,0	5,5	1,70	2,77	4,71	2,91	14,46	2,0736	0,002	0,00001	0,60	13,86	15,06
6	17,5	6,3	1,84	2,86	5,26	3,39	16,71	0,4761	2,402	0,0168	0,64	16,07	17,35
7	18,7	6,9	1,93	2,93	5,65	3,73	18,33	0,2209	8,702	0,0487	0,80	17,53	19,13
8	21,1	7,7	2,04	3,05	6,22	4,17	20,54	0,3136	27,502	0,1091	0,80	19,74	21,32
9	22,4	8,8	2,17	3,11	6,75	4,73	23,49	0,9801	44,222	0,2152	0,94	22,55	24,43
10	24,1	9,4	2,24	3,18	7,12	5,02	25,25	1,3225	68,062	0,2807	1,02	24,23	26,27
Σ			17,11	26,83	47,49	30,81		6,8498	329,96	1,5394			
П		10,2	2,32				27,42			0,3739	1,12	26,30	28,54

Запишемо економетричну модель: $y = 2{,}50248 \cdot x^{1{,}03198} + u$.

Знайдемо теоретичні значення $\hat{y}_i$:

$$\ln y_1 = \ln a_0 + a_1 \cdot \ln x_1 = 0{,}9173 + 1{,}03198 \cdot 0{,}99 = 1{,}9493,$$

$$\hat{y}_1 = \exp(\ln y_1) = 7{,}02\,.$$

3.3 Коефіцієнт еластичності

Аналіз економетричної моделі буде вважатись неповним, якщо не будуть розраховані коефіцієнти еластичності. У випадку лінійної парної регресії коефіцієнт еластичності визначається за формулою:

$$KE = a_1 \cdot \frac{\bar{x}}{\bar{y}}\ , \tag{3.18}$$

де a_1 – параметр економетричної моделі;

$\bar{x}, \bar{y}$ – середні значення ознак x та y.

При граничному переході для «миттєвих» змін аргументу x, коефіцієнт еластичності для функції $y = f(x)$ обчислюється за формулою:

$$KE = f'(x) \cdot \frac{x}{f(x)}\,, \tag{3.19}$$

Коефіцієнт еластичності показує: на скільки процентів у середньому зміниться результативний показник y, якщо незалежна змінна x зміниться на 1%. Коефіцієнт еластичності доцільно розраховувати у випадках, коли змінні x та y мають різні одиниці виміру.

Формули розрахунку коефіцієнтів еластичності для різних функцій представлено у таблиці 3.4.

Таблиця 3.4 – Коефіцієнти еластичності для деяких функцій

Вид рівняння регресії	Частинний коефіцієнт еластичності
1	2
1 $y = a_0 + a_1 x$	$KE = a_1 \frac{\bar{x}}{\bar{y}}$
2 $y = a_0 \cdot x^{a_1}$	$KE = a_1$
3 $y = a_0 + a_1 x + a_2 x^2$	$KE = \frac{x(a_1 + 2a_2 x)}{a_0 + a_1 x + a_2 x^2}$
4 $y = a_0 + \frac{a_1}{x}$	$KE = -\frac{a_1}{a_1 + a_0 x}$
5 $y = a_0 + a_1 \ln x$	$KE = x \ln a_1$
6 $y = a_0 + a_1 \cdot e^x$	$KE = \frac{a_1 x e^x}{a_1 e^x + a_0}$
7 $y = a_0 + a_1 \sqrt{x}$	$KE = \frac{a_1 x}{2a_1 x + 2a_0 \sqrt{x}}$
8 $y = a_0 + a_1 x^2$	$KE = \frac{2a_1 x^2}{a_1 x^2 + a_0}$
9 $y = a_0 + a_1 x^3$	$KE = \frac{3a_1 x^3}{a_1 x^3 + a_0}$
10 $y = \frac{1}{a_0 + a_1 x}$	$KE = -\frac{a_1 x}{a_1 x + a_0}$
11 $y = \frac{1}{a_0 + a_1 e^{-x}}$	$KE = \frac{a_1 x}{a_1 + a_0 e^{-x}}$
12 $y = a_0 \cdot a_1^x$	$KE = x \ln a_1$
13 $y = \frac{e^{a_1}}{a_0 x}$	$KE = -\frac{a_1}{x}$
14 $y = a_0 \cdot e^{a_1 x}$	$KE = a_1 x$

Вид рівняння регресії	Частинний коефіцієнт еластичності
1	2
15 $y = a_0 \cdot x^{a_1 x}$	$KE = a_1 x(\ln + 1)$
16 $y = \dfrac{x}{a_0 + a_1 x}$	$KE = \dfrac{a_0}{a_0 + a_1 x}$

Питання для самоконтролю:

1. Лінеаризація нелінійних функцій.
2. Алгоритм побудови однофакторної економетричної моделі.
3. Що являє собою кореляційне відношення?
4. Яким чином визначається значущість параметрів моделі?
5. Яким чином перевіряється адекватність моделі?
6. Що являє собою коефіцієнт еластичності? Що він показує?

Тема 4

БАГАТОФАКТОРНІ ЕКОНОМЕТРИЧНІ МОДЕЛІ

4.1 Багатофакторні лінійні економетричні моделі

В економіці, соціології, біології на величину результативного показника впливають багато різних факторів. При дослідженні впливу ряду незалежних факторів x_j на величину досліджуваного результативного показника y будують моделі множинної кореляції.

В моделях множинної кореляції залежна змінна y розглядається як функція не однієї, а декількох (в загальному випадку m) незалежних змінних x_j, а саме:

$$y = f(x_1, x_2, ..., x_j, ..., x_m) + u$$

Як і у випадку парної кореляції, одним із важливих питань моделювання множинної кореляції є питання про форму зв'язку.

В багатофакторних моделях вибір рівняння зв`язку являє собою складну задачу, оскільки дія різних факторів взаємно переплітається і відсутня можливість графічного контролю.

В даному випадку ще більшого значення набуває якісний аналіз характеру зв'язку кожного із факторів із залежним показником. Якщо цей зв`язок лінійний або близький до нього, то використовується лінійне рівняння множинної кореляції, яке для m факторів має вигляд:

$$y = a_0 + a_1 x_1 + a_2 x_2 + ... + a_j x_j + ... + a_m x_m = a_0 + \sum_{j=1}^{m} a_j x_j + u. \qquad (4.1)$$

Рівняння множинної регресії можуть бути виражені у натуральному (звичайному) та стандартизованому (нормованому) масштабі.

Коефіцієнти рівняння регресії у натуральному масштабі визначають методом найменших квадратів шляхом рішення системи нормальних рівнянь:

$$
\begin{cases}
na_0 + a_1\sum_{i=1}^{n} x_{1i} + a_2\sum_{i=1}^{n} x_{2i} + a_3\sum_{i=1}^{n} x_{3i} + \ldots + a_j\sum_{i=1}^{n} x_{ji} + \ldots + a_m\sum_{i=1}^{n} x_{mi} = \sum_{i=1}^{n} y_i \\
a_0\sum_{i=1}^{n} x_{1i} + a_1\sum_{i=1}^{n} x_{1i}^2 + a_2\sum_{i=1}^{n} x_{2i}x_{1i} + a_3\sum_{i=1}^{n} x_{3i}x_{1i} + \ldots + a_j\sum_{i=1}^{n} x_{ji}x_{1i} + \ldots + a_m\sum_{i=1}^{n} x_{mi}x_{1i} = \sum_{i=1}^{n} y_i x_{1i} \\
a_0\sum_{i=1}^{n} x_{2i} + a_1\sum_{i=1}^{n} x_{1i}x_{2i} + a_2\sum_{i=1}^{n} x_2^2 + a_3\sum_{i=1}^{n} x_{3i}x_{2i} + \ldots + a_j\sum_{i=1}^{n} x_{ji}x_{2i} + \ldots + \ldots a_m\sum_{i=1}^{n} x_{mi}x_{2i} = \sum_{i=1}^{n} y_i x_{2i} \\
a_0\sum_{i=1}^{n} x_{3i} + a_1\sum_{i=1}^{n} x_{1i}x_{3i} + a_2\sum_{i=1}^{n} x_{2i}x_{3i} + a_3\sum_{i=1}^{n} x_3^2 + \ldots + a_j\sum_{i=1}^{n} x_{ji}x_{3i} + \ldots + a_m\sum_{i=1}^{n} x_{mi}x_{3i} = \sum_{i=1}^{n} y_i x_{3i} \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
a_0\sum_{i=1}^{n} x_{ji} + a_1\sum_{i=1}^{n} x_{1i}x_{ji} + a_2\sum_{i=1}^{n} x_{2i}x_{ji} + a_3\sum_{i=1}^{n} x_{3i}x_{ji} + \ldots + a_j\sum_{i=1}^{n} x_{ji}^2 + \ldots + a_m\sum_{i=1}^{n} x_{mi}x_{ji} = \sum_{i=1}^{n} y_i x_{ji} \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
a_0\sum_{i=1}^{n} x_{mi} + a_1\sum_{i=1}^{n} x_{1i}x_{mi} + a_2\sum_{i=1}^{n} x_{2i}x_{mi} + a_3\sum_{i=1}^{n} x_{3i}x_{m1} + \ldots + a_j\sum_{i=1}^{n} x_{ji}x_{mi} + \ldots + a_m\sum_{i=1}^{n} x_{mi}^2 = \sum_{i=1}^{n} y_i x_{mi}
\end{cases} \tag{4.2}
$$

Тіснота зв'язку оцінюється за допомогою коефіцієнта множинної кореляції:

$$
R = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2 / (n-m-1)}{\sum_{i=1}^{n}(y_i - y)^2 / (n-1)}}\ , \tag{4.3}
$$

$\bar{y}_i$ – теоретичні значення результативного показника, розраховані по рівнянню регресії;

y – середнє арифметичне значення результативного показника.

Г. Тінтер запропонував таку формулу коефіцієнта множинної кореляції:

$$
R_{1,2,\ldots,m} = \sqrt{\frac{a_1 s_1 + a_2 s_2 + \ldots + a_m s_m}{s_y}}\ , \tag{4.4}
$$

$s_j - (j = 1,2,\ldots,m)$ – коваріації, які розраховуються по формулі:

$$
s_j = \overline{y x}_j - y \cdot x_j ;
$$

s_y – середньоквадратичні відхилення результативного показника;

$a_j - (j = 1,2,\ldots,m)$ – коефіцієнти регресії.

Приклад 4.1. На основі статистичних даних показника y (витрати на споживання) і факторів x_1 (рівень доходів) та x_2 (рівень заощаджень) (таблиця 4.1) знайти оцінки параметрів регресій, якщо припустимо, що стохастична залежність між факторами і результативним показником має вигляд: $y = a_0 + a_1x_1 + a_2x_2 + u$.

Розв'язок. Записуємо систему нормальних рівнянь:

$$\begin{cases} n \cdot a_0 + a_1\sum x_1 + a_2\sum x_2 = \sum y \\ a_0\sum x_1 + a_1\sum x_1^2 + a_2\sum x_2 \cdot x_1 = \sum y \cdot x_1 \\ a_0\sum x_2 + a_1\sum x_1 \cdot x_2 + a_2\sum x_2^2 = \sum y \cdot x_2 \end{cases}$$

або у матричній формі:

$$\begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{m1} \\ 1 & x_{12} & x_{22} & \dots & x_{m2} \\ 1 & x_{13} & x_{23} & \dots & x_{m3} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & \dots & x_{mn} \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ \dots \\ a_m \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \dots \\ u_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{bmatrix}$$

В таблицю 4.1 запишемо знайдені суми $\sum x_1, \sum x_2, \sum y, \sum x_1^2, \sum x_2^2, \sum x_1x_2, \sum yx_1, \sum yx_2$ і підставимо їх у систему нормальних рівнянь:

$$\begin{cases} 13a_0 + 155{,}57a_1 + 199{,}61a_2 = 2458{,}6 \\ 155{,}57a_0 + 1963{,}83a_1 + 2442{,}27a_2 = 31602{,}77 \\ 199{,}61a_0 + 2442{,}27a_1 + 3118{,}18a_2 = 39083{,}55 \end{cases}$$

Систему нормальних рівнянь розв'язуємо методом оберненої матриці: $X^{-1} \cdot Y = A$

$$A = \begin{bmatrix} 4{,}75515 & 0{,}07201 & -0{,}36080 \\ 0{,}07201 & 0{,}02071 & -0{,}02083 \\ -0{,}36080 & -0{,}02083 & 0{,}03974 \end{bmatrix} \cdot \begin{bmatrix} 2458{,}6 \\ 31602{,}77 \\ 39083{,}55 \end{bmatrix} = \begin{bmatrix} -134{,}67 \\ 17{,}41 \\ 7{,}52 \end{bmatrix}$$

Запишемо економетричну модель: $y = -134{,}67 + 17{,}41x_1 + 7{,}52x_2 + u$

Знайдемо розрахункові значення $\hat{y}_i$, залишкову і загальну дисперсії:

$$\bar{y} = \frac{2458{,}6}{13} = 189{,}123$$

$$\sigma_u^2 = \frac{\Sigma\left(y_i - \hat{y}_i\right)^2}{n - m - 1} = \frac{3451{,}73}{13 - 2 - 1} = 345{,}173$$

$$\sigma_y^2 = \frac{\Sigma\left(y_i - \bar{y}_i\right)^2}{n} = \frac{51444{,}40}{13} = 3957{,}2617$$

Таблиця 4.1 – Вхідні дані та результати розрахунків

№ Сп.	Дохід, x_{1i}	Заощадження, x_{2i}	Витрати на споживання, y_i	$x_{1i}x_{2i}$	x_{1i}^2	x_{2i}^2	$y_i x_{1i}$	$y_i x_{2i}$	Теоретичне значення $\hat{y}_i$	$(y_i - \hat{y}_i)^2$	$(y_i - \bar{y})^2$
1	7,08	12,48	100,2	88,36	50,13	155,75	709,416	1250,5	82,43	315,93	7907,31
2	8,28	15,34	112,3	127,02	68,56	235,32	929,844	1722,68	124,82	156,74	5901,79
3	10,47	15,04	128,4	157,47	109,62	226,20	1344,35	1931,14	160,69	1042,95	3687,29
4	8,66	13,25	130,2	114,75	75,00	175,56	1127,53	1725,15	115,72	209,56	3471,93
5	9,65	12,9	105,8	124,49	93,12	166,41	1020,97	1364,82	130,33	601,71	6942,74
6	11,56	13,02	180,5	150,51	133,63	169,52	2086,58	2350,11	164,49	256,41	74,36
7	13,65	17,57	250,7	239,83	186,32	308,70	3422,06	4404,8	235,08	243,92	3791,72
8	14,59	18,04	247,5	263,20	212,87	325,44	3611,03	4464,9	254,98	55,98	3407,87
9	15,35	15,06	230,7	231,17	235,62	226,80	3541,25	3474,34	245,81	228,36	1728,64
10	12,31	14,31	200,5	176,16	151,54	204,78	2468,16	2869,16	187,24	175,74	129,43
11	16,06	16,73	270,6	268,68	257,92	279,89	4345,84	4527,14	270,73	0,02	6638,49
12	14,03	18,42	260,7	258,43	196,84	339,30	3657,62	4802,09	248,09	159,05	5123,26
13	13,88	17,45	240,5	242,21	192,65	304,50	3338,14	4196,73	238,18	5,36	2639,59
Суми	155,57	199,61	2458,6	2442,27	1963,83	3118,18	31602,77	39083,55	2458,6	3451,73	51444,40

Тіснота зв'язку:

$$R = \sqrt{1 - \frac{345.173}{3957.2617}} = 0.9554$$

Оцінка адекватності моделі по F – критерію Фішера:

$$F_p = \frac{\sum(\hat{y}_i - \bar{y})^2 \; (m-1)}{\sum(y_i - \hat{y}_i)^2 \; (n-m-1)} = \frac{\sigma_y^2 - \sigma_e^2}{\sigma_e^2} = \frac{3957.262 - 345.173}{345.173} = 10.46$$

$F_{табл} = 6.93$, $F_p > F_{табл}$; модель адекватна статистичним даним.

4.2 Лінійна регресійна модель з двома незалежними змінними

В найпростішому випадку число незалежних змінних (x_1, x_2) дорівнює двом і зв'язок між ними та залежною змінною y лінійний. В цьому випадку виникає задача: розв'язати по даних n спостережень рівняння зв'язку:

$$y = a_0 + a_1 \cdot x_1 + a_2 \cdot x_2, \text{ тобто:}$$

– знайти коефіцієнти регресії a_1, a_2 та параметр a_0;

– оцінити тісноту зв'язку між y та обома ознаками x_1 і x_2;

– оцінити тісноту зв'язку між y та x_1 при постійному x_2 та між y та x_2 при постійному x_1.

Перше завдання вирішується методом найменших квадратів шляхом розв'язання системи нормальних рівнянь:

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum_{i=1}^{n} x_{1i} + a_2 \cdot \sum_{i=1}^{n} x_{2i} = \sum_{i=1}^{n} y_i \\ a_0 \cdot \sum_{i=1}^{n} x_{1i} + a_1 \cdot \sum_{i=1}^{n} x_{1i}^2 + a_2 \cdot \sum_{i=1}^{n} x_{2i} \cdot x_{1i} = \sum_{i=1}^{n} y_i \cdot x_{1i} \\ a_0 \cdot \sum_{i=1}^{n} x_{2i} + a_1 \cdot \sum_{i=1}^{n} x_{1i} \cdot x_{2i} + a_2 \cdot \sum_{i=1}^{n} x_{2i}^2 = \sum_{i=1}^{n} y_i \cdot x_{2i} \end{cases} . \quad (4.5)$$

Рішення системи (4.5) дозволить визначити параметри моделі a_0, a_1, a_2.

Параметри рівняння регресії можна визначити через парні коефіцієнти кореляції. При цьому парні лінійні коефіцієнти кореляції визначаються так, як і при парній кореляції, а саме:

$$r_{yx_1} = \frac{\overline{x_1 y} - \bar{x}_1 \cdot \bar{y}}{\sigma_{x_1} \cdot \sigma_{y}}$$

$$r_{x_1x_2} = \frac{\overline{x_1 x_2} - \bar{x}_1 \cdot \bar{x}_2}{\sigma_{x_1} \cdot \sigma_{x_2}}, \tag{4.6}$$

$$r_{yx_2} = \frac{\overline{x_2 y} - \bar{x}_2 \cdot \bar{y}}{\sigma_{x_2} \cdot \sigma_{y}}$$

де $\bar{x}_1, \bar{x}_2, \bar{y}$, $\overline{x_1 y}$, $\overline{x_1 x_2}$, $\overline{x_2 y}$ – середні арифметичні факторів x_1, x_2 та їх парних добутків.

Тіснота зв'язку ознаки y з ознаками x_1 та x_2 оцінюється вибірковим сукупним коефіцієнтом кореляції:

$$R = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2 r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}}. \tag{4.7}$$

Тіснота зв'язку між y та x_1 при постійному x_2, а також між y та x_2 при постійному x_1 оцінюється відповідно частковими вибірковими коефіцієнтами кореляції:

$$r_{yx_1(x_2)} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2) \cdot (1 - r_{x_1x_2}^2)}},$$

$$r_{yx_2(x_1)} = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2) \cdot (1 - r_{x_1x_2}^2)}}, \tag{4.8}$$

$$r_{x_1x_2(y)} = \frac{r_{x_1x_2} - r_{yx_1} \cdot r_{yx_2}}{\sqrt{(1 - r_{yx_1}^2) \cdot (1 - r_{yx_2}^2)}}.$$

Параметри регресійної моделі можна також виразити через парні лінійні коефіцієнти кореляції:

$$a_1 = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} \cdot \frac{\sigma_y}{\sigma_{x_1}} \tag{4.9}$$

$$a_2 = \frac{r_{yx_1} - r_{yx_1} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} \cdot \frac{\sigma_y}{\sigma_{x_2}}$$

або через коваріацію та дисперсію шляхом рішення системи нормальних рівнянь:

$$\begin{cases} \dfrac{\sum(y_i-\bar{y})(x_{1i}-\bar{x}_1)}{n}=\dfrac{a_1\sum(x_{1i}-\bar{x}_1)^2}{n}+\dfrac{a_2\sum(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)}{n}+\dfrac{a_3\sum(x_{1i}-x_1)(x_{3i}-x_3)}{n} \\ \dfrac{\sum(y_i-\bar{y})(x_{2i}-\bar{x}_2)}{n}=\dfrac{a_1\sum(x_{1i}-\bar{x}_1)(x_{2i}-\bar{x}_2)}{n}+\dfrac{a_2\sum(x_{2i}-\bar{x}_2)^2}{n}+\dfrac{a_3\sum(x_{2i}-\bar{x}_2)(x_{3i}-\bar{x}_3)}{n} \\ \dfrac{\sum(y_i-\bar{y}(x_{3i}-\bar{x}_3)}{n}=\dfrac{a_1\sum(x_{1i}-\bar{x}_1)(x-\bar{x}_3)}{n}+\dfrac{a_2\sum(x_{2i}-\bar{x}_2)(x_3-\bar{x}_3)}{n}+\dfrac{a_3\sum(x_{3i}-\bar{x}_3)^2}{n} \end{cases} \quad (4.10)$$

Оцінка значущості параметрів множинної регресії проводиться з використанням таких формул:

$$\sigma_{a_1}=\frac{\sigma_u}{\sigma_{x_1}\cdot\sqrt{n\cdot(1-r^2_{x_1x_2})}}\ ;\quad \sigma_{a_2}=\frac{\sigma_u}{\sigma_{x_2}\cdot\sqrt{n\cdot(1-r^2_{x_1x_2})}}\ ; \quad (4.11)$$

σ_u – залишкова дисперсія: $\sigma_u=\sqrt{\dfrac{\sum(y_i-\hat{y}_i)^2}{n-m-1}}$

4.3 Коваріація (cov)

Властивості коваріації вивчаються у коваріаційному аналізі. Коваріаційний аналіз – це одночасний аналіз сум квадратів і сум добутків відхилень двох або більше змінних від їх середніх. Він використовується при плануванні і статистичній обробці результатів дослідів як спосіб зменшення помилки експерименту, яка не піддається безпосередньому контролю (вимірюванню). Коваріаційний аналіз дозволяє встановити співвідношення між варіацією залежної змінної y і варіацією незалежної x.

У вузькому розумінні під коваріацією, яка позначається ***cov*** розуміють середній добуток відхилень двох змінних від їх середніх:

$$\mathrm{cov}=\overline{(x-\bar{x})\cdot(y-\bar{y})}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})}{n-1}. \quad (4.12)$$

Властивості коваріації:

1) коваріація ознаки x з самою собою є дисперсія:

$$\overline{(x-\bar{x})\cdot(x-\bar{x})}=\overline{(x-\bar{x})^2}=\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n}=\delta^2x\,, \quad (4.13)$$

$$\overline{(y-\bar{y})\cdot(y-\bar{y})}=\overline{(y-\bar{y})^2}=\frac{\sum_{i=1}^{n}(y_i-\bar{y})^2}{n}=\delta^2 y. \qquad (4.14)$$

2) якщо ознаки x та y функціонально залежні і пряма $y=a_0+ax$ утворює з віссю абсцис кут 45^0 або 135^0, то абсолютна величина коваріації максимальна;

3) якщо лінія регресії y по x розміщена горизонтально, а лінія регресії x по y вертикально, то коваріація дорівнює нулю, модуль коваріації знаходиться в межах:

$$0\le|\mathrm{cov}|\le\delta^2 x=\delta^2 y$$

4) знак коваріації визначається тим, що при прямому зв'язку між x та y додатні відхилення від x множаться на додатні відхилення від y, а від'ємні; при оберненому зв'язку навпаки: додатні відхилення від x множаться на від'ємні відхилення від y, а від'ємні відхилення від x на додатні відхилення від y; тому коваріація позитивна при прямому зв'язку між x і y і негативна при оберненому зв'язку.

Якщо нормувати x і y, тобто перейти до нових величин x' і y', в яких середні дорівнюють нулю, а дисперсія – одиниці:

$$x'=\frac{x-\bar{x}}{\sigma_x},\ y'=\frac{y-\bar{y}}{\sigma_y}$$

то $\mathrm{cov}(x',y')$буде дорівнювати лінійному коефіцієнту кореляції. В іншому розумінні коваріація – сума добутків відхилень двох випадкових величин від відповідних середніх, тобто: $\mathrm{cov}=\sum dxdy$

В імовірнісному вигляді:

$$\mathrm{cov}=\sum_i\sum_j(x_i-\bar{x})(y_j-\bar{y})\cdot P_{ij}, \qquad (4.15)$$

де x_i, y_j – можливі значення випадкових величин x та y;

P_{ij} – імовірність того, що випадкові величини приймуть значення x_i та y_j.

Позначивши $\mathrm{cov}=K_{ij}$, отримуємо коваріаційну матрицю:

$$K_{xy} = \begin{vmatrix} K_{11}K_{12}K_{13}\ldots K_{1n} \\ K_{21}K_{22}K_{23}\ldots K_{2n} \\ K_{31}K_{32}K_{33}\ldots K_{3n} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ K_{n1}K_{n2}K_{n3}\ldots K_{nn} \end{vmatrix}. \qquad (4.16)$$

Ділення коваріації двох ознак на добуток середньоквадратичних відхилень δx та δy дає лінійний коефіцієнт кореляції:

$$Z = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\delta_x \delta_y} = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2(y_i - \bar{y})^2}}. \qquad (4.17)$$

4.4 Багатофакторні лінійні економетричні моделі в стандартизованому масштабі

У рівнянні множин регресії у натуральному масштабі кожний фактор $x_1, x_2, ..., x_j, ..., x_m$ має свою власну одиницю виміру. Вказана обставина не дозволяє оцінювати порівняльну силу впливу кожного аргументу на функцію шляхом співставлення коефіцієнтів регресії $a_1, a_2, ..., a_j, ..., a_m$.

Тому для виміру впливу факторів на результативний показник розраховують рівняння множинної регресії в стандартизованому (сигмальному) масштабі. За початок відліку для кожної ознаки приймають значення середньої арифметичної, а за одиницю виміру – середнє квадратичне відхилення (сигму).

При використанні даного метода необхідно відібрані ознаки перевести в сигмальний масштаб за допомогою формул переводу.

$$t_{0i} = \frac{y_i - \bar{y}}{\sigma_y};\; t_{1i} = \frac{x_{1i} - \bar{x}_1}{\sigma_{x_1}};\; t_{2i} = \frac{x_{2i} - \bar{x}_2}{\sigma_{x_2}};\; \ldots;\; t_{mi} = \frac{x_{mi} - \bar{x}_m}{\sigma_{x_m}}; \qquad (4.18)$$

де $\bar{y}, \bar{x}_1, \bar{x}_2, ..., \bar{x}_m$ – середні значення відповідно $y, x_1, x_2, ..., x_m$ при довжині вибірки n;

$\sigma_1, \sigma_2, ..., \sigma_m$ – середні квадратичні відхилення відповідно $y, x_1, x_2, ..., x_m$

В стандартизованому масштабі середнє значення ознаки дорівнює нулю: $\bar{t}_m = 0$, а $\sigma_t^2 = 1$.

У відповідності з цим спрощується формула розрахунку коефіцієнта парної кореляції:

$$r_{ij}=\frac{1}{n}\cdot\sum_{i=1}^{n}\frac{(t_i-\bar{t}_i)\cdot(t_j-\bar{t}_j)}{\sigma_{t_i}\sigma_{t_j}}=\frac{1}{n}\cdot\sum_{i=1}^{n}\frac{(t_i-0)\cdot(t_j-0)}{1\cdot 1}=\frac{1}{n}\cdot\sum_{i=1}^{n}t_i\cdot t_j$$

$$r_{ij}=\frac{1}{n}\cdot\sum_{i=1}^{n}t_i\cdot t_j\,. \qquad (4.19)$$

У стандартизованому масштабі рівняння множинної лінійної регресії має вигляд:

$$\bar{t}_y=\beta_1\cdot t_1+\beta_2\cdot t_2+\ldots+\beta_m\cdot t_m, \qquad (4.20)$$

де $\beta_1,\beta_2,\ldots,\beta_m$ – стандартизовані коефіцієнти множинної регресії, які визначаються з умови:

$$\sum_{i=1}^{n}\left(t_{yi}-\bar{t}_y\right). \qquad (4.21)$$

Використовуючи цю умову, одержимо систему нормальних рівнянь:

$$\begin{cases} r_{yx_1}=\beta_1+\beta_1 r_{x_1x_2}+\beta_3 r_{x_1x_3}+\ldots+\beta_j r_{x_1x_j}+\ldots+\beta_m r_{x_1x_m}\\ r_{yx_2}=\beta_1 r_{x_2x_1}+\beta_2+\beta_3 r_{x_2x_3}+\ldots+\beta_j r_{x_2x_j}+\ldots+\beta_m r_{x_2x_m}\\ r_{yx_3}=\beta_1 r_{x_3x_1}+\beta_2 r_{x_3x_2}+\beta_3+\ldots+\beta_j r_{x_3x_j}+\ldots+\beta_m r_{x_3x_m}\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ r_{yx_j}=\beta_1 r_{x_jx_1}+\beta_2 r_{x_jx_2}+\beta_3 r_{x_jx_3}+\ldots+\beta_j+\ldots+\beta_m r_{x_jx_m}\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ r_{yx_m}=\beta_1 r_{x_mx_1}+\beta_2 r_{x_mx_2}+\beta_3 r_{x_mx_3}+\ldots+\beta_j r_{x_mx_j}+\ldots+\beta_m \end{cases}, \qquad (4.22)$$

де $r_{yx_1},r_{x_1x_2},\ldots,r_{x_1x_m}$ – коефіцієнти парної кореляції, які утворюють матрицю коефіцієнтів парної кореляції:

	y	x_1	x_2	x_3	...	x_m
y	1	r_{yx_1}	r_{yx_2}	r_{yx_3}	...	r_{yx_m}
x_1	r_{x_1y}	1	$r_{x_1x_2}$	$r_{x_1x_3}$	...	$r_{x_1x_m}$
x_2	r_{x_2y}	$r_{x_2x_3}$	1	$r_{x_2x_3}$	...	$r_{x_2x_m}$
x_3	r_{x_3y}	$r_{x_3x_1}$	$r_{x_3x_2}$	1	...	$r_{x_3x_m}$
...	...	...	...	...	...	...
x_m	r_{x_my}	$r_{x_mx_1}$	$r_{x_mx_2}$	$r_{x_mx_3}$	...	1

Матриця коефіцієнтів парної кореляції симетрична. Для запису системи нормальних рівнянь в цій матриці необхідно викреслити перший рядок; тоді перший стовпець матриці буде стовпцем вільних членів.

Система (4.22) розв'язується одним із методів. Наприклад, методом Крамера, методом повних виключень Жордана-Гауса. При рішенні системи (4.22) методом Крамера

$$\beta_j = \frac{\Delta_j}{\Delta}, \tag{4.23}$$

де

$$\Delta = \begin{vmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} \cdots & r_{x_1x_m} \\ r_{x_2x_1} & 1 & r_{x_2x_3} & r_{x_2x_m} \\ r_{x_3x_1} & r_{x_3x_2} & 1 & r_{x_3x_m} \\ r_{x_mx_1} & r_{x_mx_2} & r_{x_mx_3} & 1 \end{vmatrix}, \tag{4.24}$$

де $r_{x_ix_j}$ – парні коефіцієнти кореляції між ознаками i та j;

m – порядок матриці.

Визначник системи Δ_j отримується із Δ шляхом заміни в ній j-*го* стовпця на стовпець вільних членів системи (4.24).

Стандартизовані коефіцієнти $\beta_1, \beta_2, ..., \beta_m$ виражають швидкість зміни середнього значення результативної ознаки під впливом одного із факторів при постійних значеннях інших факторів. Оскільки всі змінні рівняння множинної регресії виражені в порівняльних одиницях вимірювання – сигмах, то стандартизовані коефіцієнти β_j показують порівняльну силу впливу зміни кожного j-*го* фактора на зміну результативної ознаки.

4.5 Рівняння множинної лінійної регресії у натуральному масштабі

Рівняння множинної лінійної регресії у стандартизованому масштабі має велике значення при проведенні економічного аналізу, так як показує порівняльну силу впливу факторіальних ознак на величину результативного показника. Для використання цього рівняння в практичних розрахунках його параметри повинні бути переведені із сигматичного у натуральний масштаб.

З цією метою використовують формулу переходу:

$$a_j = \beta_j \cdot \frac{\sigma_y}{\sigma_{x_j}}, \tag{4.25}$$

де β_j – коефіцієнти рівняння регресії в стандартизованому масштабі;

σ_y – середнє квадратичне відхилення (сигма) результативної ознаки;

σ_{x_j} – середнє квадратичне відхилення (сигма) j-ої $j=(1...m)$ факторної ознаки.

Вільний член рівняння регресії a_0 одержують із виразу:

$$a_0 = \bar{y} - a_1\bar{x}_1 - a_2\bar{x}_2 - ... - a_m\bar{x}_m. \tag{4.26}$$

Шукане рівняння регресії у натуральному масштабі має вигляд:

$$\hat{y} = a_0 + a_1x_1 + a_2x_2 + ... + a_mx_m. \tag{4.27}$$

Коефіцієнт регресії a_j показує, на скільки одиниць в середньому зміниться y із зміною x_j на одиницю при умові, що інші діючі фактори закріплені на постійному рівні.

Перехід від натурального масштабу до стандартизованого можна здійснити, використовуючи вираз (4.25):

$$\beta_j = a_j \frac{\sigma_{x_j}}{\sigma_y}$$

4.6 Коефіцієнт множинної кореляції і детермінації

У процесі багатофакторного аналізу вирішується і друга не менш важлива задача визначення тісноти зв'язку між результативною і факторіальними ознаками. Тісноту зв'язку при множинній кореляційній залежності характеризує множинний або сукупний коефіцієнт кореляції R, який розраховується по такій формулі:

$$R = \sqrt{1 - \frac{\sigma_u^2}{\sigma_y^2}}, \tag{4.28}$$

де σ_u^2 – залишкова дисперсія: $\sigma_u^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-m-1}$; (4.29)

σ_y^2 – загальна дисперсія: $\sigma_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}$. (4.30)

Множинний коефіцієнт кореляції можна також розрахувати, виходячи із стандартизованих коефіцієнтів множинної регресії β_j та коефіцієнтів парної кореляції r_{yx_j}

$$R = \sqrt{\beta_1 r_{yx_1} + \beta_2 r_{yx_2} + \ldots + \beta_m r_{yx_m}}. \tag{4.31}$$

Коефіцієнт множинної кореляції, якщо число факторних ознак більше трьох, можна також розрахувати на основі обчислень визначників, складених із парних коефіцієнтів кореляції:

$$\Delta = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & r_{yx_3} & \cdots & r_{yx_m} \\ r_{x_1y} & 1 & r_{x_1x_2} & r_{x_1x_3} & \cdots & r_{x_1x_m} \\ r_{x_2y} & r_{x_2x_1} & 1 & r_{x_2x_3} & \cdots & r_{x_2x_m} \\ r_{x_3y} & r_{x_3x_1} & r_{x_3x_2} & 1 & \cdots & r_{x_3x_m} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{x_my} & r_{x_mx_1} & r_{x_mx_2} & r_{x_mx_3} & \cdots & 1 \end{vmatrix} \tag{4.32}$$

$$\Delta^* = \begin{vmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} & \cdots & r_{x_1x_m} \\ r_{x_2x_1} & 1 & r_{x_2x_3} & \cdots & r_{x_2x_m} \\ r_{x_3x_1} & r_{x_3x_2} & 1 & \cdots & r_{x_3x_m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{x_mx_1} & r_{x_mx_2} & r_{x_mx_3} & \cdots & 1 \end{vmatrix} \tag{4.33}$$

Тоді:

$$R = \sqrt{\frac{\Delta}{\Delta^*}}, \tag{4.34}$$

Δ – визначник матриці коефіцієнтів парних кореляцій;

Δ^* – визначник, який утворюється із визначника Δ шляхом викреслення першого рядка та першого стовпця.

Коефіцієнт множинної кореляції змінюється в межах від 0 до 1 та по визначенню додатній:

$$0 <= R <= 1. \qquad (4.35)$$

Розрахований коефіцієнт R обов'язково повинен коригуватись на число спостережень, так як при малому числі спостережень значення R буде завищеним. Якщо число факторних ознак дорівнювало б числу спостережень, то гіперплощина, побудована по методу найменших квадратів, пройшла б через всі точки n і множинний коефіцієнт кореляції був би рівний 1, хоч зв'язок між результативною і факторними ознаками міг бути дуже слабким.

Величина множинного коефіцієнта кореляції коригується на основі виразу:

$$\hat{R} = \sqrt{1-(1-R^2)\cdot\frac{n-1}{n-m-1}}. \qquad (4.36)$$

Коригування R не проводиться при умові, якщо:

$$\frac{n-m}{m} => 20. \qquad (4.37)$$

Емпірично встановлено, що число спостережень повинно бути більшим за число факторних ознак приблизно у вісім разів.

Суттєвість R визначається по критерію F (Фішера):

$$F_p = \frac{R^2\cdot(n-m-1)}{(1-R^2)\cdot m}, \qquad (4.38)$$

де F_p – розрахункове значення критерія F з числом ступенів свободи $k_1 = m; k_2 = n-m-1$. Коефіцієнт R суттєвий, якщо $F_p > F_{\alpha,k_1,k_2}$.

Величина R^2 називається множинним коефіцієнтом детермінації. Вона показує, яка частина дисперсії функції пояснюється за рахунок варіації лінійної комбінації аргументів при даних значеннях коефіцієнтів регресії a_j.

Для оцінки надійності коефіцієнта кореляції розраховують його похибку:

$$\sigma_R = \frac{1-R^2}{\sqrt{n-m-1}}. \quad (4.39)$$

Тоді довірчі інтервали коефіцієнта множинної кореляції розраховують по формулі:

$$R - t_{\alpha,k} \cdot \sigma_R <= \widetilde{\rho} <= R + t_{\alpha,k} \cdot \sigma_R\,, \quad (4.40)$$

$t_{\alpha,k}$ – табличне значення t – критерія Стьюдента.

Значущість коефіцієнта множинної кореляції можна встановити із співвідношення $\frac{R}{\sigma_R} > 3$. Якщо виконується ця нерівність, то з імовірністю Р=0,99 можна вважати R значущим.

4.7 Оцінка значущості коефіцієнта регресії і перевірка адекватності моделі

Якщо правильно підібрано рівняння регресії, то хоч б приблизно виконується умова нормального розподілу відхилень фактичних значень залежної змінної від її розрахункових значень $(y_i - \hat{y}_i)$; це дозволяє побудувати довірчі інтервали для коефіцієнтів регресії. Приймемо в якості досліджуваної гіпотезу про рівність нулю коефіцієнтів регресії в генеральній сукупності. Величина $t_{a_j} = \frac{|a_j|}{\sigma_{a_j}}$ розподілена по закону Стьюдента з $k = n - m - 1$ ступенями свободи. Середньоквадратична похибка коефіцієнта регресії визначається за формулою:

$$\sigma_{a_j} = \frac{\sigma_y}{\sigma_{x_j}} \sqrt{\frac{1-R}{n-m-1}\,\frac{\Delta_{jj}}{\Delta}}, \quad (4.41)$$

Δ – визначник матриці нормальних рівнянь;

Δ_{jj} – алгебраїчне доповнення до елемента матриці нормальних рівнянь, що знаходиться на перетині *j-го* рядка та *j-го* стовпця.

Емпіричне значення t_{a_j} порівнюється з квантилем розподілу Стьюдента $t_{\alpha,k}$ для заданого рівня значущості α та заданого числа ступенів свободи $k = n - m - 1$. Гіпотеза про рівність коефіцієнта регресії нулю відкидається, якщо $t_{a_j} > t_{\alpha,k}$.

Довірчий інтервал, в якому з заданою імовірністю $P = (1 - \alpha)$ знаходиться «істинне» значення регресії a_j такий:

$$a_j - t_{\alpha,k} \cdot \sigma_{a_j} < \widetilde{a}_j < a_j + t_{\alpha,k} \cdot \sigma_{a_j}. \quad (4.42)$$

Для оцінки адекватності одержаного рівняння регресії необхідно порівняти залишкову дисперсію з факторною дисперсією залежної змінної, тобто скористуватись F – критерієм Фішера:

$$F_p = \frac{\delta_y^2}{\sigma_u^2}; \quad (4.43)$$

де $\sigma_u^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n - m - 1}$, $\delta_y^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{m}$

Емпіричне значення F_p порівнюють з табличним F_{α,k_1,k_2} для заданого рівня значущості α і заданого числа ступенів свободи $k_1 = n - 1$ та $k_2 = n - m - 1$.

При $F_p > F_{\alpha,k_1,k_2}$ відкидається нульова гіпотеза, яка полягає в тому, що вирівнювання по одержаному рівнянні регресії дає такі ж результати, як і вирівнювання по прямій $y = \bar{y}$.

Приклад 4.2. Для вхідних даних, наведених у таблиці 4.2, побудувати трьохфакторну лінійну регресійну модель у стандартизованому масштабі.

Розв'язок. По змінних y, x_1, x_2, x_3 обчислимо прості середні $\bar{y}, \bar{x}_1, \bar{x}_2, \bar{x}_3$: $\bar{y} = \frac{82}{10} = 8{,}2$; $\bar{x}_1 = \frac{55}{10} = 5{,}5$; $\bar{x}_2 = \frac{128}{10} = 12{,}8$; $\bar{x}_3 = \frac{23{,}1}{10} = 2{,}31$.

Обчислимо суми добутків $\sum y_i x_{1i}, \sum y_i x_{2i}, \sum y_i x_{3i}, \sum x_{1i} x_{2i}, \sum x_{1i} x_{3i}, \sum x_{2i} x_{3i}$, а також суми для обчислення середньоквадратичних відхилень $\sum (y_i - \bar{y})^2, \sum (x_{1i} - \bar{x}_1)^2, \sum (x_{2i} - \bar{x}_2)^2, \sum (x_{3i} - \bar{x}_3)^2$.

Таблиця 4.2 – Вхідні дані

Роки, x_{1i}	Кількість підприємств, x_{2i}	Продуктивність праці, x_{3i}	Обсяг реалізації продукції, y_i
1	4	2,2	1,2
2	7	1,8	3,6
3	9	2,3	5,8
4	10	2,4	7,4
5	14	2,0	8,5
6	16	2,1	9,5
7	15	2,5	10
8	17	2,6	12
9	16	2,4	11
10	20	2,8	13
55	128	23,1	82

Результати розрахунків зведемо в таблицю 4.3.

Таблиця 4.3 – Результати розрахунків

$y_i x_{1i}$	$y_i x_{2i}$	$y_i x_{3i}$	$x_{1i} x_{2i}$	$x_{1i} x_{3i}$	$x_{2i} x_{3i}$	$(y_i - \bar{y})^2$	$(x_{1i} - \bar{x}_1)^2$	$(x_{2i} - \bar{x}_2)^2$	$(x_{3i} - \bar{x}_3)^2$
1,2	4,8	2,64	4	2,2	8,8	49,00	20,25	7,44	0,0121
7,2	25,2	6,48	14	3,6	12,6	21,16	12,25	33,64	0,2600
17,4	52,2	13,34	27	6,9	20,7	5,76	6,25	14,44	0,0001
29,6	74,0	17,76	40	9,6	24,0	0,64	2,25	7,847	0,0081
42,5	119,0	17,00	70	10,0	28,0	0,09	0,25	1,44	0,0961
57,0	152,0	19,95	96	12,6	33,6	1,69	0,25	10,24	0,0441
70,0	150,0	25,00	105	17,5	37,5	3,24	2,25	4,84	0,0361
96,0	204,0	31,20	136	20,8	44,2	14,44	6,25	17,64	0,0841
99,0	176,0	26,40	144	21,6	38,4	7,84	12,25	10,24	0,0081
130,0	260,0	36,40	200	28,0	56,0	23,04	20,25	51,84	0,2401
549,9	1217,2	196,17	836	132,8	303,8	126,9	82,50	229,60	0,7889

Розрахуємо середні значення добутків $\overline{yx_1}, \overline{yx_2}, \overline{yx_3}, \overline{x_1x_2}, \overline{x_1x_3}, \overline{x_2x_3}$:

$$\overline{yx_1} = \frac{\Sigma\, y_i x_{1i}}{n} = \frac{549{,}9}{10} = 54{,}99; \quad \overline{yx_2} = \frac{\sum y_i x_{2i}}{n} = \frac{1217{,}2}{10} = 121{,}72;$$

$$\overline{yx_3} = \frac{\sum y_i x_{3i}}{n} = \frac{196{,}17}{10} = 19{,}617; \quad \overline{x_1x_2} = \frac{\sum x_{1i} x_{2i}}{n} = \frac{836}{10} = 83{,}6;$$

$$\overline{x_1x_3} = \frac{\sum x_{1i} x_{3i}}{n} = \frac{132{,}8}{10} = 13{,}28; \quad \overline{x_2x_3} = \frac{\sum x_{2i} x_{3i}}{n} = \frac{303{,}8}{10} = 30{,}38.$$

Розрахуємо середньоквадратичні відхилення:

$$\sigma_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n}} = \sqrt{\frac{126{,}9}{10}} = 3{,}5623;$$

$$\sigma_{x_1} = \sqrt{\frac{\Sigma(x_{1i} - \bar{x}_1)^2}{n}} = \sqrt{\frac{82{,}5}{10}} = 2{,}8723;$$

$$\sigma_{x_2} = \sqrt{\frac{\Sigma(x_{2i} - x_2)^2}{n}} = \sqrt{\frac{229{,}6}{10}} = 4{,}7917;$$

$$\sigma_{x_3} = \sqrt{\frac{\Sigma(x_{3i} - x_3)^2}{n}} = \sqrt{\frac{0{,}7889}{10}} = 0{,}2809.$$

Розраховуємо коефіцієнти парної кореляції між y та x_1, x_2, x_3 та між x_1 і x_2, x_3 та x_2 і x_3:

$$r_{yx_1} = \frac{\overline{yx_1} - \bar{y} \cdot \bar{x}_1}{\sigma_y \cdot \sigma_{x_1}} = \frac{54{,}99 - 8{,}2 \cdot 5{,}5}{3{,}5623 \cdot 2{,}8723} = 0{,}96658\,;$$

$$r_{yx_2} = \frac{\overline{yx_2} - \bar{y} \cdot \bar{x}_2}{\sigma_y \cdot \sigma_{x_2}} = \frac{121{,}72 - 8{,}2 \cdot 12{,}8}{3{,}5623 \cdot 4{,}7917} = 0{,}98185\,;$$

$$r_{yx_3} = \frac{\overline{yx_3} - \bar{y} \cdot \bar{x}_3}{\sigma_y \cdot \sigma_{x_3}} = \frac{19{,}617 - 8{,}2 \cdot 2{,}31}{3{,}5623 \cdot 0{,}2809} = 0{,}6740\,;$$

$$r_{x_1x_2} = \frac{\overline{x_1x_2} - \bar{x}_1 \cdot \bar{x}_2}{\sigma_{x_1} \cdot \sigma_{x_2}} = \frac{83{,}6 - 5{,}5 \cdot 12{,}8}{2{,}8723 \cdot 4{,}7917} = 0{,}9591\,;$$

$$r_{x_1x_3} = \frac{\overline{x_1x_3} - \bar{x}_1 \cdot \bar{x}_3}{\sigma_{x_1} \cdot \sigma_{x_3}} = \frac{13{,}28 - 5{,}5 \cdot 2{,}31}{2{,}8723 \cdot 0{,}2809} = 0{,}71267\,;$$

$$r_{x_2x_3} = \frac{\overline{x_2x_3} - \bar{x}_2 \cdot \bar{x}_3}{\sigma_{x_2} \cdot \sigma_{x_3}} = \frac{30{,}38 - 2{,}31 \cdot 12{,}8}{0{,}2809 \cdot 4{,}7917} = 0{,}60327\,.$$

Складемо матрицю коефіцієнтів парної кореляції. Матриця коефіцієнтів парної кореляції симетрична.

	y	x_1	x_2	x_3
y	1	0,96658	0,98185	0,6740
x_1	0,96658	1	0,9591	0,71267
x_2	0,98185	0,9591	1	0,60327
x_3	0,6740	0,71267	0,60327	1

Для трьох змінних x_1, x_2, x_3 система нормальних рівнянь в стандартизованому масштабі буде мати вигляд:

$$\begin{cases} r_{yx_1} = \beta_1 + \beta_2 r_{x_1x_2} + \beta_3 r_{x_1x_3} \\ r_{yx_2} = \beta_1 r_{x_2x_1} + \beta_2 + \beta_3 r_{x_2x_3} \\ r_{yx_3} = \beta_1 r_{x_3x_1} + \beta_2 x_3 x_2 + \beta_3 \end{cases} \quad (4.44)$$

Після підстановки відповідних числових значень коефіцієнтів парної кореляції **із матриці** будемо мати:

$$\begin{cases} 0{,}96658 = \beta_1 + 0{,}9591\beta_2 + 0{,}71267\beta_3 \\ 0{,}98185 = 0{,}9591\beta_1 + \beta_2 + 0{,}60327\beta_3 \\ 0{,}6740 = 0{,}71267\beta_1 + 0{,}60327\beta_2 + \beta_3 \end{cases}$$

Для рішення системи нормальних рівнянь (4.44) використаємо метод повного виключення Жордана-Гаусса. Цей метод полягає в тому, що обидві частини системи (4.44) множимо на обернену матрицю A^{-1}, в результаті чого визначається рішення:

$$X = A^{-1}\,B\,.$$

Якщо записати розширену матрицю (А/В) і використати до неї метод повного виключення, то одержимо:

$$(AA^{-1} \mid BA^{-1}) = (I \mid X),$$

де I – одинична матриця; X – вектор рішення системи; A – матриця системи нормальних рівнянь; B – вектор – стовпець вільних членів.

Якщо в розширеній матриці (А/В) перетворити матрицю А в одиночну, одночасно проводячи аналогічні перетворення матриці В, то на масці вектора – стовпця В буде вектор рішення системи X.

Вказані перетворення еквівалентні множенню системи (4.44) на A^{-1}.

Перейдемо безпосередньо до обчислень.

Розширена матриця має вигляд:

$$\left(\begin{array}{ccc|c} 1 & 0{,}9591 & 0{,}71267 & 0{,}96658 \\ 0{,}9591 & 1 & 0{,}60327 & 0{,}98185 \\ 0{,}71267 & 0{,}60327 & 1 & 0{,}67400 \end{array}\right)$$

Складемо таблицю 4.4 та зробимо відповідні перетворення.

Таким чином, $\beta_1 = 0.14787; \beta_2 = 0.78135; \beta_3 = 0.09725$.

Рівняння регресії в стандартизованому масштабі має вигляд:

$$\bar{t}_y = 0{,}14787t_1 + 0{,}78135t_2 + 0{,}097251t_3\,.$$

β – коефіцієнти дозволяють оцінювати пріоритетний вплив факторів на обсяг реалізації продукції. Найбільший вплив на обсяг реалізації виявляє фактор x_2 (кількість підприємств).

Таблиця 4.4 – Результати розрахунків

N ітерації	β_1	β_2	β_3	B	Контрольна сума
Вихідна система	[1] 0,9591 0,71267	0,9591 1 0,60327	0,71267 0,60327 1	0,96658 0,98185 0,6740	3,63839 3,54422 2,98994
1	1 0 0	0,9591 [0,080127] -0,080252	0,71267 -0,080252 0,49210	0,96658 0,054803 -0,014853	3,63839 0,05464 0,39697
2	1 0 0	0 1 0	1,67327 -1,00156 [0,41172]	0,31060 0,68395 0,04004	2,98436 0,68192 0,45170
3	1 0 0	0 1 0	0 0 1	0,14787 0,78135 0,097251	1,14787 1,78135 1,097251

Розрахуємо коефіцієнт множинної кореляції: $R = \sqrt{\beta_1 r_{yx_1} + \beta_2 r_{yx_2} + \beta_3 r_{yx_3}} =$

$\sqrt{0{,}14787 \cdot 0{,}96658 + 0{,}78135 \cdot 0{,}98185 + 0{,}097251 \cdot 0{,}674} = 0{,}9877$

Зробимо перехід від β – коефіцієнтів до коефіцієнтів регресії в натуральному масштабі:

$$a_1 = \beta_1 \frac{\sigma_y}{\sigma_{x_1}} = 0{,}14787 \cdot \frac{3{,}5623}{2{,}8723} = 0{,}18339 \quad ,$$

$$a_2 = \beta_2 \frac{\sigma_y}{\sigma_{x_2}} = 0{,}78135 \cdot \frac{3{,}5623}{4{,}7917} = 0{,}58088,$$

$$a_3 = \beta_3 \frac{\sigma_y}{\sigma_{x_3}} = 0{,}097251 \cdot \frac{3{,}5623}{0{,}2809} = 1{,}23331,$$

$$a_0 = \bar{y} - a_1\bar{x}_1 - a_2\bar{x}_2 - a_3\bar{x}_3$$

$$a_0 = 8.2 - 0{,}18339 \cdot 5{,}5 - 0{,}58088 \cdot 12{,}8 - 1{,}23331 \cdot 2{,}31 = -3{,}092855.$$

Рівняння регресії у натуральному масштабі:

$$y = -3{,}092855 + 0{,}18339x_1 + 0{,}58088x_2 + 1{,}23331x_3.$$

4.8 Метод найменших квадратів у матричній формі

Застосування матриць для визначення параметрів рівняння лінійної регресії дозволяє представити всі розрахунки в зручній і компактній формі.

Для обробки даних, представлених у матричній формі, також використовують метод найменших квадратів, оскільки в цьому випадку мінімізується сума відхилень, тобто:

$$S = (Y - X \cdot A)' \cdot (Y - X \cdot A) \to \min , \tag{4.45}$$

де $(Y - X \cdot A)$ – матриця відхилень (різниць) фактичних значень результативного показника від теоретичних; $(Y - X \cdot A)^{-1}$ – обернена матриця; X – матриця n спостережень незалежних факторів; Y – матриця фактичних значень результативного показника.

Елементи векторів $\text{A}(a_0, a_1, \ldots, a_m)$ визначають шляхом прирівнювання до нуля перших частинних похідних вказаної суми:

$$\frac{\partial s}{\partial a_0} = 0;\ \frac{\partial s}{\partial a_1} = 0;\ \frac{\partial s}{\partial a_2} = 0; \ldots; \frac{\partial s}{\partial a_m} = 0; \tag{4.46}$$

що приводить до системи нормальних рівнянь, записаної у матричній формі:

$$2 \cdot X'(Y - X \cdot A) = 0\ ; \tag{4.47}$$

звідки вектор параметрів рівняння регресії:

$$A = (X' \cdot X)^{-1}(X' \cdot Y), \tag{4.48}$$

де

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{31} & \ldots & x_{j1} & \ldots & x_{m1} \\ 1 & x_{12} & x_{22} & x_{32} & \ldots & x_{j2} & \ldots & x_{m2} \\ 1 & x_{13} & x_{23} & x_{33} & \ldots & x_{j3} & \ldots & x_{m3} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{1i} & x_{2i} & x_{3i} & \ldots & x_{ji} & \ldots & x_{mi} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{1n} & x_{2n} & x_{3n} & \ldots & x_{jn} & \ldots & x_{mn} \end{bmatrix}; Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \ldots \\ y_i \\ \ldots \\ y_n \end{bmatrix}$$

В матриці X додатково введено стовпець, всі елементи якого рівні 1, тобто умовно приймається, що в моделі є вільний член a_0.

Знайдемо матриці, що входять у вираз (4.48).

$$X' \cdot X = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & \dots & 1 \\ x_{11} & x_{12} & x_{13} & \dots & x_{1i} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2i} & \dots & x_{2n} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3i} & \dots & x_{3n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{j1} & x_{j2} & x_{j3} & \dots & x_{ji} & \dots & x_{jn} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & x_{m3} & \dots & x_{mi} & \dots & x_{mn} \end{bmatrix} \cdot \begin{bmatrix} 1 & x_{11} & x_{21} & x_{31} & \dots & x_{j1} & \dots & x_{m1} \\ 1 & x_{12} & x_{22} & x_{32} & \dots & x_{j2} & \dots & x_{m2} \\ 1 & x_{13} & x_{23} & x_{33} & \dots & x_{j3} & \dots & x_{m3} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1i} & x_{2i} & x_{3i} & \dots & x_{ji} & \dots & x_{mi} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & x_{3n} & \dots & x_{jn} & \dots & x_{mn} \end{bmatrix} =$$

$$= \begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} & \sum x_{3i} & \dots & \sum x_{ji} & \dots & \sum x_{mi} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{2i}x_{1i} & \sum x_{3i}x_{1i} & \dots & \sum x_{ji}x_{1i} & \dots & \sum x_{mi}x_{1i} \\ \sum x_{2i} & \sum x_{1i}x_{2i} & \sum x_{2i}^2 & \sum x_{3i}x_{2i} & \dots & \sum x_{ji}x_{2i} & \dots & \sum x_{mi}x_{2i} \\ \sum x_{3i} & \sum x_{1i}x_{3i} & \sum x_{2i}x_{3i} & \sum x_{3i}^2 & \dots & \sum x_{ji}x_{3i} & \dots & \sum x_{mi}x_{3i} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \sum x_{ji} & \sum x_{1i}x_{ji} & \sum x_{2i}x_{ji} & \sum x_{3i}x_{ji} & \dots & \sum x_{ji}^2 & \dots & \sum x_{mi}x_{ji} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \sum x_{mi} & \sum x_{1i}x_{mi} & \sum x_{2i}x_{mi} & \sum x_{3i}x_{mi} & \dots & \sum x_{ji}x_{mi} & \dots & \sum x_{mi}^2 \end{bmatrix}$$

$$X' \cdot Y = \begin{bmatrix} 1 & 1 & 1 & \dots & 1 & \dots & 1 \\ x_{11} & x_{12} & x_{13} & \dots & x_{1i} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2i} & \dots & x_{2n} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3i} & \dots & x_{3n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{j1} & x_{j2} & x_{j3} & \dots & x_{ji} & \dots & x_{jn} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & x_{m3} & \dots & x_{mi} & \dots & x_{mn} \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_i \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum y_i \cdot x_{1i} \\ \sum y_i \cdot x_{2i} \\ \sum y_i \cdot x_{3i} \\ \dots \\ \sum y_i \cdot x_{ji} \\ \dots \\ \sum y_n \cdot x_{mi} \end{bmatrix}$$

Із останніх двох виразів неважко одержати вже розглянуту систему нормальних рівнянь.

Для оцінки тісноти зв'язку між результативним показником та факторами x_j розрахуємо вектор залишків U, залишкову σ_u^2 та загальну σ_y^2 дисперсії:

$$U = Y - X \cdot A \tag{4.49}$$

$$\sigma_u^2 = \frac{U' \cdot U}{n-m-1} = \frac{\sum (y_i - \hat{y}_i)^2}{n-m-1}; \tag{4.50}$$

m – число незалежних змінних в моделі;

$$\sigma_y^2 = \frac{1}{n}\left(y'y - \frac{(\sum y_i)^2}{n}\right) = \frac{\sum (y_i - \bar{y})^2}{n}. \tag{4.51}$$

Коефіцієнт множинної кореляції:

$$R = \sqrt{1 - \frac{\sigma_u^2}{\sigma_y^2}}. \tag{4.52}$$

Довірчі границі та значущість коефіцієнта множинної кореляції оцінюється через дисперсію:

$$\sigma_R = \frac{1 - R_1^2}{\sqrt{n-m-1}}; \tilde{R} = R \pm t_{\alpha,k} \cdot \sigma_R; \ t_R = \frac{R}{\sigma_R}, \tag{4.53}$$

де k – число ступенів свободи;

$t_{\alpha,k}$ – нормоване значення функції Стьюдента в залежності від заданого рівня довірчої імовірності та числа ступенів свободи k.

Дисперсія коефіцієнта регресії a_j визначається по формулі:

$$\sigma_{a_j}^2 = \sigma_u^2 \cdot b_{jj}, \quad \text{або } \sigma_{a_j}^2 = \sigma_u \cdot \sqrt{b_{jj}}, \tag{4.54}$$

де b_{jj} – діагональний елемент матриці $(X' \cdot X)^{-1}$.

Тоді довірчі границі параметра моделі a_j для генеральної сукупності складуть:

$$\tilde{a}_j = a_j \pm t_{\alpha,k} \cdot \sigma_{a_j} \tag{4.55}$$

Значущість коефіцієнта регресії оцінюється співвідношенням:

$$t_{a_j} = \frac{|a_j|}{\sigma_{a_j}}. \tag{4.56}$$

Якщо $t_{a_j} > t_{таб}$, то параметр моделі a_j значущий, тобто не випадковий. В протилежному випадку фактор a_j необхідно виключити із рівняння регресії.

Довірчі границі базисних середніх значень $\hat{y}_i$ і довірчі зони рівняння визначаються по формулі:

$$\Delta y_i = t_{\alpha,k} \cdot \sigma_u \cdot \sqrt{X_i (X^{'} X)^{-1} X_i^{'}}\ , \qquad (4.57)$$

де X_i – вектор – рядок; i – номер рядка матриці X :

$$X_i = (1, x_{1i}, x_{2i}, ..., x_{ji}, ..., x_{mi})\ , \qquad (4.58)$$

а границі кожного базисного середнього:

$$\tilde{y}_i = \hat{y}_i \pm \Delta y_i\ . \qquad (4.59)$$

Середні значення прогнозного показника y_p для прогнозних значень факторів $X_p = (1, x_{1p}, x_{2p}, ..., x_{jp}, ..., x_{mp})$ розраховуються по рівнянню регресії шляхом підстановки відповідного значення X_p і обчислення:

$$y_p = a_0 + a_1 x_{1p} + ... + a_j x_{jp} + ... + a_m x_{mp} \qquad (4.60)$$

або у матричній формі:

$$Y_p = X_p \cdot A\ , \qquad (4.61)$$

де A вектор параметрів моделі.

Довірчі інтервали прогнозу обчислюють за формулою:

$$\Delta y_p = t_{\alpha,k} \cdot \sigma_u \cdot \sqrt{1 + X_p \cdot (X^{'} \cdot X)^{-1} \cdot X_p^{'}}\ . \qquad (4.62)$$

Відповідно, довірчою зоною значень прогнозу буде:

$$\hat{y}_p - \Delta y_p \le \tilde{y}_p \le \hat{y}_p + \Delta y_p\ . \qquad (4.63)$$

Необхідно зазначити, що значення коефіцієнта $a_j\ (j = \overline{1,m})$ в рівнянні регресії залежить від прийнятих одиниць вимірювання величин Y , X_j . Тому для інтерпретації коефіцієнта a_j зручно використовувати коефіцієнт еластичності, який розраховується за формулою:

$$KE_j = a_j \frac{\bar{x}_j}{\bar{y}}\ ; \quad j = \overline{1,m} \qquad (4.65)$$

Коефіцієнт еластичності показує: на скільки процентів в середньому зміниться величина y із зміною величини x_j на 1%.

Приклад 4.3. По даних таблиці 4.5 (x_{1i} – витрати праці, x_{2i} – виробничий капітал, y_i – обсяг виробництва продукції) побудувати та дослідити виробничу функцію Кобба-Дугласа виду: $y = a_0 + a_1 x_1 + .a_2 x_2 + u$. Оцінити її параметри та, користуючись методом математичної екстраполяції, знайти прогнозоване значення обсягу виробництва, якщо прогнозоване значення витрат праці $x_{1p} = 8$ млн. л/год., а виробничого капіталу $x_{2p} = 18$ млн. грн.

Таблиця 4.5 – Вхідні дані

Витрати праці, тис. людино-днів	Витрати виробничого капіталу, млн. грн.	Обсяг виробництва продукції, млн. грн.
1,8	4,7	10,2
2,4	5,6	13,4
3,2	7,1	16,2
3,6	7,0	17,4
3,5	7,8	18,5
3,9	10,1	19,1
4,6	12,5	20,4
5,5	14,0	22,5
5,4	15,2	24,0
7,0	16,0	27,0

Розв'язок. Формуємо матриці X та Y:

$$X = \begin{bmatrix} 1 & 1{,}8 & 4{,}7 \\ 1 & 2{,}4 & 5{,}6 \\ 1 & 3{,}2 & 7{,}1 \\ 1 & 3{,}6 & 7{,}0 \\ 1 & 3{,}5 & 7{,}8 \\ 1 & 3{,}9 & 10{,}1 \\ 1 & 4{,}6 & 12{,}5 \\ 1 & 5{,}5 & 14{,}0 \\ 1 & 5{,}4 & 15{,}2 \\ 1 & 7{,}0 & 16{,}0 \end{bmatrix}, \quad Y = \begin{bmatrix} 14{,}2 \\ 13{,}4 \\ 16{,}2 \\ 17{,}4 \\ 17{,}3 \\ 19{,}1 \\ 20{,}4 \\ 22{,}5 \\ 24{,}0 \\ 27{,}0 \end{bmatrix}.$$

Знаходимо добутки матриць: $X' \cdot X$ та $X' \cdot Y$.

$$X' \times X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1{,}8 & 2{,}4 & 3{,}2 & 3{,}6 & 3{,}5 & 3{,}9 & 4{,}6 & 5{,}5 & 5{,}4 & 7{,}0 \\ 4{,}7 & 5{,}6 & 7{,}1 & 7{,}0 & 7{,}8 & 10{,}1 & 12{,}5 & 14{,}0 & 15{,}2 & 16{,}0 \end{bmatrix} \times \begin{bmatrix} 1 & 1{,}8 & 4{,}7 \\ 1 & 2{,}4 & 5{,}6 \\ 1 & 3{,}2 & 7{,}1 \\ 1 & 3{,}6 & 7{,}0 \\ 1 & 3{,}5 & 7{,}8 \\ 1 & 3{,}9 & 10{,}1 \\ 1 & 4{,}6 & 12{,}5 \\ 1 & 5{,}5 & 14{,}0 \\ 1 & 5{,}4 & 15{,}2 \\ 1 & 7{,}0 & 16{,}0 \end{bmatrix} =$$

$$= \begin{bmatrix} 10 & 40{,}9 & 100 \\ 40{,}9 & 189{,}23 & 465{,}0901 \\ 100 & 465{,}0901 & 1155 \end{bmatrix}$$

$$X' \times Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1{,}8 & 2{,}4 & 3{,}2 & 3{,}6 & 3{,}5 & 3{,}9 & 4{,}6 & 5{,}5 & 5{,}4 & 7{,}0 \\ 4{,}7 & 5{,}6 & 7{,}1 & 7{,}0 & 7{,}8 & 10{,}1 & 12{,}5 & 14{,}0 & 15{,}2 & 16{,}0 \end{bmatrix} \times \begin{bmatrix} 10{,}2 \\ 13{,}4 \\ 16{,}2 \\ 17{,}4 \\ 18{,}5 \\ 19{,}1 \\ 20{,}4 \\ 22{,}5 \\ 24{,}0 \\ 27{,}0 \end{bmatrix} =$$

$$= \begin{bmatrix} 188{,}5 \\ 838{,}03 \\ 2059{,}15 \end{bmatrix}$$

Знаходимо обернену матрицю $(X' \cdot X)^{-1}$ методом Гауса

№ ітерації	Вхідна матриця			Одинична матриця			Контрольна сума
0	10	40,9	100	1	0	0	151,9
	40,9	189,23	465,09	0	1	0	696,22
	100	465,09	1155	0	0	1	1721,09
1	1	4,09	10	0,1	0	0	15,19
	0	21,949	56,09	-4,09	1	0	74,949
	0	56,09	155	-10	0	1	202,09
2	1	0	-0,45187	0,86213	-0,18634	0	1,22392
	0	1	2,555469	-0,18634	0,04556	0	3,41469
	0	0	11,66372	0,45187	-2,55547	1	10,56012
3	1	0	0	0,87964	-0,28534	0,03874	1,63304
	0	1	0	-0,28534	0,60545	-0,21910	1,10101
	0	0	1	0,03874	-0,21910	0,08574	0,90538

Запишемо обернену матрицю $(X' \cdot X)^{-1}$

$$(X' \cdot X)^{-1} = \begin{bmatrix} 0{,}87964 & -0{,}28534 & 0{,}03874 \\ -0{,}28534 & 0{,}60545 & -0{,}21910 \\ 0{,}03874 & -0{,}21910 & 0{,}08574 \end{bmatrix}.$$

Вектор параметрів моделі:

$$A = (X' \cdot X)^{-1} \cdot (X' \cdot Y) = \begin{bmatrix} 0{,}8796426 & -0{,}285348 & 0{,}03874308 \\ -0{,}285348 & 0{,}6054632 & -0{,}2190997 \\ 0{,}03874308 & -0{,}2190997 & 0{,}08573746 \end{bmatrix} \times$$

$$\times \begin{bmatrix} 188{,}5 \\ 838{,}03 \\ 2059{,}15 \end{bmatrix} = \begin{bmatrix} 6{,}460289 \\ 2{,}449219 \\ 0{,}2372437 \end{bmatrix}$$

Запишемо векторну модель, побудовану по вибіркових даних:

$$Y = 6{,}460289 + 2{,}449219x_1 + 0{,}2372437x_2 + U$$

Вектор залишків у матричній формі:

$$U = Y - \hat{Y} = Y - X \cdot A = \begin{bmatrix} 10{,}2 \\ 13{,}4 \\ 16.2 \\ 17{,}4 \\ 18{,}5 \\ 19{,}1 \\ 20{,}4 \\ 22{,}5 \\ 24{,}0 \\ 27{,}0 \end{bmatrix}, - \begin{bmatrix} 1 & 1{,}8 & 4{,}7 \\ 1 & 2{,}4 & 5{,}6 \\ 1 & 3{,}2 & 7{,}1 \\ 1 & 3{,}6 & 7{,}0 \\ 1 & 3{,}5 & 7{,}8 \\ 1 & 3{,}9 & 10{,}1 \\ 1 & 4{,}6 & 12{,}5 \\ 1 & 5{,}5 & 14{,}0 \\ 1 & 5{,}4 & 15{,}2 \\ 1 & 7{,}0 & 16{,}0 \end{bmatrix} \cdot \begin{bmatrix} 6{,}460289 \\ 2{,}449219 \\ 0{,}2372437 \end{bmatrix} = \begin{bmatrix} -0{,}7839289 \\ -0{,}2669792 \\ 0{,}217782 \\ 0{,}4618168 \\ 0{,}4169445 \\ 0{,}6915989 \\ -0{,}2922421 \\ -0{,}7524033 \\ 0{,}7078266 \\ -0{,}4007187 \end{bmatrix}.$$

Знаходимо дисперсію залишків (залишкову) та загальну дисперсію.

$$\sigma_u^2 = \frac{\sum_{i=1}^{n} u_i^2}{n-m-1} = \frac{U' \cdot U}{n-m-1} = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-m-1} =$$

$$= \frac{(-0{,}7839289)^2 + (-0{,}2669792)^2 + \ldots + 0{,}7078266^2 + (-0{,}4007187)^2}{10-2-1} = 0{,}4159697$$

Загальна дисперсія:

$$\sigma_y^2 = \frac{1}{n}\left(y' \cdot y - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} \right) = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}{n} = 20,84849$$

$$y' \cdot y = 10,2^2 + 13,4^2 + 16,2^2 + \ldots + 24,0^2 + 27,0^2$$

Тіснота зв'язку:

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n} u_i^2 \Big/ (n-m-1)}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2 \Big/ n}} = \sqrt{1 - \frac{\hat{\sigma}_u^2}{\sigma_y^2}} = \sqrt{1 - \frac{0,9899738}{20,84849}} = 0,9899738$$

Довірчі границі коефіцієнта множинної кореляції R:

$$\sigma_r = \frac{1 - R^2}{\sqrt{n-m-1}} = \frac{1 - 0,9899738}{\sqrt{10-2-1}} = \frac{0,0199518}{2,64575} = 0,00754107$$

$$t_R = \frac{R}{\sigma_R} = \frac{0,9899738}{0,00754107} = 131,27753$$

$$\Delta R = R \pm t\alpha,\kappa \cdot \sigma_R = 0,9899738 \pm 2,31 \cdot 0,00754107 = 0,9899738 \pm 0,01742$$

Оскільки $t_R > t_{\alpha,\kappa}$, то з ймовірністю $P = 0,95$ можна стверджувати, що коефіцієнт множинної кореляції значимий.

Оцінка значимості параметрів моделі по t – критерію Стьюдента:

$$\sigma_{\alpha_j} = \sqrt{\sigma_u^2 \cdot b_{jj}};$$

$$\sigma\alpha_0 = \sqrt{0,4159697 \cdot 0,8796426} = 0,6049;$$

$$\sigma_{\alpha_1} = \sqrt{0,4159697 \cdot 0,6054632} = 0,50185;$$

$$\sigma\alpha_2 = \sqrt{0,4159697 \cdot 0,08573746} = 0,18888;$$

$$t_{\alpha_0} = \frac{|a_0|}{\sigma_{a_0}} = \frac{6,460289}{0,6049} = 10,68;$$

$$t_{\alpha_1} = \frac{|a_1|}{\sigma_{a_1}} = \frac{2,449219}{0,50185} = 4,88;$$

$$t_{a_2} = \frac{|a_2|}{\sigma_{a_2}} = \frac{0,2372437}{0,18888} = 1,256;$$

$t_{a_0}, t_{a_1} > t_{\alpha,\kappa}$, а $t_{a_2} < t_{\alpha,\kappa}$; отже параметри a_0, a_1 – значимі, а a_2 – не значимі.

Оцінка адекватності моделі по F – критерію Фішера:

$$F_p = \frac{\sigma_y^2 - \sigma_u^2}{\sigma_u^2} = \frac{20,84849 - 0,4159697}{0,4159697} = 49,1202;\ F_{\alpha,\kappa_1,\kappa_2} = 5,32;$$

$$F_p > F_{\alpha,\kappa_1,\kappa_2}.$$

Отже з ймовірністю $p = 1 - \alpha = 0,95$ можна стверджувати, що економетрична модель адекватно описує математичне явище.

Обчислимо середнє значення прогнозу:

$$Y_p = X_p \cdot A = [1 \quad 8 \quad 18] \cdot \begin{bmatrix} 6,460289 \\ 2,449219 \\ 0,2372432 \end{bmatrix} = 30,32443;$$

Довірчі границі прогнозу:

$$\Delta y_p = \pm t_{\alpha,\kappa} \cdot \sigma_u \cdot \sqrt{1 + X_p \cdot (X'X)^{-1} \cdot X'_p} = 2,31 \cdot \sqrt{0,4159697} \times$$

$$\times \sqrt{1 + [1 \quad 8 \quad 18] \cdot \begin{bmatrix} 0,8796426 & -0,285348 & 0,03874308 \\ -0,285348 & 0,6054632 & -0,2190997 \\ 0,03874308 & -0,2190997 & 0,08573746 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 8 \\ 18 \end{bmatrix}} = 2,177784.$$

$$30,32443 - 2,177784 \le \tilde{y}_p \le 30,32443 + 2,177784$$

$$28,146646 \le \tilde{y}_p \le 32,502214$$

Коефіцієнти еластичності:

$$KE_1 = a_1 \cdot \frac{\bar{x}_1}{\bar{y}} = 2,449219 \cdot \frac{4,09}{18,85} = 0,5314$$

$$KE_2 = a_2 \cdot \frac{\bar{x}_2}{\bar{y}} = 0,2372437 \cdot \frac{10}{18,85} = 0,1258$$

Таким чином, при зростанні фактора x_1 (витрати праці) на 1% показник y (обсяг виробництва) зростає на 0,5314%, а при зростанні фактора x_2 (виробничий капітал) на 1% показник y (обсяг виробництва) зростає на 0,1258%.

Розрахунки параметрів багатофакторних економетричних моделей доцільно проводити у середовищі MicrosoftOfficeExcel з використанням деяких функцій користувача:

МОБР – возвращает обратную матрицу (матрица хранится в массиве);

ТРАНСП – преобразует вертикальный массив в горизонтальный, или наоборот; МУМНОЖ – возвращает произведение матриц (матрицы хранятся в массиве);

МОПРЕД – возвращает определитель матрицы (матрица хранится в массиве);

ЛИНЕЙН – возвращает параметры линейного приближения по методу наименьших квадратов;

КОРРЕЛ – возвращает коэффициенты корреляции между двумя множествами данных; ТЕНДЕНЦИЯ – возвращает значения в соответствии с линейной аппроксимацией по методу наименьших квадратов и др..

4.9 Множинна нелінійна регресія

Степенева множинна регресія

Розглянуті раніше множинні лінійні регресійні моделі є частковим випадком більш загальних нелінійних моделей.

В економічному аналізі використовують рівняння більш складного криволінійного типу, які краще описують багатофакторні економічні взаємозв'язки. Одним із таких рівнянь є степенева функція:

$$y = a_0 x_1^{a_1} \cdot \ldots \cdot x_m^{a_m} = a_0 \prod_{j=1}^{m} x_j^{a_j} . \qquad (4.66)$$

Прологарифмуємо вираз (4.66):

$$\ln y = \ln a_0 + a_1 \ln x_1 + a_2 \ln x_2 + \ldots + a_m \ln x_m$$

і зробимо заміну нелінійних елементів на лінійні:

$$x'_j = \ln x_j, (j = 1,\overline{m}); y' = \ln y \text{ та } a'_0 = \ln a_0,$$

Приклад 4.4. Визначіть параметри степеневої моделі $y = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} + u$ залежності витрат на споживання від рівня доходів x_1 та заощаджень x_2. (таблиця 4.6).

Таблиця 4.6 – Вхідні дані та результати розрахунків

Дохід, x_1	Заощадження, x_2,	Витрати на споживання, y	$\hat{y}_i$	$(y_i - \hat{y})^2$	$(y_i - \bar{y})^2$
7,08	12,48	100,2	89,44	115,839	7907,31361
8,28	15,34	112,3	119,48	51,5276	5901,78515
10,47	15,04	128,4	156,23	774,404	3687,29207
8,66	13,25	130,2	117,04	173,311	3471,92899
9,65	12,90	105,8	131,27	648,825	6942,73515
11,56	13,02	180,5	163,37	293,619	74,3574556
13,65	17,57	250,7	231,38	373,285	3791,71174
14,59	18,04	247,5	253,74	38,9086	3407,86515
15,35	15,06	230,7	246,03	235,055	1728,64053
12,31	14,31	200,5	183,41	291,983	129,434379
16,06	16,73	270,6	273,71	9,67348	6638,48899
14,03	18,42	260,7	244,80	252,96	127,2934
13,88	17,45	240,5	235,19	28,1597	2639,58822
155,57	199,61	2458,6	2445,09	3287,55	51444,4031

Розв'язок. Проводимо лінеаризацію вхідних даних:

$$\ln y = \ln a_0 + a_1 \ln x_1 + a_2 \ln x_2$$

$$y' = \ln y;\ a_0' = \ln a_0;\ x_1' = \ln x_1;\ x_2' = \ln x_2.$$

1. Формуємо матриці X, Y:

$$X = \begin{bmatrix} 1 & 1{,}95727 & 2{,}52413 \\ 1 & 2{,}11384 & 2{,}73046 \\ 1 & 2{,}34851 & 2{,}71071 \\ 1 & 2{,}15872 & 2{,}58400 \\ 1 & 2{,}26296 & 2{,}55723 \\ 1 & 2{,}44755 & 2{,}56649 \\ 1 & 2{,}61374 & 2{,}86619 \\ 1 & 2{,}68034 & 2{,}89260 \\ 1 & 2{,}73112 & 2{,}71204 \\ 1 & 2{,}51041 & 2{,}64830 \\ 1 & 2{,}77633 & 2{,}81720 \\ 1 & 2{,}64120 & 2{,}91344 \\ 1 & 2{,}63045 & 2{,}85934 \end{bmatrix} \qquad Y = \begin{bmatrix} 4{,}60717 \\ 4{,}72117 \\ 4{,}85515 \\ 4{,}86907 \\ 4{,}66155 \\ 5{,}19573 \\ 5{,}52426 \\ 5{,}51141 \\ 5{,}44112 \\ 5{,}30082 \\ 5{,}60064 \\ 5{,}56337 \\ 5{,}48272 \end{bmatrix}.$$

2. Знаходимо добутки матриць $X' \cdot X$ та $X' \cdot Y$:

$$X' \cdot X = \begin{bmatrix} 13 & 31.87644 & 35.38219 \\ 31.87644 & 78.98463 & 87.0729 \\ 35.38219 & 87.0729 & 96.52932 \end{bmatrix}, \quad X' \cdot Y = \begin{bmatrix} 67.33418 \\ 166.2392 \\ 183.7523 \end{bmatrix}.$$

3. Знаходимо обернену матрицю $(X' \cdot X)^{-1}$:

$$(X' \cdot X)^{-1} = \begin{bmatrix} 36.47272 & 3.25670 & -16.30648 \\ 3.25670 & 2.55355 & -3.49711 \\ -16.30648 & -3.49711 & 9.14189 \end{bmatrix}.$$

4. Знаходимо вектор параметрів моделі A:

$$A = (X' \cdot X)^{-1} \cdot (X' \cdot Y) = \begin{bmatrix} 0.90113 \\ 1.18524 \\ 0.50419 \end{bmatrix}, \qquad a_0 = \exp(0.90113) = 2.46239.$$

5. Запишемо економетричну модель:

$$y = 2.46239 \cdot x_1^{1.18524} \cdot x_2^{0.50419} + u.$$

6. Розрахуємо теоретичні значення $\hat{y}$:

$$\ln \hat{y}_1 = 0.90113 + 1.18524 \cdot 1.95727 + 0.50419 \cdot 2.52413 = 4.4935,$$

$$\hat{y}_1 = \exp(\ln \hat{y}_1) = 89.44;$$

$$\ln \hat{y}_2 = 0.90113 + 1.18524 \cdot 2.11384 + 0.50419 \cdot 2.73046 = 4.7831;$$

$$\hat{y}_2 = \exp(\ln \hat{y}_2) = 119.48; \text{ і т.д.}$$

Теоретичні значення $\hat{y}$ заносимо у таблицю 4.6.

7. Розраховуємо залишкову та загальну дисперсії:

$$\sigma_u^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - m - 1} = \frac{3287.55}{13 - 2 - 1} = 328.755,$$

$$\sigma_y^2 = \frac{\sum (y_i - \bar{y})^2}{n} = \frac{51444.40}{13} = 3957.262.$$

Розрахувавши дисперсії σ_u^2 та σ_y^2, знаходимо тісноту зв'язку R та проводимо оцінку адекватності моделі по F – критерію Фішера:

$$R = \sqrt{1 - \frac{328.755}{3957.2617}} = 0.95756,$$

$$F_p = \frac{\delta^2}{\sigma_u^2} = \frac{\sigma_y^2 - \sigma_u^2}{\sigma_u^2} = \frac{(3957.262 - 328.755)}{328.755} = 11.037$$

Оскільки $F_p > F_{\alpha, k_1, k_2}$, то модель адекватна статистичним даним.

Розглянемо іншу методику розрахунку параметрів моделі.

Записуємо систему нормальних рівнянь та розрахуємо відповідні суми:

$$\begin{cases} n \cdot \ln a_0 + a_1 \cdot \sum \ln x_1 + a_2 \cdot \sum \ln x_2 = \sum \ln y \\ a_0 \cdot \sum \ln x_1 + a_1 \cdot \sum (\ln x_1)^2 + a_2 \cdot \sum \ln x_2 \cdot \ln x_1 = \sum \ln y \cdot \ln x_1 \\ a_0 \cdot \sum \ln x_2 + a_1 \cdot \sum \ln x_1 \cdot \ln x_2 + a_2 \cdot \sum (\ln x_2)^2 = \sum \ln y \cdot \ln x_2 \end{cases}$$

$$\begin{cases} 13 \ln a_0 + 31.87644 a_1 + 35.38219 a_2 = 67.33418 \\ 31.87644 a_0 + 78.98463 a_1 + 87.0729 a_2 = 166.2392 \\ 35.38212 a_0 + 87.0728 a_1 + 96.52932 a_2 = 183.7523 \end{cases}$$

Систему розраховуємо методом оберненої матриці:

$$A = X^{-1} \cdot Y = \begin{bmatrix} 36.47272 & 3.256699 & -16.30646 \\ 3.256699 & 2.553548 & -3.497108 \\ -16.30648 & -3.49711 & 9.141889 \end{bmatrix} \cdot \begin{bmatrix} 67.33418 \\ 166.2392 \\ 183.7523 \end{bmatrix} = \begin{bmatrix} 0.90113 \\ 1.18524 \\ 0.50419 \end{bmatrix}$$

$$a_0 = \exp(0.901132) = 2.462388$$

Запишемо економетричну модель:

$$y = 2.46239 \cdot x_1^{1.18524} \cdot x_2^{0.50419} \cdot u$$

Розрахуємо теоретичні значення витрат на споживання:

$$\ln \hat{y}_1 = 0.90113 + 1.18524 \cdot 1.957274 + 0.50419 \cdot 2.524127 = 4.4935;$$

$$\hat{y}_1 = \exp(\ln \hat{y}_1) = 89.43713$$

$$\ln \hat{y}_2 = 0.901132 + 1.185234 \cdot 2.113842 + 0.504185 \cdot 2.730464 = 4.7831$$

$$\hat{y}_2 = \exp(4.7831) = 119.4783 \text{ і т.д.}$$

Теоретичні значення $\hat{y}_i$ заносимо у таблицю 4.6.

Розрахуємо $\sum (y_i - \hat{y}_i)^2$ та знаходимо тісноту зв'язку.

$$R = \sqrt{1 - \frac{3287.5510}{39572.617}} = 0.95756$$

Висновки. У степеневій моделі показники степеня є коефіцієнтами еластичності. Отже, при зростанні рівня доходів x_1 на 1% рівень витрат на споживання зростає на 1.1852%; при зростанні рівня заощаджень на 1% рівень споживання зростає на 0.504185%.

Показникова множинна регресія

$$y = a_0 \cdot a_1^{x_1} \cdot a_2^{x_2} \cdot \ldots \cdot a_m^{x_m}$$

Прологарифмуємо функцію:

$$\ln y = \ln a_0 + a_1 \ln x_1 + a_2 \ln x_2 + \ldots + a_m \ln x_m$$

Зробимо заміну $y' = \ln y$ та $a_0', a_1', a_2', \ldots, a_m$ відповідно на $\ln a_0, \ln a_1, \ln a_2, \ldots, \ln a_m$, одержимо лінеаризовану функцію. Для знаходження її параметрів $A = (\exp a_0, \exp a_1 \exp a_2, \ldots, \exp a_m)$ складемо та розв'язуємо систему нормальних рівнянь:

$$\begin{cases} n \cdot \ln a_0 + \ln a_1 \sum x_1 + \ln a_2 \sum x_2 + \ldots + \ln a_m \cdot \sum x_m = \sum \ln y \\ \ln a_0 \sum x_1 + \ln a_1 \sum x_1^2 + \ln a_2 \sum x_2 \cdot x_1 + \ldots + \ln a_m \sum x_m \cdot x_1 = \sum \ln y \cdot x_1 \\ \ln a_0 \cdot \sum x_2 + \ln a_1 \sum x_1 \cdot x_2 + \ln a_2 \sum x_2^2 + \ldots + \ln a_m \sum x_m \cdot x_2 = \sum \ln y \cdot x_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \ln a_0 \sum x_m + \ln a_1 \sum x_1 \cdot x_m + \ln a_2 \sum x_2 \cdot x_m + \ldots + \ln a_m \sum x_m^2 = \sum \ln y \cdot \ln x_m \end{cases}$$

Гіперболічна множинна регресія

$$y = a_0 + a_1 \sum \frac{1}{x_1} + a_2 \sum \frac{1}{x_2} + \ldots + a_m \sum \frac{1}{x_m}$$

Проведемо лінеаризацію функції шляхом заміни нелінійних елементів на лінійні: $x_1', x_2', \ldots, x_m'$ відповідно на $\frac{1}{x_1}, \frac{1}{x_2}, \ldots, \frac{1}{x_m}$, складемо систему лінійних рівнянь:

$$\begin{cases} na_0 + a_1 \sum \frac{1}{x_1} + a_2 \sum \frac{1}{x_2} + \ldots + a_m \cdot \sum \frac{1}{x_m} = \sum y \\ a_0 \sum \frac{1}{x_1} + a_1 \sum \frac{1}{x_1^2} + a_2 \sum \frac{1}{x_2 \cdot x_1} + \ldots + a_m \sum \frac{1}{x_m \cdot x_1} = \sum \frac{y}{x_{11}} \\ a_0 \cdot \sum \frac{1}{x_2} + a_1 \sum \frac{1}{x_1 x_2} + a_2 \sum \frac{1}{x_2^2} + \ldots + a_m \sum \frac{1}{x_m \cdot x_2} = \sum \frac{y}{x_2} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_0 \sum \frac{1}{x_m} + a_1 \sum \frac{1}{x_1 \cdot x_m} + a_2 \sum \frac{1}{x_2 \cdot x_m} + \ldots + a_m \sum \frac{1}{x_m^2} = \sum \frac{y}{x_m} \end{cases}$$

Параболічна множина регресія

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_m x_m + b_1 x_1^2 + b x_m^2$$

Параметри $A(a_0, a_1, a_2, \ldots, a_m, b_1, b_2, \ldots, b_m)$ знайдемо методом найменших квадратів, розв'язавши систему лінійних рівнянь:

$$
\begin{cases}
na_0 + a_1\sum x_1 + a_2\sum x_2 + \ldots + a_m\sum x_m + b_1\sum x_1^2 + b_2\sum x_2^2 + \ldots + b_m\sum x_m^2 = \sum y \\
a_0\sum x_1 + a_1\sum x_1^2 + a_2\sum x_2 x_1 + \ldots + a_m\sum x_m x_1 + b_1\sum x_1^3 + b_2\sum x_2^2\cdot x_1 + \ldots + b_m\sum x_m^2\cdot x_1 = \sum y\cdot x_1 \\
a_0\sum x_2 + a_1\sum x_1\cdot x_2 + a_2\sum x_2^2 + \ldots + a_m\sum x_m\cdot x_2 + b_1\sum x_1^2\cdot x_2 + b_2\sum x_2^3 + \ldots + b_m\sum x_m^2\cdot x_2 = \sum y\cdot x_2 \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
a_0\sum x_m + a_1\sum x_1\cdot x_m + a_2\sum x_2\cdot x_m + \ldots + a_m\sum x_m^2 + b_1\sum x_1^2\cdot x_m + b_2\sum x_2^2\cdot x_m + \ldots + b_m\sum x_m^3 = \sum y\cdot x_m \\
a_0\sum x_1^2 + a_1\sum x_1^3 + a_2\sum x_2\cdot x_1^2 + \ldots + a_m\sum x_m\cdot x_1^2 + b_1\sum x_1^4 + b_2\sum x_2^2\cdot x_1^2 + b_m\sum x_m^2\cdot x_1^2 = \sum y\cdot x_1^2 \\
a_0\sum x_2^2 + a_1\sum x_1\cdot x_2^2 + a_2\sum x_2^3 + \ldots + a_m\sum x_m\cdot x_2^2 + b_1\sum x_1^2\cdot x_2^2 + b_2\sum x_2^4 + \ldots + b_m\sum x_m^2\cdot x_2^2 = \sum y\cdot x_2^2 \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
a_0\sum x_m^2 + a_1\sum x_1\cdot x_m^2 + a_2\sum x_2\cdot x_m^2 + \ldots + a_m\sum x_m^3 + b_1\sum x_1^2\cdot x_m^2 + b_2\sum x_2^2\cdot x_m^2 + \ldots + b_m\sum x_m^4 = \sum y\cdot x_m^2
\end{cases}
$$

Логарифмічні та напівлогарифмічні функції

$$y = a_0 + a_1\ln x_1 + a_2\ln x_2 + \ldots + a_j\ln x_j + \ldots + a_m\ln x_m$$

$$\ln y = a_0 + a_1\ln x_1 + a_2\ln x_2 + \ldots + a_j\ln x_j + \ldots + a_m\ln x_m$$

$$\ln y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_j x_j + \ldots + a_m x_m$$

Лінеаризація цих функцій проводиться шляхом заміни $\ln x_j, \ln y$ на x'_j, y'.

Можлива також комбінація лінійних елементів функції з їх нелінійними елементами.

Наприклад: $y = a_0 + a_1\ln x_1 + a_2 x_2^2 + a_3\cdot x_1\cdot x_2 + \ldots + a_j/x_j + \ldots + a_m\ln x_m$

Ліва частина функції може приймати також значення $\frac{y}{x}$

Питання для самоконтролю:

1. В чому особливість лінійної багатофакторної регресійної моделі?
2. Розрахунок параметрів лінійної багатофакторної регресійної моделі.
3. В чому переваги матричного способу знаходження параметрів лінійної багатофакторної регресійної моделі?
4. Розрахунок довірчих інтервалів для значень результативного показника?
5. Яким чином оцінюється значущість параметрів моделі?
6. Яким чином розраховується прогнозне значення результативного показника?
7. Розрахунок довірчих інтервалів для прогнозного значення результативного показника?
8. Розрахунок параметрів нелінійної багатофакторної регресійної моделі.

Тема 5

МУЛЬТИКОЛІНЕАРНІСТЬ

5.1 Поняття та наслідки мультиколінеарності

Однією з умов застосування метода найменших квадратів для обробки різних експериментальних даних є відсутність мультиколінеарності між незалежними змінними.

В економетричну модель з багатьма змінними включають велику кількість факторів, які впливають на результативний показник. Серед цих факторів необхідно відібрати найбільш суттєві та виключити з моделі несуттєві. При цьому, виключенню з моделі підлягають ті фактори, які при парному корелюванні між собою дають високий лінійний коефіцієнт кореляції, який по абсолютній величині перевищує 0,85 ($r>0,85$). Наявність такого лінійного зв'язку між двома факторами називається колінеарністю, а між декількома факторами – мультиколінеарністю. Таким чином, **мультиколінеарність** означає, що в багатофакторній регресійній моделі дві або більше незалежних змінних (факторів) пов'язані між собою лінійною залежністю, тобто мають високий ступінь кореляції ($r_{x_j x_{j+s}} > 0{,}85$).

Термін «мультиколінеарність» вперше впровадив Р. Фріш.

Регресійне рівняння задовільно описує рух залежної змінної, якщо коефіцієнт множинної кореляції досить високий, а кореляція між незалежними факторами незначна. Мультиколінеарність незалежних змінних веде до зміщення оцінок параметрів і, отже, до неможливості коректної інтерпретації результатів.

Таким чином, наявність мультиколінеарності негативно впливає на кількісні характеристики економетричної моделі або робить її побудову взагалі неможливою. Так, якщо два колінеарні вектори змінюються в одному напрямку, то майже неможливо оцінити окремий вплив кожного з них на результативний показник, оскільки кожний з цих факторів виступає лінійною комбінацією інших факторів. Наприклад, вплив на продуктивність праці таких факторів як рівень механізації та рівень енергоозброєності праці, оскільки ці фактори є колінеарними.

Тема 5
МУЛЬТИКОЛІНЕАРНІСТЬ

5.1 Поняття та наслідки мультиколінеарності

Однією з умов застосування метода найменших квадратів для обробки різних експериментальних даних є відсутність мультиколінеарності між незалежними змінними.

В економетричну модель з багатьма змінними включають велику кількість факторів, які впливають на результативний показник. Серед цих факторів необхідно відібрати найбільш суттєві та виключити з моделі несуттєві. При цьому, виключенню з моделі підлягають ті фактори, які при парному корелюванні між собою дають високий лінійний коефіцієнт кореляції, який по абсолютній величині перевищує 0,85 ($r>0,85$). Наявність такого лінійного зв'язку між двома факторами називається колінеарністю, а між декількома факторами – мультиколінеарністю. Таким чином, **мультиколінеарність** означає, що в багатофакторній регресійній моделі дві або більше незалежних змінних (факторів) пов'язані між собою лінійною залежністю, тобто мають високій ступінь кореляції ($r_{x_j x_{j+s}} > 0,85$).

Термін «мультиколінеарність» вперше впровадив Р. Фріш.

Регресійне рівняння задовільно описує рух залежної змінної, якщо коефіцієнт множинної кореляції досить високий, а кореляція між незалежними факторами незначна. Мультиколінеарність незалежних змінних веде до зміщення оцінок параметрів і, отже, до неможливості коректної інтерпретації результатів.

Таким чином, наявність мультиколінеарності негативно впливає на кількісні характеристики економетричної моделі або робить її побудову взагалі неможливою. Так, якщо два колінеарні вектори змінюються в одному напрямку, то майже неможливо оцінити окремий вплив кожного з них на результативний показник, оскільки кожний з цих факторів виступає лінійною комбінацією інших факторів. Наприклад, вплив на продуктивність праці таких факторів як рівень механізації та рівень енергоозброєності праці, оскільки ці фактори є колінеарними.

Причиною мультиколінеарності є наявність трендів у динамічних рядах та використання в економетричних моделях лагових значень змінних. Наприклад, при економічному зростанні більшість базових макроекономічних показників таких як дохід, споживання, накопичення, інвестиції, зайнятість мають тенденцію до зростання з деяким лагом. Все це свідчить, що в економіці взагалі важко уникнути певного рівня залежності між показниками, тобто колінеарності.

Наслідком мультиколінеарності є велика дисперсія і коваріація оцінок параметрів, обчислених за методом найменших квадратів. При наближенні коефіцієнта кореляції до свого граничного значення дисперсії параметрів зростають із значною швидкістю; так при $r = 0{,}999$ $Var(a_1)$ у 50 разів перевищують своє значення за умови, коли немає мультиколінеарності $(r_{x_1x_2} = 0)$.

Другим практичним наслідком мультиколінеарності є збільшення інтервалу довіри. Оскільки збільшення коефіцієнта кореляції призводить до збільшення значень середньоквадратичних відхилень параметрів, то збільшується також інтервал довіри для них. За наявності високої мультиколінеарності $(r = 0{,}999)$ інтервал довіри в 500 разів більший, ніж коли її немає $(r = 0)$.

Третім практичним наслідком мультиколінеарності є незалежність t – статистики. Щоб оцінити, чи значимо параметри багатофакторної регресії відрізняються від нуля використовують t – статистику Стьюдента, тобто знаходять її фактичне значення $t_{a_j} = \frac{|a_j|}{\sigma_{a_j}}$, порівнюють його з табличним значенням t_{tab}. У випадку мультиколінеарності σ_{a_1} нескінченно зростає, а значення t_{a_1} прямує до нуля.

Мультиколінеарність незалежних змінних призводить до зміщення оцінок параметрів моделі, через що з їх допомогою не можна зробити коректні висновки про результати взаємозв'язку залежної і незалежних змінних. У випадку, якщо між незалежними змінними існує функціональний взаємозв'язок, то оцінити вплив цих змінних на залежну змінну взагалі неможливо. Тоді для оцінювання параметрів моделі метод найменших квадратів не придатний, оскільки матриця $X'X$ буде виродженою.

Тому при побудові багатофакторних економетричних моделей потрібно мати інформацію, що між незалежними змінними (факторами) не існує мультиколінеарності.

Мультиколінеарність не є проблемою, якщо єдиною метою регресійного аналізу є прогнозування; при цьому значення R^2 є високим, а параметри регресії – значимими, оскільки t – статистика висока. Чим вище значення R^2, тим точніший прогноз.

5.2 Методи визначення та способи усунення мультиколінеарності

Єдиного методу для визначення мультиколінеарності немає. Розглянемо деякі із методів тестування мультиколінеарності.

1. Високе значення R^2 і не значимість t – статистики.

Одночасна наявність цих двох факторів є «класичною» ознакою мультиколінеарності. Розглянемо m – факторну регресивну модель:

$$y = a_0 + a_1x_1 + a_2x_2 + ... + a_mx_m + u. \tag{5.1}$$

У випадку мультиколінеарності за t –статистикою Стьюдента можна визначити, що один або більше оцінених параметрів статистично не значимо відрізняються від нуля. При високому значенні R^2 ми приймаємо з великим ступенем імовірності F – критерій Фішера, оскільки він відкидає нульову гіпотезу, коли:

$$a_1 = a_2 = ...a_m. \tag{5.2}$$

Тому високе значення R^2 і статистична незначимість деяких параметрів може свідчити про наявність мультиколінеарності.

2. Високе значення парних коефіцієнтів кореляції.

Другим поширеним тестом на наявність мультиколінеарності є перевірка значень парних коефіцієнтів кореляції. Якщо значення хоча б одного коефіцієнта кореляції перевищує 0,8, то в економетричній моделі має місце мультиколінеарність. Однак високе значення парних коефіцієнтів кореляції – це достатня, але не необхідна умова наявності мультиколінеарності.

Навіть при невеликих значеннях парних коефіцієнтів кореляції мультиколінеарність може бути більшою, ніж у двохфакторній регресійній моделі.

Оцінити рівень мультиколінеарності можна за допомогою величини дисперсії інфляційного VIF фактора для кожної j-ї змінної:

$$VIF_j = \frac{1}{1-R_j^2},$$

де VIF – varience inflationary factor.

Значення $VIF_j = 10$ приймається як критичне.

Якщо $VIF_j \leq 10$, то можна стверджувати про недостатність зв'язку між j-м фактором і всіма іншими. Якщо $VIF_j \geq 10$, то це свідчить про наявність мультиколінеарності.

Існує декілька простих способів усунення мультиколінеарності.

1. Вилучення змінної (змінних) та помилки специфікації. При високій мультиколінеарності найкраще та найлегше вивести з рівняння регресії одну із незалежних змінних, для якої парний коефіцієнт кореляції найбільший. Наприклад, якщо в економетричній моделі, яка описує рівень споживання від доходу та багатства, вивести змінну, що відповідає багатству, то отримаємо регресійну модель з однією незалежною змінною – доходом.

Але вилучення змінної з моделі може призвести до помилки специфікації. Помилка специфікації виникає через некоректне визначення моделі, що використовується в аналізі. Так, якщо за економічною теорією для пояснення рівня споживання модель повинна включати доход і багатство, тоді вилучення змінної багатства створюватиме помилку специфікації. Це також може призвести до зміщення оцінок.

2. Перетворення змінних. Однією з причин мультиколінеарності даних є їх схильність змінюватись в одному напрямку, а один із шляхів зменшення такої залежності – використання перших різниць в моделі:

$$Y_t - Y_{t-1} = a_1\,(x_{1t} - x_{1,t-1}) + a_2(x_{2t} - x_{2,t-1}) + V_t, \qquad (5.3)$$

де $V = u_t - u_{t-1}$.

Наведене вище рівняння відоме як рівняння перших різниць, оскільки ми отримали регресію не з початкових змінних, а з різниць послідовних значень змінних.

3. Збільшення кількості спостережень. Оскільки мультиколінеарність змінюється у кожній вибірці, то можливо в іншій моделі вона буде відсутня. Іноді просте збільшення числа спостережень у моделі, якщо це можливо, зменшує проблему мультиколінеарності.

Серед інших способів слід виділити такі: використання первинної інформації, поєднання міжгалузевої та динамічної інформації, факторний аналіз, метод головних компонент, гребенева регресія.

5.3 Алгоритм Феррара-Глобера

Феррар і Глобер для оцінки мультиколінеарності запропонували інші статистики за допомогою яких перевіряється мультиколінеарність не тільки всього набору незалежних факторів, але кожної із пар. Мультиколінеарність встановлюється за допомогою величини χ^2 (критерія Пірсона), розрахованої на основі оцінок взаємної кореляційної матриці незалежних змінних. Для перевірки мультиколінеарності кожної із незалежних змінних обчислюються статистики F_j.

Алгоритм метода Феррара-Глобера

1. Нормалізація змінних.

Позначимо вектори незалежних змінних окремого регресійного рівняння через $x_1, x_2, \ldots, x_j, \ldots, x_m$. Елементи нормалізованих векторів обчислюємо за формулою:

$$x_{ij}^{*} = \frac{x_{ij} - \bar{x}_j}{\sqrt{n} \cdot \sigma_{x_j}} = \frac{x_{ij} - \bar{x}_j}{\sqrt{\Sigma (x_i - \bar{x})^2}}; \quad i = \overline{1,n}, \quad j = \overline{1,m} \qquad (5.4)$$

де n – число спостережень для кожного вектора; m – число незалежних змінних; $\bar{x}_j$ – середня арифметична вектора $\bar{X}_j$; σ_{x_j} – стандартна похибка змінної x_j;

$$\sigma_{x_j} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2}. \qquad (5.5)$$

Позначимо матрицю, складену з нормалізованих векторів через X*.

2. Розрахунок кореляційної матриці. Взаємна кореляційна матриця оцінюється із співвідношення:

$$R = (X^{*'} \cdot X^{*}), \tag{5.6}$$

$X^{*'}$ – матриця, транспонована до матриці X^{*}.

Це матриця коефіцієнтів парної кореляції. Вона симетрична і має розмір $m \times m$.

3. Розрахунок визначника кореляційної матриці та значення χ^2. Якщо позначити визначник кореляційної матриці через $D = R = X^{*'} \cdot X^{*}$, то:

$$\chi^2 = -\left[n - 1 - \frac{1}{6}(2m+5)\right] \cdot \ln X^{*'} \cdot X^{*}. \tag{5.7}$$

Ця величина порівнюється із табличним значенням χ^2 – квадрат критерія Пірсона для $\frac{1}{2}m(m-1)$ ступенів вільності та рівнем значимості α.

Якщо $\chi^2_{факт} > \chi^2_{табл}$, то можна зробити висновок, що в масиві змінних існує мультиколінеарність.

4. Розрахунок матриці оберненої до $(X^{*'} \cdot X^{*})$:

$$B = X^{*'} \cdot X^{*\,-1}. \tag{5.8}$$

5. Розрахунок F_j – статистик. Величина F_j розраховується за формулою:

$$F_j = (b_{jj} - 1) \cdot \frac{n-m}{m-1}, \tag{5.9}$$

де b_{jj} – діагональні елементи матриці B. Значення F_j порівнюється з критичними величинами F – критерія при $n-m$ і $m-1$ступенями свободи. Якщо $F_j < F_{\alpha, k_1, k_2}$, то ні одна із незалежних змінних не мультиколінеарна з іншими.

Коефіцієнт детермінації для кожної змінної x_j розраховується так:

$$R^2_{x_j} = 1 - \frac{1}{b_{jj}}. \tag{5.10}$$

6. Знаходження часткових коефіцієнтів кореляції.

Ці коефіцієнти розраховуються за допомогою позадіагональних елементів оберненої кореляції матриці B:

$$r^*_{ij} = \frac{-b_{ij}}{\sqrt{b_{ii} \cdot b_{jj}}}, \tag{5.11}$$

де i – номер рядка, j – номер стовпця, оберненої кореляційної матриці B.

7. Розрахунок t – статистики:

$$t_{ij} = \frac{r_{ij} \cdot \sqrt{n-m}}{\sqrt{1-r_{ij}^2}} \tag{5.12}$$

Наведені розрахунки можна інтерпретувати таким чином:

а) значення елементів матриці $(X^{*'} \cdot X^*)$ дають нам першу інформацію про мультиколінеарність незалежних факторів. Недіагональні елементи матриці дорівнюють коефіцієнтам кореляції для кожної із пар незалежних факторів і вони повинні бути меншими коефіцієнта множинної кореляції для всього рівняння;

б) мірою загальної мультиколінеарності є величина χ^2. Цей показник не повинен перевищувати табличне значення χ^2_α при заданому рівні значимості α і $k = \frac{1}{2}m(m-1)$ ступенях свободи;

в) величина F_j показує, які змінні мультиколінеарні. Якщо деякі із F_j більші табличних значень F – статистики при $n-m$ і $m-1$ ступенях свободи, то відповідні їм незалежні фактори мультиколінеарні;

г) за допомогою значень t_{ij} можна уточнити, які з пар незалежних факторів мультиколінеарні. Якщо деякі t_{ij} перевищують табличне значення критерія Стьюдента при $n-m$ ступенях свободи, то мультиколінеарність спостерігається між парою змінних x_j і x_{j+1}.

Приклад 5.1. Для побудови лінійної економетричної моделі відібрані такі показники: витрати на споживання y, рівень доходів x_1, рівень заощаджень x_2 та заробітної плати x_3. Необхідно дослідити ці показники на наявність мультиколінеарності по алгоритму Феррара-Глобера. Вхідні дані в таблиці 5.1 приведені в умовних г. о.

Таблиця 5.1 – Вхідні дані та результати розрахунків

i	x_{1i}	x_{2i}	x_{3i}	$(x_{1i}-\bar{x}_1)^2$	$(x_{2i}-\bar{x}_2)^2$	$(x_{3i}-\bar{x}_3)^2$	x^*_{1i}	x^*_{2i}	x^*_{3i}
1	4,6	11,2	29,1	6,5536	17,9776	14,669	-0,5581	-0,5181	-0,41979
2	5,6	13,3	29,3	2,4336	4,5796	13,177	-0,3401	-0,2615	-0,39787
3	6,2	12,8	33,3	0,9216	6,9696	0,1369	-0,2093	-0,3226	0,04055
4	6,9	14,6	30,9	0,0676	0,7056	4,1209	-0,0567	-0,1026	-0,22250
5	6,3	14,5	32,8	0,7396	0,8836	0,0169	-0,1875	-0,1149	-0,01425
6	7,5	17	33,2	0,1156	2,4336	0,0729	0,0741	0,1906	0,02959
7	8,6	15,6	35,5	2,0736	0,0256	6,6049	0,3139	0,0195	0,28169
8	7,7	17,9	34,6	0,2916	6,0516	2,7889	0,1177	0,3006	0,18304
9	8,7	17,1	39,2	2,3716	2,7556	39,313	0,3357	0,2028	0,68722
10	9,5	20,4	31,4	5,4756	24,6016	2,3409	0,5101	0,6060	-0,16770
Суми	71,6	154,4	329,3	21,044	66,984	83,241			
Серед. зн	7,16	15,44	32,93						

Розв'язок.

Нормалізація даних.

Елементи стандартизованих векторів розраховуємо за формулою:

$$x^*_{ij} = \frac{x_{ij}-\bar{x}}{\sqrt{n\cdot\sigma_{x_j}}} = \frac{x_{ij}-\bar{x}}{\sqrt{\sum(x_{ij}-\bar{x}_j)^2}},$$

Розраховуємо середнє арифметичне для кожної незалежної змінної $\bar{x}_j$:

$\bar{x}_1 = 7,16$; $\bar{x}_2 = 15,44$; $\bar{x}_3 = 32,93$.

Розрахунки суми квадратів відхилень:

$$\sum_{i=1}^{n}(x_{1i}-\bar{x}_1)^2 = 21,044;$$

$$\sum_{i=1}^{n}(x_{2i}-\bar{x}_2)^2 = 66,984;$$

$$\sum_{i=1}^{n}(x_{3i}-\bar{x}_3)^2 = 83,241.$$

Усі розрахунки по стандартизації змінних x_{ij} зведені у таблиці 5.1.

Матриця стандартизованих змінних X^*:

$$\begin{vmatrix} -0,55805 & -0,51806 & -0,41979 \\ -0,34006 & -0,26147 & -0,39787 \\ -0,20927 & -0,32257 & 0,04055 \\ -0,05668 & -0,10263 & -0,2225 \\ -0,18747 & -0,11485 & -0,01425 \\ 0,07412 & 0,19061 & 0,02959 \\ 0,31391 & 0,01955 & 0,28169 \\ 0,11771 & 0,,30057 & 0,18304 \\ 0,33570 & 0,20283 & 0,68722 \\ 0,51010 & 0,60603 & 0,16770 \end{vmatrix}$$

2. Кореляційна матриця:

$$R = X^{*'} \cdot X^* = \begin{bmatrix} 1 & 0,906 & 0,.634 \\ 0,906 & 1 & 0,437 \\ 0,634 & 0,437 & 1 \end{bmatrix}$$

Кожний елемент цієї матриці характеризує тісноту зв'язку однієї незалежної змінної з іншою. Оскільки діагональні елементи характеризують зв'язок кожної незалежної з цією ж змінною, то вони дорівнюють одиниці, але через неточність обчислень числові значення діагональних елементів можуть лише наближатись до одиниці. Якщо так, то вони замінюються одиницями. Інші елементи матриці R трактуються так:

$$r_{x_1x_2} = 0,906; \quad r_{x_1x_3} = 0,634; \quad r_{x_2x_3} = 0,437,$$

тобто вони є парними коефіцієнтами кореляції незалежних змінних. На основі цих коефіцієнтів можна зробити висновок, що між x_1, x_2, x_3 існує зв'язок. Але ще не можна стверджувати, що цей зв'язок є явищем мультиколінеарності і він негативно буде впливати на оцінки параметрів моделі.

3. Розрахунок визначника кореляційної матриці:

$$D = |R = |x'^* \cdot x^*| = 0,08869$$

$$\chi^2 = -\left[n - 1 - \frac{1}{6}(2m+5)\right] \cdot \ln\left|x^{*'} \cdot x^*\right| = \left[10 - 1 - \frac{1}{6}(2 \cdot 3 + 5)\right] \ln 0,08889$$

$$= -\left[9 - \frac{1}{6} \cdot 11\right] \times -2,42261 = 21,3997$$

$\chi^2_{\alpha(табл)} = 18{,}31$.

Оскільки $\chi^2 > \chi^2_\alpha$, то можна зробити висновок, що у масиві незалежних змінних існує мультиколінеарність.

4. Знаходимо матрицю, обернену до матриці R:

$$B = (X^{*'} \cdot X^*)^{-1} = \begin{vmatrix} 9{,}124 & -7{,}091 & -2{,}684 \\ -7{,}091 & 6{,}748 & 1{,}546 \\ -2{,}684 & 1{,}546 & 2{,}025 \end{vmatrix}$$

5. Використовуючи діагональні елементи матриці B, розрахуємо F – критерії:

$$F_j = (b_{jj} - 1) \cdot \frac{n-m}{m-1};$$

$$F_1 = (9{,}124 - 1) \cdot \frac{10-3}{3-1} = 8{,}124 \cdot \frac{7}{2} = 28{,}434;$$

$$F_2 = (6{,}748 - 1) \cdot \frac{10-3}{3-1} = 5{,}748 \cdot \frac{7}{2} = 20{,}118;$$

$$F_3 = (2{,}025 - 1) \cdot \frac{10-3}{3-1} = 1{,}025 \cdot \frac{7}{2} = 3{,}587.$$

При рівні значимості $\alpha = 0{,}05$ і ступенях свободи $k_1 = 7$ і $k_2 = 2$ табличне значення критерія $F_{\alpha, k_1, k_2} = 4{,}74$.

Оскільки $F_1, F_2 > F_{\alpha, k_1, k_2}$, то незалежні змінні x_1, x_2 мультиколінеарні між собою.

6. Розраховуємо часткові коефіцієнти кореляції, використовуючи елементи матриці B:

$$r_{ij} = \frac{-b_{ij}}{\sqrt{b_{ii} \cdot b_{jj}}},$$

$$r_{12.3} = \frac{-b_{12}}{\sqrt{b_{11} \cdot b_{22}}} = \frac{7{,}091}{\sqrt{9{,}124 \cdot 6{,}748}} = 0{,}9037$$

$$r_{13.2} = \frac{-b_{13}}{\sqrt{b_{11} \cdot b_{33}}} = \frac{2{,}684}{\sqrt{9{,}1248 \cdot 2{,}025}} = 0{,}624$$

$$r_{23,1} = \frac{-b_{23}}{\sqrt{b_{22} \cdot b_{33}}} = \frac{1,546}{\sqrt{6,748 \cdot 2,025}} = 0,418$$

Часткові коефіцієнти кореляції характеризують тісноту зв'язку між двома змінними за умови, що третя не впливає на цей зв'язок.

7. Визначимо t – критерії для часткових коефіцієнтів кореляції.

$$t_{ij} = \frac{r_{ij} \cdot \sqrt{n-m}}{\sqrt{1-r_{ij}^2}};$$

$$t_{12} = \frac{r_{12,3} \cdot \sqrt{n-m}}{\sqrt{1-r_{12,3}^2}} = \frac{0,903 \cdot \sqrt{7}}{\sqrt{1-0,903^2}} = 5,56;$$

$$t_{13} = \frac{r_{13,3} \cdot \sqrt{n-m}}{\sqrt{1-r_{13,3}^2}} = \frac{0,624 \cdot \sqrt{7}}{\sqrt{1-0,624^2}} = 2,12;$$

$$t_{23} = \frac{t_{23,1} \sqrt{n-m}}{\sqrt{1-r_{23,1}^2}} = \frac{0,418 \cdot \sqrt{7}}{\sqrt{1-0,418^2}} = 1,22,$$

Табличне значення t – критерія при $n-m=7$ ступенях свободи і рівнем значимості $\alpha = 0,05$ дорівнює 1,89.

$t_{12}, t_{13} > t_{\alpha,k}$ $t_{12}, t_{13} > t_{\alpha,k}$, а $t_{23} < t_{\alpha,k}$. Отже, між змінними 1,2 та 1,3 існує тісний зв'язок і ці пари змінних є мультиколінеарними. Оскільки фактичне значення t – критерія Стьюдента найбільше для пари змінних 1,2, то одну з цих змінних доцільно виключити з економетричної моделі (краще x_2, оскільки рівень доходів формує рівень заощаджень x_2).

Питання для самоконтролю:

1. У чому суть мультиколінеарності.

2. Які негативні впливи мультиколінеарності на оцінки параметрів моделі.

3. Методи виявлення мультиколінеарності.

4. Методи усунення мультиколінеарності.

5. Алгоритм Феррара-Глобера.

Тема 6

АВТОКОРЕЛЯЦІЯ

6.1 Поняття автокореляції

В економічних часових рядах наступні спостереження часто залежать від попередніх, тобто між нами існує автокореляція.

Автокореляція являє собою кореляційну залежність між наступними і попередніми членами часового ряду, тобто кореляцію між рядами $y_1, y_2, \ldots, y_n$ та $y_{p+1}, y_{p+2}, y_{p+n}$, де p показує довжину часового зміщення в ряді y_t. Таке часове зміщення часто називають лагом; p є цілим числом і його найбільше значення залежить від числа періодів ряду динаміки.

Наявність автокореляції вказує на взаємозв'язок рівнів ряду динаміки, на сильну залежність наступних рівнів від попередніх, що суттєво спотворює величину середньоквадратичних похибок коефіцієнтів регресії, утруднює побудову довірчих інтервалів для коефіцієнтів регресії, а також перевірку їх значущості.

В автокореляції значення Y_t певною мірою залежить від попередніх значень Y_{t-1}, Y_{t-2} і т.д. Рівняння, що виражає значення величини Y в момент t через її значення в моменти $t-1,\ t-2, \cdots, t-p$, називається рівнянням авторегресії. Таке рівняння має вигляд:

$$Y_t = f(Y_{t-1}, Y_{t-2}, Y_{t-p}).$$

Лінійне авторегресійне рівняння записуємо у вигляді:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2} + \ldots a_p Y_{t-p}. \qquad (\ldots$$

Коефіцієнти $a_0, a_1, a_2, \cdots, a_p$ визначають методом наймені..... квадратів.

Наявність автокореляції вказує на взаємозв'язок рівнів ряду динаміки, на значну залежність наступних рівнів від попередніх. Так як методика кореляційного аналізу передбачає, що рівні кожного з взаємопов'язаних рядів є статистично незалежними, то необхідно завжди перевіряти наявність автокореляції в досліджуваних рядах динаміки.

Наявність автокореляції може бути обумовлена декількома причинами:

1) якщо в моделі не врахований суттєвий фактор, то його вплив відображається у залишках, в силу чого останні можуть виявитися автокорельованими;

2) якщо в моделі не враховано декілька несуттєвих факторів, взаємний вплив яких є суттєвим наслідком співпадання фаз і напрямків їх зміни;

3) якщо вибрано невірний тип моделі;

4) автокореляція може бути обумовлена специфічною структурою випадкових компонентів.

Автокорельовані залишки можуть приводити до помилкового визначення істотності та довірчих границь коефіцієнтів регресії. Якщо наявність автокореляції не перевіряється, то значення результатів кореляційного і регресивного аналізу рядів динаміки є сумнівними.

Довжина часового зміщення p в ряді динаміки називається «лагом». Лаг ще можна визначити як напрямок та тривалість відставання рівнів одного із взаємопов'язаних рядів від рівнів другого ряду. Лаг є цілим числом $l = 1,2,..., p$. Його значення залежить від числа періодів у ряді динаміки і складає $m \le \frac{n}{2}$.

Приклади автокореляції: урожайність сільськогосподарських культур у визначені роки пов'язана з врожайністю попередніх років; динаміка народжуваності та укладання шлюбів; розміри залізничних перевезень та валові збори зерна і т.д.

Перевірка наявності автокореляції проводиться шляхом обчислення коефіцієнтів автокореляції – значень автокореляційної функції. Величина коефіцієнта автокореляції визначає величину лага.

Існує ряд способів виключення або зменшення автокореляції в часових рядах.

Найбільш ефективним є виключення тренда *f(t)* із часового ряду і перехід до випадкової компоненти U_t. Для усунення автокореляції можна використовувати прийом, що полягає у включенні часу в рівняння множинної регресії як аргументу (метод Фріша-Воу) та метод кінцевих різниць, коли методом найменшим квадратів обробляються не безпосередньо рівні вихідних рядів, а їх послідовні різниці між ними.

Крім автокореляції часових рядів можлива автокореляція відхилень фактичних значень ендогенної змінної (в багатофакторних рівняннях регресії) від розрахункових, обчислених по цьому рівнянні, тобто автокореляція залишків $u_i = y_i - \hat{y}_i$.

6.2 Нециклічний коефіцієнт автокореляції

Коефіцієнт автокореляції r_p характеризує щільність зв'язку між первинним рядом динаміки і цим же рядом, зсуненим на p періодів. У таблиці 6.1 наведено зсунені ряди динаміки з лагами $p = 1,2,3$.

Таблиця 6.1 – Первинний та зсунені ряди динаміки

Значення часу t	Первинний рівень ряду Y_t	$p = 1$	$p = 2$	$p = 3$
1	2	3	4	5
1	Y_1	-	-	-
2	Y_2	Y_1	-	-
3	Y_3	Y_2	Y_1	-
4	Y_4	Y_3	Y_2	Y_1
n	Y_n	Y_{n-1}	Y_{n-2}	Y_{n-3}

Із збільшенням лага p кількість пар корельованих рівнів зменшується. Так, при $p = 1$ довжина корельованих рядів менша за первинний ряд на один рівень, при $p = 2$ – на два рівні і т.д.

Нециклічний коефіцієнт автокореляції розраховується за формулою коефіцієнта парної кореляції. Так, коефіцієнт автокореляції першого порядку ($p = 1$) є не що інше, як коефіцієнт парної кореляції між двома динамічними рядами, зміщеними на один проміжок часу ($p = 1$):

$$Y_1, Y_2, Y_3, \cdots, Y_n \text{ та } Y_1, Y_2, \cdots, Y_{n-1}.$$

$$r_1 = \frac{\sum_{t=2}^{n}(Y_t - \overline{Y}_t)\cdot(Y_{t-1} - \overline{Y}_{t-1})}{\sqrt{\sum_{t=2}^{n}(Y_t - \overline{Y}_t)^2 \cdot \sum_{t=2}^{n}(Y_{t-1} - \overline{Y}_{t-1})^2}}; \tag{6.3}$$

де $\overline{Y}_t$ – середній рівень початкового ряду Y_t;

$\overline{Y}_{t-1}$ – середній рівень зміщеного на один проміжок часу ряду Y_{t-1}.

Або інші робочі формули:

$$r_1 = \frac{\sum_{t=2}^{n} Y_t \cdot Y_{t-1} - (n-1)\cdot \overline{Y}_t \cdot \overline{Y}_{t-1}}{\sqrt{\left[\sum_{t=2}^{n} Y_t^2 - (n-1)\cdot \overline{Y}_t^2\right]\cdot\left[\sum_{t=2}^{n} Y_{t-1}^2 - (n-1)\cdot \overline{Y}_{t-1}^2\right]}}; \tag{6.4}$$

$$r_1 = \frac{\overline{Y_t \cdot Y_{t-1}} - \overline{Y}_t \cdot \overline{Y}_{t-1}}{\sigma_{Y_t} \cdot \sigma_{Y_{t-1}}}.$$

Коефіцієнт парної кореляції між двома динамічними рядами, зміщеними на два проміжки часу ($p = 2$):

$$r_2 = \frac{\overline{Y_t \cdot Y_{t-2}} - \overline{Y}_t \cdot \overline{Y}_{t-2}}{\sigma_{Y_t} \cdot \sigma_{Y_{t-2}}}. \tag{6.6}$$

Коефіцієнт парної кореляції між двома динамічними рядами, зміщеними на три проміжки часу ($p = 3$):

$$r_3 = \frac{\overline{Y_t \cdot Y_{t-3}} - \overline{Y}_t \cdot \overline{Y}_{t-3}}{\sigma_{Y_t} \cdot \sigma_{Y_{t-3}}}. \tag{6.7}$$

Запишемо вирази (6.3), (6.4); (6.5) у загальному вигляді, якщо два динамічні ряди зміщені на p проміжків часу:

$$r_p = \frac{\sum_{t=p+1}^{n}(Y_t - \overline{Y}_t)\cdot(Y_{t-p} - \overline{Y}_{t-p})}{\sqrt{\sum_{t=p-1}^{n}(Y_t - \overline{Y}_t)^2 \cdot \sum_{t=p+1}^{n}(Y_{t-p} - \overline{Y}_{t-p})^2}}. \tag{6.8}$$

$$r_p = \frac{\sum_{t=p+1}^{n} Y_t \cdot Y_{t-p} - (n-p) \cdot \bar{Y}_t \cdot \bar{Y}_{t-p}}{\sqrt{\left[\sum_{t=p+1}^{n} Y_t^2 - (n-p) \cdot \bar{Y}_t^2\right] \cdot \left[\sum_{t=p+1}^{n} Y_{t-p}^2 - (n-p) \cdot \bar{Y}_{t-p}^2\right]}}; \qquad (6.9)$$

$$r_p = \frac{\overline{Y_t \cdot Y_{t-p}} - \bar{Y}_t \cdot \bar{Y}_{t-p}}{\sigma_{Y_t} \cdot \sigma_{Y_{t-p}}}. \qquad (6.10)$$

В подальшому для розрахунку коефіцієнта автокореляції будемо використовувати робочу формулу (6.10).

Розрахуємо середні значення $\bar{Y}_t$, $\bar{Y}_{t-p}$, середнє значення добутків $\overline{Y_t \cdot Y_{t-p}}$ за формулами:

$$\bar{Y}_t = \frac{1}{n-p} \cdot \sum_{t=p+1}^{n} Y_t; \; \bar{Y}_{t-p} = \frac{1}{n-p} \cdot \sum_{t=p+1}^{n} Y_{t-p}. \qquad (6.11)$$

Середні значення добутків:

$$\overline{Y_t \cdot Y_{t-p}} = \frac{1}{n-p} \cdot \sum_{t=p+1}^{n} Y_t \cdot Y_{t-p}.$$

Середньоквадратичні відхилення:

$$\sigma_{Y_t} = \sqrt{\frac{1}{n-p} \cdot \sum_{t=p+1}^{n} (Y_t - \bar{Y}_t)^2}; \; \sigma_{Y_{t-p}} = \sqrt{\frac{1}{n-p} \cdot \sum_{t=p+1}^{n} (Y_t - \bar{Y}_{t-p})^2}.$$

Для визначення часового лагу l застосовується автокореляційна функція $r_{y(l)}$, яка реалізується як множина коефіцієнтів кореляції між часовим рядом і цим же рядом, зсунутим відносно початкового положення на p проміжків часу. Нормована автокореляційна функція для часового ряду Y_t розраховується за формулою:

$$\gamma_{y(p)} = \frac{(n-p)\sum_{t=p+1}^{n} Y_t \cdot Y_{t-p} - \sum_{t=p+1}^{n} Y_t \cdot \sum_{t=p+1}^{n} Y_{t-p}}{\sqrt{\left[(n-p) \cdot \sum_{t=p+1}^{n} Y_t^2 - \left(\sum_{t=p+1}^{n} Y_t\right)^2\right] \cdot \left[(n-p) \cdot \sum_{t=p+1}^{n} Y_{t-lp} - \left(\sum_{t=p+1}^{n} Y_{t-p}\right)^2\right]}}, \qquad (6.12)$$

де n – кількість членів часового ряду;

Y_t, Y_{t-p} – рівні числового ряду.

Величину p називають зрушенням рівнів часового ряду. Зрушення, якому відповідає найбільший коефіцієнт автокореляції, називається *часовим запізненням або часовим лагом*. Графік нормованої автокореляційної функції називають корелограмом. Він наглядно показує: з яким запізненням зміна показника Y_t впливає на зміну наступних значень цього показника. Нормована автокореляційна функція може приймати значення від –1 до +1 за формулою нециклічних коефіцієнтів парної кореляції:

$$Y_{Y(p)} = \frac{\overline{Y_t \cdot Y_{t-p}} - \overline{Y}_t \cdot \overline{Y}_{t-p}}{\sigma_{Y_t} \cdot \sigma_{Y_{t-p}}}, \qquad (6.13)$$

де $\overline{Y_t}$ – середній рівень початкового ряду Y_t;

$\overline{Y}_{t-p}$ – середній рівень ряду, зміщеного на p проміжків часу;

$\overline{Y_t \cdot Y_{t-p}}$ – середнє значення добутку початкового і зміщеного на p проміжків часу рівнів ряду $Y_t \cdot Y_{t-p}$;

σ_{Y_t} – середньоквадратичне відхилення для рівнів початкового ряду Y_t;

$\sigma_{Y_{t-e}}$ – середньоквадратичне відхилення для рівнів ряду, зміщеного від початкового на l проміжків часу.

Величина та напрямок часового лага знаходяться по найбільшому модулю коефіцієнту автокореляції. Порівняння значень коефіцієнтів автокореляції показує: з якого моменту починає виявлятись вплив зміни рівнів одного часового ряду на зміну рівнів другого ряду і з якого моменту цей вплив глибшає або зовсім припиняється.

Знання автокореляційної функції має важливе значення для обґрунтування порядку рівня авторегресії.

Таким чином, послідовність значень коефіцієнтів автокореляції r_p, розрахованих для $p = 1, 2, \cdots, l$, називають автокореляційною функцією. Автокореляційна функція відображається графічно у вигляді автокорелограма з абсцисою p та ординатою r_e.

Приклад 6.1. По статистичних даних таблиці 6.2 розрахувати коефіцієнти автокореляції першого, другого та третього порядків. Вибрати найбільший по модулю коефіцієнт автокореляції, який буде визначати величину часового зміщення або лаг та порядок авторегресійної моделі.

Таблиця 6.2 –Вхідні дані та результати розрахунків

t	Y_t	Y_{t-1}	Y_{t-2}	Y_{t-3}	$Y_t \cdot Y_{t-1}$	$Y_t \cdot Y_{t-2}$	$Y_t \cdot Y_{t-3}$	$(Y_t - \overline{Y}_t)^2$	$(Y_{t-1} - \overline{Y}_{t-1})^2$	$(Y_{t-2} - \overline{Y}_{t-2})^2$	$(Y_{t-3} - \overline{Y}_{t-3})^2$
1	2,2							8.2065			
2	3,7	2,2			8.14			1.8624	5,931		
3	4,1	3.7	2.2		15.17	9.02		0.9307	0.875	5.319	
4	4	4,1	3,7	2,2	16,4	14,8	8,8	1,1336	0,287	0,650	4,382
5	4.2	4	4.1	3.7	16.8	17.22	15.54	0.7477	0.404	0.165	0,352
6	2,9	4.2	4	4.1	12.18	11.6	11.89	4.6860	0.189	0.256	0,037
7	3,5	2,9	4,2	4	10,15	14,7	14	2,4483	3,011	0,094	0,086
8	4.9	3.5	2.9	4.2	17.15	14.21	20.58	.0271	1.289	2.580	0,009
9	3,4	4,9	3,5	2,9	16,66	11,9	9,86	2,7712	0,070	1,013	1,941
10	4.3	3.4	4.9	3.5	14.62	21.7	15.05	0.5848	1.526	0.155	0,629
11	4,8	4,3	3,4	4,9	20,64	16,32	23,52	0,0701	0,112	1,224	0,368
12	5.5	4.8	4.3	3.4	26.4	23.65	18.7	0.1895	0.027	0.043	0,798
13	4.5	5.5	4.8	4.3	24.75	21.6	19.35	0.3189	0.748	0.086	0,000
14	6,6	4,5	5,5	4,8	29,7	36,3	31,68	2,3571	0,018	0,988	0,257
15	5,8	6.6	4.5	5.5	38.28	26.1	31.9	0.5407	3.860	0.000	1,456
16	7,7	5,8	6,6	4,5	44,66	50,82	34,65	6,9448	1,357	4,384	0,043
17	6.7	7.7	5.8	6.6	51.59	38.86	44.22	2.6742	9.392	1.674	5.321
18	9,5	6,7	7,7	5,8	63,65	73,15	55,1	19,672	4,263	10,200	2,270
171	86,1	78,8	72,1	64,4	426,94	401,32	354,8	47,96	33,829	28,829	17,949

Розв'язок.

Запишемо робочі формули для розрахунку коефіцієнтів автокореляції відповідно першого, другого та третього порядків:

$$r_p = \frac{\overline{Y_t \cdot Y_{t-p}} - \overline{Y}_t \cdot \overline{Y}_{t-p}}{\sigma_{Y_t} \cdot \sigma_{Y_{t-p}}};\ r_1 = \frac{\overline{Y_t \cdot Y_{t-1}} - \overline{Y}_t \cdot \overline{Y}_{t-1}}{\sigma_{Y_t} \cdot \sigma_{Y_{t-1}}}$$

$$r_2 = \frac{\overline{Y_t \cdot Y_{t-2}} - \overline{Y}_t \cdot \overline{Y}_{t-2}}{\sigma_{Y_t} \cdot \sigma_{Y_{t-2}}}\ r_3 = \frac{\overline{Y_t \cdot Y_{t-3}} - \overline{Y}_t \cdot \overline{Y}_{t-3}}{\sigma_{Y_t} \cdot \sigma_{Y_{t-3}}}.$$

Середній добуток, середні та середньоквадратичні відхилення для $p = 1,2,3$:

$p = 1$

$$\overline{Y_t \cdot Y_{t-1}} = \frac{1}{n-1} \cdot \sum_{t=2}^{n} Y_t \cdot Y_{t-1}; \qquad \overline{Y}_t = \frac{1}{n-1} \cdot \sum_{t=2}^{n} Y_t; \quad \overline{Y}_{t-1} = \frac{1}{n-1} \cdot \sum_{t=2}^{n} Y_{t-1};$$

$$\sigma_{Y_t} = \sqrt{\frac{1}{n-1} \cdot \sum_{t=2}^{n} \left(Y_t - \overline{Y}_t\right)^2}; \quad \sigma_{Y_{t-1}} = \sqrt{\frac{1}{n-1} \cdot \sum_{t=2}^{n} \left(Y_{t-1} - \overline{Y}_{t-1}\right)^2};$$

$p = 2$;

$$\overline{Y_t \cdot Y_{t-2}} = \frac{1}{n-2} \cdot \sum_{t=3}^{n} Y_t \cdot Y_{t-2}; \qquad \overline{Y}_t = \frac{1}{n-2} \cdot \sum_{t=3}^{n} Y_t; \quad \overline{Y}_{t-2} = \frac{1}{n-2} \cdot \sum_{t=3}^{n} Y_{t-2};$$

$$\sigma_{Y_t} = \sqrt{\frac{1}{n-2} \cdot \sum_{t=3}^{n} \left(Y_t - \overline{Y}_t\right)^2}; \qquad \sigma_{Y_{t-2}} = \sqrt{\frac{1}{n-2} \cdot \sum_{t=3}^{n} \left(Y_{t-2} - \overline{Y}_{t-2}\right)^2};$$

$p = 3$;

$$\overline{Y_t \cdot Y_{t-3}} = \frac{1}{n-3} \cdot \sum_{t=4}^{n} Y_t \cdot Y_{t-3}; \qquad \overline{Y}_t = \frac{1}{n-3} \cdot \sum_{t=4}^{n} Y_t; \qquad \overline{Y}_{t-3} = \frac{1}{n-3} \cdot \sum_{t=4}^{n} Y_{t-3};$$

$$\sigma_{Y_t} = \sqrt{\frac{1}{n-3} \cdot \sum \left(Y_t - \overline{Y}_t\right)^2}; \qquad \sigma_{Y_{t-3}} = \sqrt{\frac{1}{n-3} \cdot \sum \left(Y_{t-3} - \overline{Y}_{t-3}\right)^2}.$$

Тоді:

$$\overline{Y_t Y_{t-1}} = \frac{426{,}94}{17} = 25{,}11; \ \overline{Y}_t = \frac{86{,}1}{17} = 5{,}06; \ \overline{Y}_{t-1} = \frac{78{,}8}{17} = 4{,}64; \ \sigma_{Y_t} = \sqrt{\frac{47{,}96}{17}} = 1{,}68; \ \sigma_{Y_{t-1}} = \sqrt{\frac{33{,}36}{17}} = 1{,}41;$$

$$\overline{Y_t Y_{t-2}} = \frac{401{,}32}{16} = 25{,}08; \ \overline{Y}_t = \frac{82{,}4}{16} = 5{,}15; \ \overline{Y}_{t-2} = \frac{72{,}1}{16} = 4{,}51; \ \sigma_{Y_t} = \sqrt{\frac{45{,}98}{16}} = 1{,}70; \ \sigma_{Y_{t-2}} = \sqrt{\frac{28{,}83}{16}} = 1{,}34;$$

$$\overline{Y_t Y_{t-3}} = \frac{354{,}84}{15} = 23{,}66; \ \overline{Y}_t = \frac{78{,}3}{15} = 5{,}22; \ \overline{Y}_{t-3} = \frac{64{,}4}{15} = 4{,}29; \ \sigma_{Y_t} = \sqrt{\frac{44{,}80}{15}} = 1{,}73; \ \sigma_{Y_{t-3}} = \sqrt{\frac{17{,}95}{15}} = 1{,}09;$$

Коефіцієнти автокореляції:

$$r_{(1)} = \frac{25{,}11 - 5{,}06 \cdot 4{,}64}{1{,}68 \cdot 1{,}41} = 0{,}69; \quad r_{(2)} = \frac{25{,}08 - 5{,}15 \cdot 4{,}51}{1{,}70 \cdot 1{,}34} = 0{,}82;$$

$$r_{(3)} = \frac{23{,}66 - 5{,}22 \cdot 4{,}29}{1{,}73 \cdot 1{,}09} = 0{,}66.$$

Оскільки $r_{(2)} = \max$, то довжина лагу складає два місяці і прогнозування необхідно поводити по авторегресійній моделі другого порядку.

6.3 Циклічний коефіцієнт автокореляції

Наведені вище формули нециклічного коефіцієнта автокореляції свідчать, що із зростанням часового зрушення (лага) p число рівнів ряду динаміки скорочується на цю величину зрушення. Щоб позбутись цього негативного явища автокореляцію досліджують за допомогою циклічного коефіцієнта автокореляції.

Розглянемо суть циклічного коефіцієнта автокореляції. Циклічною автокореляцією називається кореляція між рядами $Y_1, Y_2, \cdots, Y_n$ та $Y_{p+1}, Y_{p+2}, \cdots, Y_n, Y_1, Y_2, \cdots, Y_p$, тобто другий ряд утворюють так, що перші p рівні першого ряду знаходяться в кінці другого ряду або останні p рівні першого ряду будуть знаходитись у кінці другого ряду. У цьому випадку:

$$\begin{cases} \sum\limits^{n} \ldots = \sum\limits_{t=p+1}^{p} Y_t \\ Y_t^2 = \sum\limits_{t=p+1}^{p} Y_t^2. \end{cases} \tag{6.14}$$

Формула циклічного коефіцієнта автокореляції має такий вигляд:

$$r_p = \frac{\sum\limits_{t=1}^{n} Y_t \cdot Y_{t-p} - \frac{\left(\sum\limits_{t=1}^{n} Y_t\right)^2}{n}}{\sum\limits_{t=1}^{n} Y_t^2 - \frac{\left(\sum\limits_{t=1}^{n} Y_t\right)^2}{n}}, \tag{6.15}$$

$$r_p = \frac{\sum\limits_{t=p+1}^{n} \left(Y_t - \overline{Y}_t\right) \cdot \left(Y_{t-1} - \overline{Y}_{t-1}\right)}{\sqrt{\sum\limits_{t=p+1}^{n} \left(Y_t - \overline{Y}_t\right) \cdot \sum\limits_{t=+1}^{n} \left(Y_{t-1} - \overline{Y}_{t-1}\right)^2}}. \tag{6.16}$$

Коефіцієнт циклічної автокореляції не придатний для визначення автокореляції між фактичними рівнями ряду динаміки, так як перші та останні рівні можуть суттєво відрізнятися і циклічні ряди динаміки дають необ'єктивну картину про динаміку явища.

Приклад 6.2. По даних таблиці 6.3 розрахувати циклічний коефіцієнт автокореляції першого порядку ($p = 1$).

Таблиця 6.3 – Вхідні дані та результати розрахунків

	t	Y_t	Y_t	Y_{t-1}	$Y_t - \overline{Y}_t$	$Y_{t-1} - \overline{Y}_{t-1}$	$(Y_t - \overline{Y}_t)\cdot(Y_{t-1} - \overline{Y}_{t-1})$	$(Y_t - \overline{Y}_t)^2$	$(Y_{t-1} - \overline{Y}_{t-1})^2$
	1	2.2							
	2	3.7	2.2	3.7	-2.43529	-1.36471	3.32346	5.93066	1.86242
	3	4.1	3.7	4.1	-0.93529	-0.96471	0.90228	0.87478	0.93066
	4	4	4.1	4	-0.53529	-1.06471	0.56993	0.28654	1.13360
	5	4.2	4	4.2	-0.63529	-0.86471	0.54934	0.40360	0.74772
	6	2.9	4.2	2.9	-0.43529	-2.16471	0.94228	0.18948	4.68595
	7	3.5	2.9	3.5	-1.73529	-1.56471	2.71522	3.01125	2.44830
	8	4.9	3.5	4.9	-1.13529	-0.16471	0.18699	1.28889	0.02713
	9	3.4	4.9	3.4	0.26471	-1.66471	-0.44066	0.07007	2.77125
	10	4.3	3.4	4.3	-1.23529	-0.76471	0.94464	1.52595	0.58478
	11	4.8	4.3	4.8	-0.33529	-0.26471	0.08875	0.11242	0.07007
	12	5.5	4.8	5.5	0.16471	0.43529	0.07170	0.02713	0.18948
	13	4.5	5.5	4.5	0.86471	-0.56471	-0.48830	0.74772	0.31889
	14	6.6	4.5	6.6	-0.13529	1.53529	-0.20772	0.01830	2.35713
	15	5.8	6.6	5.8	1.96471	0.73529	1.44464	3.86007	0.54066
	16	7.7	5.8	7.7	1.16471	2.63529	3.06934	1.35654	6.94478
	17	6.7	7.7	6.7	3.06471	1.63529	5.01170	9.39242	2.67419
	18	9.5	6.7	9.5	2.06471	4.43529	9.15758	4.26301	19.67183
Суми	171	88.3	78.8	86.1			27.84118	33.35882	47.95882

Розв'язок. Кількість інтервалів часу $n = 18$. Середні значення $\overline{Y}_t$, $\overline{Y}_{t-1}$

$$\overline{Y}_t = \frac{1}{n-1}\cdot\sum_{t=2}^{n} Y_t = \frac{78.8}{17} = 4.63529; \quad \overline{Y}_{t-1} = \frac{1}{n-1}\cdot\sum_{t=2}^{n} \overline{Y}_{t-1} = \frac{86.1}{17} = 5.06471.$$

Коефіцієнт автокореляції першого порядку:

$$r_1 = \frac{\sum_{t=2}^{n}\left(Y_t - \overline{Y}_t\right)\cdot\left(Y_{t-1} - \overline{Y}_{t-1}\right)}{\sqrt{\sum_{t=2}^{n}\left(Y_t - \overline{Y}_t\right)\cdot\sum_{t=2}^{n}\left(Y_{t-1} - \overline{Y}_{t-1}\right)^2}} = \frac{27.84118}{\sqrt{33.35882\cdot 47.95882}} = 0.696.$$

Таким же чином можна розрахувати коефіцієнти автокореляції другого, третього та інших порядків.

6.4 Критерії Дарбіна-Уотсона та Неймана

Найпоширенішим методом перевірки автокореляції є критерій Дарбіна-Уотсона:

$$d = \frac{\sum_{t=p-1}^{n}(Y_{t-1} - Y_t)^2}{\sum_{t=p-1}^{n} Y_t^2}; \tag{6.17}$$

Можливі значення критерія знаходяться в інтервалі 0-4. Якщо ряд не містить автокореляції, то значення критерія коливаються навколо 2. Емпіричне значення d порівнюють з табличним значенням.

Можливі такі випадки:

1) $d < d_1$ – ряд містить автокореляцію;

2) $d > d_2$ – ряд не містить автокореляції;

3) $d_1 < d < d_2$ – необхідні подальші дослідження;

де d_1 і d_2 – нижня і верхня межа критерія.

Так як при від'ємній автокореляції значення d знаходяться в інтервалі 2-4, то для перевірки автокореляції необхідно визначити величину $d' = 4 - d$.

Критерій Неймана:

$$Q = \frac{\sigma^2}{S_u^2}; \tag{6.18}$$

де $\sigma^2 = \frac{1}{n-1} \cdot \sum_{t=2}^{n}(Y_t - Y_{t-1})^2$, $S_u^2 = \frac{1}{n} \cdot \sum_{t=1}^{n}(Y_t - \bar{Y}_t)^2$.

Теоретичне використання цих критеріїв базується на тому, що в динамічних рядах як самі спостереження, так і відхилення від них розподіляються в хронологічному порядку. При числі спостережень n число значень Y_{t-1} буде дорівнювати $n-1$.

При умові, якщо відхилення рівнів від тенденції (залишки) випадкові, то значення d, що лежать в інтервалі 0-4, завжди будуть знаходитись ближче до 2. Якщо автокореляція додатня, то d<2, якщо від'ємна, то $2 \le d \le 4$. Таким чином, оцінки, отримані по критерію

Дарбіна-Уотсона, є не точковими, а інтервальними. Їх значення для трьох рівнів значущості σ=0.01, σ=0.025 та σ=0.05 з урахуванням числа спостережень наведені у додатку.

Критерій Неймана дає тільки точкову оцінку. У додатку значення Q наведені для додатньої (Q_1) та для від'ємної автокореляції (Q_2). Якщо отримана в результаті обробки динамічного ряду величина Q буде мати значення нижче табличного, то автокореляція додатня, а якщо вище – то від'ємна. Якщо отримане значення критерія лежить в інтервалі $Q_1 - Q_2$, то автокореляція відсутня.

Приклад 6.3. Перевірити наявність автокореляції між рівнями ряду динаміки за критеріями Дарбіна-Уотсона та Неймана.

Таблиця 6.4 – Вхідні дані та результати розрахунків

	t	Y_t	Y_t^2	Y_{t-1}	$(Y_t - Y_{t-1})^2$	$(Y_t - \overline{Y}_t)^2$
1	2	3	4	5	6	7
	1	2.2	4.84			7.320031
	2	3.7	13.69	2.2	2.25	1.453364
	3	4.1	16.81	3.7	0.16	0.64892
	4	4	16	4.1	0.01	0.820031
	5	4.2	17.64	4	0.04	0.497809
	6	2.9	8.41	4.2	1.69	4.022253
	7	3.5	12.25	2.9	0.36	1.975586
	8	4.9	24.01	3.5	1.96	3.09E-05
	9	3.4	11.56	4.9	2.25	2.266698
	10	4.3	18.49	3.4	0.81	0.366698
	11	4.8	23.04	4.3	0.25	0.011142
	12	5.5	30.25	4.8	0.49	0.353364
	13	4.5	20.25	5.5	1	0.164475
	14	6.6	43.56	4.5	4.41	2.871142
	15	5.8	33.64	6.6	0.64	0.800031
	16	7.7	59.29	5.8	3.61	7.80892
	17	6.7	44.89	7.7	1	3.220031
	18	9.5	90.25	6.7	7.84	21.10892
Суми	171	88.3	484.03	78.8	28.77	55.70944

Розв'язок.

Критерій Дарбіна-Уотсона:

$$d = \frac{\sum_{t=p-1}^{n}(Y_{t-1} - Y_t)^2}{\sum_{t=p+1}^{n} Y_t^2} = \frac{28,77}{484,03} = 0,059438 \;;$$

Табличне значення критерія для n=18 та числа пояснюючих змінних κ=1 при рівні значимості p =0,95 складає: d_L=1,16; d_V=1,39. Оскільки $d < d_L$, то ряд містить автокореляцію; оскільки $d < d_U$, то для виявлення автокореляції необхідно провести подальші дослідження.

Оцінимо наявність автокореляції за критерієм Неймана:

$$\sigma^2 = \frac{1}{n-1} \cdot \sum_{t=2}^{n} \left(Y_t - Y_{t-1}\right)^2 = \frac{28,77}{17} = 1,692353;$$

$$S_u = \frac{1}{n-1} \cdot \sum_{t=1}^{n} \left(Y_t - \bar{Y}_t\right)^2 = \frac{55,70944}{18} = 3,094969.$$

$$Q = \frac{\sigma^2}{S_u^2} = \frac{1,692353}{3,094969} = 0,546808 \;;$$

Табличне значення $Q_{табл}$ – критерія для додатнього коефіцієнта автокореляції r >0 та p=0,95 і n=18 складає 1,3405. Оскільки $Q<Q_{табл}$, то існує позитивна автокореляція.

Питання для самоконтролю:

1. Сутність автокореляції. Навести приклади.
2. Причини автокореляції.
3. Способи усунення автокореляції в часових рядах.
4. Нециклічний коефіцієнт автокореляції.
5. Циклічний коефіцієнт автокореляції.
6. Оцінка автокореляції за критерієм Дарбіна-Уотсона.
7. Оцінка автокореляції за критерієм Неймана.

Тема 7

МОДЕЛІ АВТОРЕГРЕСІЇ. УЗАГАЛЬНЕНИЙ МЕТОД НАЙМЕНШИХ КВАДРАТІВ (МЕТОД ЕЙТКЕНА)

7.1 Поняття гомо- та гетероскедастичності

Одним з основних припущень моделі класичної лінійної регресії є припущення про сталість дисперсії кожної випадкової величини u_i. Це явище називається гомоскедастичністю.

Формалізовано це припущення записується у вигляді:

$$\operatorname{var}(u_i) = M(uu') = \sigma_u^2 \cdot I = \sigma_u^2 = const' \qquad (7.1)$$

Якщо це припущення не задовольняється у якомусь окремому випадку, то має місце гетероскедастичність.

$$\operatorname{var}(u_i) = \sigma_u^2 \neq const$$

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

Графічно випадок гомоскедастичності для простої лінійної регресії показано випадковою дисперсією u у межах сталої відстані навколо лінії регресії (рисунок 7.1 а).

Якщо σ_x^2 не є сталою, а її значення залежить від значення x, можемо записати $\sigma_u^2 = f(x_{1i}, x_{2i}, \ldots x_{mi})$ У цьому випадку маємо справу з гетероскедастичність. Графічна форма розкиду спостережень залежить від форми гетероскедастичності, тобто форми зв`язку між σ_u^2 та x_i. На рисунку 7.1 (б, в, г) показано три різні форми гетероскедастичності.

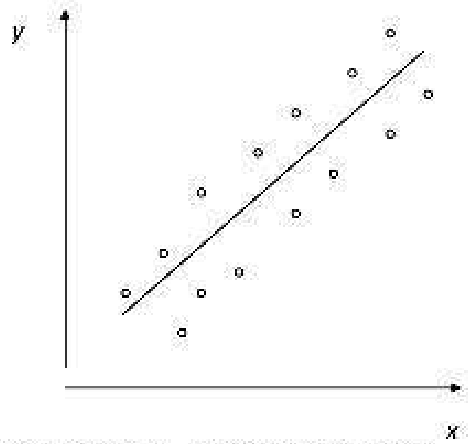


Рисунок 7.1 а – Гомоскедастичність

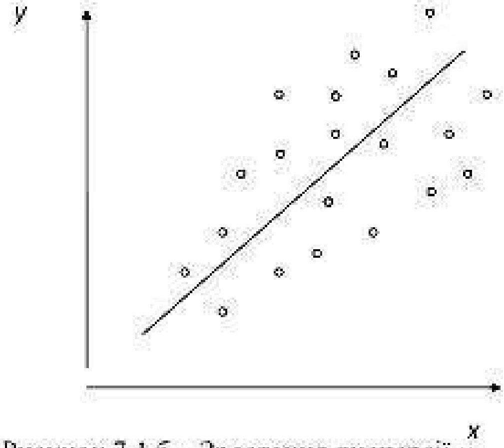


Рисунок 7.1 б – Зростання дисперсії u

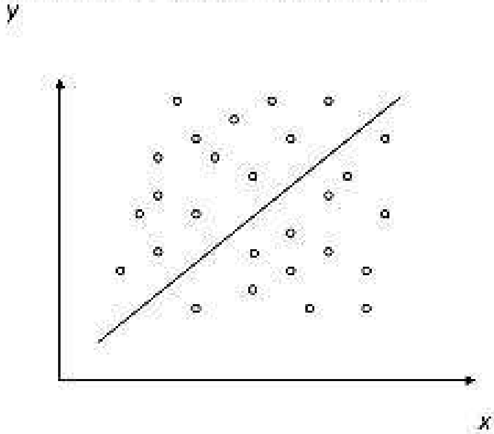


Рисунок 7.1 в – Спадання дисперсії u

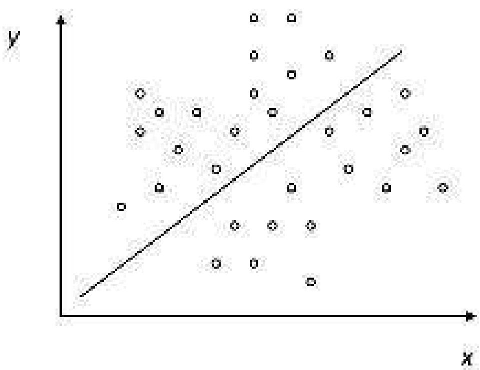


Рисунок 7.1 г – Складний випадок

Таким чином, можна дати такі визначення гомо- та гетероскедастичності.

Якщо дисперсія залишків стала для кожного спостереження, тобто $M(uu') = \sigma_u^2 \cdot I$, то ця її властивість називається гомоскедастичністю.

Якщо дисперсія залишків змінюється для кожного спостереження або групи спостережень, тобто $M(uu') = \sigma_u^2 \cdot S$, то це явище називається гетероскедастичністю.

Якщо існує гетероскедастичність залишків, то це спричиняється до того, що оцінки параметрів моделі 1МНК будуть незміщеними, обґрунтованими, але не ефективними. При цьому формулу для стандартної помилки оцінки застосовувати не можна.

Виявити наявність гетероскедастичності можна як графічними, так і аналітичними методами. Аналітичним найпростішим методом є тест рангової кореляції Спірмена, який можна використовувати як до малих, так і до великих вибірок.

Запишемо коефіцієнт рангової кореляції Спірмена:

$$r_s = 1 - 6 \cdot \left[\frac{\sum_{i=1}^{n} di}{n \cdot (n^2 - 1)} \right]. \tag{7.2}$$

де d – різниця між рангами, що приписуються двом характеристикам *i-го* об'єкта; n – кількість об'єктів, що ранжуються.

Розглянемо методику використання коефіцієнта рангової кореляції для визначення гетероскедастичності.

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

Етап 1. Будуємо регресію для даних $x\max$ і розраховуємо відхилення u_i.

Етап 2. Нехтуючи знаком u_i, тобто беручи абсолютні значення $|u_i|$ ранжуємо $|u_i|$ та x у зростаючому чи спадаючому порядку і підраховуємо коефіцієнт рангової кореляції Спірмена.

Етап 3. Перевіряємо значимість отриманого коефіцієнта рангової кореляції за t – критерієм Стьюдента. Для цього побудуємо статистику:

$$t = \frac{r_s \cdot \sqrt{n-2}}{\sqrt{1 - r_s^2}},$$

де n – кількість спостережень і, $k = (n-2)$ – кількість ступенів свободи. По таблицях знаходимо $t_{\alpha.k}$. Якщо розрахункове значення t перевищує критичне $t_{\alpha.k}$ $(t > t_{\alpha.k})$, то підтверджується гіпотеза про гетероскедастичність. Якщо $(t < t_{\alpha.k})$, то в регресійній моделі правильним є припущення про гомоскедастичність.

При великих вибірках перевірка гетероскедастичності здійснюється за тестами Гольдфельда та Квандта, тестом Глейсера та ін.

Коли на базі будь-якого тесту встановлено гетероскедастичність, то для її вилучення змінюють початкову модель таким чином, щоб помилки мали постійну дисперсію. Далі невідомі параметри трансформованої моделі розраховуються за методом найменших квадратів. Трансформація моделі зводиться до зміни первісної форми моделі, причому необхідно враховувати $\sigma_{u_i}^2$ та значеннями незалежних змінних:

$$\sigma_{u_i}^2 = f(x_i). \tag{7.4}$$

Розглянемо можливі випадки трансформації моделі на прикладі простої лінійної функції. Припустимо, що початкова модель $y_i = a_0 + a_1 x_i + u_i$, де випадкова величина u_i гетероскедастична, але відповідає всім іншим припущенням лінійної регресії.

Випадок 1. Припустимо, що гетероскедастичність має форму $E(u_i)^2 = \sigma_{u_i}^2 = k^2 x^2$ (де k – скінченна константа), тобто дисперсія u зростає пропорційно до x^2. Виражаючи коефіцієнт пропорційності k^2, маємо $k^2 = \sigma_{u_i}^2 / x^2$. Це означає, що трансформація моделі полягає у діленні початкової моделі на $\sqrt{x^2} = x$. Трансформована таким чином модель має вигляд:

$$\frac{y_i}{x_i} = \frac{a_0}{x_i} + a_1 + \frac{u_i}{x_i}. \tag{7.5}$$

Нове змінене значення випадкової величини $\frac{u_i}{x_i}$ є гомоскедастичним, оскільки:

$$E\left(\frac{u_i}{x_i}\right)^2 = \frac{1}{x_i^2} E(u_i)^2 = \frac{1}{x_i^2} \sigma_{u_i}^2. \tag{7.6}$$

Таким чином, ми можемо застосувати класичний МНК для розрахунку невідомих параметрів трансформованої моделі.

Випадок 2. Припустимо, що гетероскедастичність виражена формулою:

$$E(u_i^2)^2 = \sigma_{u_i}^2 = k^2 x_i. \tag{7.7}$$

Допустима трансформація полягає в діленні початкової моделі на $\sqrt{x}$, тобто:

$$\frac{y_i}{\sqrt{x_i}} = \frac{a_0}{\sqrt{x_i}} + a_1\frac{x_i}{\sqrt{x_i}} + \frac{u_i}{\sqrt{x_i}}. \tag{7.8}$$

Випадкова величина $\frac{u_i}{\sqrt{x_i}}$ є гомоскедастична із сталою дисперсією k^2.

Випадок 3. Припустимо, що форма гетероскедастичності:

$$E(u_i)^2 = \sigma_{u_i}^2 = k^2(a_0 + a_1x_1)^2. \tag{7.9}$$

Допустима трансформація полягає в діленні початкової моделі на $\sqrt{(a_0 + a_1x_1)^2} = (a_0 + a_1x)$, тобто:

$$\frac{y_i}{a_0 + a_1x_i} = a_0\cdot\frac{1}{a_0 + a_1x_i} + a_1\frac{x_i}{a_0 + a_1x_i} + \frac{u_i}{a_0 + a_1x_i}. \tag{7.10}$$

Нова випадкова величина є гомоскедастичною із сталою дисперсією, рівною k^2.

7.2 Розрахунок параметрів рівняння авторегресії

Параметри авторегресійної моделі

$$Y_t = a_0 + a_1Y_{t-1} + a_2Y_{t-2} + ... + a_pY_{t-p} + u$$

розрахуємо за методом найменших квадратів, який базується на мінімізації залишкової дисперсії: $\sum_{t=p+1}^{n}(Y_{t-p} - \hat{Y}_{t-p})^2 \to \min$.

Ця вимога приводить до системи нормальних рівнянь:

$$\begin{cases}
(n-p)\cdot a_0 + a_1\cdot\sum\limits_{t=p+1}^{n} Y_{t-1} + a_2\cdot\sum\limits_{t=p+1}^{n} Y_{t-2} + \dots + a_p\cdot\sum\limits_{t=p+1}^{n} Y_{t-p} = \sum\limits_{t=p+1}^{n} Y_t \\
a_0\cdot\sum\limits_{t=p+1}^{n} Y_{t-1} + a_1\cdot\sum\limits_{t=p+1}^{n} Y_{t-1}^2 + a_2\cdot\sum\limits_{t=p+1}^{n} Y_{t-2}\cdot Y_{t-1} + \dots + a_p\cdot\sum\limits_{t=p+1}^{n} Y_{t-p}\cdot Y_{t-1} = \sum\limits_{t=p+1}^{n} Y_t\cdot Y_{t-1} \\
a_0\cdot\sum\limits_{t=p+1}^{n} Y_{t-2} + a_1\cdot\sum\limits_{t=p+1}^{n} Y_{t-1}\cdot Y_{t-2} + a_2\cdot\sum\limits_{t=p+1}^{n} Y_{t-2}^2 + \dots + a_p\cdot\sum\limits_{t=p+1}^{n} Y_{t-p}\cdot Y_{t-2} = \sum\limits_{t=p+1}^{n} Y_t\cdot Y_{t-2} \\
\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\
a_0\cdot\sum\limits_{t=p+1}^{n} Y_{t-p} + a_1\cdot\sum\limits_{t=p+1}^{n} Y_{t-1}\cdot Y_{t-p} + a_2\cdot\sum\limits_{t=p+1}^{n} Y_{t-2}\cdot Y_{t-p} + \dots + a_p\cdot\sum\limits_{t=p+1}^{n} Y_{t-p}^2 = \sum\limits_{t=p+1}^{n} Y_t\cdot Y_{t-p}
\end{cases}$$

(7.11)

Якщо в рівнянні авторегресії відсутній вільний член a_0, то для отримання системи нормальних регресій в системі (7.11) необхідно викреслити перший рядок та перший стовпець.

$$\begin{cases}
a_1\cdot\sum\limits_{t=p+1}^{n} Y_{t-1}^2 + a_2\cdot\sum\limits_{t=p+1}^{n} Y_{t-2}\cdot Y_{t-1} + \dots + a_p\cdot\sum\limits_{t=p+1}^{n} Y_{t-p}\cdot Y_{t-1} = \sum\limits_{t=p+1}^{n} Y_t\cdot Y_{t-1} \\
a_1\cdot\sum\limits_{t=p+1}^{n} Y_{t-1}\cdot Y_{t-2} + a_2\cdot\sum\limits_{t=p+1}^{n} Y_{t-2}^2 + \dots + a_p\cdot\sum\limits_{t=p+1}^{n} Y_{t-p}\cdot Y_{t-2} = \sum\limits_{t=p+1}^{n} Y_t\cdot Y_{t-2}^n \\
\dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \quad \dots \\
a_1\cdot\sum\limits_{t=p+1}^{n} Y_{t-1}\cdot Y_{t-p} + a_2\cdot\sum\limits_{t=p+1}^{n} Y_{t-2}\cdot Y_{t-p} + \dots + a_p\cdot\sum\limits_{t=p+1}^{n} Y_{t-p}^2 = \sum\limits_{t=p-1}^{n} Y_t\cdot Y_{t-p}
\end{cases} \quad (7.12)$$

Визначивши по вихідному динамічному ряду всі суми, що входять у систему нормальних рівнянь, розрахуємо коефіцієнти авторегресії.

Запишемо системи нормальних рівнянь:

а) першого порядку ($p=1$) з вільним членом $Y_t = a_0 + a_1\cdot Y_{t-1}$:

$$\begin{cases}
(n-1)\cdot a_0 + a_1\cdot\sum\limits_{t=2}^{n} Y_{t-1} = \sum\limits_{t=2}^{n} Y_t \\
a_0\cdot\sum\limits_{t=2}^{n} Y_{t-1} + a_1\cdot\sum\limits_{t=2}^{n} Y_{t-1}^2 = \sum\limits_{t=2}^{n} Y_t\cdot Y_{t-1}
\end{cases}$$

без вільного члена $Y_t = a_1 \cdot Y_{t-1}$

$$a_1 \cdot \sum_{t=2}^{n} Y_{t-1}^2 = \sum_{t=2}^{n} Y_t \cdot Y_{t-1}; \qquad (7.14)$$

б) другого порядку ($p=2$) з вільним членом $Y_t = a_0 + a_1 \cdot Y_{t-1} + a_2 \cdot \sum Y_{t-2}$:

$$\begin{cases} (n-2)\cdot a_0 + a_1 \cdot \sum_{t=3}^{n} Y_{t-1} + a_2 \cdot \sum_{t=3}^{n} Y_{t-2} = \sum_{t=3}^{n} Y_t \\ a_0 \cdot \sum_{t=3}^{n} Y_{t-1} + a_1 \cdot \sum_{t=3}^{n} Y_{t-1}^2 + a_2 \cdot \sum_{t=3}^{n} Y_{t-2} \cdot Y_{t-1} = \sum_{t=3}^{n} Y_t \cdot Y_{t-1} \\ a_0 \cdot \sum_{t=3}^{n} Y_{t-2} + a_1 \cdot \sum_{t=3}^{n} Y_{t-1} \cdot Y_{t-2} + a_2 \cdot \sum_{t=3}^{n} Y_{t-2}^2 = \sum_{t=3}^{n} Y_t \cdot Y_{t-2} \end{cases} \qquad (7.15)$$

без вільного члена $Y_t = a_1 \cdot Y_{t-1} + a_2 \cdot Y_{t-2}$

$$\begin{cases} a_1 \cdot \sum_{t=3}^{n} Y_{t-1}^2 + a_2 \cdot \sum_{t=3}^{n} Y_{t-2} \cdot Y_{t-1} = \sum_{t=3}^{n} Y_t \cdot Y_{t-1} \\ a_1 \cdot \sum_{t=3}^{n} Y_{t-1} \cdot Y_{t-2} + a_2 \cdot \sum_{t=3}^{n} Y_{t-2}^2 = \sum_{t=3}^{n} Y_t \cdot Y_{t-2} \end{cases} \qquad (7.16)$$

в) третього порядку ($p=3$) $Y_t = a_0 + a_1 \cdot Y_{t-1} + a_2 \cdot Y_{t-2} + a_3 \cdot Y_{t-3}$:

$$\begin{cases} (n-3)\cdot a_0 + a_1 \cdot \sum_{t=4}^{n} Y_{t-1} + a_2 \cdot \sum_{t=4}^{n} Y_{t-2} + a_3 \cdot \sum_{t=4}^{n} Y_{t-3} = \sum_{t=4}^{n} Y_t \\ a_0 \cdot \sum_{t=4}^{n} Y_{t-1} + a_1 \cdot \sum_{t=4}^{n} Y_{t-1}^2 + a_2 \cdot \sum_{t=4}^{n} Y_{t-2} \cdot Y_{t-1} + a_3 \cdot \sum_{t=4}^{n} Y_{t-3} \cdot Y_{t-1} = \sum_{t=4}^{n} Y_t \cdot Y_{t-1} \\ a_0 \cdot \sum_{t=4}^{n} Y_{t-2} + a_1 \cdot \sum_{t=4}^{n} Y_{t-1} \cdot Y_{t-2} + a_2 \cdot \sum_{t=4}^{n} Y_{t-2}^2 + a_3 \cdot \sum_{t=4}^{n} Y_{t-3} \cdot Y_{t-2} = \sum_{t=4}^{n} Y_t \cdot Y_{t-2} \\ a_0 \cdot \sum_{t=4}^{n} Y_{t-3} + a_1 \cdot \sum_{t=4}^{n} Y_{t-1} \cdot Y_{t-3} + a_2 \cdot \sum_{t=4}^{n} Y_{t-2} \cdot Y_{t-3} + a_3 \cdot \sum_{t=4}^{n} Y_{t-3}^2 = \sum_{t=4}^{n} Y_t \cdot Y_{t-3} \end{cases} \qquad (7.17)$$

без вільного члена $Y_t = a_1 \cdot Y_{t-1} + a_2 \cdot Y_{t-2} + a_3 \cdot Y_{t-3}$:

$$\begin{cases} a_1 \cdot \sum_{t=4}^{n} Y_{t-1}^2 + a_2 \cdot \sum_{t=4}^{n} Y_{t-2} \cdot Y_{t-1} + a_3 \cdot \sum_{t=4}^{n} Y_{t-3} \cdot Y_{t-1} = \sum_{t=4}^{n} Y_t \cdot Y_{t-1} \\ a_1 \cdot \sum_{t=4}^{n} Y_{t-1} \cdot Y_{t-2} + a_2 \cdot \sum_{t=4}^{n} Y_{t-2}^2 + a_3 \cdot \sum_{t=4}^{n} Y_{t-3} \cdot Y_{t-2} = \sum_{t=4}^{n} Y_t \cdot Y_{t-2} \\ a_1 \cdot \sum_{t=4}^{n} Y_{t-1} \cdot Y_{t-3} + a_2 \cdot \sum_{t=4}^{n} Y_{t-2} Y_{t-3} + a_3 \cdot \sum_{t=4}^{n} Y_{t-3}^2 = \sum_{t=4}^{n} Y_t \cdot Y_{t-3} \end{cases} \quad (7.18)$$

Порядок рівняння авторегресії можна визначити так. Обчислимо параметри одночленної, двочленної, трьохчленної, p – членної моделі авторегресії та співставимо між собою залишкові дисперсії $\sigma_u^2 = \sum (Y_t - \hat{Y}_t)^2 / (n - 2p - 1)$. Мінімальна залишкова дисперсія $\sigma_{u(p)}^2$ відповідає періоду запізнення (p).

Таким чином, процес авторегресії, що вивчається, можна описати рівнянням p-*го* порядку.

Порядок рівняння авторегресії можна оцінити також за автокореляційною функцією; порядок рівняння авторегресії буде відповідати найбільшому по модулю значенню коефіцієнта автокореляції для часового зміщення p. Треба відзначити також, що найбільшому по модулю коефіцієнту автокореляції для часового зміщення p відповідає найменше значення залишкової дисперсії $\sigma_{u(p)}^2$.

7.3 Прогнозування по рівняннях авторегресії

Рівняння авторегресії є аналітичною основою для форму прогнозу з періодом випередження p часових одиниць (днів, місяців, років і т.п.).

Прогнозування змінної Y_{t+p} здійснюється в момент часу t на p кроків вперед і являє собою багатоетапну процедуру. На кожному етапі визначають величину показника на наступний одиничний відрізок часу. Послідовність розрахунків така. Спочатку встановлюють теоретичні значення змінної $\hat{Y}_t$ для $t = n+1$. При виконанні розрахунків в рівняння авторегресії підставляють p останніх членів динамічного ряду. Потім визначають величину показника $\hat{Y}_t$ для $t = n+2$. При цьому як аргумент в рівнянні авторегресії використовують зміну $\hat{Y}_t = n+1$, знайдену на попередньому етапі. В ході наступних обчислень визначають $\hat{Y}_t = n+3$, $\hat{Y}_t = n+4$, ..., $\hat{Y}_t = n+p$.

Точність прогнозу в значній мірі залежить від величини випереджання l. Чим більше l, тим вища можливість помилки прогнозу, тим значніші розходження між розрахунковими та фактичними значеннями змінної. Адекватність рівняння авторегресії можна оцінити за допомогою показника абсолютного середнього відхилення $\left|\nu_{сер}\right| = \left[\sum_{t=p+l-1}^{n}\left(Y_t - \hat{Y}_t\right)\right] / (n-p-l+1)$. Обчислення $\left|\nu_{сер}\right|$ необхідно проводити для передпрогнозного періоду часу шляхом співставлення розрахункових і фактичних рівнів ряду динаміки.

Приклад 7.1. По статистичних даних таблиці 7.1 розрахувати коефіцієнти автокореляції, вибрати найбільший по модулю, який буде визначати порядок авторегресійної моделі та побудувати авторегресійну модель відповідного порядку. Методом прогнозної екстраполяції розрахувати прогнозні значення показника Y_t на 19, 20, 21 місяці по авторегресійній моделі і порівняти їх з прогнозом по однофакторній моделі. Побудувати графік тренда авторегресійної моделі.

Розв'язок.

Розрахуємо суми, середні та середньоквадратичні відхилення (результати розрахунків у таблиці 7.1).

$$\overline{Y_t} = \frac{1}{(n-2)} \cdot \sum_{t=3}^{n} Y_t = \frac{82,4}{16} = 5,15; \; \overline{Y_{t-2}} = \frac{1}{(n-2)} \cdot \sum_{t=3}^{n} Y_{t-2} = \frac{72,1}{16} = 4,5063;$$

$$\overline{Y_t \cdot Y_{t-2}} = \frac{1}{(n-2)} \cdot \sum_{t=3}^{n} Y_t \cdot Y_{t-2} = \frac{401,32}{16} = 25,0825;$$

$$\sigma_{Y_t} = \sqrt{\frac{\sum_{t=3}^{n}\left(Y_t - \overline{Y}_t\right)^2}{n-2}} = \sqrt{\frac{45,98}{16}} = 1,6952;$$

$$\sigma_{Y_{t-2}} = \sqrt{\frac{\sum_{t=3}^{n}\left(Y_{t-2} - \overline{Y}_{t-2}\right)^2}{n-2}} = \sqrt{\frac{28{,}82938}{16}} = 1{,}34233.$$

Таблиця 7.1 – Вхідні дані та результати розрахунків

t	Y_t	Y_{t-1}	Y_{t-2}	$Y_t \cdot Y_{t-1}$	$Y_t \cdot Y_{t-2}$	$Y_{t-1} \cdot Y_{t-2}$	Y_{t-1}^2	Y_{t-2}^2	$\left(Y_t - \overline{Y}_t\right)^2$	$\left(Y_{t-2} - \overline{Y}_{t-2}\right)^2$	$\hat{Y}_t$
1	2	3	4	5	6	7	8	9	10	11	12
1	2,2										
2	3,7	2,2		8,14			4,84		2,10250		
3	4,1	3,7	2,2	15,17	9,02	8,14	13,69	4,84	1,10250	5,31879	2,88
4	4	4,1	3,7	16,4	14,8	15,17	16,81	13,69	1,32250	0,65004	4,25
5	4,2	4	4,1	16,8	17,22	16,4	16	16,81	0,90250	0,16504	4,54
6	2,9	4,2	4	12,18	11,6	16,8	17,64	16	5,06250	0,25629	4,53
7	3,5	2,9	4,2	10,15	14,7	12,18	8,41	17,64	2,72250	0,09379	4,24
8	4,9	3,5	2,9	17,15	14,21	10,15	12,25	8,41	0,06250	2,58004	3,38
9	3,4	4,9	3,5	16,66	11,9	17,15	24,01	12,25	3,06250	1,01254	4,36
10	4,3	3,4	4,9	14,62	21,07	16,66	11,56	24,01	0,72250	0,15504	4,99
11	4,8	4,3	3,4	20,64	16,32	14,62	18,49	11,56	0,12250	1,22379	4,07
12	5,5	4,8	4,3	26,4	23,65	20,64	23,04	18,49	0,12250	0,04254	4,98
13	4,5	5,5	4,8	24,75	21,6	26,4	30,25	23,04	0,42250	0,08629	5,64
14	6,6	4,5	5,5	29,7	36,3	24,75	20,25	30,25	2,10250	0,98754	5,87
15	5,8	6,6	4,5	38,28	26,1	29,7	43,56	20,25	0,42250	0,00004	5,78
16	7,7	5,8	6,6	44,66	50,82	38,28	33,64	[illegible].56	6,50250	4,38379	7,22
17	6,7	7,7	5,8	51,59	38,86	44,66	59,2[illegible]	[illegible].64	2,40250	1,67379	7,23
18	9,5	6,7	7,7	63,65	73,15	51,59	44,[illegible]	[illegible]9,29	18,92250	10,20004	8,44
171	82,4	76,6	72,1	418,8	401,32	363,29	39[illegible]	353,73	45,98000	28,82938	82,40
	5,15	4,7875	4,5063	26,175	25,0825		[illegible]	[illegible]	2,87375	1,80184	
19											8,59
20											9,71
21											10,22

Розраховуємо коефіцієнт автокор[illegible]того порядку:

$$r_2 = \frac{\overline{Y_t \cdot Y_{t-2}} - \overline{Y}_t \cdot \overline{Y}_{t-2}}{\sigma_{Y_t} \cdot \sigma_{Y_{t-2}}} = [illegible] \frac{2[illegible]5 \cdot 4{,}5063}{[illegible].69[illegible]34233} = 0{,}824.$$

Система нормальних рівнянь другого п[illegible]ку:

$$\begin{cases} (n-2)\cdot a_0 + a_1 \cdot \sum_{t=3}^{n} Y_{t-1} + a_2 \cdot \sum_{t=3}^{n} Y_{t-2} = \sum_{t=3}^{n} Y_t \\ a_0 \cdot \sum_{t=3}^{n} Y_{t-1} + a_1 \cdot \sum_{t=3}^{n} Y_{t-1}^2 + a_2 \cdot \sum_{t=3}^{n} Y_{t-2}\cdot Y_{t-1} = \sum_{t=3}^{n} Y_t \cdot Y_{t-1} \\ a_0 \cdot \sum_{t=3}^{n} Y_{t-2} + a_1 \cdot \sum_{t=3}^{n} Y_{t-1}\cdot Y_{t-2} + a_2 \cdot \sum_{t=3}^{n} Y_{t-2}^2 = \sum_{t=3}^{n} Y_t \cdot Y_{t-2} \end{cases},$$

або у матричній формі:

$Y' \cdot A = Y$. Звідки $A = Y'^{-1} \cdot Y$.

Сформуємо матрицю Y' та Y:

$$Y' = \begin{bmatrix} 16 & 76,6 & 72,1 \\ 76,6 & 393,78 & 363,29 \\ 72,1 & 363,29 & 353,73 \end{bmatrix}, \quad Y = \begin{bmatrix} 82,4 \\ 418,8 \\ 401,32 \end{bmatrix}.$$

Обернена матриця Y'^{-1}:

$$Y'^{-1} = \begin{bmatrix} 1,01101 & -0,12478 & -0,07792 \\ -0,12478 & 0,06378 & -0,04007 \\ -0,07792 & -0,04007 & 0,05986 \end{bmatrix}.$$

Вектор параметрів авторегресійної моделі:

$$A = Y'^{-1} \cdot Y = \begin{bmatrix} 1,01101 & -0,12478 & -0,07792 \\ -0,12478 & 0,06378 & -0,04007 \\ -0,07792 & -0,04007 & 0,05986 \end{bmatrix} \cdot \begin{bmatrix} 82,4 \\ 418,8 \\ 401,32 \end{bmatrix} = \begin{bmatrix} -0,22136 \\ 0,34824 \\ 0,82201 \end{bmatrix}$$

Автокореляційна модель:

$$Y_t = -0,22136 + 0,34824 Y_{t-1} + 0,82201 Y_{t-2}$$

Розрахуємо теоретичні та прогнозні значення рівнів часового ряду $\hat{Y}_t$:

$$\hat{Y}_3 = -0,22136 + 0,34824 \cdot 3,7 + 0,8220 \cdot 2,2 = 2,88;$$

$$\hat{Y}_4 = -0,22136 + 0,34824 \cdot 4,1 + 0,8220 \cdot 3,7 = 4,25;$$

… … … … … … … … …

$$\hat{Y}_{18} = -0,22136 + 0,34824 \cdot 6,7 + 0,8220 \cdot 7,7 = 8,44;$$

$$\hat{Y}_{19} = -0,22136 + 0,34824 \cdot 9,5 + 0,8220 \cdot 6,7 = 8,59;$$

$$\hat{Y}_{20} = -0,22136 + 0,34824 \cdot 8,59 + 0,8220 \cdot 9,5 = 9,71;$$

$$\hat{Y}_{21} = -0,22136 + 0,34824 \cdot 9,71 + 0,8220 \cdot 8,59 = 10,22.$$

Будуємо графік тренда авторегресійної моделі.

7.4 Метод Ейткена

Якщо за допомогою критеріїв Дарбіна-Уотсона, Неймана або циклічного коефіцієнта автокореляції виявлено наявність автокореляції в залишках, то необхідно знайти адекватний спосіб перерахунку параметрів моделі або покращити специфікацію моделі.

Найбільш ефективним способом оцінювання параметрів в регресійних рівняннях з автокорельованими залишками є узагальнений метод найменших квадратів, відомий як метод Ейткена.

Допустимо, що коваріаційна матриця залишків позитивно визначена і має вигляд:

$$E(uu') = S = R'R, \tag{7.19}$$

де R – не вироджена квадратна матриця порядку n.

Помножимо рівняння регресії:

$$y = X \cdot A + u \tag{7.20}$$

зліва на R^{-1}, отримаємо:

$$R^{-1} \cdot y = R^{-1} \cdot X \cdot A + u \tag{7.21}$$

Тоді коваріаційна матриця помилок R^{-1} буде діагональною:

$$E\left(R^{-1} uu' \cdot \left(R^{-1}\right)' \right) = R^{-1} \cdot S \left(R^{-1}\right)' = I \tag{7.22}$$

Таким чином, до перетвореного рівняння (7.21) можна використати звичайний метод найменших квадратів. Оцінку параметрів A тоді отримаємо із співвідношення:

$$A = \left(X' \cdot S^{-1}X\right)^{-1} X' \cdot S^{-1}Y \tag{7.23}$$

Нажаль, апріорі ми не знаємо ні значень помилок, ні коваріаційної матриці S. Якщо залишки u_t утворюють стаціонарний авторегресійний процес першого порядку, то автоковаріаційна матриця помилок буде мати вигляд:

$$S = \sigma_u^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 \ldots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \rho^3 \ldots & \rho^{n-2} \\ \rho^2 & \rho^1 & \rho & \rho^2 \ldots & \rho^{n-3} & \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \rho^{n-5} \ldots & 1 \end{bmatrix} \tag{7.24}$$

У цій симетричній матриці ρ^S виражає коефіцієнт автокореляції *s-го* порядку для залишків u_t. Очевидно, що коефіцієнт автокореляції нульового порядку дорівнює 1.

Оскільки коваріація залишків ρ^S при $s > 2$ часто наближається до нуля, то матриця, обернена до матриці S, матиме такий вигляд:

$$S^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} 1 & -\rho & 0 & 0 & 0 \ldots 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 \ldots 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 \ldots 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 \ldots 1 \end{bmatrix} \tag{7.25}$$

Таку матрицю іноді можна використовувати при оцінюванні параметрів моделі з автокорельованими залишками за методом Ейткена.

Коваріаційну матрицю S для вектора залишків $u = \left(u_1, u_2, u_3, \ldots u_n\right)'$ можна записати в загальній формі так:

$$S = \begin{bmatrix} u_1 & u_2 & u_3 & \dots & u_t & \dots & u_n \\ \sigma_{u_1}^2 & \sigma_{u_1u_2} & \sigma_{u_1u_3} & \dots & \sigma_{u_1u_t} & \dots & \sigma_{u_1u_n} \\ \sigma_{u_2u_1} & \sigma_{u_2}^2 & \sigma_{u_2u_3} & \dots & \sigma_{u_2u_t} & \dots & \sigma_{u_2u_n} \\ \sigma_{u_3u_1} & \sigma_{u_3u_2} & \sigma_{u_3}^2 & \dots & \sigma_{u_3u_t} & \dots & \sigma_{u_3u_n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \sigma_{u_tu_1} & \sigma_{u_tu_2} & \sigma_{u_tu_3} & \dots & \sigma_{u_t}^2 & \dots & \sigma_{u_tu_n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \sigma_{u_nu_1} & \sigma_{u_nu_2} & \sigma_{u_nu_3} & \dots & \sigma_{u_nu_t} & \dots & \sigma_{u_n}^2 \end{bmatrix} \tag{7.26}$$

Матрицю (7.26) можна записати як:

$$S = \sigma_u^2 \cdot \Omega \tag{7.27}$$

На головній діагоналі коваріаційної матриці *t-й* елемент $\sigma_{u_t}^2$ являє собою дисперсію $\mathrm{var}(u_t)$ відхилень випадкової змінної для u_t в *t-му* рівнянні $y_t = a_1x_{1t} + \ldots + a_jx_{jt} + \ldots + a_mx_{mt} + u_t$.

Поза головною діагоналлю цієї коваріаційної матриці в t^**-му* рядку *t-го* стовпця є елемент $\sigma_{u_{t^*}u_t}$ $(t^* \neq t)$. Він виражає коваріацію відхилень t^*-*го* стовпця і *t-го* рівняння. Матриця S симетрична.

Якщо $\sigma_{u_{t^*}u_t}$ додатня, то обидві змінні відхилень ...ріюють в одному напрямку. Це значить, що змінні u_{t^*} і u_t додатньо корельовані.

Якщо $\sigma_{u_{t^*}u_t}$ від'ємна, то обидві змінні варіюють у протилежних напрямках. В цьому випадку змінні u_{t^*} і u_t від'ємно корельовані.

Якщо $\sigma_{u_{t^*}u_t}$ рівні нулю, то відхилення u_{t^*} і u_t не корельовані.

В класичній регресійній моделі $(n \times n)$ – мірна коваріаційна матриця вектора відхилень U розміром $(n \times 1)$ м… такий вигляд:

$$S = \sigma_U^2 \cdot I = \begin{bmatrix} \sigma_U^2 & 0 & 0 \ldots 0 \ldots 0 \\ 0 & \sigma_U^2 & 0 \ldots 0 \ldots 0 \\ 0 & 0 & \sigma_U^2 \ldots 0 \ldots 0 \\ \ldots & \ldots & \ldots \\ 0 & 0 & 0 \ldots \sigma_U^2 \ldots 0 \\ \ldots & \ldots & \ldots \\ 0 & 0 & 0 \ldots 0 \ldots \sigma_U^2 \end{bmatrix} \qquad (7.28)$$

При цьому I є одиничною матрицею порядку n. Порівнюючи вирази (7.26) і (7.28), можна зауважити, що в класичній регресійній моделі має місце рівність $\Omega = I$, тобто в класичній регресійній моделі досить розрахувати значення одного елемента, а саме дисперсію відхилень σ_U^2, щоб знати всі елементи матриці S, які є невідомими згідно умови. В цьому випадку проблема зводиться до знаходження $\chi \sigma_u^2$ і ρ. Їх оцінки можна отримати таким чином:

1) розрахувати параметри рівняння регресії за допомогою звичайного метода наймени х квадратів (1МНК);

2) підставляючи ці оі івняння регресії $\hat{y} = X \cdot A + U$, знайдемо значення залишків;

3) по формулах $\sigma_u^2 = \frac{}{n-\quad -1} \cdot u' \cdot u$ та $r' = \dfrac{\frac{1}{1-n} \cdot \sum_{t=2}^{n} u_t \cdot u_{t-1}}{\frac{1}{n} \sum_{t=1}^{n} u_t^2}$ знайдемо

оцінки σ_u^2 і r' та перевіримо зал и на наявність автокореляції;

4) оскільки параметр r' має щення, то при формуванні матриці S необхідно скорегувати коефі т автокореляції r' на величину зміщення:

$$r'_{скор} = \frac{u_t \cdot u_{t-1}}{u^2} \cdot \frac{n}{n-1} + \frac{m+1}{n}, \qquad (7.29)$$

де $\frac{m+1}{n}$ – величина зміщення кі кість незалежних змінних;

5) формування матриці коваріації залишків *(S)* та знаходження оберненої матриці *(S^{-1})*;

6) оцінка параметрів методом Ейткена, тобто за формулою:

$$A = \left(X' \cdot S^{-1} \cdot X\right)^{-1} \cdot X' \cdot S^{-1} \cdot Y \tag{7.30}$$

Крім узагальненого методу найменших квадратів (УМНК) Ейткена для усунення автокореляції залишків використовують ітераційний метод, запропонований Д. Кокрейном та Ж. Оркуттом, який базується на авторегресійному перетворенні змінних. Проблема автокорекції також вирішується за допомогою перетворень Койка.

Приклад 7.2. По даних таблиці 7.2 побудувати економетричну модель, що характеризує залежність величини пропозиції від ціни на окремий вид товару.

Таблиця 7.2 – Вхідні дані

Рік	1	2	3	4	5	6	7	8	9	10
Ціна	2.5	2.7	3.2	3.5	3.7	4.0	4.3	4.6	5.0	5.5
Пропозиція	4.2	4.5	4.8	5.2	5.4	5.8	6.5	6.9	7.7	8.1

Розв'язок.

Позначимо величину пропозиції через y, а ціну – через x. Тоді загальний вигляд економетричної моделі:

$y_t = a_0 + a_1 x + u_t$,

де a_0, a_1 – оцінки параметрів моделі;

u_t – залишки.

Крок 1. Визначимо параметри моделі a_0, a_1 на основі методу найменших квадратів, припускаючи, що залишки u_t не корельовані: $A = (X'X)^{-1} \cdot (X'Y)$,

де X' – матриця, транспонована до матриці X.

Формуємо матриці X та Y:

$$x = \begin{bmatrix} 1 & 2,5 \\ 1 & 2,7 \\ 1 & 3,2 \\ 1 & 3,5 \\ 1 & 3,7 \\ 1 & 4,0 \\ 1 & 4,3 \\ 1 & 4,6 \\ 1 & 5,0 \\ 1 & 5,5 \end{bmatrix}; \qquad y = \begin{bmatrix} 4,2 \\ 4,5 \\ 4,8 \\ 5,2 \\ 5,4 \\ 5,8 \\ 6,5 \\ 6,9 \\ 7,7 \\ 8,1 \end{bmatrix}$$

$$X'X = \begin{bmatrix} 10,0 & 39,0 \\ 39,0 & 160,6 \end{bmatrix}; \ (X'X)^{-1} = \begin{bmatrix} 1,88521 & -0,45775 \\ -0,45775 & 0,11737 \end{bmatrix}; \ X'Y = \begin{bmatrix} 59,10 \\ 242,13 \end{bmatrix};$$

$$A = \begin{bmatrix} -0,58183 \\ 1,36620 \end{bmatrix}$$

$$a_0 = -0,58183; \qquad a_1 = 1,36620; \quad \sigma_u^2 = 0,04949; \qquad R^2 = 0,98429 .$$

Економетрична модель має вигляд: $\hat{y}_t = -0,58183 + 1,36620x + u_t$.

Крок 2. На основі економетричної моделі $\hat{y}_t = -0,58183 + 1,36620x + u_t$ знайдемо розрахункові значення пропозиції $\hat{y}_t$ і порівняємо їх з фактичними. Тоді $u_t = y_t - \hat{y}_t$.

Таблиця 7.3 – Результати розрахунків

Рік	y_t	$\hat{y}_t$	u_t	u_t^2	$u_t \cdot u_{t-1}$	$(u_t - u_{t-1})^2$	$u_t \cdot u_{t-1}$
1	2	3	4	5	6	7	8
1	4,2	3,99732	0,20268	0,041079	—	—	—
2	4,5	4,27056	0,22944	0,052643	0,02676	0,000518	0,046503
3	4,8	4,93566	-0,15366	0,023611	-0,3831	0,146766	-0,03526
4	5,2	5,36352	-0,16352	0,026739	-0,00986	0,0000972	0,025126
5	5,4	5,63676	-0,23676	0,056055	-0,07324	0,005364	0,038715
6	5,8	6,04662	-0,24662	0,060821	-0,00986	0,0000972	0,05839
7	6,5	6,45648	0,04352	0,001894	0,29014	0,084181	-0,01073
8	6,9	6,86634	0,03366	0,001133	-0,00986	0,0000972	0,001465
9	7,7	7,41282	0,28718	0,082472	0,25352	0,064272	0,009666
10	8,1	8,09592	0,00408	0,0000166	-0,2831	0,080146	0,001172
				0,346465		0,381539	0,135049

Знайдемо критерій Дарбіна-Уотсона:

$$d=\frac{\sum_{t=2}^{10}\left(u_t-u_{t-1}\right)^2}{\sum_{t=1}^{10}u_t^2}=\frac{0{,}381539}{0{,}346465}=1{,}1$$

Оскільки критерій d менше двох, то можна стверджувати про існування додатньої або прямої автокореляції. Зробити порівняння фактичного значення критерію d з табличними неможливо, бо критичні значення подані, починаючи з n=25 (в даному випадку n=10).

Критерій Неймана $Q=\frac{10}{10-1}\cdot d=\frac{10}{9}\cdot 1{,}1=1{,}22$. Це значення порівнюється з табличним. $Q_{табл}=1{,}18$ при n=10 і рівні значимості $\alpha=0{,}05$.

Так як $Q_{факт}>Q_{табл}$, то існує додатня автокореляція залишків.

Циклічний коефіцієнт автокореляції:

$$r'=\frac{\frac{1}{n-1}\cdot\sum_{t=2}^{n}u_t\cdot u_{t-1}}{\frac{1}{n}\cdot\sum_{t=1}^{n}u_t^2}=\frac{\frac{1}{9}\cdot 0{,}135049}{\frac{1}{10}\cdot 0{,}346465}=0{,}433$$

$r'>r_{табл}$, таким чином існує автокореляція залишків u_t

Крок 3. Формування матриці S.

Оскільки параметр r' має зміщення, то при формуванні матриці S необхідно скоригувати коефіцієнт автокореляції r' на величину зміщення.

$$r'_{скор}=\frac{\sum_{t=2}^{n}u_t\cdot u_{t-1}}{\sum_{t=1}^{n}u_t^2}\cdot\frac{n}{n-1}+\frac{m+1}{n}=\frac{0{,}135040}{0{,}3464659}\cdot\frac{10}{9}+\frac{1+1}{10}=0{,}433+0{,}2=0{,}633$$

$$\text{або } \rho=r'_{скор}=r'+\frac{m+1}{n}=0.433+0.2=0.633$$

Матриця коваріації S буде мати вигляд:

$$S = \sigma_u^2 \cdot \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 & \rho^7 & \rho^8 & \rho^9 \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 & \rho^7 & \rho^8 \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 & \rho^7 \\ \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^8 & \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^9 & \rho^8 & \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

$$\sigma_u^2 = \frac{\sum_{t=1}^{n} u_t^2}{n} = \frac{0.346465}{10} = 0.03346465$$

$$a_0 = 1.991; \quad a_1 = 1.028$$

Записуємо рівняння регресії: $\hat{y}_t^* = 1.991 + 1.028x$

Знайдемо розрахункові значення $\hat{y}_t^*$ та визначимо залишки W_t (таблиця7.4)

Таблиця 7.4 – Результати розрахунків

Рік	y_t	$\hat{y}_t^*$	w_t	w_t^2	$w_t - w_{t-1}$	$(w_t - w_{t-1})^2$
1	4,2	4,561	-0,361	0,130321	—	—
2	4,5	4,7666	-0,2666	0,071076	0,0944	0,0088911
3	4,8	5,2806	-0,4806	0,230976	-0,214	0,045796
4	5,2	5,589	-0,389	0,151321	0,0916	0,008391
5	5,4	5,7946	-0,3946	0,155709	-0,0056	0,0000314
6	5,8	6,103	-0,303	0,091809	0,0916	0,008391
7	6,5	6,4114	0,0886	0,00785	0,3916	0,153351
8	6,9	6,7198	0,1802	0,032472	0,0916	0,008391
9	7,7	7,131	0,569	0,323761	0,3888	0,151165
10	8,1	7,645	0,455	0,207025	-0,114	0,012996
				1,40232		0,397422

Розрахуємо критерії Дарбіна-Уотсона і Неймана:

$$d=\frac{\sum_{t=2}^{n}\left(w_t-w_{t-1}\right)^2}{\sum_{t=1}^{n}w_t^2}=\frac{0,397422}{1,40232}=0,2834 \qquad Q=\frac{n}{n-1}\cdot d=\frac{10}{9}\cdot 0,2834=0,3149$$

Циклічний коефіцієнт автокореляції:

$$r'=\frac{\frac{1}{n-1}\cdot\sum_{t=2}^{n}w_t\cdot w_{t-1}}{\frac{1}{n}\cdot\sum_{t=1}^{n}w_t^2}=\frac{\frac{1}{9}0,397422}{\frac{1}{10}\cdot 1,40232}=\frac{10}{9}\cdot 0,2834=0,31489$$

Оскільки $Q<Q_{табл}$, а $r'<r_{табл}$ при числі спостережень n=10 та рівні значимості $\alpha=0,05$, то автокореляція відсутня.

Питання для самоконтролю:

1. Поняття гомоскедастичності та гетероскедастичності.
2. Методи виявлення гетероскедастичності.
3. Розрахунок параметрів авторегресійної моделі.
4. Обґрунтування порядку рівняння авторегресії.
5. Розрахунок теоретичних та прогнозних рівнів часового ряду.
6. Метод Ейткена.

Тема 8

ЕКОНОМЕТРИЧНІ МОДЕЛІ НА ОСНОВІ СИСТЕМИ СТРУКТУРНИХ РІВНЯНЬ

8.1 Система одночасних рівнянь

Задачі, що наводились раніше, зводились до розгляду односторонніх стохастичних причинних залежностей між економічними явищами, які моделювались однією регресією $y = f(x_1, x_2, \ldots, x_j, x_m)$, тобто вивчався вплив декількох факторів на один показник, наприклад, залежність обсягу виробництва продукції від витрат праці та капіталу, залежність попиту від факторів, що його формують.

В економіці найчастіше мають місце задачі, що відображають багатосторонні одночасні зв'язки між економічними явищами, які описуються системою регресій. Системи регресій, що відображають присутність одночасних багатосторонніх зв'язків між економічними явищами, називається системою одночасних регресій.

Розглянемо просту кейнсіанську модель формування доходів.

$$C_t = \alpha + \beta Y_t + u_t, \tag{8.1}$$

$$Y_t = C_t + I_t, \tag{8.2}$$

де Y – сукупний дохід, C – обсяг споживання, I – інвестиції.

Кейнсіанська модель складається з двох рівнянь, одне з яких є регресійним, а друге – тотожність.

Необхідно оцінити параметри рівняння функції споживання.

Підставимо вираз (8.1) у вираз (8.2), отримаємо:

$$Y_t = \alpha + \beta Y_t + u_t + I_t$$

$$(Y_t - \beta Y_t) = \alpha + \dot{I}_t + u_t$$

$$Y_t(1 - \beta) = \alpha + \dot{I}_t + u_t$$

$$Y_t = \frac{\alpha}{1 - \beta} + \frac{\dot{I}_t}{1 - \beta} + \frac{u_t}{1 - \beta}$$

Системи одночасних структурних рівнянь, як правило, включають лінійні рівняння. Запишемо економетричну модель на основі системи одночасних рівнянь:

$$\begin{cases} y_{1t} = b_{11}y_{1t} + \ldots + b_{1k}y_{kt} + a_{10}\cdot x_{0t} + \ldots + a_{1m}x_{mt} + u_{1t} \\ y_{2t} = b_{21}y_{1t} + \ldots b_{2k}y_{kt} + a_{20}\cdot x_{0t} + \ldots + a_{2m}x_{mt} + u_{2t} \\ y_{kt} = b_{k1}y_{1t} + \ldots + b_{kk}y_{kt} + a_{k0}\cdot x_{0t} + \ldots + a_{km}x_{mt} + u_{kt} \end{cases} \quad (8.3)$$

У цій моделі $x_{0t} = 1$; окремі коефіцієнти $b_{11}, b_{kk}, a_{10}, \ldots, a_{kt}$ можуть дорівнювати нулю, якщо відповідна змінна не входить до рівняння. Залишки u_{st}, де $s = 1,2,\ldots,k$, також можуть дорівнювати нулю, якщо відповідне рівняння є тотожністю.

Систему (8.3) можна записати у матричній формі:

$$Y = BY + AX + u, \quad (8.4)$$

де Y – вектор ендогенних залежних змінних; X – матриця ендогенних пояснювальних змінних; u – вектор залишків; A – матриця коефіцієнтів при змінних Y розміром $k \times k$; B – матриця коефіцієнтів при змінних X розміром $k \times m$.

В процесі оцінювання параметрів рівняння економетричної моделі важливо виділити ендогенні та екзогенні змінні.

Так, обсяг споживання C та сукупний дохід Y в моделі є ендогенними змінними, які приймають свої значення в рівнянні функції споживання і в тотожності для сукупного доходу.

Екзогенними є змінні, значення яких визначаються поза межами моделі і тому беруться як задані. В моделі формування доходу розмір інвестицій I – екзогенна змінна. Модель не пояснює, як одержується значення цієї змінної, вона просто використовується як наперед задана.

Екзогенні величини визначають ендогенні величини, але самі від них не залежать, тобто між ними існує одностороння стохастична залежність.

При побудові моделей ендогенна величина може залежати від екзогенної величини або ендогенної величини за попередні періоди.

Економетрична модель (8.3), яка безпосередньо відображає структуру зв'язків між змінними називається структурною формою економетричної моделі.

8.2 Система незалежних регресій

Можливі випадки, що кожна з регресій економетричної моделі має тільки одну ендогенну величину, яка підлягає обчисленню і яка не залежить від ендогенних величин інших регресій та не здійснює вплив на інші ендогенні величини. Така економетрична модель називається системою незалежних регресій. У загальному вигляді систему незалежних регресій можна записати таким чином:

$$\begin{aligned} y_1 &= a_{10} + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1j}x_j + \ldots + a_{1m}x_m + u_1 \\ y_2 &= a_{20} + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2j}x_j + \ldots + a_{2m}x_m + u_2 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n &= a_{n0} + a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nj}x_j + \ldots + a_{nm}x_m + u_{n2} \end{aligned} \qquad (8.5)$$

або у матричній формі:

$$Y = A \cdot X + u. \qquad (8.6)$$

Параметри системи незалежних регресій оцінюються методом найменших квадратів (МНК). Для цього формуємо матриці емпіричних значень X та Y.

$$X = \begin{bmatrix} 1 & X_{11} & X_{21} \ldots X_{j1} \ldots X_{m1} \\ 1 & X_{12} & X_{22} \ldots X_{j2} \ldots X_{m2} \\ 1 & X_{13} & X_{23} \ldots X_{j3} \ldots X_{m3} \\ \ldots & \ldots & \ldots\ldots\ldots\ldots\ldots\ldots \\ 1 & X_{1i} & X_{2i} \ldots X_{ji} \ldots X_{mi} \\ \ldots & \ldots & \ldots\ldots\ldots\ldots\ldots\ldots \\ 1 & X_{1n} & X_{2n} \ldots X_{jn} \ldots X_{mn} \end{bmatrix} \qquad (8.7)$$

$$Y = \begin{bmatrix} Y_{11} & Y_{21} \ldots Y_{k1} \\ Y_{12} & Y_{22} \ldots Y_{k2} \\ Y_{13} & Y_{23} \ldots Y_{k3} \\ \ldots & \ldots\ldots\ldots \\ Y_{1i} & Y_{2i} \ldots Y_{ki} \\ \ldots & \ldots\ldots\ldots \\ Y_{1n} & Y_{2n} \ldots Y_{kn} \end{bmatrix} \qquad (8.8)$$

Якщо $(X'X) \neq 0$, то вектор параметрів моделі дорівнює:

$$A = (X'X)^{-1} \cdot (X'Y). \qquad (8.9)$$

Прикладом економетричної моделі, яка описується системою незалежних регресій може бути модель попиту та пропозиції, якщо попит та пропозиція формуються під впливом одних і тих же екзогенних величин X_j або факторів. Наприклад, з ціни певного виду товару X_1, рівня насиченості ринку цим товаром X_2, рівня доходів X_3 та рівня заощаджень населення X_4 та ін.

Приклад 8.1. Побудувати економетричну модель попиту і пропозиції системи незалежних регресій та знайти точку рівноваги:

Розв'язок.

Допускаємо, що залежить між ціною, попитом та пропозицією описується параболою:

$$y_1 = a_0 + a_1 x + a_2 x^2 + u_1,$$

$$y_2 = b_0 + b_1 x_1 + b_2 x^2 = u_2.$$

Проводимо лінеаризацію вхідних даних $x' = x^2$ та формуємо матриці X та Y:

$$X = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \\ 1 & 6 & 36 \\ 1 & 7 & 49 \\ 1 & 8 & 64 \\ 1 & 9 & 81 \\ 1 & 10 & 100 \\ 1 & 11 & 121 \end{bmatrix}; \quad Y = \begin{bmatrix} 10{,}20 & 2{,}82 \\ 9{,}60 & 3{,}10 \\ 8{,}70 & 3{,}42 \\ 7{,}80 & 3{,}62 \\ 7{,}20 & 3{,}95 \\ 6{,}50 & 4{,}80 \\ 6{,}10 & 5{,}60 \\ 5{,}30 & 6{,}30 \\ 4{,}51 & 7{,}50 \\ 3{,}80 & 8{,}90 \\ 2{,}70 & 9{,}30 \end{bmatrix}.$$

Знаходимо добуток матриць $X' \cdot X$ та $X' \cdot Y$:

$$X' \cdot X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 1 & 4 & 9 & 16 & 25 & 36 & 49 & 64 & 81 & 100 & 121 \end{bmatrix} \cdot \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \\ 1 & 6 & 36 \\ 1 & 7 & 49 \\ 1 & 8 & 64 \\ 1 & 9 & 81 \\ 1 & 10 & 100 \\ 1 & 11 & 121 \end{bmatrix} = \begin{bmatrix} 11{,}0 & 66{,}0 & 506{,}0 \\ 66{,}0 & 506{,}0 & 4356{,}0 \\ 506{,}0 & 4356{,}0 & 39974{,}0 \end{bmatrix}$$

$$Y' \cdot X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 1 & 4 & 9 & 16 & 25 & 36 & 49 & 64 & 81 & 100 & 121 \end{bmatrix} \cdot \begin{bmatrix} 10{,}2 & 2{,}82 \\ 9{,}60 & 3{,}10 \\ 8{,}70 & 3{,}42 \\ 7{,}80 & 3{,}62 \\ 7{,}20 & 3{,}95 \\ 6{,}50 & 4{,}80 \\ 6{,}10 & 5{,}60 \\ 5{,}30 & 6{,}30 \\ 4{,}51 & 7{,}50 \\ 3{,}80 & 8{,}90 \\ 2{,}70 & 9{,}30 \end{bmatrix} = \begin{bmatrix} 72{,}41 & 59{,}31 \\ 355{,}09 & 430{,}71 \\ 2375{,}81 & 3675{,}87 \end{bmatrix}$$

Знаходимо обернену матрицю $(X' \cdot X)^{-1}$:

$$(X' \cdot X)^{-1} = \begin{bmatrix} 1{,}20606 & -0{,}41818 & 0{,}03030 \\ -0{,}41818 & 0{,}17692 & -0{,}01399 \\ 0{,}03030 & -0{,}01399 & 0{,}00117 \end{bmatrix}.$$

Вектор параметрів моделі:

$$A = (X'X)^{-1} \cdot (X'Y) =$$

$$= \begin{bmatrix} 1{,}20606 & -0{,}41818 & 0{,}03030 \\ -041818 & 0{,}17692 & -0{,}01399 \\ 0{,}03030 & -0{,}01399 & 0{,}00117 \end{bmatrix} \cdot \begin{bmatrix} 72{,}41 & 59{,}31 \\ 355{,}09 & 430{,}71 \\ 2375{,}81 & 3675{,}87 \end{bmatrix} = \begin{bmatrix} 10{,}83291 & 2{,}80636 \\ -0{,}68504 & -0{,}01059 \\ -0{,}00304 & 0{,}05759 \end{bmatrix}.$$

Запишемо моделі попиту та пропозиції:

$$y_1 = 10{,}83291 - 0{,}68504x - 0{,}00304x^2 + u_1,$$

$$y_2 = 2{,}80636 - 0{,}01059x + 0{,}05759x^2 + u_2$$

Знаходимо розрахункові значення попиту та пропозиції $\hat{y}_{1i}, \hat{y}_{2i}$ та заносимо їх у таблицю 8.1.

Залишкова дисперсія:

$$\sigma_{u_1}^2 = \frac{u_1' \cdot u_1}{n-m-1} = 0{,}03683; \ \sigma_{u_2}^2 = \frac{u_2' \cdot u_2}{n-m-1} = 0{,}06279.$$

Загальна дисперсія:

$$\sigma_{y_1}^2 = \frac{\sum_{i=1}^{n}\left(y_{1i} - \overline{y}_1\right)^2}{n} = \frac{57{,}53483}{11} = 5{,}23044;$$

$$\sigma_{y_2}^2 = \frac{\sum_{i=1}^{n}\left(y_{2i} - \overline{y}_2\right)^2}{n} = \frac{54{,}21696}{11} = 4{,}92881.$$

Коефіцієнт множинної кореляції:

$$R_1 = \sqrt{1 - \frac{0{,}03683}{5{,}23044}} = 0{,}99647; \ R_2 = \sqrt{1 - \frac{0{,}06279}{4{,}9881}} = 0{,}99361.$$

Перевірка адекватності моделі по F – критерію Фішера:

$$F_{p1} = \frac{\left(\sum_{i=1}^{n}\left(y_{1i} - \overline{y}_1\right)^2 - \sum_{i=1}^{n}\left(y_{1i} - \hat{y}_{1i}\right)^2\right)(m-1)}{\sum_{i=1}^{n}\left(y_{1i} - \hat{y}_{1i}\right)^2 (n-m-1)} = 142{,}01242;$$

$$F_{p2} = \frac{\left(\sum_{i=1}^{n}\left(y_{2i} - \overline{y}_2\right)^2 - \sum_{i=1}^{n}\left(y_{2i} - \hat{y}_{2i}\right)^2\right)(m-1)}{\sum_{i=1}^{n}\left(y_{2i} - \hat{y}_{2i}\right)^2 (n-m-1)} = 78{,}49380.$$

$F_{табл}$=4,96, $F_{p1}, F_{p2} > F_{табл.}$;отже з імовірністюP=0,95 можна стверджувати, що моделі попиту та пропозиції адекватні статистичним даним.

Знаходимо точку рівноваги попиту та пропозиції (рівноважну ціну). Для цього прирівнюємо $\hat{y}_1 = \hat{y}_2$:

$10{,}83291 - 0{,}68504x - 0{,}00304x^2 = 2{,}80636 - 0{,}01059x + 0{,}05759x^2$;

$0{,}06063x^2 + 0{,}67504x - 8{,}02655 = 0$

$$x_1, x_2 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-0.67504 \pm \sqrt{0.67504^2 + 4 \cdot 0.06063 \cdot 8.02655}}{2 \cdot 0.06063} = \frac{-0.67504 \pm 1.54993}{0.12126};$$

$x_1 = 7{,}215.$

Отже, при ціні $x_1 = 7{,}215$ умовних г.о. наступає рівновага між попитом та пропозицією.

Знаходимо коефіцієнти еластичності попиту та пропозиції:

$$KE_1 = \frac{a_1 \cdot x_1 + 2 \cdot a_2 \cdot x_1^2}{\hat{y}_1(x_1)} = \frac{-0{,}68504 \cdot 7{,}215 + 2 \cdot (-0{,}00304) \cdot 7{,}215^2}{5{,}73209} = -0{,}92;$$

$$KE_2 = \frac{b_1 \cdot x_1 + 2 \cdot b_2 \cdot x_1^2}{\hat{y}_2(x_1)} = \frac{-0{,}01059 \cdot 7{,}215 + 2 \cdot 0{,}05759 \cdot 7{,}215^2}{5{,}72787} = 1{,}03.$$

Отже, при зростанні ціни на 1% попит падає на 0,89%, а пропозиція зростає на 0,51%.

Будуємо графіки попиту та пропозиції.

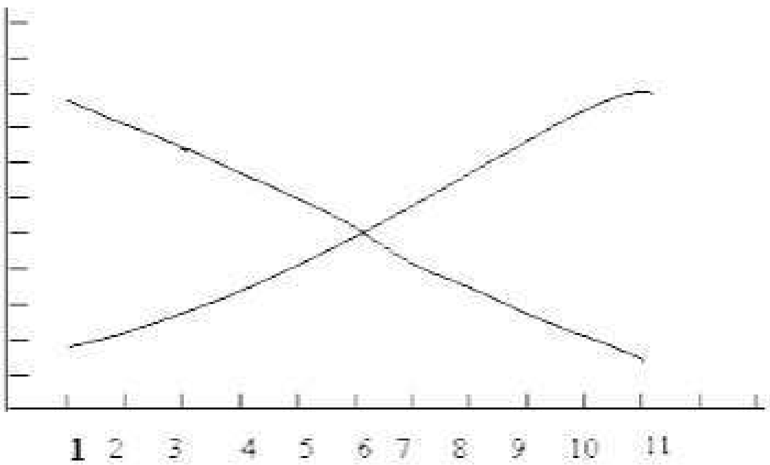


Рисунок 8.1 – Графік попиту та пропозиції

8.3 Ідентифікація економетричної моделі

Параметри структурної економетричної моделі можуть бути представлені у структурній та у зведеній (прогнозній) формах. Якщо параметри структурної форми однозначно виражаються через параметри прогнозної форми системи регресії, то така економетрична модель називається ідентифікованою. Якщо число оцінюваних параметрів структурної форми регресії більше числа оцінюваних параметрів прогнозної форми, тобто число оцінюваних параметрів прогнозної форми регресії більше числа рівнянь, то таку систему називають неідентифікованою.

Система структурних регресій буде ідентифікованою, якщо для кожної регресії виконується умова $m+n-(n_i+m_i)\ge n-1$, n – число ендогенних величин та число регресій у системі; m – число екзогенних величин у системі регресій; n_i – число ендогенних величин в *і-й* регресії, m_i – число екзогенних величин в *і-й* регресії.

Запишемо систему структурних регресій:

$$\begin{aligned} Y_1 &= b_{12}Y_2 + a_{11}X_2 + U_1 \\ Y_2 &= b_{21}Y_1 + a_{22}X_2 + U_2 \end{aligned} \qquad (8.10)$$

Для першої регресії маємо таке співвідношення: $2+2-(2+1)=2-1$

Аналогічно для другої регресії: $2+2-(2+1)=2-1$

Умова ідентифікації регресії виконана.

8.4 Рекурсивні системи

Якщо в економетричній моделі матриця параметрів B при внутрішніх змінних Y_i має трикутний вигляд, то така система рівнянь називається рекурсивною.

Розглянемо економетричну модель, яка складається із системи регресій, в якій матриця B має трикутну форму:

$$\begin{aligned} &Y_1 = a_{10} + a_{11}X_1 + \ldots + a_{1m}X_m + U_1 \\ &Y_2 = b_{21}Y_1 + a_{20} + a_{21}X_1 + \ldots + a_{2m}X_m + U_2 \\ &Y_3 = b_{31}Y_1 + b_{32}Y_2 + a_{30} + a_{31}X_1 + \ldots + a_{3m}X_m + U_3 \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ &Y_n = b_{n1}Y_1 + b_{n2}Y_2 + \ldots + b_{n,n-1}Y_{n-1} + a_{n0} + a_{n1}X_1 + \ldots + a_{nm}X_m + U_n \end{aligned} \qquad (8.11)$$

Для цієї системи регресій матриця параметрів B має вигляд:

$$B = \begin{bmatrix} -1 & 0 & 0\ldots0 & 0 \\ b_{21} & -1 & 0\ldots0 & 0 \\ b_{31} & b_{32} & -1\ldots0 & 0 \\ \ldots & \ldots & \ldots & \ldots \\ b_{n1} & b_{n2} & b_{n3}\ldots b_{n,n-1} & 1 \end{bmatrix} \qquad (8.12)$$

В рекурсивній моделі спостерігається одностороння залежність між внутрішніми змінними, оскільки Y_2 залежить від Y_1, але Y_1 не залежить від Y_2 і т.д. Звісно, що не всі регресії рекурсивної моделі будуть ідентифікованими. Якщо рекурсивна економетрична модель ідентифікована, то використовують такий алгоритм оцінювання параметрів рекурсивної моделі:

1) використовуючи МНК оцінюють параметри першої регресії (оскільки в правій частині цього рівняння знаходяться тільки екзогенні величини X_j $\left(j=\overline{1,m}\right)$;

2) обчислюють значення $\hat{Y}_{1i}$ $\left(i=\overline{1,n}\right)$, величину $\hat{Y}_{1i}$ вводять в регресію як пояснюючу, і, вводячи значення $\hat{Y}_{1i}$, оцінюють параметри другої регресії;

3) для третьої регресії приймається, що $\hat{Y}_{1i}, \hat{Y}_{2i}$ наперед визначені, і оцінюють параметри регресії МНК.

Як рекурсивну модель можна навести модель німецького економіста М. Вольфганга.

Якщо позначити через ендогенні величини: Y_1 – грошові доходи населення; Y_2 – особисте споживання; Y_3 – споживання; і через екзогенні величини: X_1 – національний дохід; X_2 – особисте споживання за попередній рік; X_3 – чисельність населення; X_4 – збереження на кінець попереднього року; X_5 – суспільний фонд споживання, то модель буде мати вигляд:

$$
\begin{aligned}
&Y_1 = a_{10} + a_{11}X_1 = U_1, \\
&Y_2 = b_{21}Y_1 + a_{20} + a_{22}X_2 + a_{23}X_3 + a_{24}X_4 = U_2, \\
&Y_3 = Y_2 + X_3
\end{aligned}
\tag{8.13}
$$

Рекурсивна модель М. Вольфганга складається з двох регресій та однієї тотожності. У приведеній економетричній моделі матриця B має трикутну форму:

$$
B = \begin{bmatrix} -1 & 0 & 0 \\ b_{21} & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix}
\tag{8.14}
$$

Ендогенні величини Y_1 та Y_2, то Y_2 (особисте споживання) залежать від грошових доходів населення Y_1, але грошові доходи Y_1 не залежать від особистого споживання Y_2, тобто в даній економетричній моделі спостерігається односторонній зв'язок між ендогенними величинами.

Сформуємо матриці екзогенних змінних X та ендогенних змінних Y для регресії, що описує грошові доходи населення $Y_1 = a_{10} + a_{11}X_1 + u_1$:

$$X^* = \begin{bmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{13} \\ \ldots & \ldots \\ 1 & x_{1i} \\ \ldots & \ldots \\ 1 & x_{1n} \end{bmatrix}; Y_1 = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \ldots \\ y_i \\ \ldots \\ y_n \end{bmatrix} \qquad (8.15)$$

Цій регресії відповідає система нормальних рівнянь:

$$X^{*'}X = \begin{bmatrix} 1 & 1 & 1 \ldots 1 \ldots 1 \\ x_{11} & x_{12} & x_{13} \ldots x_{1i} \ldots x_{1n} \end{bmatrix} \times \begin{bmatrix} 1 & x_{11} \\ 1 & x_{12} \\ 1 & x_{13} \\ \ldots & \ldots \\ 1 & x_{1i} \\ \ldots & \ldots \\ 1 & x_{1n} \end{bmatrix} = \begin{bmatrix} n & \sum x_{1i} \\ \sum x_{1i} & \sum x_{1i}^2 \end{bmatrix} \qquad (8.16)$$

$$X^{*'}Y = \begin{bmatrix} 1 & 1 & 1 \ldots 1 \ldots 1 \\ x_{11} & x_{12} & x_{13} \ldots x_{1i} \ldots x_{1n} \end{bmatrix} \times \begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ \ldots \\ y_{1i} \\ \ldots \\ y_{1n} \end{bmatrix} = \begin{bmatrix} \sum y_{1i} \\ \sum y_{1i}x_{1i} \end{bmatrix} \qquad (8.17)$$

Використовуючи МНК знайдемо вектор параметрів моделі:

$$A = (X'X)^{-1} \cdot (X'Y)$$

Після обчислення розрахункових значень $\hat{y}_1$ та вважаючи значення $\hat{y}_2$, передвизначеними, формуємо матрицю екзогенних X'' та ендогенних Y_2 змінних:

$$X^{**} = \begin{bmatrix} 1 & x_{21} & x_{31} & x_{41} & \hat{y}_{11} \\ 1 & x_{22} & x_{32} & x_{42} & \hat{y}_{12} \\ 1 & x_{23} & x_{33} & x_{43} & \hat{y}_{13} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{2i} & x_{3i} & x_{4i} & \hat{y}_{1i} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{2n} & x_{3n} & x_{4n} & \hat{y}_{1n} \end{bmatrix}; \quad Y_2 = \begin{bmatrix} y_{21} \\ y_{22} \\ y_{23} \\ \dots \\ y_{2i} \\ \dots \\ y_{2n} \end{bmatrix} \tag{8.18}$$

Тоді:

$$X^{**'} \cdot X^{**} = \begin{bmatrix} 1 & 1 & 1 \dots & 1 \dots & 1 \\ x_{21} & x_{22} & x_{23} \dots & x_{2i} \dots & x_{2n} \\ x_{31} & x_{32} & x_{33} \dots & x_{3i} \dots & x_{3n} \\ x_{41} & x_{42} & x_{43} \dots & x_{4i} \dots & x_{4n} \\ \hat{y}_{11} & \hat{y}_{12} & \hat{y}_{13} \dots & \hat{y}_{1i} \dots & \hat{y}_{1n} \end{bmatrix} \cdot \begin{bmatrix} 1 & x_{21} & x_{31} & x_{41} & \hat{y}_{11} \\ 1 & x_{22} & x_{32} & x_{42} & \hat{y}_{12} \\ 1 & x_{23} & x_{33} & x_{43} & \hat{y}_{13} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{2i} & x_{3i} & x_{4i} & \hat{y}_{1i} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{2n} & x_{3n} & x_{4n} & \hat{y}_{1n} \end{bmatrix} =$$

$$= \begin{bmatrix} n & \sum x_{2i} & \sum x_{3i} & \sum x_{4i} & \sum \hat{y}_{1i} \\ \sum x_{2i} & \sum x_{2i}^2 & \sum x_{3i}x_{2i} & \sum x_{4i}x_{2i} & \sum \hat{y}_{1i}x_{2i} \\ \sum x_{3i} & \sum x_{2i}x_{3i} & \sum x_{3i}^2 & \sum x_{3i}x_{4i} & \sum \hat{y}_{1i}x_{3i} \\ \sum x_{4i} & \sum x_{2i}x_{4i} & \sum x_{3i}x_{41} & \sum x_{4i}^2 & \sum \hat{y}_{1i}x_{4i} \end{bmatrix}; \tag{8.19}$$

$$X^{**'} \cdot Y_2 = \begin{bmatrix} 1 & 1 & 1 \dots & 1 \dots & 1 \\ x_{21} & x_{22} & x_{23} \dots & x_{2i} \dots & x_{2n} \\ x_{31} & x_{32} & x_{33} \dots & x_{3i} \dots & x_{3n} \\ x_{41} & x_{42} & x_{43} \dots & x_{4i} \dots & x_{4n} \\ \hat{y}_{11} & \hat{y}_{12} & \hat{y}_{13} \dots & \hat{y}_{1i} \dots & \hat{y}_{1n} \end{bmatrix} \cdot \begin{bmatrix} y_{21} \\ y_{22} \\ y_{23} \\ \dots \\ y_{2i} \\ \dots \\ y_{2n} \end{bmatrix} = \begin{bmatrix} \sum y_{2i} \\ \sum y_{2i}x_{2i} \\ \sum y_{2i}x_{3i} \\ \sum y_{2i}x_{4i} \\ \sum y_{2i}\hat{y}_{1i} \end{bmatrix} \tag{8.20}$$

Використовуючи МНК, знайдемо оцінки параметрів моделі витрат на особисте споживання.

$$A = \left(X^{**'} \cdot X^{**}\right)^{-1} \cdot \left(X^{**'} \cdot Y\right) \tag{8.21}$$

Обчислюємо значення $\hat{y}_{2i}$. Для третьої моделі (тотожності) приймаємо, що $\hat{y}_{1i}$ та $\hat{y}_{2i}$ передвизначені, оцінюємо $\hat{y}_{3i}$ як суму $\hat{y}_{3i} = \hat{y}_{1i} + \hat{y}_{2i}$.

8.5 Непрямий метод найменших квадратів (НМНК) оцінки параметрів системи двох регресій

Розглянемо систему взаємопов'язаних регресій, якщо між ендогенними та екзогенними величинами існує лінійна залежність:

$$\begin{aligned} Y_{1t} &= b_{12}Y_{2t} + a_{10} + a_{11}X_{1t} + U_{1t} \\ Y_{2t} &= b_{21}Y_{1t} + a_{20} + a_{22}X_{2t} + U_{2t} \end{aligned} \tag{8.22}$$

Нехай ендогенні величини: Y_{1t}; Y_{2t} та екзогенні величини: X_{1t}; X_{2t}.

Запишемо систему регресій у прогнозній (приведеній) формі:

$$\begin{aligned} Y_{1t} &= c_{10} + c_{11}x_{1t} + c_{12}x_{1t} + e_{1t} \\ Y_{2t} &= c_{20} + c_{21}x_{1t} + c_{22}x_{2t} + e_{2t} \end{aligned} \tag{8.23}$$

Сформуємо матриці X та Y:

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \cdots & \cdots & \cdots \\ 1 & x_{1t} & x_{2t} \\ \cdots & \cdots & \cdots \\ 1 & x_{1n} & x_{2n} \end{bmatrix}; Y = \begin{bmatrix} y_{11} & y_{21} \\ y_{12} & y_{22} \\ \cdots & \cdots \\ y_{1t} & y_{2t} \\ \cdots & \cdots \\ y_{1n} & y_{2n} \end{bmatrix}; \tag{8.24}$$

$$X' \cdot X = \begin{bmatrix} 1 & 1 \ldots 1 \ldots 1 \\ x_{11} & x_{12} \ldots x_{1t} \ldots x_{1n} \\ x_{21} & x_{22} \ldots x_{2t} \ldots x_{2n} \end{bmatrix} \cdot \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \cdots & \cdots & \cdots \\ 1 & x_{1t} & x_{2t} \\ \cdots & \cdots & \cdots \\ 1 & x_{1n} & x_{2n} \end{bmatrix} = \begin{bmatrix} n & \sum x_{1t} & \sum x_{2t} \\ \sum x_{1t} & \sum x_{1t}^2 & \sum x_{2t}x_{1t} \\ \sum x_{2t} & \sum x_{1t}x_{2t} & \sum x_{2t}^2 \end{bmatrix}; \tag{8.25}$$

$$X' \cdot Y = \begin{bmatrix} 1 & 1 \ldots 1 \ldots 1 \\ x_{11} & x_{12} \ldots x_{1t} \ldots x_{1n} \\ x_{21} & x_{22} \ldots x_{2t} \ldots x_{2n} \end{bmatrix} \cdot \begin{bmatrix} y_{11} & y_{21} \\ y_{12} & y_{22} \\ \cdots & \cdots \\ y_{1t} & y_{2t} \\ \cdots & \cdots \\ y_{1n} & y_{2n} \end{bmatrix} = \begin{bmatrix} \sum y_{1t} & \sum y_{2t} \\ \sum y_{1t}x_{1t} & \sum y_{2t}x_{1t} \\ \sum y_{1t}x_{2t} & \sum y_{2t}x_{2t} \end{bmatrix}; \tag{8.26}$$

Звідки вектор параметрів прогнозної системи регресій:

$$C = \left(X' \cdot X\right)^{-1} \cdot \left(X' \cdot Y\right) = \begin{bmatrix} c_{10} & c_{20} \\ c_{11} & c_{21} \\ c_{12} & c_{22} \end{bmatrix} \tag{8.27}$$

Таким чином, матриця C оцінок параметрів прогнозної форми буде мати вигляд:

$$C = \begin{bmatrix} c_{10} & c_{20} \\ c_{11} & c_{21} \\ c_{12} & c_{22} \end{bmatrix} \quad (8.28)$$

Якщо параметри структурної форми (8.22) однозначно виражаються через параметри прогнозної форми (8.23) системи регресії, то така економетрична модель називається ідентифікованою. Якщо число оцінюваних параметрів структурної форми регресії більше числа оцінюваних параметрів прогнозної форми, тобто число оцінюваних параметрів прогнозної форми регресії більше числа рівнянь, то систему називають не ідентифікованою.

Параметри прогнозної форми є комбінацією всіх параметрів структурної форми регресії. По параметрах прогнозної форми не можна робити висновок про взаємозалежність ендогенних величин, оскільки при переході із структурної форми до прогнозної форми регресії вони розподіляються на екзогенні величини і відхилення. З іншого боку, структурна форма не придатна для визначення прогнозних значень ендогенних величин, тому що в правій частині регресії знаходяться значення ендогенних величин.

Розглянемо методику розрахунку параметрів структурної форми регресії через параметри прогнозної форми. Економетричну модель перепишемо у такому вигляді:

$$\begin{aligned} -Y_{1t} + b_{12}Y_{2t} + a_{10} + a_{11}X_{1t} + U_{1t} = 0 \\ -Y_{2t} + b_{21}Y_{1t} + a_{20} + a_{22}X_{2t} + U_{2t} = 0 \end{aligned} \quad (8.29)$$

або у матричній формі:

$$B \cdot Y + A \cdot X + U = 0 \quad (8.30)$$

$$B = \begin{bmatrix} -1 & b_{12} \\ b_{22} & -1 \end{bmatrix}; \quad A = \begin{bmatrix} a_{10} & a_{11} & 0 \\ a_{20} & 0 & a_{22} \end{bmatrix}; \quad (8.31)$$

$$X=\begin{bmatrix}1 & x_{11} & x_{21}\\ 1 & x_{12} & x_{22}\\ \cdots & \cdots & \cdots\\ 1 & x_{1t} & x_{2t}\\ \cdots & \cdots & \cdots\\ 1 & x_{1n} & x_{2n}\end{bmatrix};\ Y=\begin{bmatrix}y_{11} & y_{21}\\ y_{12} & y_{22}\\ \cdots & \cdots\\ y_{1t} & y_{2t}\\ \cdots & \cdots\\ y_{1n} & y_{2n}\end{bmatrix};\ U=\begin{bmatrix}u_{11} & u_{21}\\ u_{12} & u_{22}\\ \cdots & \cdots\\ u_{1t} & u_{2t}\\ \cdots & \cdots\\ u_{1n} & u_{2n}\end{bmatrix}$$

де B – матриця параметрів при ендогенних величинах структурної форми системи регресій; A – матриця параметрів при екзогенних величинах структурної форми системи регресій; X – матриця екзогенних величин; Y –матриця ендогенних величин.

Тоді прогнозна форма системи регресій матиме вигляд:

$$Y=C\cdot X+E \tag{8.32}$$

або у розгорнутій формі:

$$\begin{aligned}Y_1&=c_{10}+c_{11}x_1+c_{12}x_1+E_1\\ Y_2&=c_{20}+c_{21}x_1+c_{22}x_2+E_2\end{aligned} \tag{8.33}$$

Запишемо структурну форму регресій у вигляді матричного рівняння:

$$\begin{bmatrix}-1 & b_{12}\\ b_{22} & -1\end{bmatrix}\cdot\begin{bmatrix}c_{10} & c_{11} & c_{12}\\ c_{20} & c_{21} & c_{22}\end{bmatrix}=-\begin{bmatrix}a_{10} & a_{11} & 0\\ a_{20} & 0 & a_{22}\end{bmatrix}; \tag{8.34}$$

Знайдемо добуток матриць B і C та прирівняємо його до матриці A:

$$\begin{aligned}&-c_{10}+b_{12}c_{20}=-a_{10},\\ &-c_{11}+b_{12}c_{21}=-a_{11},\\ &-c_{12}+b_{12}c_{22}=0,\\ &b_{21}c_{10}-c_{20}=-a_{20},\\ &b_{21}c_{11}-c_{21}=0,\\ &b_{21}c_{12}-c_{12}=-a_{22}\end{aligned} \tag{8.35}$$

З цієї системи знайдемо оцінки елементів матриць B і A:

$$b_{12}=\frac{c_{12}}{c_{22}};\ \ b_{21}=\frac{c_{21}}{c_{11}}, \tag{8.36}$$

$$a_{10}=c_{10}-\frac{c_{12}\cdot c_{20}}{c_{22}};\ \ a_{11}=c_{11}-\frac{c_{12}\cdot c_{21}}{c_{22}};$$

$$a_{20}=c_{20}-\frac{c_{21}\cdot c_{10}}{c_{11}};\ \ a_{22}=c_{22}-\frac{c_{21}\cdot c_{12}}{c_{11}}$$

Приклад 8.2. На основі статистичних даних за 10 років ендогенних величин (Y_1 – експорт, Y_2 – імпорт (в умовних г. о.)) та екзогенних величин (X_1 – національний дохід, X_2 – середній обсяг зовнішньої торгівлі країн СНД) використовуючи непрямий метод найменших квадратів оцінити параметри структурної системи регресій зовнішньої торгівлі *k-ї* країни.

Використовуючи прогнозну форму системи регресій, оцінити значення прогнозу експорту та імпорту, якщо структурна система регресій має вигляд:

$$\begin{aligned} Y_1 &= b_{12}Y_2 + a_{10} + a_{11}X_1 + U_1 \\ Y_2 &= b_{21}Y_1 + a_{20} + a_{22}X_2 + U_2 \end{aligned} \tag{8.37}$$

Таблиця 8.2 – Вхідні дані

X_1	X_2	Y_1	Y_2
70,00	5,50	55	73
75,00	6,10	63	86
77,00	6,90	71	94
80,00	7,40	80	103
85,00	7,90	93	111
92,00	8,30	103	117
99,00	8,80	117	122
130,00	9,10	128	130
107,00	9,80	139	135
112,00	10,20	151	139
115,00	11,00		

Розв'язок.

1. Перевіримо умову ідентифікації регресії. Система структурних регресій називається ідентифікованою, якщо з матричного рівняння *BC=-A* елементи матриці *A* і *B* однозначно визначаються через елементи матриці *C*. Звідси випливає, що система структурних регресій буде ідентифікованою, якщо для кожної регресії виконується умова: $m + n - (n_i + m_i) \geq n - 1$, де n – число ендогенних величин і число регресій у системі, m – число екзогенних величин у системі регресій, n_i – число ендогенних величин в *i-й* регресії, m_i – число екзогенних величин в *i-й* регресії. Для першої регресії маємо таке співвідношення: $2 + 2 - (2 + 1) = 2 - 1$

Аналогічно для другої регресії: $2+2-(2+1)=2-1$

Умова ідентифікації регресії виконана, тому оцінка параметрів структурної системи регресій можна знайти непрямим методом найменших квадратів.

Перепишемо структурну систему регресій (8.37) у вигляді:

$$-Y_1 + b_{12}Y_2 + a_{10} + a_{11}X_1 + U_1 = 0$$
$$-Y_2 + b_{21}Y_1 + a_{20} + a_{22}X_2 + U_2 = 0$$

або у матричній формі: $B \cdot Y + A \cdot X + U = 0$

де $B = \begin{bmatrix} -1 & b_{12} \\ b_{22} & -1 \end{bmatrix}$; $A = \begin{bmatrix} a_{10} & a_{11} & 0 \\ a_{20} & 0 & a_{22} \end{bmatrix}$; $X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \dots & \dots & \dots \\ 1 & x_{1t} & x_{2t} \\ \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} \end{bmatrix}$; $Y = \begin{bmatrix} y_{11} & y_{21} \\ y_{12} & y_{22} \\ \dots & \dots \\ y_{1t} & y_{2t} \\ \dots & \dots \\ y_{1n} & y_{2n} \end{bmatrix}$;

Запишемо прогнозну систему регресій у матричній формі:

$$Y = C \cdot X + U$$

або в розгорнутій формі:

$$Y_1 = c_{10} + c_{11}x_1 + c_{12}x_1 + E_1$$
$$Y_2 = c_{20} + c_{21}x_1 + c_{22}x_2 + E_2$$

2. Для оцінки параметрів прогнозної системи регресій використовуємо 1 МНК, формуємо матрицю екзогенних величин X та ендогенних Y.

$$X = \begin{bmatrix} 1 & 70 & 5{,}5 \\ 1 & 75 & 6{,}1 \\ 1 & 77 & 6{,}9 \\ 1 & 80 & 7{,}4 \\ 1 & 85 & 7{,}9 \\ 1 & 92 & 8{,}3 \\ 1 & 99 & 8{,}8 \\ 1 & 103 & 9{,}1 \\ 1 & 107 & 9{,}8 \\ 1 & 112 & 10{,}2 \end{bmatrix}; \quad Y = \begin{bmatrix} 55 & 73 \\ 63 & 86 \\ 71 & 94 \\ 80 & 103 \\ 93 & 111 \\ 103 & 117 \\ 117 & 122 \\ 128 & 130 \\ 139 & 135 \\ 151 & 139 \end{bmatrix}$$

Знаходимо добуток матриць $X'X$ та $X'Y$

$$X'X=\begin{bmatrix}10{,}0 & 900{,}0 & 80{,}0\\ 900{,}0 & 82946{,}0 & 7400{,}0\\ 80{,}0 & 7400{,}0 & 661{,}5\end{bmatrix};\; X'Y=\begin{bmatrix}1000{,}0 & 1110{,}0\\ 94375{,}0 & 10218{,}0\\ 8455{,}1 & 9182{,}3\end{bmatrix}$$

3. Знаходимо оцінки параметрів прогнозної системи регресій:

$$C=\left(X'\cdot X\right)^{-1}\cdot\left(X'\cdot Y\right)=\begin{bmatrix}6{,}21743 & -0{,}20500 & 1{,}54154\\ -0{,}20500 & 0{,}01340 & -0{,}12517\\ 1{,}54154 & -0{,}12517 & 1{,}21549\end{bmatrix}\cdot\begin{bmatrix}1000{,}0 & 1110{,}0\\ 94375{,}0 & 102718{,}0\\ 8455{,}1 & 9182{,}3\end{bmatrix}=$$

$$=\begin{bmatrix}-95{,}30806 & -0{,}67470\\ 1{,}67736 & -0{,}06661\\ 5{,}54316 & 14{,}70868\end{bmatrix}$$

Таким чином: $C=\begin{bmatrix}-95{,}30806 & -0{,}67470\\ 1{,}67736 & -0{,}06661\\ 5{,}54316 & 14{,}70868\end{bmatrix}$

Запишемо прогнозну систему регресій у розгорнутій формі:

$$Y_1=-95{,}30806+1{,}67736X_1+5{,}54316X_2+U_1$$
$$Y_2=-0{,}67470-0{,}06661X_1+14{,}70868X_2+U_2$$

4. Знайдемо розрахункові значення експорту $\hat{Y}_{1t}$ та імпорту $\hat{Y}_{2t}$, користуючись прогнозною формою системи регресій:

$$Y=X\cdot C=\begin{bmatrix}1 & 70 & 5{,}5\\ 1 & 75 & 6{,}1\\ 1 & 77 & 6{,}9\\ 1 & 80 & 7{,}4\\ 1 & 85 & 7{,}9\\ 1 & 92 & 8{,}3\\ 1 & 99 & 8{,}8\\ 1 & 103 & 9{,}1\\ 1 & 107 & 9{,}8\\ 1 & 112 & 10{,}2\end{bmatrix}\cdot\begin{bmatrix}-95{,}30806 & -0{,}67470\\ 1{,}67736 & -0{,}06661\\ 5{,}54316 & 14{,}70868\end{bmatrix}$$

Результати обчислень запишемо у таблицю 8.3. Розрахуємо також вектори залишків U_1 та U_2 як різниці між $y-\hat{y}$ та запишемо їх також у таблиці 8.3.

Залишкові дисперсії: $\sigma^2_{u_1}=4{,}47$; $\sigma^2_{u_2}=6{,}87797$.

Загальні дисперсії: $\sigma^2_{y_1}=671{,}56$; $\sigma^2_{y_2}=914{,}49$.

Таблиця 8.3 – Результати розрахунків

$\hat{y}_{1t}$	$\hat{y}_{2t}$	U_{1t}	U_{2t}
52,59482	75,56046	2,40518	-2,56046
64,30753	84,05263	-1,30753	1,94737
72,09679	95,68636	-1,09679	-1,68636
79,90046	102,84087	0,09954	0,15913
91,05886	109,86217	1,94114	1,13783
105,01768	115,27939	-2,01768	1,72061
119,53081	122,16747	-2,53081	-0,16747
127,90321	126,31364	0,09679	3,68636
138,49288	136,34329	0,50712	-1,34329
149,09696	141,89372	1,90304	-2,89372

Коефіцієнти множинної кореляції:

$$R_1 = 0.99666;\ R_2 = 0.99623$$

Оцінка адекватності моделі по F – критерію Фішера:

$$F_{1p} = 150.12798;\ F_{2p} = 132.95937;$$

$$F_{1p} > F_{табл};\ F_{2p} > F_{табл}.$$

Отже, прийняту економетричну модель з імовірністю $P=0,95$ можна вважати адекватною статистичним даним.

5. Знаходимо точкові оцінки прогнозу, використовуючи систему регресій у прогнозній формі: для прогнозних величин $X_{1p} = 115,\ X_{2p} = 11$ прогнозні ендогенні величини приймуть значення $Y_{1p} = 158.56,\ Y_{2p} = 153.46$ умовних г.о.

6. Перейдемо від прогнозної форми системи регресій до структурної:

$$b_{12} = \frac{c_{12}}{c_{22}} = \frac{5.54316}{14.70868} = 0.37686;\ b_{21} = \frac{c_{21}}{c_{11}} = \frac{-0.06661}{1.67736} = -0.03971;$$

$$a_{10} = c_{10} - \frac{c_{12} \cdot c_{20}}{c_{22}} = -95.30806 - \frac{5.54316 \cdot (-0.67470)}{14.70868} = -95.05379$$

$$a_{11} = c_{11} - \frac{c_{12} \cdot c_{21}}{c_{22}} = 1.67736 - \frac{5.54316 \cdot (-0.06661)}{14.70868} = 1.702246$$

$$a_{20} = c_{20} - \frac{c_{21} \cdot c_{10}}{c_{11}} = -0.67470 - \frac{(-0.6661) \cdot (-95.30806)}{1.67736} = -4.45950;$$

$$a_{22} = c_{22} - \frac{c_{21} \cdot c_{12}}{c_{11}} = 14.70868 - \frac{(-0.06661) \cdot 5.54316}{1.67736} = 14.9288$$

Запишемо структурну систему регресій:

$$Y_1 = 0.37686Y_2 - 95.05379 + 1.70246X_1 + U_1;$$
$$Y_2 = -0.03971Y_1 - 4.45950 + 14.9288X_2 + U_2$$

Із структурної форми системи регресій випливає:

Зі збільшенням Y_2 (імпорту) на одиницю при незмінних інших величинах, Y_1 (експорт) збільшиться на 0,37686 умовних г.о. Зі збільшенням X_1 (національного доходу) на одиницю, при незмінних інших величинах, Y_1 (експорт) збільшиться на 1,70246 умовних г.о. Зі збільшенням Y_1 (експорту) на одиницю при незмінних інших величинах, Y_2 (імпорт) зменшиться на 0,03971 умовних г.о. Зі збільшенням X_2 (середнього товарообігу країн СНД) при незмінних інших величинах, Y_2 (імпорт) збільшиться на 14,9288 умовних г.о.

8.6 Непрямий метод найменших квадратів для системи з n регресій

Розглянемо певну економетричну модель. Система регресій називається повною, якщо:

1) вона має стільки регресій, скільки в ній ендогенних величин;

2) вона має всі змінні, які мають суттєвий вплив на сумісно залежні ендогенні величин;

3) визначник матриці, складеної з коефіцієнтів при ендогенних величинах системи регресій у структурній формі, відмінний від нуля, тобто систему можна розв`язати відносно ендогенних величин.

Нехай повна система регресій структурної форми n ендогенних величин та m екзогенних величин:

$$\begin{array}{l}
y_{1t} = b_{12}y_{2t} + b_{13}y_{3t} + \ldots + b_{1n}y_{nt} + a_{10} + a_{11}x_{1t} + \ldots + a_{1m}x_{mt} + u_{1t} \\
y_{2t} = b_{21}y_{1t} + b_{23}y_{3t} + \ldots + b_{2n}y_{nt} + a_{20} + a_{21}x_{1t} + \ldots + a_{2m}x_{mt} + u_{2t} \\
y_{3t} = b_{31}y_{1t} + b_{32}y_{2t} + \ldots + b_{3n}y_{nt} + a_{30} + a_{31}x_{1t} + \ldots + a_{3m}x_{mt} + u_{3t} \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
y_{nt} = b_{n1}y_{1t} + b_{n2}y_{2t} + \ldots + b_{n-n\,n}y_{n-1} + a_{n0} + a_{n1}x_{1t} + \ldots + a_{nm}x_{mt} + u_{nt}
\end{array} \qquad (8.38)$$

Якщо економетрична модель ідентифікована, то для оцінки параметрів приведеної системи регресій можна застосувати непрямий метод найменших квадратів (НМНК).

Економетрична модель буде ідентифікованою, якщо буде ідентифікованою кожна регресія системи регресій, тобто буде виконуватись умова:

$$n+m-(n_i+m_i)\ge n-1,$$

або

$$m-m_i\ge n-1$$

Припустимо, що система регресій ідентифікована. Для оцінки параметрів цієї системи регресій можна застосувати НМНК. Запишемо структурну систему регресій у матричній формі:

$$B=\begin{bmatrix} -1 & b_{12} & b_{13}\dots b_{1n} \\ b_{21} & -1 & b_{23}\dots b_{2n} \\ b_{31} & b_{32} & -1\ \dots b_{3n} \\ \dots & \dots & \dots \\ b_{n1} & b_{n2} & b_{n3}\dots -1 \end{bmatrix};\ A=\begin{bmatrix} a_{10} & a_{11}\dots a_{1m} \\ a_{20} & a_{21}\dots a_{2m} \\ a_{30} & a_{31}\dots a_{3m} \\ \dots & \dots \\ a_{n0} & a_{n1}\dots a_{nm} \end{bmatrix};$$

$$X=\begin{bmatrix} 1 & x_{11} & x_{21} & x_{31}\dots x_{m1} \\ 1 & x_{12} & x_{22} & x_{32}\dots x_{m2} \\ 1 & x_{13} & x_{23} & x_{33}\dots x_{m3} \\ \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & x_{3n}\dots x_{mn} \end{bmatrix};\ Y=\begin{bmatrix} y_{11} & y_{21}\dots y_{n1} \\ y_{12} & y_{22}\dots y_{n2} \\ y_{13} & y_{23}\dots y_{n3} \\ \dots & \dots \\ y_{1n} & y_{2n}\dots y_{nn} \end{bmatrix}$$

Запишемо структурну систему регресій у матричній формі:

$$B\cdot Y+A\cdot X+U=0 \qquad (8.39)$$

Якщо визначник матриці $B\neq 0$, то систему регресій можна розв'язати відносно ендогенних величин. Для цього запишемо систему регресій у прогнозній (приведеній) формі:

$$\begin{aligned} y_{1t}&=c_{10}+c_{11}x_{1t}+c_{12}x_{2t}+\dots+a_{1m}x_{mt}+u_{1t},\\ y_{2t}&=c_{20}+c_{21}x_{1t}+c_{22}x_{2t}+\dots+c_{2m}x_{mt}+u_{2t},\\ y_{3t}&=c_{30}+c_{31}x_{1t}+c_{32}x_{2t}+\dots+c_{3m}x_{mt}+u_{3t},\\ &\dots\dots\dots\dots\dots\dots\\ y_{nt}&=c_{n0}+c_{n1}x_{1t}+c_{n2}x_{2t}+\dots+c_{nm}x_{mt}+u_{nt} \end{aligned} \qquad (8.40)$$

Для оцінки параметрів прогнозної системи регресій (8.40) використовуємо метод найменших квадратів. Отримаємо матрицю оцінок:

$$C=\left(X'\cdot X\right)^{-1}\cdot\left(X'\cdot Y\right),\quad C^{*}=\begin{bmatrix} c_{10} & c_{20} & c_{30}\dots c_{n0}\\ c_{11} & c_{21} & c_{31}\dots c_{n1}\\ c_{12} & c_{22} & c_{32}\dots c_{n2}\\ \dots & \dots & \dots\\ c_{1m} & c_{2m} & c_{3m}\dots c_{nm}\end{bmatrix}$$

Після оцінки параметрів матриці C прогнозної форми МНК знаходимо параметри матриць A та B структурної форми:

$$-B^{-1}\cdot A=C\Rightarrow A=-B\cdot C \tag{8.41}$$

$$\begin{bmatrix} a_{10} & a_{11}\dots a_{1m}\\ a_{20} & a_{21}\dots a_{2m}\\ a_{30} & a_{31}\dots a_{3m}\\ \dots & \dots\\ a_{n0} & a_{n1}\dots a_{nm}\end{bmatrix}=-\begin{bmatrix} -1 & b_{12} & b_{13}\dots b_{1n}\\ b_{21} & -1 & b_{23}\dots b_{2n}\\ b_{31} & b_{32} & -1\ \dots b_{3n}\\ \dots & \dots & \dots\\ b_{n1} & b_{n2} & b_{n3}\dots -1\end{bmatrix}\cdot\begin{bmatrix} c_{10} & c_{20} & c_{30}\dots c_{n0}\\ c_{11} & c_{21} & c_{31}\dots c_{n1}\\ c_{12} & c_{22} & c_{32}\dots c_{n2}\\ \dots & \dots & \dots\\ c_{1m} & c_{2m} & c_{3m}\dots c_{nm}\end{bmatrix} \tag{8.42}$$

Розписавши добуток матриць, отримаємо *n-й* рівень, звідки розрахуємо елементи матриць B та A.

8.7 Двокроковий метод найменших квадратів (ДМНК)

У тих випадках, коли система одночасних регресій не ідентифікована, то для оцінки параметрів регресій використовують двокроковий метод найменших квадратів (ДМНК), який є аналогією оцінювання параметрів рекурсивної моделі.

Розглянемо структурну систему регресій:

$$\begin{cases} y_1=b_{12}y_2+b_{13}y_3+\dots+b_{1n}y_{n}+a_{10}+a_{11}x_1+\dots+a_{1m}x_m+u_1\\ y_2=b_{21}y_1+b_{23}y_3+\dots+b_{2n}y_n+a_{20}+a_{21}x_1+\dots+a_{2m}x_m+u_2\\ y_3=b_{31}y_1+b_{32}y_2+\dots+b_{3n}y_n+a_{30}+a_{31}x_1+\dots+a_{3m}x_m+u_3\\ \dots\\ y_n=b_{n1}y_1+b_{n2}y_2+\dots+b_{n+n\,n}y_{n-1}+a_{n0}+a_{n1}x_1+\dots+a_{mm}x_m+u_n\end{cases} \tag{8.43}$$

На основі статистичних даних потрібно оцінити параметри неідентифікованої структурної системи регресій.

Систему (8.43) запишемо у матричній формі: $B\cdot Y+A\cdot X+U=0$

Якщо визначник матриці $B = 0$, то структурну систему регресій представимо в приведеній (прогнозній) формі: $Y = C \cdot X + E$,

де $C = -[B]^{-1} \cdot A$; $E = -[B]^{-1} \cdot U$

Матрицю C можна записати у розгорнутому вигляді:

$$C = \begin{bmatrix} c_{10} & c_{11} & c_{12} \cdots c_{1m} \\ c_{20} & c_{21} & c_{22} \cdots c_{2m} \\ c_{30} & c_{31} & c_{32} \cdots c_{3m} \\ \cdots & \cdots & \cdots \\ c_{n0} & c_{n1} & c_{n2} \cdots c_{nm} \end{bmatrix}$$

Алгоритм двокрокового МНК.

Перший крок ДМНК.

1. Записуємо приведену форму структурних рівнянь:

$$Y = C \cdot X + E$$

2. Використовуємо МНК для кожної з регресій, отримаємо оцінки матриць C при умові, якщо $\det(X'X) \neq 0$.

$$C = \left(X' \cdot X\right)^{-1} \cdot \left(X' \cdot Y\right)$$

3. Використовуючи матрицю спостережень над екзогенними величинами:

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} & x_{31} \cdots x_{m1} \\ 1 & x_{12} & x_{22} & x_{32} \cdots x_{m2} \\ 1 & x_{13} & x_{23} & x_{33} \cdots x_{m3} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{1n} & x_{2n} & x_{3n} \cdots x_{mn} \end{bmatrix};$$

і матрицю оцінок параметрів C за формулою $Y = C \cdot X$ знаходимо ендогенну матрицю розрахункових значень (матрицю значень ендогенних величин):

$$Y = \begin{bmatrix} \hat{y}_{11} & \hat{y}_{21} \cdots \hat{y}_{k1} \\ \hat{y}_{12} & \hat{y}_{22} \cdots \hat{y}_{k2} \\ \hat{y}_{13} & \hat{y}_{23} \cdots \hat{y}_{k3} \\ \cdots & \cdots \\ \hat{y}_{1n} & \hat{y}_{2n} \cdots \hat{y}_{kn} \end{bmatrix}$$

4. Приймаючи величини y_i наперед визначеними після заміни їх на $\hat{y}_i$ методом найменших квадратів (МНК) знаходяться оцінки параметрів для кожної регресії окремо.

Після знаходження матриці $\hat{y}$ складаємо таку систему регресій:

$$\begin{cases} y_1 = b_{12}\hat{y}_2 + b_{13}\hat{y}_3 + \ldots + b_{1n}\hat{y}_n + a_{10} + a_{11}x_1 + \ldots + a_{1m}x_m + E_1 \\ y_2 = b_{21}\hat{y}_1 + b_{23}\hat{y}_3 + \ldots + b_{2n}\hat{y}_n + a_{20} + a_{21}x_1 + \ldots + a_{2m}x_m + E_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = b_{n1}\hat{y}_1 + b_{n2}\hat{y}_2 + \ldots + b_{n\,n-1}\hat{y}_{n-1} + a_{n0} + a_{n1}x_1 + \ldots + a_{nm}x_m + E_n \end{cases}$$

5. Для оцінки матриць A і B до кожного з рівнянь (8.43) застосовується МНК, наприклад, для першої регресії, що дасть змогу отримати оцінки параметрів першої регресії.

Питання для самоконтролю:

1. Система одночасних рівнянь.
2. Система незалежних регресій.
3. Оцінка параметрів системи незалежних регресій.
4. Поняття ідентифікованих і не ідентифікованих систем.
5. Рекурсивні системи рівнянь.
6. Оцінка параметрів рекурсивних систем.
7. Непрямий метод найменших квадратів.
8. Непрямий метод найменших квадратів n регресій.
9. Двокроковий метод найменших квадратів.

Тема 9

НЕПАРАМЕТРИЧНІ МЕТОДИ ОЦІНКИ ТІСНОТИ ЗВ'ЯЗКУ

9.1 Коефіцієнти рангової кореляції Спірмена і Кендалла, коефіцієнт конкордації

Серед непараметричних (емпіричних) методів оцінки тісноти зв'язку розраховують рангові коефіцієнти Спірмена ρ і Кендалла τ.

Ці коефіцієнти можуть бути використанні для визначення тісноти зв'язку як між кількісними, так і між якісними ознаками при умові, якщо значення цих показників можуть бути впорядковані або проранговані по спаданню або зростанню ознаки.

Для визначення рангового коефіцієнта кореляції ранжують (тобто записують у зростаючому або спадаючому порядку) всі значення факторної ознаки x_i і разом з тим записують відповідні значення результативної ознаки y_i (номер/ранг кожної ознаки в рангових рядах).

Ступінь тісноти зв'язку між ознаками визначається ранговим **коефіцієнтом кореляції Спірмена** по формулі:

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot \left(n^2 - 1\right)}, \qquad (9.1)$$

де d_i^2 – квадрати різниці рангів зв'язаних величин x_i і y_i;

n – число спостережень (число пар рангів).

У випадку відсутності зв'язку $\rho = 0$; при прямому зв'язку коефіцієнт ρ додатній, а при оберненому зв'язку – від'ємний.

Приклад 9.1. За даними таблиці 9.1 визначити, чи існує залежність між стажем роботи робітника та його виробітком.

Таблиця 9.1 – Вхідні дані

№ п/п	Стаж роботи робітників, x	Виробіток на 1 робітника, y
1	2,5	222
2	2,5	223
3	1	200
4	1	202
5	1	205
6	5	244
7	5	250
8	3	234
9	4,5	241
10	4,4	244
11	2,7	230

Розв'язок. Фактори x і y_i ранжуємо (впорядкуємо) в порядку зростання (спадання) їх значень (таблиця 9.2).

Таблиця 9.2 – Ранжування та результати розрахунків

x_1	y_1	Ранг ознаки x	Ранг ознаки y	Рангова різниця, d_i	d_i^2
1	200	4	3	1	1
1	202	4	4	0	0
1	205	4	5	-1	1
2,5	222	4,5	1	0,5	0,25
2,5	223	4,5	2	-0,5	0,25
2,7	230	6	11	0	0
3	234	7	8	0	0
4,4	241	8	9	1	1
4,5	244	9	8	1	1
5	244	6,5	8	-1,5	2,25
5	250	6,5	7	-0,5	0,25
Всього				3,5-3,5=0	7

2. Визначаємо ранги по обох ознаках, тобто номер кожної ознаки в рангованих рядах. Для рівних значень факторів x та y ранг визначаємо шляхом ділення суми рангів, що приходяться на неї, на число рівних значень.

3. Знаходимо рангову різницю d_i та d_i^2.

4. Розрахуємо коефіцієнт кореляції рангів Спірмена:

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot (n^2 - 1)} = 1 - \frac{6 \cdot 7}{11 \cdot (11^2 - 1)} = 0{,}968$$

Розрахунок рангового коефіцієнта Кендалла здійснюється по формулі:

$$\tau = \frac{2S}{n(n-1)}, \qquad (9.2)$$

де n – число спостережень; S – сума додатніх та від'ємних балів по одній із зв'язаних величин, ранги якої розміщені у відповідності з впорядкованими рангами другої.

Приклад 9.2. По даних десяти підприємств про чисельність робітників x та випуск продукції y (таблиця 9.3) розрахувати коефіцієнти рангової кореляції.

Розв'язок. Впорядковуємо ранг по ознаці x та y, (графи 4, 5 таблиці 9.3).

Таблиця 9.3 – Вхідні дані та результати розрахунків

Підприємство	Чисельність робітників x	Випуск продукції y	Ранг ознаки x	Ранг ознаки y	d^2	Бали для рангу y		
						від'ємні	додатні	Всього
1	2	3	4	5	6	7	8	9
1	345	23	1	1	0	0	9	9
2	485	42	4	5.5	2.25	0	4	4
3	515	37	5	3	4	1	6	7
4	622	40	6	4	4	1	5	6
5	417	30	2	2	0	3	5	8
6	450	45	3	7	16	0	3	3
7	655	42	7	5.5	2.25	1	3	4
8	815	64	8	9	1	0	1	1
9	925	73	10	10	0	0	0	0
10	878	50	9	8	1	2	0	2
Всього					30.5	8	35	44

1. Підраховуємо бали, починаючи з першого рангу ознаки y, рівного одиниці.

Число рангів, попередніх йому і більших його дорівнює нулю (від'ємні бали), а наступних за ним і більших його дорівнює дев'яти (додатні бали). Аналогічний розрахунок балів проводиться по всім рангам (графи 6, 7);

3. Знаходимо суму додатніх та від'ємних балів та загальну суму балів (графи 6, 7, 8);

4. Тоді $\tau = \frac{2 \cdot S}{n \cdot (n-1)} = \frac{2 \cdot 44}{10 \cdot 9} = 0{,}98$;

5. Розрахуємо коефіцієнт рангової кореляції Спірмена по даних таблиці 9.3.

$$\rho = 1 - \frac{6 \cdot \sum_{i=1}^{n} d_i^2}{n \cdot \left(n^2 - 1\right)} = 1 - \frac{6 \cdot 30{,}5}{10 \cdot (100 - 1)} = 0{,}815$$

Для визначення тісноти зв'язку між довільним числом рангових ознак використовують **множний коефіцієнт рангової кореляції (коефіцієнт конкордації)** W, який розраховується по формулі:

$$W = \frac{12 \cdot S}{m^2 \cdot \left(n^3 - n\right)}, \tag{9.3}$$

де m – кількість факторів; n – число спостережень; S – різниця між сумою квадратів сум по рядках і середнім квадратом суми сум рядків.

Приклад 9.3. Визначити по коефіцієнту W тісноту зв'язку між обсягом реалізованої продукції, сумою накладних витрат на реалізацію, собівартістю одиниці продукції і середньою заробітною платою робітників десяти однотипних підприємств (таблиця 9.4).

Таблиця 9.4 – Вхідні дані

Реалізація продукції, y	Накладні витрати, x	Собівартість одиниці продукції, z	Середня заробітна плата робітника, V
12,0	462	68,8	168,5
18,8	939	70,2	158,7
11,0	506	71,4	171,7
29,0	1108	78,5	188,9
17,5	872	66,9	160,4
23,4	765	69,7	165,2
35,6	1368	72,3	175,0
15,4	1002	77,5	170,4
26,1	998	65,2	162,7
20,7	804	70,7	163,0

Розв'язок. Проводимо ранжування факторів y, x, z, V (таблиця 9.5).

Таблиця 9.5 – Ранжування та результати розрахунків

R_y	R_x	R_z	R_V	Сума рядків	Квадрати сум
2	1	3	6	12	144
5	6	5	1	17	289
1	2	7	8	18	324
9	9	10	10	38	1444
4	5	2	2	13	169
7	3	4	5	19	361
10	10	8	9	37	1369
3	8	9	7	27	729
8	7	1	3	19	361
6	4	6	4	20	400
Всього				220	5590

$$S = 5590 - \frac{(220)^2}{10} = 750$$

Згідно формули (9.3):

$$W = \frac{12 \cdot 750}{16 \cdot (1000 - 10)} = 0{,}568$$

Значущість множинного коефіцієнта рангової кореляції перевіряємо по критерію x^2 Пірсона: $\chi^2 = \dfrac{S}{m \cdot n(n-1)}$

Розрахункове значення критерія:

$$\chi^2_{розр} = \frac{750}{4 \cdot 10 \cdot 9} = 2{,}083$$

Табличне значення χ^2 для імовірності $P=0,95$ складає $\chi^2_{табл} = 3{,}325$; оскільки $\chi^2_p < \chi^2_{табл}$, то значущість W підтверджується.

9.2 Коефіцієнт Фехнера

Одим із найпростіших показників кореляційної залежності пов'язаний з іменем відомого німецького вченого психофізика Фехнера. **Коефіцієнт Фехнера** ґрунтується на застосуванні перших ступенів відхилень всіх значень взаємозв'язаних ознак від середньої величини по кожній ознаці.

Коефіцієнт Фехнера вимірює тісноту зв'язку по формулі:

$$K = \frac{\sum C - \sum H}{\sum C + \sum H}, \tag{9.4}$$

де $\sum C$, $\sum H$ – число співпадінь та неспівпадінь знаків відхилень значень фактичної і результативної ознак від свої середніх, тобто $\bar{x}$, $\bar{y}$. При цьому фіксуються співпадання та неспівпадання знаків у відхиленнях від середньої y різних пар значень ознак.

Коефіцієнт Фехнера K змінюється в межах від -1 до +1. Якщо зв'язок між ознаками обернений, то K від'ємний; у випадку прямого зв'язку – додатній. Чим ближче K до ±1, тим зв'язок тісніший.

Приклад 9.4. Розрахувати тісноту зв'язку за коефіцієнтом Фехнера (таблиця 9.6).

Таблиця 9.6 – Вхідні дані

Стаж роботи, x	$x-\bar{x}$	Виробіток на 1 робітника, y	$y-\bar{y}$	Співпадання чи не співпадання знаків
2,5	-	222	-	С
2,5	-	223	-	С
1	-	200	-	С
1	-	202	-	С
1	-	205	-	С
5	+	244	+	С
5	+	250	+	С
3	+	234	+	С
4,5	+	241	+	С
4,5	+	244	+	С
2,7	-	230	+	Н
$\bar{x} = 2{,}96$		$\bar{y} = 226{,}82$		

Розв'язок. Коефіцієнт Фехнера $K = \frac{10-1}{10+1} = 0{,}818$

Величина K досить близька до величини коефіцієнта рангової кореляції Спірмена, що свідчить про тісний зв'язок між ознаками x і y.

9.3 Коефіцієнти асоціації і контингенції

Для визначення тісноти зв'язку двох якісних ознак, кожна з яких складається із двох груп, використовують коефіцієнти асоціації і контингенції. Для їх розрахунку будується чотирьохклітинна таблиця кореляції, яка виражає зв'язок між двома явищами, кожне із них в свою чергу повинно бути альтернативним, тобто складається із двох якісно відмінних одне від одного значень ознаки (наприклад, добрий / поганий).

Наприклад, при вивченні залежності врожайності від кількості внесених в грунт добрив виділимо по врожайності і по кількості внесених добрив лише по дві групи. При цій умові можна побудувати таку чотирьохклітинну таблицю.

Таблиця 9.7 – Чотирьохклітинна таблиця кореляції

Удобрено / Урожайність	Добре	Погано	Всього
Висока	a	$в$	$a+b$
Низька	c	d	$c+d$
Всього	$a+c$	$b+d$	

Числа на перетині рядків і граф a, $в$, c, d показують, скільки дільниць зустрічаються з тією або іншою кількістю добрив, що внесені в грунт, з тією або іншою врожайністю.

Коефіцієнт асоціації Юла і коефіцієнт контингенції розраховуються по таких формулах 9.5 і 9.6.

Коефіцієнт асоціації Юла:

$$KA = \frac{a \cdot d - в \cdot c}{a \cdot d + в \cdot c} \quad . \tag{9.5}$$

Коефіцієнт контингенції:

$$K = \frac{a \cdot d - в \cdot c}{\sqrt{(a+в)(в+d)(a+c)(c+d)}} \quad . \tag{9.6}$$

де a, $в$, c, d – кількісні характеристики досліджувальних груп.

Коефіцієнт контингенції завжди менший коефіцієнта асоціації Юла. Зв'язок вважається підтвердженим, якщо:

$A \geq 0{,}5$ або $K \geq 0{,}3$.

Приклад 9.5. Дослідити зв'язок між виконанням норм виробітку молодими робітниками і закінченням ними професійно технічного училища (ПТУ) за даними таблиці 9.8.

Таблиця 9.8 – Вхідні дані

Групи робітників	Виконують норму	Не виконують норму	Всього
Закінчили ПТУ	78	22	100
Не закінчили ПТУ	32	68	100
Всього	110	90	200

Розв'язок.

$$KA = \frac{78 \cdot 68 - 32 \cdot 22}{78 \cdot 68 + 32 \cdot 22} = 0.766; \qquad 0.766 > 0.5;$$

$$K = \frac{78 \cdot 68 - 32 \cdot 22}{\sqrt{(78+22)(22+68)(78+32)(32+68)}} = 0,46; \qquad 0,46 > 0,3;$$

Між досліджувальними ознаками спостерігається тісний зв'язок, що підтверджується досить високими значеннями коефіцієнтів асоціації і контингенції.

Приклад 9.6. Проведено групування студентів по зросту та вазі. Для цього вибрано ценз: по зросту – 167 см і по вазі – 67 кг. Будемо умовно вважати студентів «низькими», зріст яких нижче 167 см, і «легкими» – студентів вагою меншою 67 кг. Результати групування об'єднанні в таблицю 9.9.

Таблиця 9.9 – Розподіл 500 студентів по вазі і зросту

<table>
<tr><td colspan="2" rowspan="2">Ознака А
Ознака В</td><td colspan="2">Число студентів по вазі</td><td rowspan="2">Всього</td></tr>
<tr><td>«легкі»
(до 67 кг)</td><td>«важкі»
(більше 67 кг)</td></tr>
<tr><td rowspan="2">Число студентів по зросту</td><td>«низькі»
(до 67 кг)</td><td>304 (а)</td><td>17 (в)</td><td>321 (а+в)</td></tr>
<tr><td>«важкі»
(більше 67 кг)</td><td>112 (с)</td><td>67 (d)</td><td>179 (c+d)</td></tr>
<tr><td colspan="2">Всього</td><td>411 (а+с)</td><td>84 (в+d)</td><td>500 (а+в+с+d)</td></tr>
</table>

Коефіцієнт асоціації Юла: $KA = \frac{a \cdot d - в \cdot c}{a \cdot d + в \cdot c} = \frac{304 \cdot 67 - 112 \cdot 17}{304 \cdot 676 + 112 \cdot 17} = 0,83$

Коефіцієнт асоціації близький до одиниці, що свідчить про тісний зв'язок між зростом та вагою студентів.

9.4 Коефіцієнт взаємної спряженості Пірсона і Чупрова

Якщо кожна із якісних ознак складається більше ніж із двох груп, то для визначення тісноти зв'язку можна використати коефіцієнт взаємної спряженості Пірсона, який розраховується по такій формулі:

$$c = \sqrt{\frac{q^2}{1+q}} \qquad (9.7)$$

де q^2 – показник взаємної спряженості.

Коефіцієнт Чупрова:

$$K = \frac{q^2}{\sqrt{(K_1 - 1)(K_2 - 1)}}; \qquad (9.8)$$

де K_1, K_2 – число груп по кожній із ознак.

$$q^2 = \sum_{j_1=1}^{K_1} \sum_{j_2=1}^{K_2} \frac{n_{j_1}\ f_2}{n_{j_1} \cdot n_{j_2}}$$

Розрахунок коефіцієнта взаємної спряженості проводиться по такій схемі:

Таблиця 9.10 – Групи ознак

Групи ознаки А	Групи ознак В			Разом
	B_1	B_2	B_3	
A_1	f_1	f_2	f_3	n_1
A_2	f_4	f_5	f_6	n_2
A_3	f_7	f_8	f_9	n_3
Разом	m_1	m_2	m_3	

Розрахунок q^2:

по першому рядку $\left(\frac{f_1^2}{m_1} + \frac{f_2^2}{m_2} + \frac{f_3^2}{m_3}\right) : n_1 = Z_1$;

по другому рядку $\left(\frac{f_4^2}{m_1} + \frac{f_5^2}{m_2} + \frac{f_6^2}{m_3}\right) : n_2 = Z_2$;

по третьому рядку $\left(\frac{f_7^2}{m_1} + \frac{f_8^2}{m_2} + \frac{f_9^2}{m_3}\right) : n_3 = Z_3$

$q^2 = Z_1 + Z_2 + Z_3 - 1 = \sum Z_i - 1$

Приклад 9.7. В таблиці 9.11 приведені згруповані дані вартості основних виробничих засобів (x) і обсягу реалізації продукції y. По кожній ознаці утворено три групи. По основних засобах: перша група <2,5; друга група – 2,5-3,5 і третя група >3,5 умовних г.о. По обсягу реалізації продукції: перша група <6,5; друга група – 6,5-9,5 і третя група >9,5 умовних г.о.

Таблиця 9.11 – Вхідні дані

Групи підприємств по обсягу реалізації продукції, г.о., y	Групи підприємств по вартості основних виробничих засобів, г.о., x			Разом
	1,5-2,5	2,5-3,5	3,5-4,5	
3,5-6,5	48	18	3	69
6,5-9,5	15	30	13	58
9,5-12,5	-	1	7	8
Разом	63	49	23	

Розв'язок. Розрахуємо q^2:

по першому рядку $\left(\frac{48^2}{63}+\frac{18^2}{49}+\frac{3^2}{23}\right):69=0{,}631;$

по другому рядку $\left(\frac{15^2}{63}+\frac{30^2}{49}+\frac{13^2}{23}\right):58=0{,}505;$

по третьому рядку $\left(\frac{1^2}{49}+\frac{7^2}{23}\right):8=0{,}269;$

$q^2=0{,}631+0{,}505+0{,}269-1=0{,}405$.

Підставляємо q^2 у відповідні формули і знаходимо:

коефіцієнт Пірсона: $c=\sqrt{\frac{0{,}405}{1{,}405}}=0{,}54$;

коефіцієнт Чупрова: $K=\frac{0{,}405}{\sqrt{(3-1)(3-1)}}=0{,}20$

Приклад 9.8. В таблиці 9.12 приведені згруповані дані накладних видатків (x) та собівартості продукції (y). Дослідити зв'язок між собівартістю продукції та накладними витратами за коефіцієнтом взаємної спряженості.

Таблиця 9.12 – Вхідні дані

Накладні видатки	Собівартість			Разом
	нижня	середня	висока	
Нижні	19	12	9	40
Середні	7	18	15	40
Високі	4	10	26	40
Разом	30	40	50	

Розв'язок. Розрахуємо q^2:

по першому рядку $\left(\frac{19^2}{30}+\frac{12^2}{40}+\frac{9^2}{50}\right):40=0{,}431$;

по другому рядку $\left(\frac{7^2}{30}+\frac{18^2}{40}+\frac{15^2}{50}\right):40=0{,}356$;

по третьому рядку $\left(\frac{4^2}{30}+\frac{10^2}{40}+\frac{26^2}{50}\right):40=0{,}414$;

$q^2=0{,}431+0{,}356+0{,}414-1=0{,}204$

Підставляємо q^2 у відповідні формули і знаходимо:

коефіцієнт Пірсона: $c=\sqrt{\frac{q^2}{1+q}}=c=\sqrt{\frac{0{,}204}{1{,}204}}=0{,}41$;

коефіцієнт Чупрова: $K=\frac{0{,}204}{\sqrt{(3-1)(3-1)}}=0{,}102$.

Досить високе значення c вказує на наявність зв'язку між собівартістю продукції та накладними видатками на реалізацію.

Непараметричні методи вимірювання зв'язку використовуються для перевірки умов використання метода найменших квадратів, незалежності розподілу ознак, однорідності вибірок, наявності тренда в рядах динаміки.

Питання для самоконтролю:

1. Поняття ранга, ранжування.
2. Коефіцієнт рангової кореляції Спірмена.
3. Коефіцієнт рангової кореляції Кендала.
4. Множинний коефіцієнт рангової кореляції (коефіцієнт конкордації).
5. Коефіцієнт Фехнера.
6. Коефіцієнт асоціації Юла і контингенції.
7. Коефіцієнт взаємної спряженості Пірсона і Чупрова.

Тема 10

ПРИКЛАДНІ ЕКОНОМЕТРИЧНІ МОДЕЛІ

10.1 Моделювання сезонних коливань економічних процесів рядами Фур'є

Значна частина економічних процесів, наприклад, попит на товари та послуги, пасажиропотік, виробництво цукру та консерв носить сезонний характер.

Для аналізу та прогнозування економічних процесів, яким притаманні сезонні коливання, використовують ряд Фур'є.

В загальному вигляді ряд Фур'є можна записати так:

$$Y_t = a_0 + \sum_{k=1}^{m} (a_k \cos kt + b_k \sin kt), \qquad (10.1)$$

де a_0, a_k, b_k – параметри моделі, t – фактор часу, k – порядковий номер гармоніки, m – кількість порядкових номерів гармоніки.

В економетричних дослідженнях кількість гармонік ряду Фур'є приймають не більше 4 і визначають, яка із гармонік найбільш адекватно описує сезонні коливання економічних явищ.

Параметри ряду Фур'є a_0, a_k, b_k визначають за методом найменших квадратів.

Запишемо формули для розрахунку цих параметрів:

$$a_0 = \frac{\sum y_i}{n} = \bar{y},\ a_k = \frac{2 \cdot \sum y_i \cos kt_i}{n},\ b_k = \frac{2 \cdot \sum y_i \sin kt}{n}; \qquad (10.2)$$

де n – кількість періодів часу за які розглядається явище (місяців, днів, кварталів, років).

Для побудови економічної моделі необхідно зробити перехід фактора часу від натурального масштабу до радіанного або градусного.

Перехід від натурального масштабу до радіанного або градусного здійснюється за формулою:

$$\frac{2\pi}{n} \cdot n', \qquad (10.3)$$

де n – кількість спостережень (або кількість інтервалів часу, за який аналізується сезонне явище), n' – натуральний ряд чисел від 0 до n-1 (0, 1, 2,..., n-1).

Запишемо місяці у радіанній формі:

січень	0
лютий	$\pi/6$
березень	$\pi/3$
квітень	$\pi/2$
травень	$2\pi/3$
червень	$5\pi/6$
липень	π
серпень	$7\pi/6$
вересень	$4\pi/3$
жовтень	$3\pi/2$
листопад	$5\pi/3$
грудень	$11\pi/6$

Для визначення параметрів моделі a_k та b_k необхідно скласти таблицю значень тригонометричних функцій $\begin{matrix} \cos t, \ \cos 2t, \ldots \\ \sin t, \ \cos 2t, \ldots \end{matrix}$

Таблиця 10.1 – Значення тригонометричних функцій

t_i	t_i, рад.	$\cos t_i$	$\cos 2t_i$	$\sin t_i$	$\sin 2t_i$
1	0	1	1	0	0
2	$\pi/6$	0,866	0,5	0,5	0,866
3	$\pi/3$	0,5	-0,5	0,866	0,866
4	$\pi/2$	0	-1	1	0
5	$2\pi/3$	-0,5	-0,5	0,866	-0,866
6	$5\pi/6$	-0,866	0,5	0,5	-0,866
7	π	-1	1	0	0
8	$7\pi/6$	-0,866	0,5	-0,5	0,866
9	$4\pi/3$	-0,5	-0,5	-0,866	0,866
10	$3\pi/2$	0	-1	-1	0
11	$5\pi/3$	0,5	-0,5	-0,866	-0,866
12	$11\pi/6$	0,866	0,5	-0,5	-0,866

Приклад 10.1. Описати сезонні коливання реалізації одягу рядами Фур'є та вибрати гармоніку, яка найбільш адекватно описує ці сезонні коливання за даними таблиці 10.2.

Розв'язок. Ми отримали статистику для визначення параметрів моделі для першої гармоніки.

$$k = 1$$

$$a_0 = \frac{\sum y_i}{n} = \frac{552}{12} = 46;$$

$$a_1 = \frac{2 \cdot \sum y_i \cos kt_i}{n} = \frac{2 \cdot (-66{,}24)}{12} = -11{,}04;$$

$$b_1 = \frac{2 \cdot \sum y_i \sin kt}{n} = \frac{2 \cdot 34{,}43}{12} = 5{,}74$$

Записуємо ряд Фур'є для першої гармоніки:

$$y_1 = 46 - 11{,}04 const + 5{,}74 \sin t + u_1$$

Знаходимо теоретичні значення $\hat{y}_1$. Замість значення $const$ та $\sin t$ у ряд Фур'є підставимо їх значення із таблиці 10.1

Розраховуємо тісноту зв'язку (кореляційне відношення) для першої гармоніки. Для цього нам потрібно організувати стовпці: $(y_i - \hat{y}_{1i})^2$, $(y_i - \bar{y})^2$.

Кореляційне відношення для *1-ої* гармоніки:

$$\eta_1 = \sqrt{1 - \frac{\sum (y_i - \hat{y}_{1i})^2}{\sum (y_i - \bar{y}_{1i})^2}} = \sqrt{1 - \frac{286{,}88}{12{,}18}} = 0{,}8774$$

Розрахуємо параметри ряду Фур'є по другій гармоніці; для цього необхідно сформувати два стовпчики: $y_i \cos 2t_i$ та $y_i \sin 2t_i$.

Таким чином, маємо дані розрахунку для параметрів a_2 та b_2.

$$a_2 = \frac{2 \cdot \sum y_i \cdot \cos 2t_i}{n} = \frac{2 \cdot 17{,}5}{12} = 2{,}9;$$

$$a_2 = \frac{2 \cdot \sum y_i \cdot \sin 2t_i}{n} = \frac{2 \cdot -7{,}78}{12} = -1{,}3$$

Записуємо ряд Фур'є для другої гармоніки:

$$y_2 = 46 - 11{,}04 cont + 5{,}74 \sin t + 2{,}9 \cos 2t - 1{,}3 \sin 2t + u_2$$

Таблиця 10.2 – Вхідні дані та результати розрахунків

Місяці	t_i, рад	y_i	$y_i \cos t_i$	$y_i \sin t_i$	$\hat{y}_{1t}$	$(y_i - \hat{y}_{1t})^2$	$(y_i - \bar{y})^2$	$y_i \cos 2t_i$	$y_i \sin 2t_i$	$\hat{y}_{2t}$	$(y_i - \hat{y}_{2t})^2$
січень	0	37	37,000	0,000	34,96	4,161	81	37	0,000	37,88	0,769
лютий	$\pi/6$	40	34,641	20,000	39,31	0,479	36	20	34,641	39,64	0,128
березень	$\pi/3$	44	22,000	38,105	45,45	2,101	4	-22	38,105	42,87	1,286
квітень	$\pi/2$	52	0,000	52,000	51,74	0,069	36	-52	0,000	48,82	10,104
травень	$2\pi/3$	46	-23,000	39,837	56,49	110,023	0	-23	-39,837	56,16	103,142
червень	$5\pi/6$	70	-60,622	35,000	58,43	133,869	576	35	-60,622	61,01	80,764
липень	π	60	-60,000	0,000	57,04	8,762	196	60	0,000	59,96	0,002
серпень	$7\pi/6$	48	-41,569	-24,000	52,69	22,013	4	24	41,569	53,03	25,252
вересень	$4\pi/3$	46	-23,000	-39,837	46,55	0,303	0	-23	39,837	43,97	4,132
жовтень	$3\pi/2$	38	0,000	-38,000	40,26	5,116	64	-38	0,000	37,35	0,429
листопад	$5\pi/3$	36	18,000	-31,177	35,51	0,239	100	-18	-31,177	35,18	0,677
грудень	$11\pi/6$	35	30,311	-17,500	33,57	2,044	121	17,5	-30,311	36,15	1,331
Σ		552	-66,239	34,428	552,0	289,180	1218	17,5	-7,794	552,0	228,014

Записуємо тісноту зв'язку (кореляційне відношення) для другої гармоніки:

$$\eta_2 = \sqrt{1 - \frac{\sum (y_i - \hat{y}_{1i})^2}{\sum (y_i - \bar{y}_{1i})^2}} = \sqrt{1 - \frac{287{,}9}{1218}} = 0{,}874$$

Оскільки $\eta_1 > \eta_2$, то ряд Фур'є по першій гармоніці більш адекватно описує сезонні коливання попиту.

Будуємо графік сезонних коливань продажу одягу.

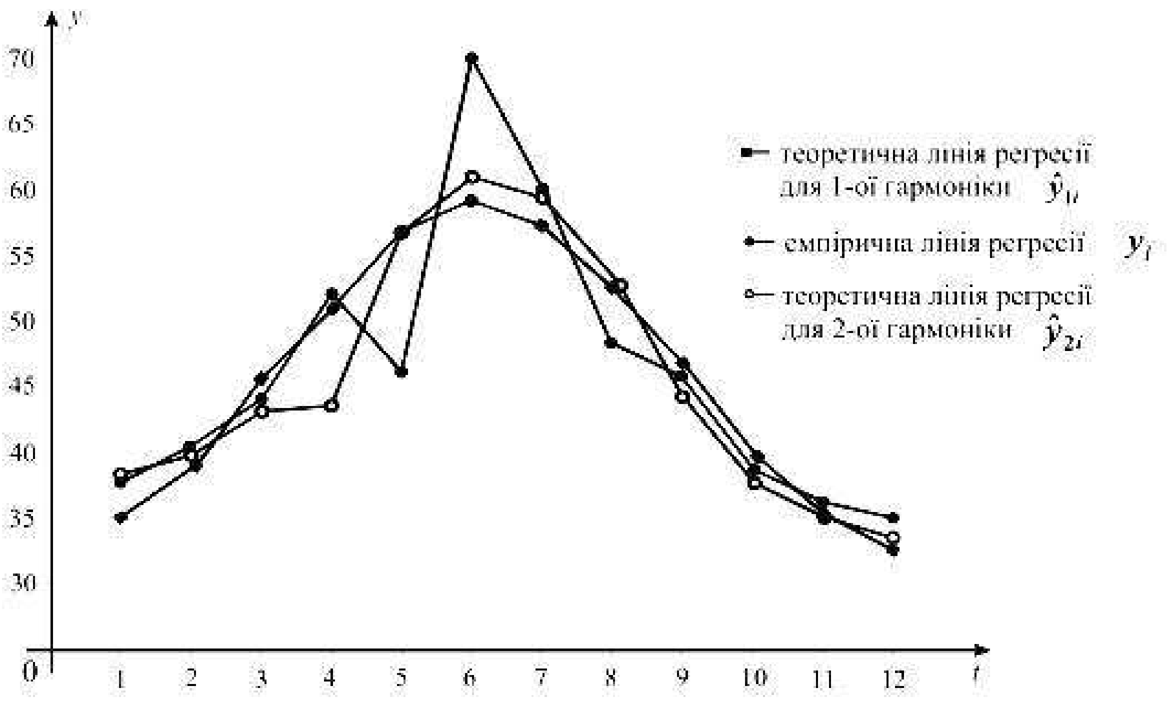


Рисунок 10.1 – Сезонні коливання попиту

10.2 Обернені функції (гіпербола)

Узагальнена обернена функція має вигляд:

$$y_i = \beta_0 + \beta_1 \left(\frac{1}{x_i} \right) + U_i \tag{10.4}$$

Вона нелінійна за зміною x, але лінійна за параметрами β_0, β_1, тому є лінійною регресійною моделлю.

Позначивши $x_i' = \frac{1}{x_i}$, отримаємо:

$$y_i = \beta_0 + \beta_1 x_i' + U_i . \tag{10.5}$$

Вибіркова обернена модель має вигляд:

$$y = a_0 + a_1 \cdot \frac{1}{x} + u, \tag{10.6}$$

де a_0, a_1 – невідомі параметри моделі, які необхідно знайти.

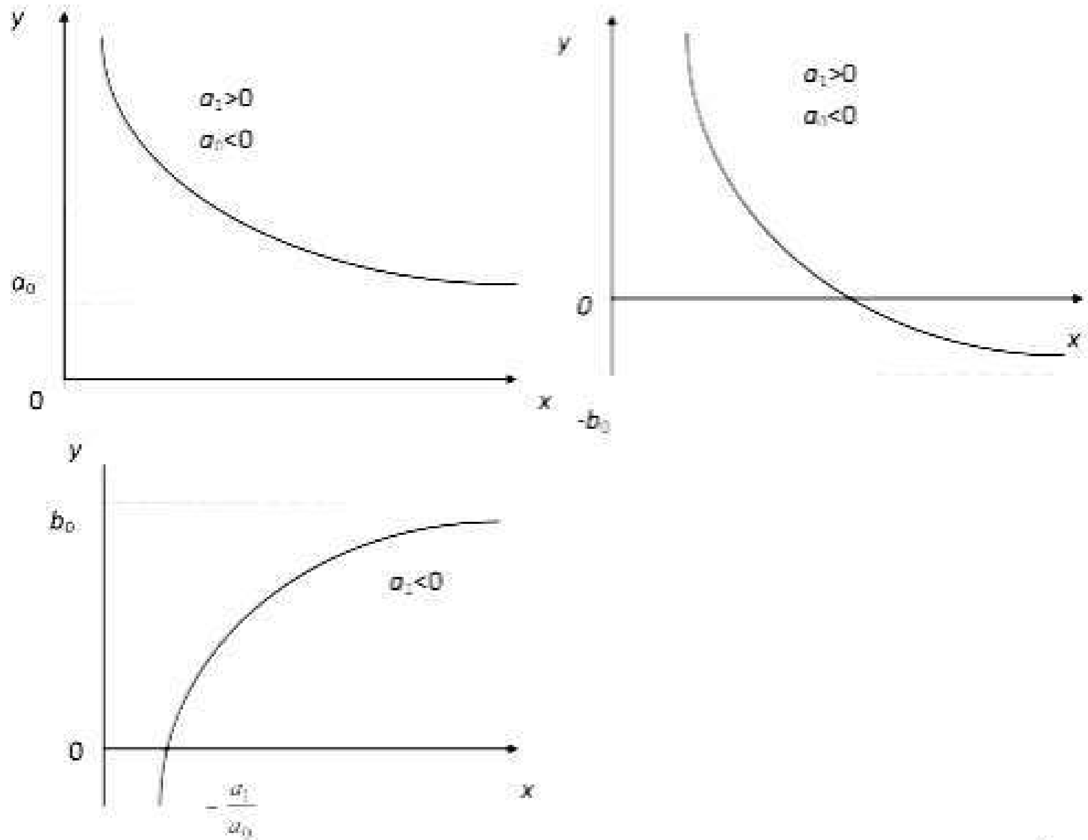


Рисунок. 10.2 – Графіки оберненої функції: $y = a_0 + a_1 \cdot \frac{1}{x}$

Обернена модель має особливість: якщо x прямує до нескінченності, то величина $a_1 \cdot \frac{1}{x}$ прямує до нуля, а y прямує до граничного значення.

Вигляд моделі (10.6) залежить від знаку параметрів a_0 і a_1.

Нахил моделі $\frac{dy}{dx} = -b \cdot \left(\frac{1}{x_i}\right)$. Він є додатнім, якщо $b_1 < 0$, від'ємним, якщо $b_1 > 0$.

Обернена функція широко використовується при моделюванні в макроекономіці. Прикладом цього може бути крива Філіпса (рисунок 10.3).

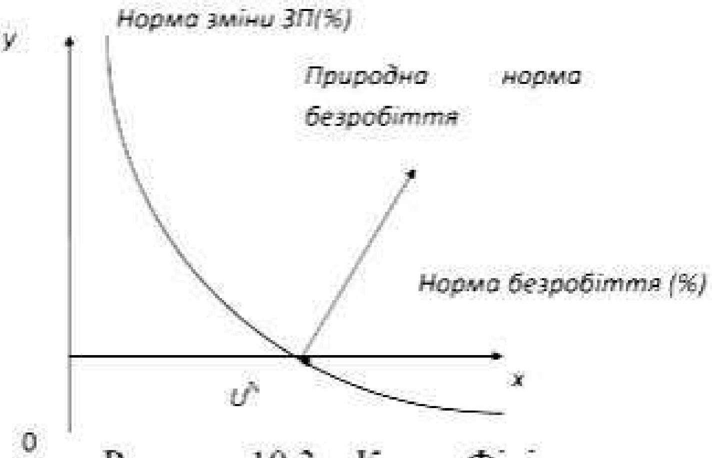
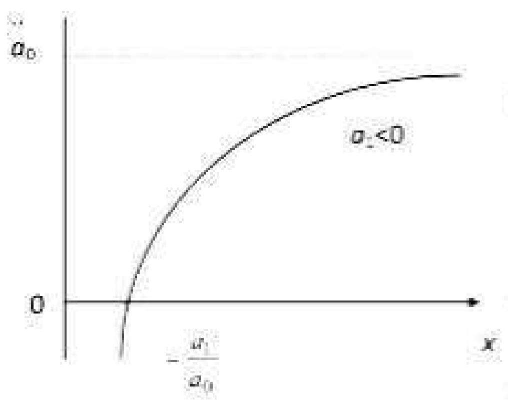


Рисунок 10.3 – Крива Філіпса

Крива Філіпса описує залежність норми процента зміни заробітної плати від процентна безробіття асимптота (границя зміни заробітної плати) пов'язана з параметром a_0. Також $U^{\wedge}$ є значенням природної норми безробіття, якщо $x < U^{\wedge}$, то норма зміни заробітної плати додатня; якщо значення x стає більшим, ніж природна норма безробіття, y (норма зміни заробітної плати) буде від'ємною.

Крива Філіпса дає змогу розрахувати мінімальну заробітну плату, компенсацію за безробіття.

Крива Енгеля описує залежність витрат на споживання від загальних витрат або доходу. Позначимо через y – витрати на споживання, а через x – дохід, тоді крива Енгеля для певного виду товару буде мати такі особливості:



а) Критичний рівень доходу, нижче від якого товар не буде куплено – це значення $-\frac{a_1}{a_0}$

б) межа насичення, яку не можна збільшити, як би не зростав дохід – це значення a_0.

Рисунок 10.4 – Крива Енгеля

10.3 Квадратичні функції

Квадратичні функції широко використовуються для опису залежності виробітку робітника від віку.

У загальному випадку квадратична функція має вигляд:

$$y = a_0 + a_1 x + a_2 x^2.$$

або

$$y = a_0 - a_1 t - a_2 t^2.$$

Виходячи із значень a_0, a_1, a_2, крива може відображати різну еволюцію у часі *t*.

Якщо $a_2 > 0$, то маємо параболу на *min*, при $a_2 < 0$ – параболу на *max*. Екстремум функції досягається у праці $t = -a_1 / 2a_2$.

10.4 Експоненційна модифікована крива

Модифікована експонента широко використовується при дослідженні ринку. Вона має вигляд:

$$y = a_0 + a_1^t + c$$

Графічне зображення модифікованої експоненти наведено на рисунку10.5.

Значення модифікованої експоненти знизу обмежене значенням *c*.

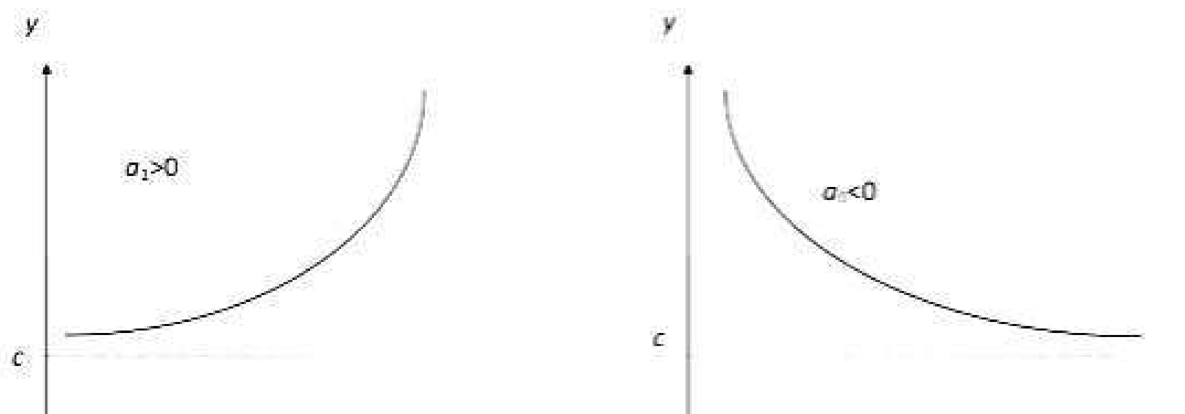


Рисунок 10.5 – Модифікована експонента

Якщо відоме значення параметра c, то для оцінки параметрів експоненти виконуємо перетворення та логарифмування правої та лівої її частин:

$$y - c = a_0 \cdot a_1^x$$

$$\ln(y - c) = \ln a_0 + x \ln a_1$$

В цьому випадку модифікована експонента приводиться до експоненційної кривої, яку після логарифмування зводимо до простої лінійної регресії.

10.5 Крива Гомперца

Крива Гомперца використовується в основному для опису процесів з насиченням і дуже поширена в демографії, маркетингу, досліджені ринку, дослідженні збуту продукції. Вона має вигляд:

$$y = e^{a_0 \cdot a_1^x} + c, \text{ де } 0 < a_1 < 1$$

Шляхом логарифмування і подальшої заміни змінних криву Гомперца зводимо до модифікованої експоненти:

$$\ln y = a_0 \cdot a_1^x + c\,.$$

Крива Гомперца має такі властивості. Якщо $a_0 > 0$, то функція опукла. Якщо $a < 0$, крива Гомперца є S-кривою та має точку перегину X', якщо $1 + a_0 \cdot a_1^a = 0$, тобто $X' = \dfrac{\ln(-1\ a_0)}{\ln a_1}$.

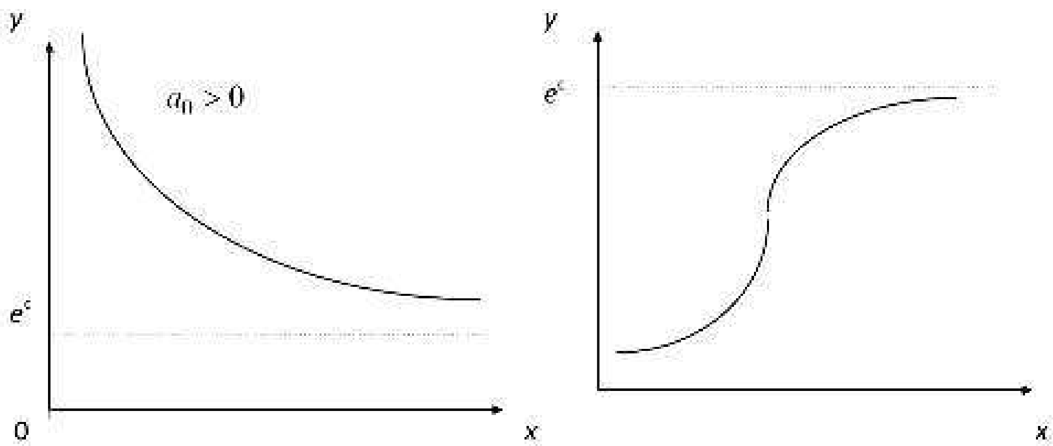


Рисунок 10.6 – Крива Гомперца

У цій точці $y = e^{-1+c}$. Точка перетину входить до інтервалу спостережень $[1;x]$, якщо $a_0, a_1 < -1$.

Якщо $a_0 > 0$, то крива спадає асимптотично до e^c, а якщо $a_0 < 0$, то вона має форму *S*-кривої, тобто спочатку зростає швидко, а потім повільно. Такою функцією можна описати життєвий цикл продажу товару.

10.6 Логістична крива

Логістична крива є оберненою функцією до модифікованої експоненти. Вона має вигляд:

$$y = \frac{1}{a_0 \cdot a_1^x + c}, \tag{10.7}$$

де $0 < a_1 \le 1, a_0 > 0, c > 0$

Логістичну криву можна представити у вигляді:

$$y = \frac{a}{1 + \left(\frac{a}{y_0} - 1\right) \cdot e^{-akt}}, \tag{10.8}$$

де a – рівень насиченості.

Якщо рівень насиченості прийняти за 100% (a=1), то функція має вигляд:

$$y = \frac{a}{1 + \left(\frac{a}{y_0} - 1\right) \cdot e^{-kt}}, \tag{10.9}$$

де y_0 – початкове насичення на період $t = 0$. З рівності (10.9) можна визначити коефіцієнт пропорційності k.

Точка $\left(t'; \frac{a}{2}\right)$ є точкою перегину графіка функції (рисунок 10.7)

$$t' = \frac{1}{ak} \ln\left(\frac{a}{y_0 - 1}\right)$$

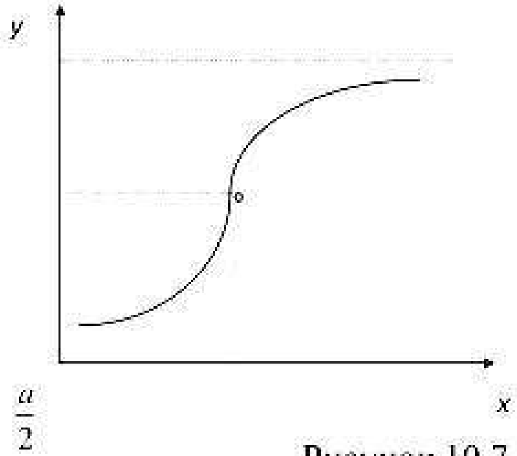


Рисунок 10.7 – Графік логістичної кривої

Наприклад, якщо на початок сезону послугою масового попиту було забезпечено 12% населення, а на кінець першої декади – 20%, то маємо залежність:

$$y = \frac{1}{\left(1 + \frac{22}{3} \cdot e^{-0{,}0606}\right)}$$

Максимальний попит буде в час $t' = \ln \frac{22}{3} = 32871$, тобто на 33-й день,

а 90% населення буде забезпечено в час $t = \frac{1}{-0{,}0606} \cdot \frac{\ln \frac{1}{0{,}9} - 1}{22\ 3} = 69{,}21$, тобто на 70-й день.

Якщо ж є в наявності n-спостережень t_i, y_i, то залежність (10.9) будується як кореляційна. Заміною змінних $x = t,\ y = \ln\left(\frac{1}{y} - 1\right)$ вона лінеаризується.

10.7 Експоненціальна змінна рівня забезпеченості

Якщо рівень забезпеченості y наближається до насиченості, то вважають, що:

$$\frac{dy}{dx} = k\left(a - y\right) \tag{10.10}$$

і при початкових умовах $y(t_0) = y^{(0)}$, взятих із (10.10) або із сезону минулого року, одержимо:

$$y = \left(a - \left(a - y^{(0)}\right)\right) \cdot e^{-k(t - t_0)}, \tag{10.11}$$

змістивши відлік часу в точку $t_0 = 0$ і заклавши $a = 1$, замість (10.11) матимемо:

$$y = 1 - \left(1 - y^{(0)}\right) \cdot e^{-kt}. \tag{10.12}$$

10.8 Функція Лаффера

Крива Лаффера дозволяє встановити залежність між податковими ставками і обсягом податкових надходжень.

Нехай Y – податкові надходження, X – податкова ставка. Тоді криву Лаффера запишемо у вигляді:

$$Y = a \cdot e^{b(x-c)}. \tag{10.13}$$

Для оцінки параметрів цієї регресії прологарифмуємо вираз (10.13).

$$\ln Y = \ln a + b \cdot X^2 - 2 \cdot b \cdot c \cdot X + b \cdot c^2$$

Після заміни: $\hat{Y}_1 = \ln \hat{Y}$, $a_0 = \ln a + b \cdot c^2$, $d = -2bc$

отримаємо таку регресію: $\hat{Y}_1 = a_0 + dX + bX^2$ із системи нормальних рівнянь:

$$\begin{cases} a_0 \cdot n + d \cdot \sum\limits_{i=1}^{n} X_i + b \cdot X_i^2 = \sum\limits_{i=1}^{n} Y_{1i} \\ a_0 \cdot \sum\limits_{i=1}^{n} X_i + d \cdot \sum\limits_{i=1}^{n} X_i^2 + b \cdot \sum\limits_{i=1}^{n} X_i^3 = \sum\limits_{i=1}^{n} Y_{1i} \cdot X_i \\ a_0 \cdot \sum\limits_{i=1}^{n} X_i^2 + d \cdot \sum\limits_{i=1}^{n} X_i^3 + b \cdot \sum\limits_{i=1}^{n} X_i^4 = \sum\limits_{i=1}^{n} Y_{1i} \cdot X_i^2 \end{cases}$$

Отримаємо оцінки параметрів регресії a_0, d, b з рівняння $-2b \cdot c = d$ отримаємо оцінку параметра c:

$$c = -\frac{d}{2 \cdot b}.$$

Відповідно з рівняння $a_0 = \ln a + b \cdot c^2$ отримаємо:

$$\ln a = a_0 - b \cdot c^2$$

$$a = \exp\left(a_0 - \frac{d^2}{4 \cdot b}\right)$$

Якщо $b < 0$, то графік кривої Лаффера буде мати вигляд:

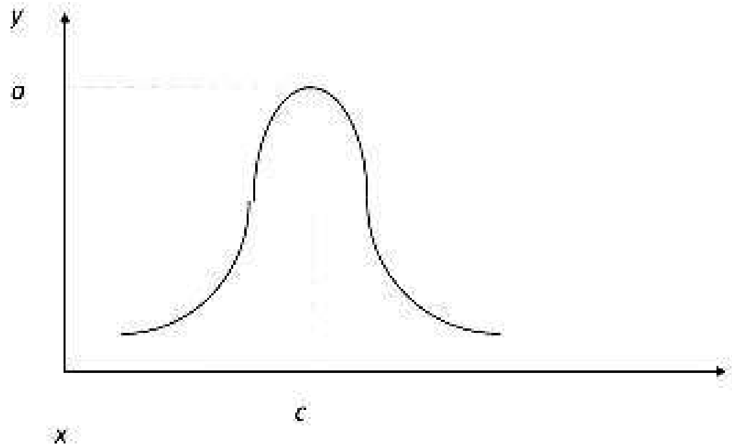


Рисунок 10.8 – Крива залежності обсягу податкових надходжень від податкової ставки (крива Лаффера)

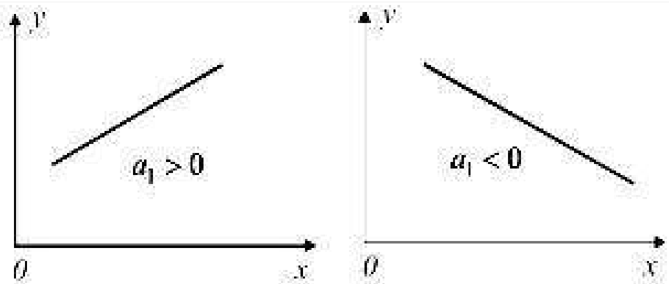
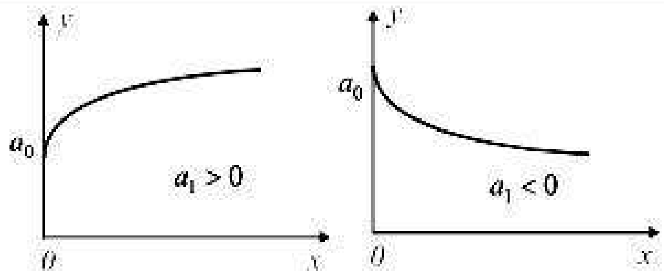
Отже, можемо оцінити вплив рівня податкових ставок та обсяг податкових надходжень. При збільшенні податків до певного рівня податкові надходження (величина a) зростають, а подальше підвищення податків веде до зменшення обсягу податкових надходжень.

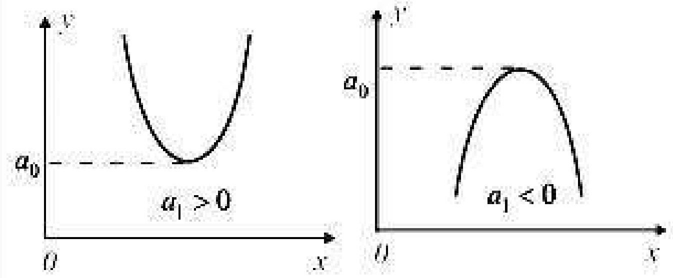
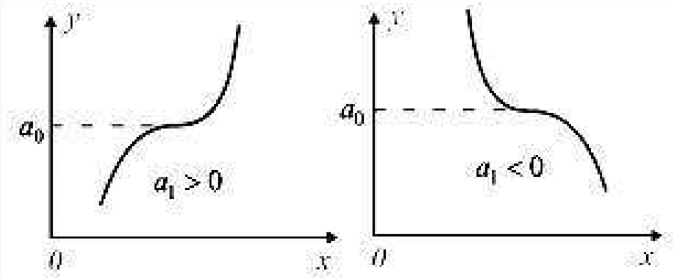
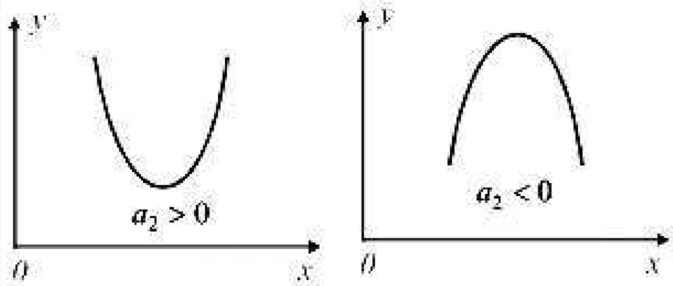
Питання для самоконтролю:

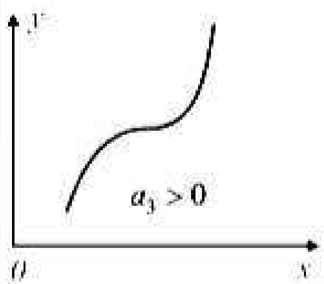
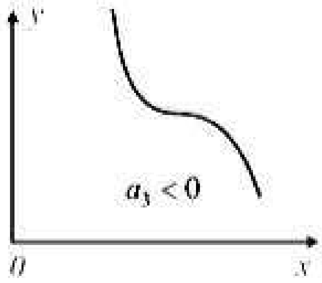
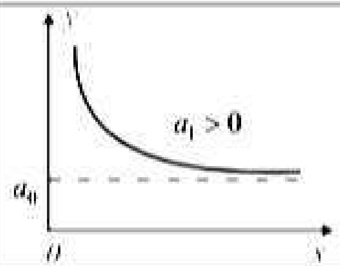
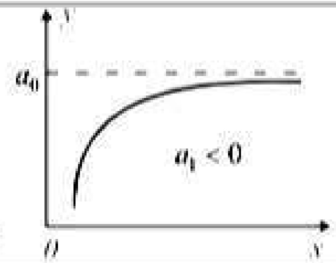
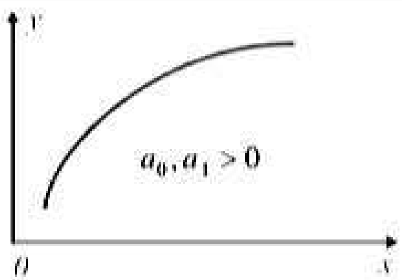
1. Які економічні явища вважаються сезонними?
2. Якими економетричнами моделями описуються сезонні явища?
3. Алгоритм розрахунку параметрів моделі рядами Фур'є.
4. За яким критерієм проводиться вибір порядку гармоніки?
5. Що описує крива Філіпса?
6. Які залежності описує крива Енгеля?
7. Які явища описує квадратична функція?
8. Особливості модифікованої експоненти.
9. Крива Гомперца, її особливості та застосування.
10. Логістична крива, її особливості та застосування.
12. Функція Лаффера та сфера її застосування.

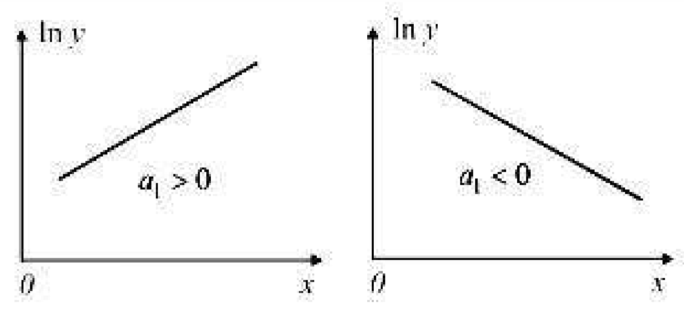
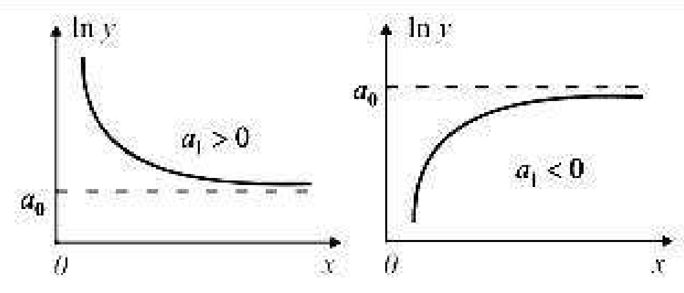
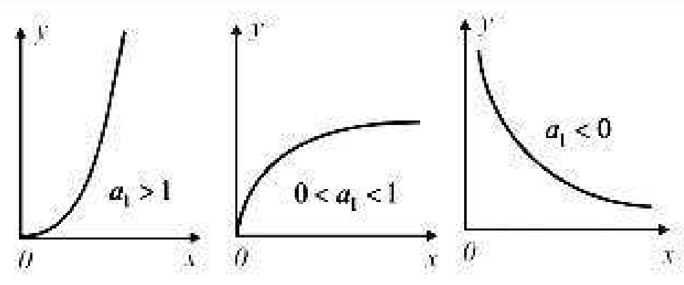
ДОДАТКИ

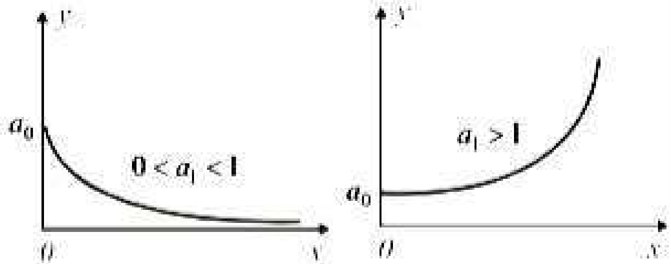
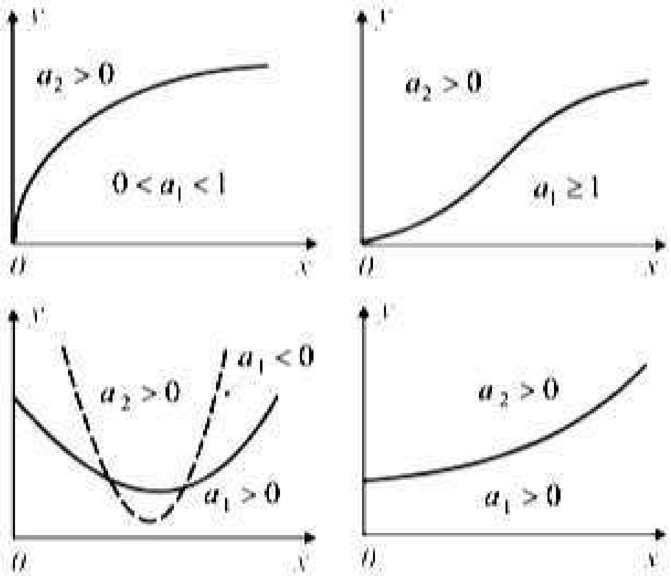
Функції зростання

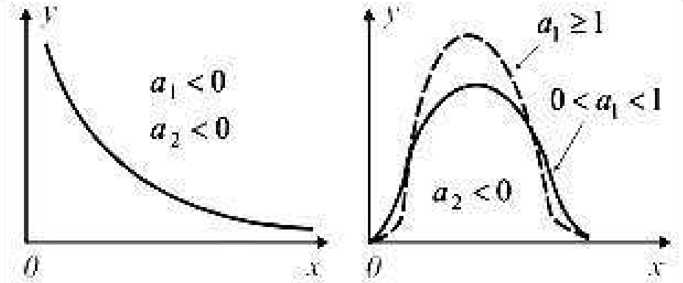
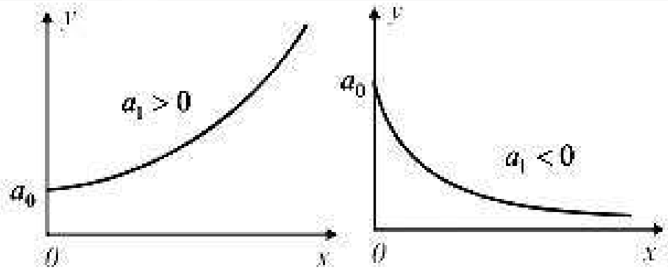
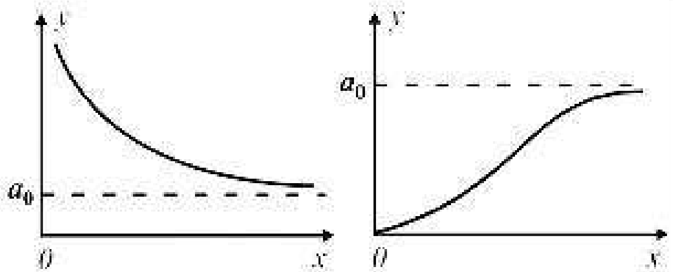
Вид рівняння регресії	Схематичний графік	Формула лінеарізації	Системи нормальних рівнянь по методу найменших квадратів
1	2	3	4
$y = a_0 + a_1 x$			$\begin{cases} na_0 + a_1\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ a_0\sum_{i=1}^{n} x_i + a_1\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i \end{cases}$
$y = a_0 + a_1\sqrt{x}$		$X = \sqrt{x}$, $Y = y$	$\begin{cases} na_0 + a_1\sum_{i=1}^{n} \sqrt{x_i} = \sum_{i=1}^{n} y_i \\ a_0\sum_{i=1}^{n} \sqrt{x_i} + a_1\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i\sqrt{x_i} \end{cases}$

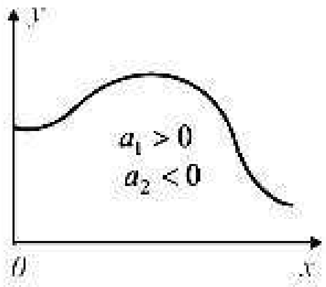
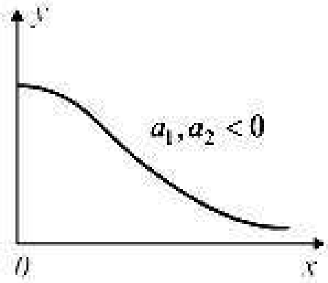
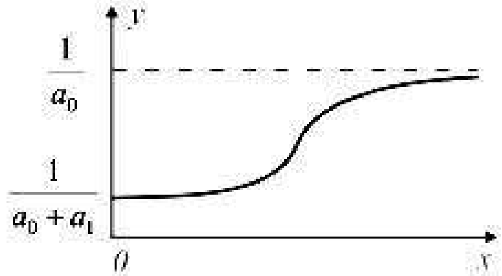
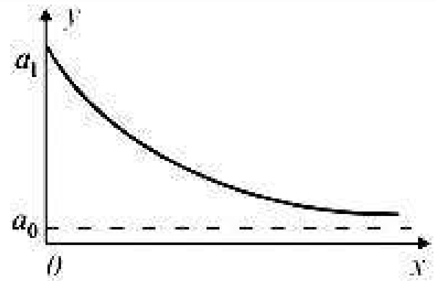
$y = a_0 + a_1 x^2$		$X = x^2$, $Y = y$	$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i \\ a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} y_i x_i^2 \end{cases}$
$y = a_0 + a_1 x^3$		$X = x^3$, $Y = y$	$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} y_i \\ a_0 \sum_{i=1}^{n} x_i^3 + a_1 \sum_{i=1}^{n} x_i^6 = \sum_{i=1}^{n} y_i x_i^3 \end{cases}$
$y = a_0 + a_1 x + a_2 x^2$		$X_1 = x$, $X_2 = x^2$, $Y = y$	$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} x_i + a_2 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i \\ a_0 \sum x_i + a_1 \sum_{i=1}^{n} x_i^2 + a_2 \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} y_i x_i \\ a_0 \sum x_i^2 + a_1 \sum_{i=1}^{n} x_i^3 + a_2 \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} y_i x_i^2 \end{cases}$

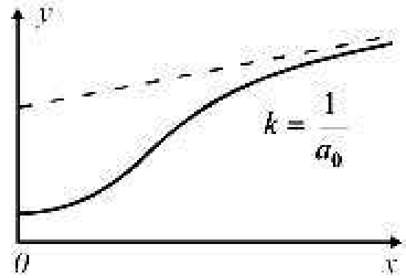
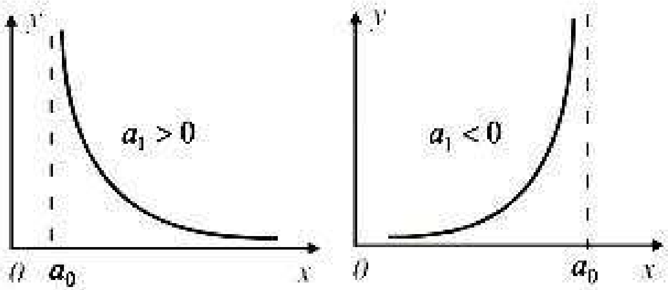
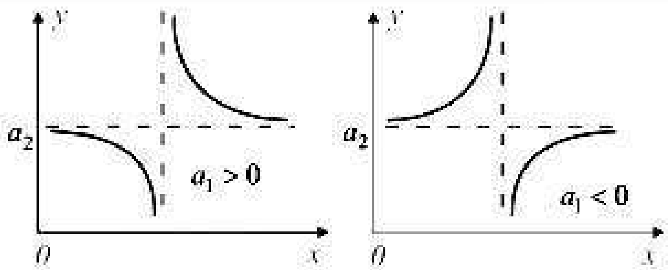
$y=a_0+a_1x+a_2x^2+a_3x^3$	 	$X_1=x_1,$ $X_2=x^2,$ $X_3=x^3$ $Y=y$	$\begin{cases} na_0+a_1\sum_{i=1}^{n}x_i+a_2\sum_{i=1}^{n}x_i^2+a_3\sum_{i=1}^{n}x_i^3=\sum_{i=1}^{n}y_i \\ a_0\sum_{i=1}^{n}x_i+a_1\sum_{i=1}^{n}x_i^2+a_2\sum_{i=1}^{n}x_i^3+a_3\sum_{i=1}^{n}x_i^4=\sum_{i=1}^{n}x_iy_i \\ a_0\sum_{i=1}^{n}x_i^2+a_1\sum_{i=1}^{n}x_i^3+a_2\sum_{i=1}^{n}x_i^4+a_3\sum_{i=1}^{n}x_i^5=\sum_{i=1}^{n}x_i^2y_i \\ a_0\sum_{i=1}^{n}x_i^3+a_1\sum_{i=1}^{n}x_i^4+a_2\sum_{i=1}^{n}x_i^5+a_3\sum_{i=1}^{n}x_i^6=\sum_{i=1}^{n}x_i^3y_i \end{cases}$
$y=a_0+\frac{a_1}{x}$	 	$X=\frac{1}{x},$ $Y=y$	$\begin{cases} na_0+a_1\sum_{i=1}^{n}\frac{1}{x_i}=\sum_{i=1}^{n}y_i \\ a_0\sum_{i=1}^{n}\frac{1}{x_i}+a_1\sum_{i=1}^{n}\frac{1}{x_i^2}=\sum_{i=1}^{n}\frac{y_i}{x_i} \end{cases}$
$y=a_0+a_1\ln x$		$X=\ln x,$ $Y=y$	$\begin{cases} na_0+a_1\sum_{i=1}^{n}\ln x_i=\sum_{i=1}^{n}y_i \\ a_0\sum_{i=1}^{n}\ln x_i+a_1\sum_{i=1}^{n}(\ln x_i)^2=\sum_{i=1}^{n}y_i\ln x_i \end{cases}$

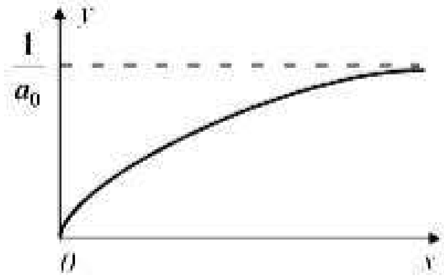
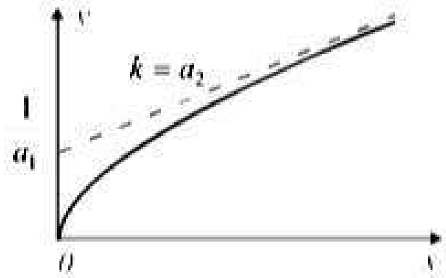
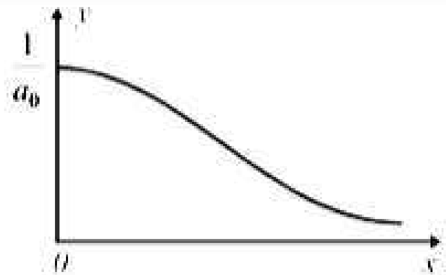
$\ln y = a_0 + a_1 x$		$X = x,$ $Y = \ln y$	$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \ln y_i \\ a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i \ln y_i \end{cases}$
$\ln y = a_0 + \frac{a_1}{x}$		$X = \frac{1}{x},$ $Y = \ln y$	$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} \frac{1}{x_i} = \sum_{i=1}^{n} \ln y_i \\ a_0 \sum_{i=1}^{n} \frac{1}{x_i} + a_1 \sum_{i=1}^{n} \frac{1}{x_i^2} = \sum_{i=1}^{n} \frac{\ln y_i}{x_i} \end{cases}$
$y = a_0 x^{a_1}$		$X = \ln x,$ $Y = \ln y$	$\begin{cases} n \ln a_0 + a_1 \sum_{i=1}^{n} \ln x_i = \sum_{i=1}^{n} \ln y_i \\ \ln a_0 \sum_{i=1}^{n} \ln x_i + a_1 \sum_{i=1}^{n} (\ln x_i)^2 = \sum_{i=1}^{n} \ln y_i \ln x \end{cases}$

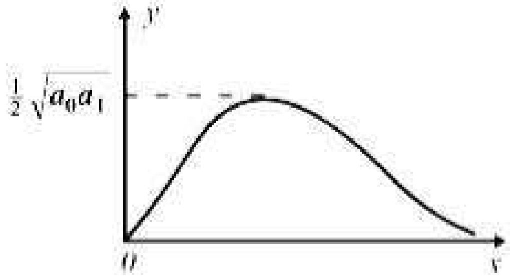
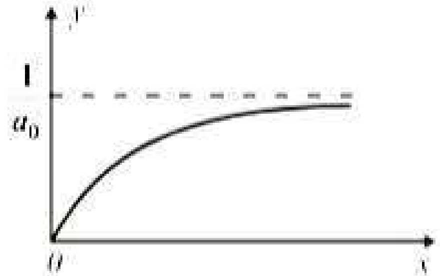
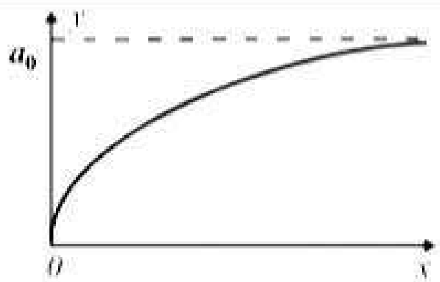
$y = a_0 \cdot a_1^{x}$		$X = x,$ $Y = \ln y$	$\begin{cases} n\ln a_0 + \ln a_1 \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} \ln y_i \\ \ln a_0 \sum\limits_{i=1}^{n} x_i + \ln a_1 \sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} x_i \ln y_i \end{cases}$
$y = a_0 \cdot x^{a_1} \cdot e^{a_2 x}$		$X_1 = x,$ $X_2 = \ln x,$ $Y = \ln y$	$\begin{cases} n\ln a_0 + a_1 \sum\limits_{i=1}^{n} x_i + a_2 \sum\limits_{i=1}^{n} \ln x_i = \sum\limits_{i=1}^{n} \ln y_i \\ \ln a_0 \sum\limits_{i=1}^{n} x_i + a_1 \sum\limits_{i=1}^{n} (x_i)^2 + a_2 \sum\limits_{i=1}^{n} x_i \ln x_i = \sum\limits_{i=1}^{n} x_i \ln y_i \\ \ln a_0 \sum \ln x_i + a_1 \sum\limits_{i=1}^{n} x_i \ln x_i + a_2 \sum\limits_{i=1}^{n} (\ln x_i)^2 = \sum\limits_{i=1}^{n} \ln x_i \ln \end{cases}$

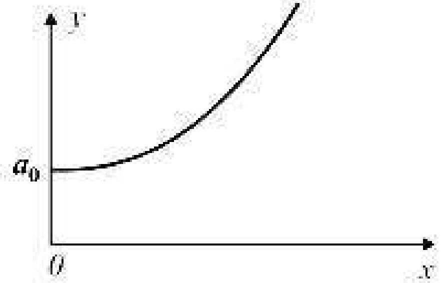
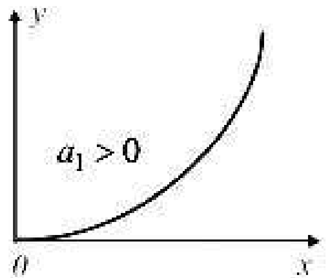
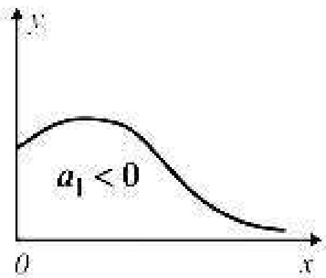
			
$y = a_0 e^{a_1 x}$		$X = x$, $Y = \ln y$	$\begin{cases} n \ln a_0 + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \ln y_i \\ \ln a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i \ln y_i \end{cases}$
$y = a_0 \cdot e^{\frac{a_1}{x}}$		$X = \frac{1}{x}$, $Y = \ln y$	$\begin{cases} n \ln a_0 + a_1 \sum_{i=1}^{n} \frac{1}{x_i} = \sum_{i=1}^{n} \ln y_i \\ \ln a_0 \sum_{i=1}^{n} \frac{1}{x_i} + a_1 \sum_{i=1}^{n} \frac{1}{x_i^2} = \sum_{i=1}^{n} \frac{\ln y_i}{x_i} \end{cases}$

$y = a_0 \cdot e^{a_1 x + a_2 x^2}$	 	$X_1 = x$, $X_2 = x^2$, $Y = \ln y$	$\begin{cases} n\ln a_0 + a_1\sum_{i=1}^{n} x_i + a_2\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} \ln y_i \\ \ln a_0\sum_{i=1}^{n} x_i + a_1\sum_{i=1}^{n} x_i^2 + a_2\sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} x_i \ln \\ \ln a_0\sum x_i^2 + a_1\sum_{i=1}^{n} x_i^3 + a_2\sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} x_i^2 \ln \end{cases}$
$y = \dfrac{1}{a_0 + a_1 e^{-x}}$		$X = e^{-x}$, $Y = \dfrac{1}{y}$	$\begin{cases} na_0 + a_1\sum_{i=1}^{n} e^{-x_i} = \sum_{i=1}^{n} \dfrac{1}{y_i} \\ a_0\sum_{i=1}^{n} e^{-x_i} + a_1\sum_{i=1}^{n} \left(e^{-x_i}\right)^2 = \sum_{i=1}^{n} \dfrac{e^{-x_i}}{y_i} \end{cases}$
$y = a_0 + a_1 \cdot e^{-x}$		$X = e^{-x}$, $Y = y$	$\begin{cases} na_0 + a_1\sum_{i=1}^{n} e^{\ X_i} = \sum_{i=1}^{n} y_i \\ a_0\sum_{i=1}^{n} e^{-X_i} + a_1\sum_{i=1}^{n} \left(e^{-X_i}\right)^2 = \sum_{i=1}^{n} y_i e^{-X_i} \end{cases}$

$y = \dfrac{x}{a_0 + a_1 e^{-x}}$		$X = e^{-x}$, $Y = \dfrac{x}{y}$	$\begin{cases} na_0 + a_1 \sum\limits_{i=1}^{n} e^{-x_i} = \sum\limits_{i=1}^{n} \dfrac{x_j}{y_i} \\ a_0 \sum\limits_{i=1}^{n} e^{-x_i} + a_1 \sum\limits_{i=1}^{n} \left(e^{-x_i}\right)^2 = \sum\limits_{i=1}^{n} \dfrac{x_i}{y_j} e^{-x_i} \end{cases}$
$y = \dfrac{1}{a_0 + a_1 x}$		$X = x$, $Y = \dfrac{1}{y}$	$\begin{cases} na_0 + a_1 \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} \dfrac{1}{y_i} \\ a_0 \sum\limits_{i=1}^{n} x_i + a_1 \sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} \dfrac{x_i}{y_i} \end{cases}$
$y = \dfrac{1}{a_0 + a_1 x} + a_2$		$X = x$, $Y = \dfrac{1}{y - a_2}$, a_2 підбирається, $\eta(a_2) \to \min$	$\begin{cases} na_0 + a_1 \sum\limits_{i=1}^{n} x_i = \sum\limits_{i=1}^{n} \dfrac{1}{y_i - a_2} \\ a_0 \sum\limits_{i=1}^{n} x_i + a_1 \sum\limits_{i=1}^{n} x_i^2 = \sum\limits_{i=1}^{n} \dfrac{x_i}{y_i - a_2} \end{cases}$

$y = \dfrac{x}{a_0 + a_1 x}$		$X = x$, $Y = \dfrac{x}{y}$	$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \dfrac{x_i}{y_i} \\ a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} \dfrac{x_i^2}{y_i} \end{cases}$
$y = \dfrac{x}{a_0 + a_1 x} + a_2 x$		$X = x$, $Y = \dfrac{x}{y - a_2 x}$, a_2 підбирається, $\eta(a_2) \to \min$	$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \dfrac{x_i}{y_i - a_2 x_i} \\ a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} \dfrac{x_i^2}{y_i - a_2 x_i} \end{cases}$
$y = \dfrac{1}{a_1 x^2 + a_0}$		$X = x^2$, $Y = \dfrac{1}{y}$	$\begin{cases} na_0 + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} \dfrac{1}{y_i} \\ a_0 \sum_{i=1}^{n} x_i^2 + a_1 \sum_{i=1}^{n} x_i^4 = \sum_{i=1}^{n} \dfrac{x_i^2}{y_i} \end{cases}$

$y=\dfrac{x}{a_1x^2+a_0}$		$X=x^2$ $Y=\dfrac{x}{y}$	$\begin{cases} na_0+a_1\sum\limits_{i=1}^{n}x_i^2=\sum\limits_{i=1}^{n}\dfrac{x_i}{y_i} \\ a_0\sum\limits_{i=1}^{n}x_i^2+a_1\sum\limits_{i=1}^{n}x_i^4=\sum\limits_{i=1}^{n}\dfrac{x_i^3}{y_i} \end{cases}$
$y=\dfrac{1}{a_0+\dfrac{a_1}{x}}$		$X=\dfrac{1}{x}$, $Y=\dfrac{1}{y}$	$\begin{cases} na_0+a_1\sum\limits_{i=1}^{n}\dfrac{1}{x_i}=\sum\limits_{i=1}^{n}\dfrac{1}{y_i} \\ a_0\sum\limits_{i=1}^{n}\dfrac{1}{x_i}+a_1\sum\limits_{i=1}^{n}\dfrac{1}{x_i^2}=\sum\limits_{i=1}^{n}\dfrac{1}{y_ix_i} \end{cases}$
$y=\dfrac{a_0x}{x+a_1}$			$a_0=\dfrac{1}{n}\sum\limits_{i=1}^{n}y_i-a_1\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i$, $a_1=\dfrac{\sum\limits_{i=1}^{n}x_iy_i-\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i\sum\limits_{i=1}^{n}y_i}{\sum\limits_{i=1}^{n}(x_i^2-nx_i^2)}$

$y = a_0 + a_1 e^x$		$X = e^x$ $Y = y$,	$\begin{cases} na_0 + a_1\sum\limits_{i=1}^{n} e^{x_i} = \sum\limits_{i=1}^{n} y_i \\ a_0\sum\limits_{i=1}^{n} e^{x_i} + a_1\sum\limits_{i=1}^{n}\left(e^{x_i}\right) = \sum\limits_{i=1}^{n} y_i e^{x_i} \end{cases}$
$y = a_0 x^{a_1 x}$		$X = x\ln x$, $Y = \ln y$	$\begin{cases} n\ln a_0 + a_1\sum\limits_{i=1}^{n} x_i \ln x_i = \sum\limits_{i=1}^{n} \ln y_i \\ a_0\sum\limits_{i=1}^{n} x_i \ln x_i + a_1\sum\limits_{i=1}^{n}(x_i \ln x_i)^2 = \sum\limits_{i=1}^{n} x_i \ln x_i \ln \end{cases}$

Таблиця F – розподілу для ймовірності $p = 0,95$

	Число ступенів свободи $k_1 = m$												
k_2	1	2	3	4	5	6	7	8	9	10	11	12	13
1	162	200	216	225	230	234	237	239	241	242	243	244	245
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,20	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26
9	5,12	4,26	3,68	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,22
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,20
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20	2,18
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18	2,15
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,12
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13	2,10
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	2,09
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,08
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89	1,86
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88	1,84
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,90	1,86	1,83
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85	1,82
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,87	1,83	1,80
140	3,91	3,06	2,67	2,44	2,28	2,16	2,08	2,01	1,95	1,90	1,86	1,82	1,79
160	3,90	3,05	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,81	1,78
180	3,89	3,05	2,65	2,42	2,26	2,15	2,06	1,99	1,93	1,88	1,84	1,81	1,77
200	3,88	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80	1,77
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,79	1,75	1,72

Таблиця F – розподілу для ймовірності $p = 0{,}95$ (продовження)

Число ступенів свободи $k_1 = m$													
k_2	14	15	16	17	18	19	20	30	40	50	100	200	∞
1	245	246	246	247	247	248	248	250	251	252	253	254	254
2	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
3	8,71	8,70	8,69	8,68	8,67	8,67	8,66	8,62	8,60	8,58	8,55	8,54	8,53
4	5,87	5,86	5,84	5,83	5,82	5,81	5,80	5,75	5,72	5,70	5,66	5,65	5,63
5	4,64	4,62	4,60	4,59	4,58	4,57	4,56	4,50	4,46	4,44	4,41	4,39	4,37
6	3,96	3,94	3,92	3,91	3,90	3,88	3,87	3,81	3,77	3,75	3,71	3,69	3,67
7	3,53	3,51	3,49	3,48	3,47	3,46	3,44	3,38	3,34	3,32	3,27	3,25	3,23
8	3,24	3,22	3,20	3,19	3,17	3,16	3,15	3,08	3,04	3,02	2,97	2,95	2,93
9	3,03	3,01	2,99	2,97	2,96	2,95	2,94	2,86	2,83	2,80	2,76	2,73	2,71
10	2,86	2,85	2,83	2,81	2,80	2,79	2,77	2,70	2,66	2,64	2,59	2,56	2,54
11	2,74	2,72	2,70	2,68	2,67	2,66	2,65	2,57	2,53	2,51	2,46	2,43	2,40
12	2,64	2,62	2,60	2,58	2,57	2,56	2,54	2,47	2,43	2,40	2,36	2,32	2,30
13	2,55	2,53	2,51	2,50	2,48	2,47	2,46	2,38	2,34	2,31	2,26	2,23	2,21
14	2,48	2,46	2,44	2,43	2,41	2,40	2,39	2,31	2,27	2,24	2,19	2,16	2,13
15	2,42	2,40	2,38	2,37	2,35	2,34	2,33	2,25	2,20	2,18	2,12	2,10	2,07
16	2,37	2,35	2,33	2,32	2,30	2,29	2,28	2,19	2,15	2,12	2,07	2,04	2,01
17	2,33	2,31	2,29	2,27	2,26	2,24	2,23	2,15	2,10	2,08	2,02	1,99	1,96
18	2,29	2,27	2,25	2,23	2,22	2,20	2,19	2,11	2,10	2,08	1,98	1,95	1,92
19	2,26	2,23	2,21	2,20	2,18	2,17	2,16	2,07	2,03	2,00	1,94	1,91	1,88
20	2,22	2,20	2,18	2,17	2,15	2,14	2,12	2,04	1,99	1,97	1,91	1,88	1,84
21	2,20	2,18	2,16	2,14	2,12	2,11	2,10	2,01	1,96	1,94	1,88	1,84	1,81
22	2,17	2,15	2,13	2,11	2,10	2,08	2,07	1,98	1,94	1,91	1,85	1,82	1,78
23	2,15	2,13	2,11	2,09	2,08	2,06	2,05	1,96	1,91	1,88	1,82	1,79	1,76
24	2,13	2,11	2,09	2,07	2,05	2,04	2,03	1,94	1,89	1,86	1,80	1,77	1,73
25	2,11	2,09	2,07	2,05	2,04	2,02	2,01	1,92	1,87	1,84	1,78	1,75	1,71
26	2,09	2,07	2,05	2,03	2,02	2,00	1,99	1,90	1,85	1,82	1,76	1,73	1,69
27	2,08	2,06	2,04	2,02	2,00	1,99	1,97	1,88	1,84	1,81	1,74	1,71	1,67
28	2,06	2,04	2,02	2,00	1,99	1,97	1,96	1,87	1,82	1,79	1,73	1,69	1,65
29	2,05	2,03	2,01	1,99	1,97	1,96	1,94	1,85	1,81	1,77	1,71	1,67	1,64
30	2,04	2,01	1,99	1,98	1,96	1,95	1,93	1,84	1,79	1,76	1,70	1,66	1,62
40	1,95	1,92	1,90	1,89	1,87	1,85	1,84	1,74	1,69	1,66	1,63	1,60	1,51
50	1,89	1,87	1,85	1,83	1,81	1,80	1,78	1,69	1,63	1,60	1,59	1,55	1,44
60	1,86	1,84	1,82	1,80	1,78	1,76	1,75	1,65	1,59	1,56	1,48	1,44	1,39
70	1,84	1,81	1,79	1,77	1,75	1,74	1,72	1,62	1,57	1,53	1,45	1,40	1,35
80	1,82	1,79	1,77	1,75	1,73	1,72	1,70	1,60	1,54	1,51	1,43	1,38	1,32
90	1,80	1,78	1,76	1,74	1,72	1,70	1,69	1,59	1,53	1,49	1,41	1,36	1,30
100	1,79	1,77	1,75	1,73	1,71	1,69	1,68	1,57	1,52	1,48	1,39	1,34	1,28
120	1,78	1,75	1,73	1,71	1,69	1,67	1,66	1,55	1,50	1,46	1,37	1,32	1,25
140	1,76	1,74	1,72	1,70	1,68	1,66	1,65	1,54	1,48	1,44	1,35	1,30	1,23
160	1,75	1,73	1,71	1,69	1,67	1,65	1,64	1,53	1,47	1,43	1,34	1,28	1,21
180	1,75	1,72	1,70	1,68	1,66	1,64	1,63	1,52	1,46	1,42	1,33	1,27	1,20
200	1,74	1,72	1,69	1,67	1,66	1,64	1,62	1,52	1,46	1,41	1,32	1,26	1,19
∞	1,69	1,67	1,64	1,62	1,60	1,59	1,57	1,46	1,39	1,35	1,24	1,17	1,00

Критичне значення t (критерій Стьюдента)

Число ступенів свободи k	Рівні значущості			Число ступенів свободи k	Рівні значущості		
	0,05	0,01	0,005		0,05	0,01	0,005
1	12,71	63,66	127,32	9	2,26	3,25	3,69
2	4,30	9,92	14,09	10	2,23	3,17	3,58
3	3,18	5,84	7,45	20	2,09	2,05	3,15
4	2,78	4,60	5,60	30	2,04	2,75	3,03
5	2,57	4,03	4,77	40	2,02	2,70	3,97
6	2,45	3,71	4,32	60	2,00	2,66	2,91
7	2,36	3,50	4,03	120	1,98	2,62	2,86
8	2,31	3,36	3,83	∞	1,96	2,58	2,81

Значення χ^2 в залежності від числа ступенів свободи k та довірчої ймовірності p

k	α			k	α		
	0,05	0,01	0,001		0,05	0,01	0,001
1	3,84	6,63	10,83	16	26,30	32,00	39,25
2	5,99	9,21	13,81	17	27,59	33,41	40,79
3	7,81	11,34	16,27	18	28,87	34,80	42,31
4	9,49	13,28	18,46	19	30,14	36,19	43,82
5	11,07	15,09	20,52	20	31,41	37,57	45,31
6	12,59	16,81	22,46	21	32,67	38,93	46,80
7	14,07	18,47	24,32	22	33,92	40,29	48,27
8	15,51	20,09	26,12	23	35,17	41,63	49,73
9	16,92	21,67	27,88	24	36,41	42,98	51,18
10	18,31	23,21	29,59	25	37,65	44,31	52,62
11	19,67	24,72	31,26	26	38,88	45,64	54,05
12	21,03	26,22	32,91	27	40,11	46,96	55,48
13	22,37	27,69	34,53	28	41,34	48,28	56,89
14	23,68	29,14	36,12	29	42,56	49,59	58,30
15	25,00	30,58	37,70	30	43,77	50,89	59,70

Критичні значення циклічного коефіцієнта автокореляції r^0

Кількість спостережень	Додатня автокореляція		Від'ємна автокореляція	
	$\alpha = 0{,}05$	$\alpha = 0{,}01$	$\alpha = 0{,}05$	$\alpha = 0{,}01$
10	0,360	0,525	0,564	0,705
15	0,328	0,475	0,462	0,597
20	0,299	0,432	0,399	0,524
25	0,276	0,398	0,356	0,473
30	0,257	0,370	0,325	0,433

Критичні значення d_l і d_u для коефіцієнта автокореляції

по критерію Дарбіна-Уотсона для $p = 0{,}95$

Число спостережень	Число факторів									
	$m=1$		$m=2$		$m=3$		$m=4$		$m=5$	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
6	0,61	1,40								
7	0,70	1,36	0,47	1,90						
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13	0,30	2,59		
10	0,88	1,32	0,70	1,64	0,53	2,02	0,38	2,41	0,24	2,82
11	0,93	1,32	0,76	1,60	0,60	1,93	0,44	2,28	0,32	2,65
12	0,97	1,33	0,81	1,58	0,66	1,86	0,51	2,18	0,38	2,51
13	1,01	1,34	0,86	1,56	0,72	1,82	0,57	2,09	0,45	2,39
14	1,05	1,35	0,91	1,55	0,77	1,78	0,63	2,03	0,51	2,30
15	1,08	1,36	0,95	1,54	0,81	1,75	0,69	1,98	0,56	2,22
16	1,11	1,37	0,98	1,54	0,86	1,73	0,73	1,94	0,62	2,16
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,66	2,10
18	1,16	1,39	1,05	1,54	0,93	1,70	0,82	1,87	0,71	2,05
19	1,18	1,40	1,07	1,54	0,97	1,69	0,86	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,89	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,95
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,65	1,04	1,77	0,95	1,89
26	1,30	1,46	1,23	1,55	1,14	1,65	1,06	1,76	0,98	1,87
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,75	1,00	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,37	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,53	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,52	1,74	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,58	1,72	1,55	1,75	1,53	1,77
90	1,64	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,65	1,69	1,62	1,71	1,60	1,73	1,58	1,76	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78
150	1,72	1,75	1,71	1,76	1,69	1,77	1,68	1,79	1,67	1,80
200	1,76	1,78	1,75	1,79	1,74	1,80	1,73	1,81	1,72	1,82

Критичні значення d_l і d_u для коефіцієнта автокореляції по критерію Дарбіна-Уотсона для $p = 0.95$ (продовження)

Число спосте-режень	Число факторів									
	$m = 6$		$m = 7$		$m = 8$		$m = 9$		$m = 10$	
	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u	d_l	d_u
11	0.12	2.89								
12	0,16	2,67	0.11	3.06						
13	0,21	2,49	0.14	2.64	0.09	3.18				
14	0,26	2,35	0.18	2.67	0.12	2.96	0.08	3,29		
15	0,30	2.24	0.23	2.53	0.16	2.82	0.11	3,10	0,07	3,37
16	0.35	2.15	0.27	2.42	0.20	2.68	0.14	2.94	0,09	3,20
17	0,39	2,08	0.31	2.31	0.24	2.57	0.18	2,81	0,13	3,05
18	0,44	2,02	0.37	2.24	0.28	2.47	0.22	2,70	0,16	2,93
19	0,48	1,96	0.40	2.17	0.32	2.36	0.26	2,60	0,20	2,81
20	0,52	1,92	0.44	2.11	0.36	2.31	0.29	2,51	0,23	2,71
21	0.55	1.88	0.47	2.06	0.40	2.24	0.33	2,43	0,27	2,63
22	0.57	1.85	0.51	2.02	0.44	2.19	0.37	2,37	0,30	2,55
23	0.62	1.82	0.55	1.98	0.47	2.14	0.40	2,31	0,34	2,48
24	0,65	1,79	0,58	1.94	0.51	2.10	0.44	2,26	0,38	2,42
25	0,68	1,78	0.61	1.92	0.54	2.06	0,47	2,21	0,41	2,37
26	0.71	1.76	0.64	1.89	0.57	2.03	0.51	2.17	0,44	2,31
27	0.74	1.74	0.67	1.87	0.60	2.00	0.54	2.13	0,47	2,27
28	0.76	1.73	0.70	1.85	0.66	1.97	0.57	2.10	0.50	2,23
29	0.79	1.72	0.72	1.83	0.69	1.95	0.60	2.07	0.53	2,19
30	0,81	1,71	0,76	1,82	0.69	1.93	0.62	2.04	0.56	2.16
35	0,91	1.67	0.86	1.76	0.80	1.85	0.74	1.94	0,69	2,04
40	1.00	1.65	0.95	1.72	0.90	1.80	0.84	1.88	0,79	1,96
45	1,07	1.64	1.02	1.71	0.97	1.77	0.93	1.83	0,86	1,90
50	1.12	1.64	1.06	1.69	1.04	1.75	1.00	1.81	0.96	1,86
55	1.17	1.64	1.13	1.69	1.10	1.73	1.06	1.79	1.02	1,84
60	1.21	1.64	1.18	1.68	1.14	1.72	1.11	1.77	1.07	1,82
65	1.25	1.64	1.22	1.68	1.19	1.72	1.15	1.76	1.12	1,80
70	1.28	1.65	1.25	1.68	1.22	1.72	1.19	1.75	1.16	1,79
75	1.31	1.65	1.28	1.68	1.26	1.71	1.23	1.75	1.20	1,79
80	1.34	1.65	1.31	1.68	1.29	1.71	1.26	1.75	1.23	1,78
85	1.36	1.66	1.34	1.69	1.31	1.71	1.29	1.74	1.26	1.77
90	1.38	1.66	1.36	1.69	1.34	1.71	1.31	1.74	1,29	1,77
95	1.40	1.67	1.38	1.69	1.36	1.72	1.34	1.74	1.31	1,77
100	1.42	1.67	1.40	1.69	1.38	1.72	1.36	1.74	1.34	1,77
150	1.54	1.71	1.53	1.72	1.52	1.74	1.50	1.75	1.49	1,77
200	1,61	1.74	1.60	1.75	1.59	1.76	1.58	1.77	1,57	1,78

Критичні точки розподілу Неймана

n	Величина Q для $r > 0$		Величина Q для $r < 0$	
	$\alpha = 0{,}01$	$\alpha = 0{,}05$	$\alpha = 0{,}05$	$\alpha = 0{,}01$
4	0,8341	1,0406	4,2927	4,4992
5	0,6724	1,0255	3,9745	4,3276
6	0,6738	1,0682	3,7318	4,1262
7	0,7163	1,0919	3,5748	3,9504
8	0,7575	1,1228	3,4486	3,8139
9	0,7974	1,1524	3,3476	3,7025
10	0,8353	1,1803	3,2642	3,6091
11	0,8706	1,2062	3,1938	3,5294
12	0,9033	1,2301	3,1335	3,4603
13	0,9336	1,2521	3,0812	3,3996
14	0,9618	1,2725	3,0352	3,3458
15	0,9880	1,2914	2,9943	3,2977
16	1,0124	1,3090	2,9577	3,2543
17	1,0352	1,3253	2,9247	3,2148
18	1,0556	1,3405	2,8948	3,1787
19	1,0766	1,3547	2,8675	3,1456
20	1,0954	1,3680	2,8425	3,1151
21	1,1131	1,3805	2,8195	3,0869
22	1,1298	1,3923	2,7982	3,0607
23	1,1456	1,4035	2,7784	3,0362
24	1,1606	1,4141	2,7599	3,0133
25	1,1748	1,4241	2,7426	2,9919
26	1,1883	1,4336	2,7264	2,9718
27	1,2012	1,4426	2,7112	2,9528
28	1,2135	1,4512	2,6969	2,9348
29	1,2252	1,4594	2,6834	2,9177
30	1,2363	1,4672	2,6707	2,9016
35	1,2940	1,5014	2,6163	2,8324
40	1,3266	1,5304	2,5722	2,7760
45	1,3620	1,5552	2,5357	2,7289
50	1,3907	1,5716	2,5064	2,6908
55	1,4156	1,5923	2,4819	2,6585
60	1,4384	1,6082	2,4596	2,6294

При $r > 0$ – критична область для $Q_{стат} < Q_{табл}$;

при $r < 0$ – критична область для $Q_{стат} > Q_{табл}$.

ЛІТЕРАТУРА

1. Вайну Я. Я. Кореляция рядов динамики / Я. Я. Вайну. – М.: Статистика, 1977. – 119 с.

2. Винн Р. Введение в прикладной эконометрический анализ / Р. Винн, К. Холден. – М.: Финансы и статистика, 1981. – 294 с.

3. Грубер И. Эконометрия / Й. Грубер. – К.: Астарта, 1996. – 397 с.

4. Гренджер К. Спектральный анализ временных рядов в экономике / К. Гренджер, М. Хатанака. – М.: Статистика; Пер. с англ., 1972. – 312 с.

5. Джонсон Дж. Эконометрические методы / Дж. Джонсон. – М.; Статистика, 1980. – 444с.

6. Диха М.В. Концептуальні засади макроекономічного моделювання соціально-економічних процесів / М.В. Диха // Вісник Хмельницького національного університету. – 2012. – № 6, Т. 1. – С. 215–223.

7. Королев Ю. Г. Метод наименших квадратов в социально-экономических исследованиях / Ю. Г. Королев. – М.: Статистика, 1980. – 112 с.

8. Крамер Г. Математические методы статистики / Г. Крамер. – М.: Изд. иностр. лит.; Пер. с англ., 1948. – 631 с.

9. Корольов О. А. Економетрика: Навч. посіб., 2-ге вид., випр. та скор. / О. А. Корольов. – К.: Книга, 2005. – 416 с.

10. Кулінич О. І. Економетрика: Навчальний посібник / О. І. Кулінич. – Хмельницький: Видавництво «Поділля», 2003. – 215 с.

11. Лук'яненко І. Г. Економетрика: Підручник / І. Г. Лук'яненко, Л. І. Краснікова. – К.: Тов. «Знання», КОО, 1998. – 494 с.

12. Маленво Э. Статистические методы в эконометрии / Э. Маленво. – М.: Статистика, Вып. 1, 1975. – 423 с.

13. Маленво Э. Статистические методы в эконометрии / Э. Маленво. – М.: Статистика, Вып. 2, 1976. – 326 с.

14. Миллс Ф. Статистические методы / Ф. Миллс. – М.: Госстатиздат, Пер. с англ., 1958. – 799 с.

15. Мороз В. С. Економетрика: Навчальний посібник / В. С. Мороз, В. В. Мороз. – Хмельницький: ТУП, 2000. – 166 с.

16. Мороз В. С. Економетрика. Програми, завдання та методичні вказівки до виконання самостійної і контрольної робіт для студентів денної і заочної форм навчання економічних спеціальностей / В. С. Мороз. – Хмельницький: ТУП, 2000. – 130 с.

17. Наконечний С. І. Економетрія: Навчальний посібник / С. І. Наконечний, Т. О. Терещенко, Т. П. Романюк. – К.: КНЕУ, 1997. – 352 с.

18. Наконечний С.І. Економетрика: Підручник. Вид. 2-ге, допов. та перероб. / С. І. Наконечний, Т. О. Терещенко, Т. П. Романюк. – К.: КНЕУ, 2000. – 236 с.

19. Себер Дж. Линейный регрессионный аналіз / Дж. Себер. – М.: Мир, Пер. с англ., 1980. – 456 с.

20. Тинтнер Г. Введение в эконометрику / Г. Тинтнер. – М.: Статистика, Пер. с нем., 1965. – 361 с.

21. Толбатов Ю. А. Загальна теорія статистики засобами Excel. Навчальний посібник / Ю. А. Толбатов. – К.: Четверта хвиля, 1999. – 224 с.

22. Толбатов Ю. А. Економетрика: Підручник для екон. спец. вищ. навч. закл. / Ю. А. Толбатов. – К.: Четверта хвиля, 1997. – 320 с.

23. Фишер Р. Проблема идентификации в эконометрии / Р. Фишер. – М.: Госстатиздат, Пер. с англ., 1978. – 223 с.

24. Farrar D. E., Glouber R. R. Multicollinerity in regresion analysis; the problem revisited. – «Review of Economics and Statistic», 1967. – Vol. 49, N1. – P. 92–107.

НАВЧАЛЬНЕ ВИДАННЯ

Марія Василівна ДИХА
Віктор Степанович МОРОЗ

ЕКОНОМЕТРІЯ

НАВЧАЛЬНИЙ ПОСІБНИК

Підписано до друку 04.11.2021 р. Формат 60x84 1/16.
Друк лазерний. Папір офсетний. Гарнітура Times New Roman.
Обл. вид. арк. 17,80. Тираж 100 прим.

ТОВ «Видавництво «Центр учбової літератури»
вул. Електриків, 23 м. Київ 04176
тел./факс 044-425-01-34
тел.: 044-425-20-63; 425-04-47; 451-65-95
800-501-68-00 (безкоштовно в межах України)

Свідоцтво про внесення суб'єкта видавничої справи до державного реєстру видавців, виготівників і розповсюджувачів видавничої продукції
ДК № 4162 від 21.09.2011 р.

Диха Марія Василівна – доктор економічних наук, професор кафедри економіки підприємства і підприємництва Хмельницького національного університету.
Автор понад 100 опублікованих наукових праць, трьох одноосібних монографій.
Коло наукових інтересів: проблеми соціально-економічного розвитку України, забезпечення конкурентоспроможності та інвестиційної привабливості національної економіки, інтеграції України у світовий економічний простір, моделювання та прогнозування економічних явищ та процесів.

Мороз Віктор Степанович – кандидат економічних наук, доцент кафедри економіки підприємства і підприємництва Хмельницького національного університету, член національної спілки краєзнавців України.
Автор понад 90 публікацій, навчальних посібників з економетрії та організації виробництва.
Коло наукових інтересів: економіко-математичне моделювання та прогнозування, організація та управління діяльністю підприємства.

WWW.CUL.COM.UA