

МІЖРЕГІОНАЛЬНА
АКАДЕМІЯ УПРАВЛІННЯ ПЕРСОНАЛОМ



МАУП

**О. Л. Лещинський, В. В. Рязанцева,
О. О. Юнькова**

ЕКОНОМЕТРІЯ

*Рекомендовано
Міністерством освіти і науки України
як навчальний посібник для студентів
вищих навчальних закладів*

2-ге видання, стереотипне

Київ
ДП «Видавничий дім «Персонал»
2008

ББК 65в6я73
Л54

Рецензенти: *О. А. Корольов*, д-р екон. наук, проф.
Ю. В. Крак, д-р фіз.-мат. наук, проф.

Схвалено Вченою радою Міжрегіональної Академії управління персоналом (протокол № 8 від 28.11.02)

Рекомендовано Міністерством освіти і науки України (лист № 14/18.2-1513 від 22.09.03)

Лещинський О. Л.

Л54 Економетрія: Навч. посіб. для студ. вищ. навч. закл. / О. Л. Лещинський, В. В. Рязанцева, О. О. Юнькова. — 2-ге вид., стереотип. — К.: ДП «Вид. дім «Персонал», 2008. — 208 с.: іл. — Бібліогр.: с. 203–205.

ISBN 978-966-608-951-2

У навчальному посібнику “Економетрія” розглядаються основні завдання економетричних досліджень і методи їх розв’язання, визначається місце та значення цієї наукової дисципліни, її зв’язок з економікою, статистикою та математикою. Основним методом оцінювання параметрів регресійних моделей обрано метод найменших квадратів (МНК). Розглянуто особливості оцінювання параметрів у разі порушення основних передумов застосування класичного МНК. Описано альтернативні методи оцінювання параметрів моделей (метод головних компонентів, узагальнений МНК тощо). Моделювання динаміки економічних процесів розглянуто на прикладах дистрибутивно-лагових і авторегресійних моделей. Для моделювання економічних процесів з прямими та зворотними зв’язками використано системи одночасних рівнянь. Описано способи дослідження та застосування моделей з якісними незалежними змінними.

Для студентів економічних спеціальностей, які вивчали вищу математику, лінійну алгебру, теорію ймовірностей та математичну статистику.

ББК 65в6я73

- © О. Л. Лещинський, В. В. Рязанцева, О. О. Юнькова, 2003
- © О. Л. Лещинський, В. В. Рязанцева, О. О. Юнькова, 2008, стереотип.
- © Міжрегіональна Академія управління персоналом (МАУП), 2008
- © ДП «Видавничий дім «Персонал», 2008

ISBN 978-966-608-951-2

ВСТУП

Відтоді як економіка стала серйозною самостійною наукою, дослідники намагаються спрогнозувати ту чи іншу ситуацію, передбачити майбутні значення економічних показників, запропонувати інструменти зміни ситуації в бажаному напрямку. Політики або керуючі виробництвом, обираючи одну з можливих стратегій, отримують певний результат. Поганий він чи гарний і чи можна було досягти кращого результату, перевірити дуже важко. Економічна ситуація практично ніколи не повторюється в точності, отже, неможливо застосувати дві стратегії за тих самих умов з метою порівняння кінцевого результату. Тому одним із основних завдань економічного аналізу є моделювання розвитку економічних явищ і процесів при створенні тих чи інших умов. Зрозумівши глибинні рушійні сили досліджуваного процесу, можна навчитися раціонально керувати ним.

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

Критерієм істини для будь-якої теорії є практика. Зокрема, практика економічної діяльності відображається у статистичній інформації. Поєднання економічної теорії з практичними результатами є наріжним каменем економетрії.

Економетрія як наукова дисципліна, її зв'язок з іншими економічними дисциплінами

Економетрія — це порівняно новий напрямок економічної науки, що утворився від поєднання теоретичної економіки, математики та статистики.

Слово “економетрія” (у деяких джерелах “економетрика”) буквально означає “вимірювання в економіці”, що дає підстави під цим терміном розуміти все, що пов’язано з вимірюваннями в економіці. Однак таке тлумачення надзвичайно широке і не відображає особливостей цієї галузі знань. З іншого боку, через необхідність застосування математико-статистичних методів інколи економетрії дають вужче тлумачення, а саме розглядають її лише як певний набір математико-статистичних засобів, якими кількісно досліджують взаємозв’язки певних рядів статистичних даних. Тому точнішим є таке визначення [1]:

Економетрія — це самостійна наукова дисципліна, яка об’єднує сукупність теоретичних результатів, засобів, прийомів, методів і моделей, призначених для того, щоб на базі економічної теорії, економічної статистики та математико-статистичного інструментарію надавати конкретних кількісних значень загальним (якісним) закономірностям, обґрунтованим економічною теорією.

Стосовно даного визначення слід мати на увазі, що завдання економічної теорії в межах економетрії полягають не лише в тому, щоб виявляти закони та зв’язки, які об’єктивно існують в економіці, а й описувати їх математичними методами. Економічна статистика акумулює всю інформацію про економічні процеси, що відбуваються в реальній економіці, та уособлює той практичний досвід, який має підтвердити чи спростувати відповідні економічні теорії. А під математико-статистичним інструментарієм розуміють не всю математичну статистику, а лише окремі її розділи: лінійні моделі регресійного аналізу, аналіз часових рядів, побудову та аналіз систем одночасних рівнянь, перевірку статистичних гіпотез.

Саме “приземлення” економічної теорії на базу конкретної економічної статистики та отримання за допомогою відповідних математичних методів кількісних взаємозв’язків між економічними показниками є сутністю економетрії.

Зазначені в такий спосіб ключові моменти у визначенні економетрії забезпечують її розмежування з такими дисциплінами, як математична економіка, описова економічна статистика та математична статистика. Математична економіка — це математично сформульована економічна теорія, що вивчає зв’язки між економічними змінними на загальному (некількісному) рівні. Вона стає економетрією, коли символічно подані в рівняннях коефіцієнти замінюють конкретними числовими оцінками, отриманими на базі відповідних статистичних даних (даних описової статистики) методами математичної статистики.

Отже, *економетрія* — це прикладна економіко-математична дисципліна, яка вивчає методи кількісного вимірювання взаємозв'язків між економічними показниками та напрямки їх застосування в економічних дослідженнях і практичній економічній діяльності [17].

Об'єкт, предмет, мета і завдання економетрії

Об'єктом економетрії є економічні системи та простори різного рівня складності: від окремого підприємства чи фірми до економіки галузей, регіонів, держави й світу загалом.

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

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

Основне завдання економетрії — оцінити параметри моделей з урахуванням особливостей вхідної економічної інформації, перевірити відповідність моделей досліджуваному явищу і спрогнозувати розвиток економічних процесів.

Основні етапи економетричного аналізу

Процес економетричного моделювання складається з таких кроків:

- 1) вибір конкретної форми аналітичної залежності між економічними показниками (специфікація моделі) на підставі відповідної економічної теорії;
- 2) збирання та підготовка статистичної інформації;
- 3) оцінювання параметрів моделей;
- 4) перевірка адекватності моделі та достовірності її параметрів;
- 5) застосування моделі для прогнозування розвитку економічних процесів з метою подальшого керування ними.

Економічні задачі, які розв'язують за допомогою економетричних методів

Застосування різноманітних економетричних моделей на різних рівнях економічної діяльності дає змогу розв'язувати економічні проблеми різного рівня складності.

На рівні макроекономіки економетричними засобами досліджують закономірності у виробництві, розподілі, перерозподілі та кінцевому використанні валового внутрішнього продукту, у яких суттєву роль відіграють державний бюджет, податкова політика, страхування, кредит, ощадна справа. Узгодженість усіх галузей фінансово-кредитної системи визначає ефективність розподільчих відносин, збалансованість доходів і витрат у народному господарстві, забезпечення процесів відтворення грошових ресурсів, фінансової захищеності державного, колективного та особистого майна від інфляції та інших негативних явищ.

На мікрорівні економетричні дослідження передбачають наукове обґрунтування управлінських рішень, що приймаються на підприємствах різних форм власності й мають ураховувати постійний вплив зовнішнього середовища.

Моделі можуть використовуватися для аналізу економічних і соціально-економічних показників, що характеризують відповідну економічну систему, для прогнозування їх подальшого змінювання або для імітації можливих сценаріїв соціально-економічного розвитку досліджуваної системи за умови, що деякі показники можна змінювати цілеспрямовано.

За рівнем ієрархії виокремлюють: макрорівень (країна загалом), мезорівень (регіони, галузі, корпорації) та мікрорівень (сім'я, підприємство, фірма).

Засобами економетричного моделювання вивчають проблеми ринку, інвестицій, фінансової чи соціальної політики, ціноутворення, попиту та пропозиції тощо.

Особливого значення економетричні дослідження набувають в макроекономіці, де взаємозв'язки величин часто неочевидні та мінливі. Не виключені ситуації, коли модель раптом перестає "працювати" через появу або активізацію якогось фактора. Саме такі ситуації зумовлюють розвиток макроекономічної теорії. З іншого боку, саме економетричний аналіз дає змогу обґрунтувати та уточнити форму залежностей в макроекономічних моделях, краще зрозуміти механізми взаємозв'язку макроекономічних показників.

Отже, поєднуючи в собі економічну теорію та математико-статистичні методи, економетричне моделювання широко застосовується при прийнятті практичних рішень в економічній діяльності (у бізнесі, банківській справі, прогнозуванні, державному регулюванні економіки), а також є потужною базою для отримання нових знань з економіки.

Місце курсу серед дисциплін фундаментальної підготовки бакалаврів з економічних спеціальностей

Економетрія — одна з основних дисциплін у підготовці бакалаврів з економічних спеціальностей. Вона будується на основі математичних та економічних знань.

Методи економетрії є найсучаснішими засобами аналізу та дослідження різних соціально-економічних систем. За допомогою економетричних методів можна відхилити деякі економічні гіпотези або показати неможливість застосування їх у конкретних умовах. Хоча засоби економетрії не дають змоги довести теоретичні твердження, але з допомогою її методів можна показати, що те чи інше твердження не суперечить даним спостережень. Оволодівши елементарним інструментарієм економетрії, можна обгрунтовано прогнозувати розвиток цих систем, оцінювати вплив рішень чи урядових постанов щодо зміни цін, податків тощо на стан справ будь-якого підприємства, розробляти шляхи ефективного керування ними, приймати ефективні управлінські рішення.

Мета вивчення курсу “Економетрія” — навчитися аналізувати інформаційні потоки в соціально-економічних системах, прогнозувати їх поведінку, оцінювати та будувати економетричні моделі різного рівня.

Вивчення курсу передбачає відповідну математичну та економічну підготовку. Проте для того, щоб ознайомитися з проблемами, які вивчає економетрія і з якими стикаються ті, хто використовує економетричні методи, не потрібно бути спеціалістом з усіх розділів математики та економіки. Знання певних розділів математики, зокрема основ лінійної алгебри, теорії матриць, теорії ймовірностей, математичної статистики та основ економіки, можуть виявитися достатніми для вивчення курсу економетрії.

Структура курсу

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

У цьому курсі вивчається здебільшого матеріал, що входить до першої частини, тобто економетричні методи. Їх можна умовно розбити на *чотири групи*. *Перша група* — це методи оцінювання параметрів класичної економетричної моделі, насамперед метод наймен-

ших квадратів (МНК). *Друга група* — це методи оцінювання параметрів узагальненої моделі, коли порушуються деякі передумови використання методу найменших квадратів. До *третьої групи* входять методи оцінювання параметрів динамічних економетричних моделей. *Четверта група* охоплює методи оцінювання параметрів економетричних моделей, які побудовані на основі системи одночасних структурних рівнянь.

Коротка історична довідка

На початку ХХ ст. у деяких країнах були спроби скласти так звані “барометри розвитку”. Найвідоміший з них “гарвардський барометр”, за допомогою якого в 20-ті роки намагалися передбачити поведінку товарного і грошового ринку.

Гарвардська школа вважалася на той час центром економічних досліджень. Тут уперше почали системно вивчати ряди економічних показників з урахуванням взаємозв'язку між ними і на основі цих показників досліджувати тенденції та цикли економічних процесів. Криза 1929–1933 рр. змусила критично переглянути методи аналізу, які застосовувалися на той час в економіці.

Лише після того як в економічних дослідженнях почали враховувати випадкові аспекти економічних явищ, стало можливим формування економетрії як галузі економічної науки.

Сучасні методи математичної статистики почали застосовувати в біології. Наприкінці ХІХ ст. англійський біолог К. Пірсон досліджував криві розподілу деяких числових показників людського організму. Пізніше він та його школа почали вивчати кореляції в біології та будувати лінійні регресії. Підходи, запропоновані біологами, були застосовані в економіці. У 1897 р. з'явилася праця В. Парето, у якій досліджувалися доходи населення в різних країнах. У ній вперше була застосована так звана крива Парето, параметри якої було отримано статистичними методами.

На початку ХХ ст. вийшло кілька праць англійського статистика Гукера, у яких за допомогою кореляційно-регресійних методів, започаткованих школою Пірсона, вивчалися взаємозалежності між економічними показниками, зокрема вплив банкрутств на товарній біржі на ціну зерна. Пізніше з'явилося багато праць як з розвитку теорії математичної статистики та її прикладних елементів, так і з практичного застосування цих методів в економічному аналізі. Насамперед,

праці Мура, які вийшли друком протягом 1914–1917 рр. У 1928 р. було опубліковано дослідження Ч. Кобба і П. Дугласа про виробничу функцію, яка ввійшла в економетрію як класичний приклад і досі є важливим інструментом економетричного аналізу. Саме ці праці заклали підвалини сучасної економетрії.

Економетрія як окрема галузь науки відома під такою назвою лише з 1930 р. Саме тоді було засновано економетричне товариство, яке визначало себе так: “Міжнародне товариство для розвитку економічної теорії і її зв'язку зі статистикою та математикою”. Зауважимо, що термін “економетрія” вперше запровадив львівський учений П. Чомпа, опублікувавши у Львові в 1910 р. книгу “Нариси економетрії і природної теорії бухгалтерії, яка ґрунтується на політичній економії”. Однак це поняття не набуло поширення, оскільки на той час не було фундаментальних праць у цій галузі науки.

Засновниками економетрії вважають Р. Фріша, Е. Шумпетера, Я. Тінбергена — послідовників неокласичної економічної школи і кейнсіанства. Вони одними з перших цілеспрямовано намагалися поєднати економічну теорію з математичними та статистичними методами. Спочатку вчені обмежувалися вивченням деяких моделей попиту і пропозиції. Лише після Другої світової війни вони почали вивчати комплексні економетричні моделі на макрорівні, у яких основна увага приділялася попиту, фінансовому стану й податкам, прибутку, цінам тощо. Основним внеском цих учених в економетричну науку є розробка економетричних моделей прийняття рішень, за яку в 1969 р. Р. Фріш та Я. Тінберген були відзначені Нобелівською премією.

Пізніше Нобелівську премію в галузі економіки отримали Т. К. Купманс (1975) за розробку лінійних економетричних моделей і розвиток статистичних методів у економетрії, Л. Р. Клейн (1980) за розробку складних економетричних моделей та їх застосування для аналізу кон'юнктурних коливань і економічної політики, Т. Хаавелмо (1989) за розробку та застосування теоретико-імовірнісних методів, особливо для аналізу взаємозалежних економетричних структур.

Значний внесок у розробку економічних моделей зробили також нобелівські лауреати В. Леонт'єв (1973), якому належать розробки в галузі балансових моделей для моделювання взаємозв'язків з великою кількістю змінних, Л. Канторович (1975), який досліджував виробничі моделі, Г. Дебрю (1983), який працював у галузі математики економічної теорії. Хоча їхні праці безпосередньо не пов'язані з економетричними дослідженнями, усе ж вони значною мірою впли-

нули на подальший розвиток не лише економетрії, а й економічної науки загалом. У 2000 р. Дж. Хекман і Д. Макфеден відзначені Нобелівською премією за розробку мікроеконометрії та методів статистичного аналізу.

Контрольні запитання

1. У чому полягає основний зміст проблематики економетрії?
2. Як пов'язані між собою математична економіка, описова статистика та математична статистика?
3. Визначте основне завдання економетричного дослідження.
4. Розкрийте зміст основних етапів економетричного аналізу.

МАУП

Розділ 1. Математичне моделювання як метод наукового пізнання економічних явищ і процесів

1.1. Загальні принципи моделювання в економіці

1.1.1. Поняття математичної моделі

При вивченні складних економічних процесів та явищ часто застосовується моделювання. Модель — це спеціально створений об'єкт, на якому відтворюються певні характеристики досліджуваного явища, а моделювання — це конкретне відтворення цих характеристик, що дає змогу вивчати можливу поведінку явища без проведення експериментів над ним.

Моделювання є важливим інструментом наукової абстракції, що допомагає виокремити, уособити та проаналізувати суттєві для даного об'єкта характеристики (властивості, взаємозв'язки, структурні та функціональні параметри).

Для економіки, де неможливе будь-яке експериментування, особливого значення набуває математичне моделювання. Завдяки застосуванню потужного математичного апарату воно є найефективнішим і найдосконалішим методом. У свою чергу, математичні методи не можуть застосовуватися безпосередньо щодо дійсності, а лише щодо математичних моделей того чи іншого кола явищ.

Прикладами економічних моделей є моделі споживчого вибору, моделі фірми, моделі економічного зростання, моделі рівноваги на товарних, факторних і фінансових ринках тощо.

Поведінка й значення будь-якого економічного показника залежать практично від безлічі факторів, усі їх урахувати нереально. Але в цьому й немає потреби. Звичайно лише обмежена кількість фак-

торів насправді істотно впливає на досліджуваний економічний показник. Вплив інших факторів настільки незначний, що їх ігнорування не може призвести до істотних відхилень у поведінці досліджуваного об'єкта. Виокремлення й урахування в моделі лише обмеженої кількості реально домінуючих факторів і є важливою передумовою якісного аналізу, прогнозування й керування ситуацією.

Математична модель, аби бути ефективним інструментом вивчення економічних процесів, насамперед має відповідати таким вимогам: будуватися на основі економічної теорії й відбивати об'єктивні закономірності процесів;

правильно відтворювати функцію та (чи) структуру реальної економічної системи;

відповідати певним математичним умовам (мати розв'язок, узгоджені розмірності тощо).

Природно, результати досліджень будь-якої моделі можуть мати практичну цінність, якщо модель адекватна явищу, що вивчається, тобто досить добре відтворює реальну ситуацію.

1.1.2. Етапи побудови економічної моделі

Процес побудови моделі складається з таких етапів:

- 1) формулюються предмет і мета дослідження;
- 2) у досліджуваній економічній системі виокремлюються структурні чи функціональні елементи, що відповідають поставленій меті, визначаються найважливіші якісні характеристики цих елементів;
- 3) словесно, якісно описуються взаємозв'язки між елементами моделі;
- 4) вводяться символічні позначення для відповідних характеристик економічного об'єкта та формалізуються, наскільки можливо, взаємозв'язки між ними, тим самим формалізується (описується мовою математики) математична модель;
- 5) виконуються розрахунки за математичною моделлю та аналізуються отримані результати.

Зауважимо, що різні за природою економічні явища можуть мати однаковий математичний вираз, хоча економічна інтерпретація моделі та результати розрахунків будуть різними.

За визначенням, будь-яка економічна модель є абстрактною, а отже, неповною. Це пов'язано з тим, що для виокремлення закономірностей функціонування економічного об'єкта потрібно абстрагуватися

від інших факторів, які хоч і мають незначний вплив, однак у сукупності можуть визначати не лише відхилення в поведінці об'єкта, а й його поведінку. Звичайно вважають, що всі фактори, невраховані явно в моделі, мають незначний результуючий вплив на процес чи явище, що досліджується. Склад урахованих факторів і їх структура коригуються в процесі вдосконалення моделі.

1.1.3. Класифікація моделей

Математичні моделі, що використовуються в економіці, можна поділити на класи за рядом ознак. Залежно від особливостей об'єкта моделювання та застосованого математичного інструментарію виокремлюють такі моделі: макро- та мікроекономічні, теоретичні та прикладні, статичні та динамічні, детерміновані та стохастичні, оптимізаційні та моделі рівноваги тощо.

Макроекономічні моделі описують економіку загалом, пов'язуючи між собою узагальнені матеріальні та фінансові показники: ВВП, споживання, інвестиції, зайнятість, процентну ставку, кількість грошей тощо. Мікроекономічні моделі описують взаємодію структурних і функціональних складових економіки або поведінку окремої складової в ринковому середовищі. Завдяки різноманіттю типів економічних елементів і форм їх взаємодії на ринку мікроекономічне моделювання становить основну частину економіко-математичної теорії. Останніми роками найсуттєвіші теоретичні результати в мікроекономічному моделюванні отримано в процесі дослідження стратегічної поведінки фірм в умовах олігополії.

Теоретичні моделі дають змогу вивчати загальні властивості економіки та її характерних елементів і отримувати нові результати на підставі формальних припущень. За допомогою прикладних моделей можна оцінити певні економічні показники, надати їм конкретних значень виходячи з відповідної статистичної інформації.

У статичних моделях описується стан економічного об'єкта в певний момент чи період часу, а динамічні моделі вивчають взаємозв'язки економічних змінних у часі. Змінні, що вивчаються в динаміці, у статичних моделях мають фіксоване значення. Однак динамічна модель не зводиться до простої суми статичних моделей, а описує взаємодію сил, що рухають економіку.

Детерміновані моделі передбачають жорсткі функціональні зв'язки між змінними моделі, а стохастичні — припускають наявність випадкових впливів на досліджувані показники.

У моделюванні ринкової економіки важливе місце належить моделям рівноваги. Вони описують такий стан економіки, коли всі сили, що намагаються вивести її з рівноваги, мають нульову сумарну дію. Оптимізаційні моделі найчастіше застосовують на мікрорівні: вони дають змогу визначати найкращі рішення в умовах обмежених можливостей.

Предметом економетричного дослідження є прикладні стохастичні економічні моделі, тобто загальні економічні моделі, у яких модельні коефіцієнти набувають конкретних числових значень залежно від використаної статистичної інформації.

1.2. Кореляційно-регресійний аналіз в економіці

У багатьох задачах потрібно встановити та оцінити залежність деякого економічного показника від одного чи кількох інших показників. Очевидно, будь-які економічні показники, зазвичай, перебувають під впливом випадкових факторів, а тому з математичної точки зору інтерпретуються як випадкові величини.

З теорії ймовірностей відомо, що випадкові величини можуть бути пов'язані функціональною чи статистичною залежністю або ж узагалі бути незалежними. Звичайно, співвідношення між незалежними змінними тут не розглядаються. Строга функціональна залежність реалізується в економіці рідко. Частіше спостерігається так звана статистична залежність.

Нагадаємо, що *статистичною* називають залежність, коли зі змінюванням однієї випадкової величини змінюється закон розподілу ймовірностей іншої. Зокрема, статистична залежність виявляється в тому, що зі змінюванням однієї величини змінюється середнє значення іншої. Така залежність називається *кореляційною*.

Наприклад, у землеробстві з однакових за площею ділянок землі при рівних кількостях внесених добрив збирають різний врожай. Звичайно, немає строгої функціональної залежності між урожайністю землі та кількістю внесених добрив. Це пояснюється впливом випадкових факторів (опаді, температура повітря, розташування ділянки тощо). Водночас, як показує досвід, середній врожай залежить від кількості внесених добрив, тобто зазначені показники, напевне, пов'язані кореляційною залежністю.

Можна зазначити два типи взаємозв'язку змінних. В одному випадку невідомо, яка зі змінних незалежна, а яка — залежна, тобто вони

рівноправні й зв'язок можна розглядати як в один, так і в інший бік. У другому випадку змінні нерівноправні, тобто змінювання лише однієї з них впливає на змінювання іншої, а не навпаки. У цьому разі при розгляді зв'язку між двома змінними величинами важливо встановити на основі логічного міркування, яка з ознак є причиною, а яка — наслідком. Наприклад, урожайність залежить від родючості землі, а не навпаки, тобто економічна оцінка землі є незалежною змінною, а врожайність — залежною.

Варто мати на увазі, що статистичний аналіз залежностей сам по собі не розкриває сутності причинних зв'язків між явищами, тобто він не вирішує питання, з яких причин одна змінна впливає на іншу. Розв'язок такої задачі є результатом якісного (змістовного) вивчення зв'язків, що обов'язково має або передувати статистичному аналізу, або супроводжувати його.

Нехай з певних економічних міркувань встановлено, що деякий економічний показник x є причиною змінювання іншого показника y . Статистичні дані по кожному з показників інтерпретуються як деякі реалізації випадкових величин X і Y . Як відомо з курсу теорії ймовірностей, математичним сподіванням випадкової величини називається її середнє (арифметичне чи зважене) значення. А залежність середнього значення від іншої випадкової величини зображується за допомогою умовного математичного сподівання.

Кореляційну залежність між ними або залежність в середньому в загальному випадку можна подати у вигляді співвідношення

$$M(Y | x) = f(x), \quad (1.1)$$

де $M(Y | x)$ — умовне математичне сподівання.

Функція $f(x)$ називається *функцією регресії* Y на X . При цьому X називається *незалежною (пояснюючою) змінною (регресором)*, Y — *залежною (пояснюваною) змінною (регресандом)*. Розглядаючи залежність двох випадкових величин, говорять про *парну регресію*.

Залежність Y від кількох змінних, що описується функцією

$$M(Y | x_1, x_2, \dots, x_m) = F(x_1, x_2, \dots, x_m), \quad (1.2)$$

називають *множинною регресією*.

Термін “регресія” (рух назад, повернення до попереднього стану) увів Френсіс Галтон наприкінці XIX ст., проаналізувавши залежність

між зростом батьків і зростом дітей. Він помітив, що зріст дітей у дуже високих батьків у середньому менший, ніж середній зріст батьків. У дуже низьких батьків, навпаки, середній зріст дітей вищий. В обох випадках середній зріст дітей прямує (повертається) до середнього зросту людей у даному регіоні. Звідси й вибір терміна, що відбиває таку залежність.

Однак реальні значення залежної змінної не завжди збігаються з її умовним математичним сподіванням, тому аналітична залежність (у вигляді функції $y = f(x)$) має бути доповнена випадковою складовою u , що, власне, і вказує на стохастичну сутність залежності.

Означення 1.1. Зв'язки між залежною та незалежною (незалежними) змінними, що описуються співвідношеннями

$$y = f(x) + u, \quad (1.3)$$

$$y = F(x_1, x_2, \dots, x_m) + u, \quad (1.4)$$

називають *регресійними рівняннями (моделями)*.

Виникає питання про причини обов'язкової присутності в регресійних моделях випадкового фактора (відхилення). Серед таких причин виокремимо найістотніші.

1. *Уведення в модель не всіх пояснюючих змінних.* Будь-яка регресійна (зокрема, економетрична) модель — це спрощення реальної ситуації. Остання завжди є складною композицією різних факторів, багато з яких у моделі не враховуються, що призводить до відхилення реальних значень залежної змінної від її модельних значень. Наприклад, попит на товар визначається його ціною, цінами на товари-замінники, на товари, що його доповнюють, прибутком споживачів, їхніми смаками, уподобаннями тощо. Безумовно, перелічити всі пояснюючі змінні практично неможливо. Зокрема, неможливо врахувати такі фактори, як традиції, національні чи релігійні особливості, географічне положення району, погоду та багато інших, вплив яких призводить до деяких відхилень реальних спостережень від модельних. Ці відхилення можуть бути описані як випадкова складова моделі.

У деяких випадках заздалегідь невідомо, які фактори за умов, що склалися, насправді є визначальними, а якими можна знехтувати. Крім того, інколи безпосередньо врахувати якийсь фактор неможливо через відсутність статистичних даних. Наприклад, обсяг заощаджень домогосподарств може визначатися не лише прибутками їх членів, а й станом здоров'я останніх, інформація про яке в цивілі-

зованих країнах становить лікарську таємницю. У деяких ситуаціях ряд факторів має принципово випадковий характер, що додає неоднозначності певним моделям, наприклад погода в моделях, що прогнозують обсяг врожаю.

2. *Неправильний вибір функціональної форми моделі.* Через слабку вивченість досліджуваного процесу або через його мінливість може бути неправильно дібрано функцію, що його моделює. Це, безумовно, спричинить відхилення моделі від реальності, що позначиться на величині випадкової складової. Наприклад, виробнича функція (Y) одного фактора (X) може моделюватися функцією $Y = a + bX$, хоча мала б використовуватися інша модель: $Y = aX^b$ ($0 < b < 1$), що враховує закон спадної ефективності. Крім того, неправильним може бути добір пояснюючих змінних.

3. *Агрегування змінних.* У багатьох моделях розглядаються залежності між факторами, що самі є складною комбінацією інших, простіших змінних. Наприклад, при вивченні сукупного попиту аналізується залежність, у якій пояснювана змінна (сукупний попит) є складною композицією індивідуальних попитів, що також може виявитися причиною відхилення реальних значень від модельних.

4. *Помилки вимірювань.* Якою б якісною не була модель, помилки вимірювання змінних впливатимуть на розбіжності між модельними та емпіричними даними, що також позначиться на величині випадкового члена.

5. *Обмеженість статистичних даних.* Найчастіше будуються моделі, що описуються неперервними функціями. А для оцінювання параметрів моделі використовується набір даних, що має дискретну структуру. Ця невідповідність знаходить відображення у випадковому відхиленні.

6. *Непередбачуваність людського фактора.* Ця причина може “зіпсувати” найякіснішу модель. Дійсно, при правильному виборі форми моделі, скрупульозному доборі пояснюючих змінних неможливо спрогнозувати поведінку кожного індивідуума.

Сукупність методів, за допомогою яких досліджуються та узагальнюються взаємозв'язки кореляційно пов'язаних змінних, називається кореляційно-регресійним аналізом.

Зазначеними методами розв'язують дві основні задачі:

1) знаходження загальної закономірності, що характеризує залежність двох (чи більше) кореляційно пов'язаних змінних, тобто розробка математичної моделі зв'язку (задача регресійного аналізу);

2) визначення тісноти зв'язку (задача кореляційного аналізу).

Здебільшого процедура аналізу зв'язку між змінними дає змогу встановити його природу, тобто визначити форму залежності між змінними.

Побудова якісного рівняння регресії, що відповідає емпіричним даним і цілям досліджень, є досить складним процесом. Його можна поділити на *три етапи*:

- 1) вибір форми рівняння регресії;
- 2) визначення параметрів обраного рівняння;
- 3) аналіз якості рівняння та перевірка адекватності рівняння емпіричним даним, удосконалення рівняння.

Вибір форми зв'язку змінних називається *специфікацією* моделі регресії.

У випадку парної регресії вибір формули звичайно здійснюється за графічним зображенням реальних статистичних даних у вигляді точок у декартовій системі координат, що називається *кореляційним полем* (діаграмою розсіювання) (рис. 1.1).

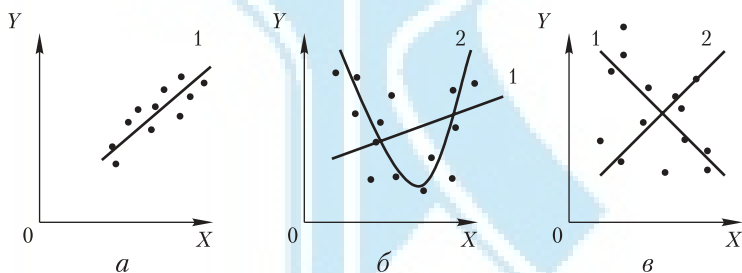


Рис. 1.1

На рис. 1.1 проілюстровано три ситуації.

На графіку 1.1, *а* взаємозв'язок між *X* і *Y* близький до лінійного, і пряма 1 досить добре узгоджується з емпіричними точками. Тому щоб описати залежність між *X* і *Y*, доцільно вибрати лінійну функцію $Y = b_0 + b_1X$.

На графіку 1.1, *б* реальний взаємозв'язок між *X* і *Y*, найімовірніше, описується квадратичною функцією $Y = aX^2 + bX + c$ (лінія 2).

На графіку 1.1, *в* явний взаємозв'язок між *X* і *Y* відсутній. Тому щоб краще вибрати форму зв'язку, необхідно, можливо, збільшити

кількість спостережень – точок кореляційного поля або скористати-ся іншими способами вимірювання показників.

У випадку множинної регресії визначити форми залежності ще складніше.

Якщо природа зв'язку невідома, то співвідношення між показниками описують за допомогою наближених спрощених форм залежностей, насамперед лінійних.

Наприклад, Кейнс запропонував лінійну формулу залежності індивідуального споживання C від доходу Y : $C = c_0 + bY$, де $c_0 > 0$ – величина автономного споживання; b – гранична схильність до споживання, $0 < b < 1$.

Однак поки не обчислено кількісні значення коефіцієнтів c_0 і b й не перевірено надійність отриманих результатів, зазначена формула залишається лише гіпотезою.

1.3. Економетрична модель та її елементи

Економетрична модель – це логічний (звичайно математичний) опис того, що економічна теорія вважає особливо важливим при дослідженні певної проблеми.

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

Отже, сформулюємо таке означення економетричної моделі.

Означення 1.2. Економетрична модель – це функція чи система функцій, що описує кореляційно-регресійний зв'язок між економічними показниками, причому залежно від причинних зв'язків між ними один чи кілька із цих показників розглядаються як залежні змінні, а інші – як незалежні.

У загальному випадку рівняння в економетричній моделі має вигляд

$$Y = f(x_1, x_2, \dots, x_m, u),$$

де Y – результат, або залежна змінна, змінування якої описує дане рівняння; x_1, x_2, \dots, x_m – фактори, або незалежні змінні, що визначають поведінку Y . Змінна u містить ту частину руху Y , що не пояснюється змінними x_1, x_2, \dots, x_m , і має випадковий характер. Символ f відображує аналітичний вид зв'язку між досліджуваними змінними.

Означення 1.3. Процес опису явища чи процесу, тобто вибір аналітичної форми моделі, називається специфікацією моделі. Іншими словами, *специфікація моделі* – це аналітична форма залежності між економічними показниками.

Незалежні змінні x_1, x_2, \dots, x_m , що задані заздалегідь чи за межами моделі, називаються екзогенними змінними (регресорами). Залежна змінна Y , що визначається як розв'язок рівняння, називається ендогенною змінною (регресандом). Функція f у кожному конкретному випадку окрім змінних x_1, x_2, \dots, x_m і u містить ще щонайменше деякі коефіцієнти, що поєднують змінні у певних співвідношеннях і визначають структуру рівняння. Ці коефіцієнти називаються параметрами моделі.

Означення 1.4. Визначення значень коефіцієнтів (параметрів) обраної форми статистичного зв'язку змінних на підставі відповідних статистичних даних називається *параметризацією* рівняння регресії або *оцінюванням параметрів*.

Існує відмінність між змінними та параметрами моделі. Змінні – це економічні величини, що можуть набувати певних значень з деякої множини допустимих величин. Параметри – це сталі коефіцієнти. Хоча вони не завжди відомі, та все ж у будь-якій ситуації вони мають фіксоване значення. Параметри можна назвати “незмінними” (інколи відомими, інколи невідомими), що пов'язують змінні в рівняннях. Ці рівняння, а отже, і параметри визначають структуру моделі: вони вказують на характер припустимих співвідношень між змінними.

Параметри чимось подібні до незалежних (заданих ззовні) змінних, однак між ними є важливі відмінності. Припускається, що параметри залишаються незмінними протягом усього періоду спостереження, а екзогенні змінні, безумовно, мають змінюватися з ча-

сом. Саме змінювання екзогенних змінних приводить модель у рух, зумовлює перехід системи до нового стану.

Зауважимо, що в багатьох економетричних моделях є такі екзогенні змінні, які можуть бути змінені керівними органами (державним регулюванням чи керівництвом фірми). Ці керовані змінні, наприклад державні витрати та податки, є політичними інструментами. Якщо відомо структуру економічного процесу, то державні органи, змінюючи значення таких змінних, могли б робити заданими ендогенні змінні, тобто впливати на подальший розвиток процесу.

Економетричні моделі можуть бути статичними та динамічними. У статичних моделях зв'язки розглядаються у фіксований момент часу і часові зміни в них ролі не відіграють. У динамічній моделі, навпаки, взаємозв'язки вивчаються в розвитку й час є необхідним фактором змін.

Моделі розрізняють також за рівнем агрегування змінних (мікрочи макроекономічні показники), за способом відображення змінних (у постійних чи поточних цінах, у абсолютних значеннях чи приростах показників), за кількістю змінних (одно- чи багатофакторні моделі), за кількістю рівнянь (одне чи кілька), за часом спостережень (річні, кварталні чи місячні дані).

Класифікують моделі також за призначенням та метою використання (аналітичні, імітаційні, прогностичні).

1.4. Статистична база економетричних досліджень

Будь-яке економетричне дослідження завжди поєднує теорію (математичні моделі) і практику (статистичні дані). За допомогою моделей описують і пояснюють процеси, що вивчаються, а статистичні дані використовують для побудови та обґрунтування моделей. Без конкретних кількісних даних, що характеризують функціонування економічного об'єкта, не завжди можна визначити практичну значущість певної моделі.

Економічні дані звичайно поділяють на два види: перехресні дані та часові ряди. Перехресними є дані за якимось економічним показником, що отримані для різних однотипних об'єктів (фірм, регіонів). Причому дані отримано в один і той самий момент часу або часова приналежність несуттєва. Часові ряди характеризують один і той

самий об'єкт, але в різні моменти часу. Наприклад, дані бюджетних досліджень населення в певний момент часу є перехресними, а динаміка рівня інфляції за певний період відображується часовими рядами. Послідовні значення часових рядів можуть бути пов'язані між собою певними залежностями: спостерігаються деякі закономірності у відхиленнях від загальної тенденції розвитку чи виявляються часові зсуви показників (часові лаги). Тому методи обробки таких даних дещо відрізняються від методів, що застосовуються для обробки перехресних даних.

Метою збирання статистичних даних є побудова інформаційної бази для прийняття рішень. Природно, що аналіз даних і прийняття рішень здійснюються на підставі деякої інтуїтивної (неявної) або кількісної (явної) економічної моделі. Тому збирають саме дані, що стосуються певної моделі. Їх можна отримати опитуванням, анкетуванням, інтерв'юванням або із джерел офіційної статистичної звітності. Кожний показник, отриманий одним із зазначених способів, називається спостереженням.

Будь-які економічні дані є кількісними характеристиками економічних об'єктів. Вони формуються під дією багатьох факторів, які не завжди можна проконтролювати ззовні. Неконтрольовані фактори можуть набувати випадкових значень з деякої множини допустимих значень і тим самим зумовлювати випадковість даних. Стохастична природа економічних даних вимагає застосування спеціальних адекватних їм статистичних методів для їх аналізу та обробки.

При підготовці статистичних даних для роботи з певною моделлю необхідно забезпечити відповідність цих даних моделі та спільну методичну базу для їх оцінювання. Дані мають утворювати взаємно узгоджений набір, тобто якщо вимірювання здійснюється в грошових одиницях, то це мають бути поточні або фіксовані (одного й того самого року) ціни. Реальним об'ємним показникам (тобто у фіксованих цінах) мають відповідати реальні відносні показники (наприклад, процентні ставки слід скоригувати відносно темпу інфляції). Залежно від поставлених завдань вибирають узагальнені показники: валовий внутрішній продукт, валові внутрішні збереження тощо. Відсутні статистичні дані здебільшого можуть бути розраховані за іншими показниками, якщо між ними існує певна функціональна залежність. Наприклад, інфляція розраховується за даними про дефлятор, і навпаки.

Отже, формуючи сукупність спостережень, слід забезпечити порівнянність даних у просторі та часі. Це означає, що дані вхідної сукупності повинні мати:

- однаковий ступінь агрегування;
- однорідну структуру одиниць сукупності;
- одні й ті самі методи розрахунку показників у часі чи просторі;
- однакову періодичність обліку окремих змінних;
- порівнянні ціни та однакові інші зовнішні економічні умови.

Висновки, які можна зробити в результаті економетричного моделювання, цілком зумовлені якістю вхідних даних, а саме їх повнотою та достовірністю.

1.5. Особливості математичного моделювання економічних систем

В економіко-математичному аналізі інформація формується, як правило, у результаті спостереження за об'єктом дослідження. При отримванні, оцінюванні та використанні цієї інформації слід мати на увазі важливі специфічні риси джерела даних.

Суттєве значення мають стохастичні (випадкові) фактори, які виявляються у впливі на економіку як з боку природи та суспільства, так і у внутрішньоекономічних зв'язках. Через складність і динамічність техніко-економічних, особливо соціально-економічних, процесів попередній розрахунок економічних показників можливий лише з певним рівнем довіри.

Водночас величезні масштаби економічної системи, розгалуженість зв'язків між її елементами та відома інерційність значною мірою зумовлюють майбутній її стан попереднім. Тому розвиток системи можна передбачити з великою мірою впевненості.

В означеній ситуації найприйнятнішими методами дослідження є методи математичної статистики, адаптовані до економічних явищ. Саме ці методи дають змогу будувати економетричні моделі та оцінювати їх параметри, перевіряти гіпотези стосовно властивостей економічних показників і форм зв'язку між ними. Однак особливість економетричного підходу до моделювання економічних об'єктів полягає не у використанні економічної термінології, а насамперед у детальному дослідженні відповідності вибраної моделі явищу, що вивчається, а також в аналізі якості статистичної інформації, що є основою параметризації (оцінювання параметрів) моделей.

Контрольні запитання

1. Що таке математична модель економічного об'єкта?
2. Назвіть типи математичних моделей і відмінності між ними.
3. До якого типу математичних моделей належить економетрична модель?
4. Поясніть сутність кореляційних зв'язків між економічними показниками.
5. З яких елементів складається економетрична модель?
6. Які вимоги висуваються до статистичної бази економетричних досліджень?

МАУП

Розділ 2. МОДЕЛІ ПАРНОЇ РЕГРЕСІЇ ТА ЇХ ДОСЛІДЖЕННЯ

2.1. Приклади парних зв'язків в економіці

Економічна теорія виявила й дослідила значну кількість сталих і стабільних зв'язків між різними показниками. Наприклад, добре вивчено залежності споживання від рівня доходу, попиту — від цін на товари, залежність між процентною ставкою та інвестиціями, обмінним курсом валюти та обсягом чистого експорту, між рівнями безробіття та інфляції, залежність обсягу виробництва від окремих факторів (розміру основних фондів, їх віку, підготовки персоналу тощо); залежність між продуктивністю праці та рівнем механізації, а також багато інших залежностей.

Здебільшого залежність між показниками можна відобразити за допомогою лінійних співвідношень.

Наприклад, для моделювання залежності індивідуального споживання C від наявного прибутку Y Кейнс запропонував лінійне рівняння

$$C = c_0 + bY,$$

де c_0 — величина автономного споживання; b — гранична схильність до споживання ($0 < b \leq 1$).

Однак припущення щодо лінійної залежності між певними показниками економічного явища чи процесу може не підтверджуватися даними спостережень цих показників. І це природно, оскільки в деяких випадках залежність є суттєво нелінійною. Наприклад, залежність між рівнем безробіття x і рівнем інфляції y відображається так званою кривою Філіпса:

$$y = \frac{a}{x - b},$$

де $a > 0$, $b > 0$ – параметри моделі, а змінні x і y вимірюються у процентах.

При незмінній річній дисконтній (обліковій) ставці r і початковому внеску a через x років у банку наявна сума грошей обчислюватиметься за формулою

$$y = a(1+r)^x,$$

де a , y – параметри моделі.

При маркетингових і ринкових дослідженнях, при дослідженні збуту продукції та в демографії застосовують так звану криву Гомперця:

$$y = e^{ab^x+c},$$

де параметри a та c можуть набувати будь-яких значень, а b перебуває в таких межах: $0 < b < 1$.

Зв'язок між обсягом виробленої продукції y та основними виробничими ресурсами, а саме обсягом витраченого капіталу C і обсягом витрат праці L , також має нелінійний характер:

$$y = dC^a, \quad y = cL^b,$$

a , b , c , d – числові параметри; c , $d > 0$, a , $b \geq 0$.

Нелінійні зв'язки, як правило, певними перетвореннями (заміною змінних чи логарифмуванням) зводять до лінійного вигляду або апроксимують (наближують) лінійними функціями.

Отже, модель лінійної регресії (лінійне рівняння) є найпоширенішим (і найпростішим) видом залежності між економічними змінними. Крім того, побудоване лінійне рівняння може слугувати початковою точкою в разі складних (суттєво нелінійних) залежностей.

2.2. Лінійна модель з двома змінними

У загальному випадку *парна лінійна регресія* є лінійною функцією між залежною змінною Y і однією пояснюючою змінною X :

$$Y = a_0 + a_1X. \tag{2.1}$$

Співвідношення (2.1) називається *теоретичною лінійною регресійною моделлю*; a_0 і a_1 – *теоретичні параметри (теоретичні коефіцієнти) регресії*.

Зазначимо, що принциповою в цьому разі є лінійність за параметрами a_0 і a_1 рівняння (2.1).

Щоб визначити значення теоретичних коефіцієнтів регресії, необхідно знати й використовувати всі значення змінних X і Y генеральної сукупності, що практично неможливо. Тому за вибіркою обмеженого обсягу будують так зване *емпіричне рівняння регресії*, у якому коефіцієнтами є оцінки теоретичних коефіцієнтів регресії:

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X, \quad (2.2)$$

де \hat{a}_0 і \hat{a}_1 – оцінки невідомих параметрів a_0 і a_1 .

Через розбіжність статистичної бази для генеральної сукупності та вибірки оцінки \hat{a}_0 і \hat{a}_1 практично завжди відрізняються від дійсних значень коефіцієнтів a_0 і a_1 , що призводить до розбіжності емпіричної та теоретичної ліній регресії. Різні вибірки з однієї й тієї самої генеральної сукупності звичайно зумовлюють різні оцінки.

Можливе співвідношення між теоретичним і емпіричним рівняннями регресії схематично зображено на рис. 2.1.

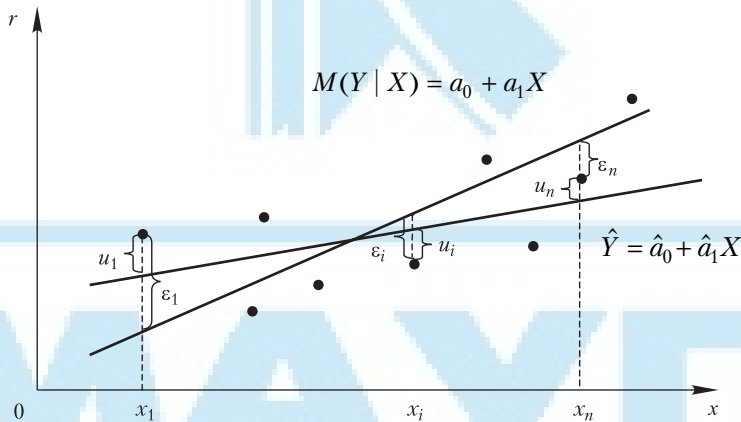


Рис. 2.1

Задачі лінійного регресійного аналізу полягають у тому, щоб за навчними статистичними даними (x_i, y_i) , $i = 1, 2, \dots, n$, для змінних X і Y :

а) отримати найкращі оцінки \hat{a}_0 , \hat{a}_1 невідомих параметрів a_0 і a_1 ;

б) перевірити статистичні гіпотези про параметри моделі;

в) перевірити, чи досить добре модель узгоджується зі статистичними даними (адекватність моделі даним спостережень).

Для відображення того факту, що кожне індивідуальне значення y_i відхиляється від відповідного умовного математичного сподівання, у модель вводять випадковий доданок u_i :

$$y_i = M(Y | X = x_i) + u_i = a_0 + a_1 x_i + u_i.$$

Отже, індивідуальні значення y_i подають у вигляді суми двох компонент — систематичної ($a_0 + a_1 x_i$) і випадкової (u_i). Причина появи останньої досить докладно розглядалася раніше.

Таким чином, регресійне рівняння набуває вигляду

$$Y = a_0 + a_1 X + u. \quad (2.3)$$

Завдання полягає в тому, щоб за конкретною вибіркою (x_i, y_i) , $i = 1, 2, \dots, n$, знайти такі значення оцінок невідомих параметрів a_0 і a_1 , щоб побудована лінія регресії була найкращою в певному розумінні серед усіх інших прямих. Іншими словами, побудована пряма має бути “найближчою” до точок спостережень за їх сукупністю.

Мірою якості знайдених оцінок можуть бути визначені композиції відхилень u_i , $i = 1, 2, \dots, n$. Наприклад, коефіцієнти a_0 і a_1 рівняння регресії можуть бути оцінені за умови мінімізації однієї з таких сум:

$$1) \sum_{i=1}^n u_i = \sum_{i=1}^n (y_i - \hat{y}_i) = \sum_{i=1}^n (y_i - \hat{a}_0 - \hat{a}_1 x_i); \quad (2.4)$$

$$2) \sum_{i=1}^n |u_i| = \sum_{i=1}^n |y_i - \hat{y}_i| = \sum_{i=1}^n |y_i - \hat{a}_0 - \hat{a}_1 x_i|; \quad (2.5)$$

$$3) \sum_{i=1}^n u_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - \hat{a}_0 - \hat{a}_1 x_i)^2. \quad (2.6)$$

Однак перша сума не може бути мірою якості знайдених оцінок через те, що існує безліч прямих (зокрема, $Y = \bar{y}$), для яких $\sum_{i=1}^n u_i = 0$ (доведення цього твердження вноситься як вправа).

Метод визначення оцінок коефіцієнтів за умови мінімізації другої суми називається *методом найменших модулів* (МНМ).

Найпоширенішим і теоретично обґрунтованим є метод визначення коефіцієнтів, при якому мінімізується третя сума. Він дістав назву *методу найменших квадратів* (МНК).

Останній метод оцінювання параметрів найпростіший з обчислювальної точки зору. Крім того, оцінки коефіцієнтів регресії, знайдені за МНК при визначених передумовах, мають ряд оптимальних властивостей (незміщеність, ефективність, обґрунтованість).

Серед інших методів визначення оцінок коефіцієнтів регресії виокремимо метод моментів (ММ) і метод максимальної правдоподібності (ММП).

2.3. Метод найменших квадратів

Нехай за вибіркою (x_i, y_i) , $i = 1, 2, \dots, n$, потрібно визначити оцінки \hat{a}_0 і \hat{a}_1 емпіричного рівняння регресії (2.2), тобто підібрати такі значення коефіцієнтів рівняння, щоб сума квадратів відхилень була мінімальною (рис. 2.2).

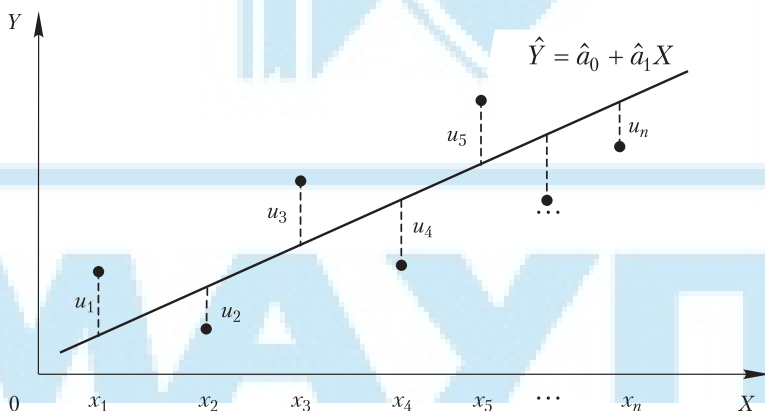


Рис. 2.2

У цьому разі $\sum_{i=1}^n u_i^2$ є квадратичною функцією двох параметрів \hat{a}_0

і \hat{a}_1 , оскільки $x_i, y_i (i = 1, 2, \dots, n)$ – відомі дані спостережень:

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

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

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

$$\begin{cases} \frac{\partial Q}{\partial \hat{a}_0} = -2 \sum (y_i - \hat{a}_0 - \hat{a}_1 x_i) = 0; \\ \frac{\partial Q}{\partial \hat{a}_1} = -2 \sum (y_i - \hat{a}_0 - \hat{a}_1 x_i) x_i = 0; \end{cases} \Rightarrow \quad (2.7)$$

$$\begin{cases} n \hat{a}_0 + \hat{a}_1 \sum x_i = \sum y_i; \\ \hat{a}_0 \sum x_i + \hat{a}_1 \sum x_i^2 = \sum x_i y_i. \end{cases} \quad (2.8)$$

Поділивши обидва рівняння системи (2.8) на n , отримаємо

$$\begin{cases} \hat{a}_0 + \hat{a}_1 \bar{x} = \bar{y}; \\ \hat{a}_0 \bar{x} + \hat{a}_1 \bar{x}^2 = \overline{xy} \end{cases} \Rightarrow \begin{cases} \hat{a}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - \bar{x}^2}; \\ \hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}. \end{cases} \quad (2.9)$$

Тут $\bar{x} = \frac{1}{n} \sum x_i$, $\bar{x}^2 = \frac{1}{n} \sum x_i^2$, $\bar{y} = \frac{1}{n} \sum y_i$, $\overline{xy} = \frac{1}{n} \sum x_i y_i$.

У наступних формулах для спрощення знаки сум ($\sum_{i=1}^n$) записуватимемо без індексів, допускаючи, що додавання виконується від $i=1$ до $i=n$. Також для змінних з індексом i розумітимемо, що $i = 1, 2, \dots, n$ (якщо не зазначено інше).

Отже, згідно з МНК оцінки параметрів \hat{a}_0 та \hat{a}_1 визначаються за формулами (2.9).

Неважко помітити, що \hat{a}_1 можна обчислити за формулою

$$\hat{a}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{S_{xy}}{S_x^2}, \quad (2.10)$$

де $S_{xy} = \text{cov}(x, y) = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})$ – вибірковий кореляційний момент випадкових величин X і Y ; $S_x^2 = \frac{1}{n} \sum (x_i - \bar{x})^2 = \frac{1}{n} \sum x_i^2 - \bar{x}^2$ – вибіркова дисперсія X ; $S_x = \sqrt{S_x^2}$ – стандартне відхилення X .

Тоді

$$\hat{a}_1 = \frac{S_{xy}}{S_x^2} = \frac{S_{xy}}{S_x S_y} \frac{S_y}{S_x} = r_{xy} \frac{S_y}{S_x}, \quad (2.11)$$

де r_{xy} – вибірковий коефіцієнт кореляції; S_y – стандартне відхилення Y . Отже, коефіцієнт регресії пропорційний коефіцієнту кореляції, а коефіцієнти пропорційності використовують для зіставлення різних величин X і Y .

Таким чином, якщо коефіцієнт кореляції r_{xy} уже розрахований, то за формулою (2.11) неважко знайти коефіцієнт \hat{a}_1 парної регресії.

Якщо окрім рівняння регресії Y на X ($\hat{Y} = \hat{a}_0 + \hat{a}_1 X$) для тих самих емпіричних даних знайдено рівняння регресії X на Y ($\hat{X} = \hat{b}_0 + \hat{b}_1 Y$), то добуток коефіцієнтів \hat{a}_1 та \hat{b}_1 дорівнює r_{xy}^2 :

$$\hat{a}_1 \cdot \hat{b}_1 = r_{xy} \frac{S_y}{S_x} \cdot r_{xy} \frac{S_x}{S_y} = r_{xy}^2.$$

Зазначимо, що коефіцієнти \hat{b}_0 і \hat{b}_1 обчислюються за формулами, аналогічними формулам (2.9):

$$\begin{cases} \hat{b}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{y^2 - \bar{y}^2}; \\ \hat{b}_0 = \bar{x} - \hat{b}_1 \bar{y}. \end{cases} \quad (2.12)$$

Властивості оцінок параметрів

Отримані результати, зокрема формули (2.9) і (2.12), дають змогу зробити ряд висновків.

1. Оцінки МНК є функціями від вибірки.
2. Оцінки МНК є точковими оцінками теоретичних коефіцієнтів регресії.
3. Відповідно до другої формули співвідношення (2.9) емпірична пряма регресії обов'язково проходить через точку (\bar{x}, \bar{y}) .

4. Емпіричне рівняння регресії побудоване в такий спосіб, що сума відхилень $\sum_{i=1}^n u_i$, а також середнє значення відхилення

$$\bar{u} = \frac{1}{n} \sum_{i=1}^n u_i \text{ дорівнюють нулю (показати самостійно).}$$

5. Випадкові відхилення u_i некорельовані зі спостереженими значеннями y_i залежної змінної Y .

Для підтвердження цього висновку необхідно показати, що коваріація між Y і u дорівнює нулю, тобто $S_{yu} = 0$.

6. Випадкові відхилення u_i некорельовані зі спостереженими значеннями x_i незалежної змінної X і з оціненими за лінійною регресійною моделлю значеннями залежної змінної \hat{Y} .

Щоб підтвердити даний висновок, необхідно показати, що коваріація між X і u дорівнює нулю, тобто $S_{xu} = 0$, $S_{\hat{y}u} = 0$.

(Доведення пп. 5 і 6 виконати самостійно.)

Зауважимо, що в класичній лінійній економетричній моделі змінна u розглядається як випадкова змінна з нульовим математичним сподіванням і сталою дисперсією. Оскільки u охоплює вплив багатьох неврахованих факторів, які можна вважати незалежними, то на підставі центральної граничної теореми теорії ймовірностей роблять висновок, що ця випадкова величина підпорядкована нормальному закону розподілу (закону Гаусса).

Доведено (теорема Гаусса), що застосування методу найменших квадратів можливе лише тоді, коли залишки розподілені нормально з параметрами $M(U) = 0$, $D[U] = \sigma_u^2 = \text{const}$.

Для ілюстрації МНК розглянемо такий приклад.

Приклад. Для аналізу залежності обсягу споживання Y (у. о.) домогосподарства від наявного прибутку X (у. о.) обрано вибірку обсягу $n = 12$ (щомісячно впродовж року), результати якої наведені в табл. 2.1. Необхідно визначити вид залежності; за МНК оцінити параметри рівняння регресії Y і X ; оцінити силу лінійної залежності між X і Y ; а також спрогнозувати споживання при прибутку $X = 160$.

Таблиця 2.1

i	1	2	3	4	5	6	7	8	9	10	11	12
x_i	107	109	110	113	120	122	123	128	136	140	145	150
y_i	102	105	108	110	115	117	119	125	132	130	141	144

Для визначення виду залежності побудуємо кореляційне поле (рис. 2.3).

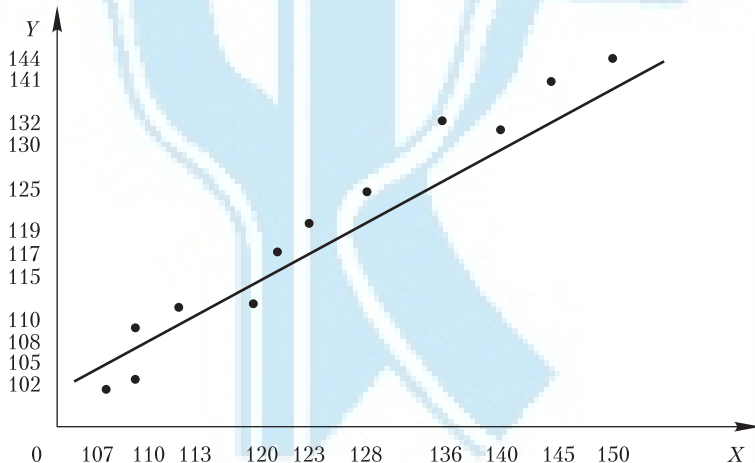


Рис. 2.3

За розміщенням точок на кореляційному полі припускаємо, що залежність між X і Y лінійна: $\hat{Y} = \hat{a}_0 + \hat{a}_1 X$; \hat{a}_0, \hat{a}_1 – оцінки невідомих параметрів моделі.

Для наочності розрахунків за МНК складемо табл. 2.2.

Згідно з МНК маємо

$$\begin{cases} \hat{a}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{15298,08 - 125,25 \cdot 120,67}{15884,75 - (125,25)^2} = \frac{184,1625}{197,1875} = 0,9339; \\ \hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x} = 120,67 - 0,9339 \cdot 125,25 = 3,699. \end{cases}$$

Отже, рівняння парної лінійної регресії має вигляд $\hat{Y} = 3,699 + 0,9339X$. Зобразимо цю пряму регресії на кореляційному полі. За наведеним рівнянням розрахуємо \hat{y}_i , а також $u_i = y_i - \hat{y}_i$.

Для аналізу сили лінійної залежності обчислимо коефіцієнт кореляції:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{x^2 - \bar{x}^2} \cdot \sqrt{y^2 - \bar{y}^2}} = \frac{184,1625}{14,04 \cdot 13,23} = 0,9914.$$

Отримане значення коефіцієнта кореляції дає змогу зробити висновки про сильну (пряму) лінійну залежність між змінними X і Y . Це також підтверджується розміщенням точок на кореляційному полі.

Таблиця 2.2

i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2	\hat{y}_i	e_i	e_i^2
1	107	102	11449	10914	10404	103,63	-1,36	2,66
2	109	105	11881	11445	11025	105,49	-0,49	0,24
3	110	108	12100	11880	11664	106,43	1,57	2,46
4	113	110	12769	12430	12100	109,23	0,77	0,59
5	120	115	14400	13800	13225	115,77	-0,77	0,59
6	122	117	14884	14274	13689	117,63	-0,63	0,40
7	123	119	15129	14637	14161	118,57	0,43	0,18
8	128	125	16384	16000	15625	123,24	1,76	3,10
9	136	132	18496	17952	17424	130,71	1,29	1,66
10	140	130	19600	18200	16900	134,45	-4,45	19,8
11	145	141	21025	20445	19881	139,11	1,89	3,57
12	150	144	22500	21600	20736	143,78	0,22	0,05
Сума	1503	1448	190617	183577	176834	—	$\approx 0^{**}$	35,3
Середнє*	125,25	120,67	15884,75	15298,08	14736,17			

* Значення округлюються до сотих.

** Ураховуються похибки округлень.

Прогнозоване споживання при доступному доході $X = 160$ за даною моделлю становить $\hat{y}(160) \approx 153,12$.

Побудоване рівняння регресії в будь-якому разі потребує певної інтерпретації та аналізу.

Інтерпретація, тобто словесний опис отриманих результатів, необхідна для того, щоб побудована залежність набула якісного економічного змісту.

У нашому прикладі коефіцієнт \hat{a}_1 може розглядатися як гранична схильність до споживання. Фактично він показує, на яку величину зміниться обсяг споживання, якщо доступний дохід збільшиться на одиницю. На графіку (рис. 2.3) коефіцієнт \hat{a}_1 визначає тангенс кута нахилу прямої регресії відносно додатного напрямку осі абсцис (пояснюючої змінної). Тому часто він називається кутовим коефіцієнтом.

Вільний член \hat{a}_0 рівняння регресії визначає прогнозоване значення Y при величині наявного прибутку X , що дорівнює нулю (тобто автономне споживання). Однак тут необхідна певна обережність. Важливо, наскільки віддалені дані спостережень за пояснюючою змінною від осі ординат (залежної змінної), тому що навіть при вдалому виборі рівняння регресії для досліджуваного інтервалу немає гарантії, що вона залишиться такою самою й віддалік від вибірки. У нашому випадку значення $\hat{a}_0 = 3,699$ (у. о.). Цей факт можна пояснити для окремого домогосподарства (воно може витратити накопичені або позичені кошти), однак для комплексу домогосподарств він втрачає сенс. У будь-якому разі значення коефіцієнта \hat{a}_0 визначає точку перетину прямої з віссю ординат і характеризує зсув лінії регресії вздовж осі Y .

Необхідно пам'ятати, що емпіричні коефіцієнти регресії \hat{a}_0 і \hat{a}_1 є лише оцінками теоретичних коефіцієнтів a_0 та a_1 , а саме рівняння відображає лише загальну тенденцію в поведінці розглянутих змінних. Індивідуальні значення змінних з різних причин можуть відхилитися від модельних значень. У нашому прикладі ці відхилення виражені через значення u_i , які є оцінками відповідних відхилень для генеральної сукупності.

Однак за певних умов рівняння регресії є незамінним і дуже якісним інструментом аналізу та прогнозування. Ці теми обговорюватимуться в наступних розділах.

Контрольні запитання

1. Що таке функція регресії?
2. Чим регресійна модель відрізняється від функції регресії?

3. Назвіть основні причини наявності в регресійній моделі випадкового відхилення.
4. Назвіть основні етапи регресійного аналізу.
5. У чому полягає відмінність між теоретичним та емпіричним рівняннями регресії?
6. Дайте визначення теоретичної регресійної моделі.
7. У чому суть методу найменших квадратів?
8. Наведіть формули розрахунку коефіцієнтів емпіричного парного лінійного рівняння регресії за МНК.
9. Як пов'язані емпіричні коефіцієнти лінійної регресії з вибірковою коефіцієнтом кореляції між змінними рівняння регресії?
10. Які висновки можна зробити про оцінки коефіцієнтів регресії та випадкового відхилення, отриманих за МНК?
11. Проінтерпретуйте коефіцієнти емпіричного парного лінійного рівняння регресії.

Вправи та завдання

1. Чи існує, на вашу думку, залежність між такими показниками:
 - а) ВВП і обсягом чистого експорту;
 - б) обсягом інвестувань і відсотковою ставкою;
 - в) видатками на оборону та видатками на освіту;
 - г) оцінками в школі та оцінками в університеті;
 - д) обсягом імпорту та прибутком на душу населення;
 - е) ціною на каву та ціною на чай?

У разі ствердної відповіді оцініть напрямок залежності (пряма чи обернена), а також зазначте, яка із змінних буде пояснюючою, а яка – залежною.

2. У наступній вибірці подано дані щодо ціни P деякого блага й кількості Q цього блага, яке домогосподарство купує щомісяця впродовж року.

Місяць	1	2	3	4	5	6	7	8	9	10	11	12
P	10	20	15	25	30	35	40	35	25	40	45	40
Q	110	75	100	80	60	55	40	80	60	30	40	30

- а. Побудуйте кореляційне поле і за його виглядом визначте формулу залежності між P і Q .
- б. Оцініть за МНК параметри рівняння лінійної регресії.
- в. Оцініть вибіркового коефіцієнта кореляції r_{pq} .
- г. Проінтерпретуйте результати.

3. Дано таблицю тижневого прибутку (X) і тижневого споживання (Y) для 60 домашніх господарств:

X	Y							
	60	65	75	85	90	100	120	150
100	60	65	75	85	90	100		
120	70	70	80	85	90	100		
140	90	95	95	100	100	120		
160	100	110	115	120	125	125	130	
180	110	120	120	130	135	140	150	150
200	120	125	130	135	140	150	160	165
220	120	140	145	145	155	165	180	
240	150	160	170	190	200			
260	140	160	180	210	220			
280	180	210	230					

а. Для кожного рівня прибутку розрахуйте середнє значення споживання, що є оцінкою умовного математичного сподівання $M(Y|X = x_i)$.

б. Побудуйте кореляційне поле для даної вибірки.

в. Складіть емпіричне лінійне рівняння регресії, використовуючи всі дані.

г. Складіть емпіричне лінійне рівняння регресії, використовуючи тільки середні значення споживання для кожного рівня прибутку.

д. Порівняйте складені рівняння. Яке з них, на ваш погляд, ближче до теоретичного?

е. Розрахуйте вибірко́вий коефіцієнт кореляції для в) і г). Чи буде лінійний зв'язок між даними змінними суттєвим? Відповідь обґрунтуйте.

4. За 10 парами спостережень отримано такі результати:

$$\sum x_i = 100; \quad \sum y_i = 200; \quad \sum x_i y_i = 21000;$$

$$\sum x_i^2 = 12000; \quad \sum y_i^2 = 45000.$$

За МНК оцініть коефіцієнти рівнянь регресії Y на X і X на Y . Оцініть коефіцієнт кореляції r_{xy} .

5. Дано таку емпіричну регресійну модель, побудовану за МНК:

$$y_t = b_0 + b_1 x_t + e_t, \quad t = 1, 2, \dots, T.$$

Доведіть, що $\sum_{t=1}^T e_t = 0$; $\sum_{t=1}^T e_t x_t = 0$.

6. Для вибірки обсягу $n = 10$ отримано такі дані:

$$\sum x_i = 993,4; \quad \sum y_i = 531,3; \quad \sum x_i y_i = 53196,61;$$

$$\sum x_i^2 = 105004,5; \quad r_{xy} = 0,75.$$

Розрахуйте оцінки коефіцієнтів регресії Y на X і X на Y .

7. Дано дві регресії, розраховані за 25 річними спостереженнями:

а) $y_t = -30 + 0,18x_t$ (y_t – витрати на житло, x_t – прибуток);

б) $y_t = 50 + 4,5t$ (y_t – витрати на житло, t – час).

Дайте економічну інтерпретацію побудованих регресій. Чи взаємопов'язані вони між собою?

МАУП

Розділ 3. ЗАГАЛЬНА ЛІНІЙНА ЕКОНОМЕТРИЧНА МОДЕЛЬ

3.1. Багатофакторні економетричні моделі та їх специфікація

У багатьох дослідженнях виявляється, що деяка результативна ознака змінюється під впливом не одного, а кількох факторів. Зокрема, аналізуючи економічну діяльність підприємства та прогнозуючи його подальший розвиток, досліджують такі функції:

1) виробничу функцію, що визначає залежність між обсягом виробленої продукції та витраченими для цього ресурсами, наприклад основним капіталом і працею;

2) функцію ціни, що дає змогу дослідити, як зміниться ціна товару, якщо зміниться обсяг поставок та ціни конкуруючих товарів;

3) функцію попиту, що дає змогу встановити, як зміниться попит на продукцію, якщо змінюватимуться ціна товару, ціни товарів-конкурентів і доходи споживачів;

4) функцію витрат, що описує залежність середніх витрат на виробництві від ціни та кількості виробничих ресурсів;

5) функцію чутливості ринку, яка визначає залежність обсягу збуту продукції від витрат на рекламу та індексу “чистоти” виробленої продукції (“екологічного індикатора”);

6) рівняння стратегії підприємства, у якому відображається залежність рентабельності підприємства від питомої ваги на ринку товарів, подібних до тих, які виробляє підприємство, а також від якості товарів, витрат на маркетинг і наукові дослідження, від інвестиційних витрат тощо.

Розглянемо детальніше першу з цих функцій.

Будь-яка виробнича система характеризується залежністю між кількістю виробленої в ній продукції та спожитими для цього ресур-

сами. Причому певні показники цієї залежності мають деякі випадкові коливання. Залежність між ними, формалізовану у відповідний спосіб у вигляді регресійного рівняння, називають *виробничою функцією*.

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

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

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

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

На мікроекономічному рівні за допомогою таких функцій, наприклад, описують зв'язок між величиною використаного ресурсу протягом року та річним обсягом випуску продукції одного підприємства, однієї галузі чи міжгалузевого виробничого комплексу. Якщо виробничою системою є регіон чи країна загалом, то маємо виробничу функцію для макроекономічного рівня.

Приклад. Нехай виробничу функцію задано у вигляді $f(x) = ax^b$, де x – величина витраченого ресурсу (наприклад, робочого часу), $f(x)$ – обсяг випущеної продукції (наприклад, кількість готових виробів). Величини a та b – параметри виробничої функції $f(x)$. Причому a та b – додатні числа, а $b \leq 1$. Задана функція $f(x)$ за малих значень аргументу дає значний приріст, якщо x збільшується на одиницю; за великих значень аргументу таке саме збільшення аргументу зумовлює значно менший приріст функції. Ця властивість $f(x)$ відбиває фундаментальне положення економічної теорії, яке називається законом спадної ефективності, а сама функція є типовим представником однофакторних виробничих функцій.

У реальних ситуаціях обсяг випуску продукції визначається, як правило, не одним, а багатьма факторами, тому частіше застосовують багаторесурсні або багатофакторні виробничі функції. Найпошире-

нішою серед них є виробнича функція Кобба – Дугласа, яка описує залежність між обсягом виробленої продукції Y і витратами праці L та капіталу F :

$$Y = aF^\alpha L^\beta$$

Множник a і показники степеня α та β – параметри цієї моделі. Задана в такому вигляді виробнича функція є мультиплікативною (нелінійною відносно параметрів). Логарифмуванням її можна звести до адитивного (лінійного відносно параметрів) вигляду:

$$\ln Y = a + \alpha \ln F + \beta \ln L.$$

Значена функція має такі властивості:

1) коефіцієнт α показує, на скільки відсотків зміниться обсяг випуску продукції, якщо витрати праці зміняться на 1 %, а витрати капіталу залишаться незмінними. Такий показник називається коефіцієнтом еластичності випуску за витратами праці;

2) коефіцієнт β є коефіцієнтом еластичності випуску за витратами капіталу;

3) сума параметрів $\alpha + \beta$ описує масштаб виробництва.

Якщо ця сума дорівнює одиниці, маємо постійний масштаб виробництва. А це означає, що зі збільшенням обох виробничих ресурсів на одиницю обсяг продукції також зростає на одиницю. Якщо сума менша одиниці, то масштаб виробництва спадний, тобто темпи зростання обсягу продукції нижчі за темпи зростання обсягу ресурсів. Якщо сума перевищує одиницю, маємо зростаючий масштаб: темпи зростання обсягу продукції перевищують темпи зростання обсягу виробничих ресурсів.

Параметр a у функції Кобба – Дугласа залежить від одиниць вимірювання Y , F та L і також визначається ефективністю виробничого процесу.

Отже, економетрична модель виробничої функції дає змогу проаналізувати виробничу діяльність, щоб визначити шляхи підвищення її ефективності. Обґрунтованість такого аналізу цілком залежить від достовірності моделі та її адекватності відповідному реальному процесу.

Вплив багатьох чинників на результативну змінну може бути описаний лінійною моделлю

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m + u, \quad (3.1)$$

де y — досліджувана (залежна, пояснювана) змінна, або регресанд;
 x_1, x_2, \dots, x_m — незалежні, пояснюючі змінні, або регресори;
 a_1, a_2, \dots, a_m — параметри моделі; u — випадкова складова регресійно-го рівняння.

Функція (3.1) є лінійною відносно незалежних змінних і параметрів моделі, але саме лінійність за параметрами є більш суттєвою, оскільки це пов'язано з методами оцінювання параметрів. Випадкова складова u є результативною дією всіх неконтрольованих випадкових факторів, що зумовлюють відхилення реальних значень досліджуваного показника y від аналітичних (обчислених на підставі обраної регресійної залежності).

Зрозуміло, що лінійні зв'язки не вичерпують усіх можливих форм залежності між показниками. Тому при дослідженні конкретного економічного явища першочерговим завданням є пошук найточнішої аналітичної форми опису статистичного зв'язку між його показниками. Певна форма залежності повинна мати відповідне економічне обґрунтування. Якщо вигляд залежності встановити важко, то за перше наближення до моделі все ж обирають лінійну залежність.

Звичайним математичним підходом до розв'язання задач є виокремлення специфічних класів задач або зведення задач до деякого класу і застосування відповідних методів розв'язування. Оскільки дослідження лінійних функцій має незаперечні переваги перед іншими класами функцій, то нелінійні функції намагаються передусім звести до лінійних. Наприклад, степенева функція

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

після логарифмування набуває вигляду

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

і після заміни $\ln a_0 = a$ є лінійною відносно параметрів a, a_1, \dots, a_m

Показникова функція

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

після логарифмування набуває вигляду

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

і після заміни $\ln a_i = b_i$, $i = 0, 1, 2, \dots, m$, є лінійною відносно параметрів b_i .

Гіперболічна

$$y = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + \dots + \frac{a_m}{x_m}$$

і квадратична

$$y = a_0 + a_1x_1^2 + a_2x_2^2 + \dots + a_mx_m^2$$

функції заміною змінних $z_i = \frac{1}{x_i}$ або $z_i = x_i^2$, $i = 1, 2, \dots, m$, зводяться до лінійного вигляду:

$$y = a_0 + a_1z_1 + a_2z_2 + \dots + a_mz_m.$$

Зуважимо, що в сучасному економічному аналізі існують залежності, які не зводяться до лінійних елементарними перетвореннями, однак їх параметри можна легко розрахувати спеціальними спрощеними методами [13].

Оскільки найпоширенішими в економетричному моделюванні є лінійні функції, обґрунтування економетричних методів розглядають, як правило, на базі лінійних моделей.

Отже, предметом наших досліджень буде узагальнена багатofакторна лінійна регресійна модель (3.1).

Як зазначалося, узагальнена регресійна модель справджується для всієї генеральної сукупності, а похибка регресії має певний закон розподілу.

На практиці мають справу з вибірковою моделлю, тобто з такою, яка побудована для деякої вибірки. Параметри вибіркової моделі є випадковими величинами, а їх математичне сподівання дорівнює параметрам узагальненої моделі. Щоб визначити параметри узагальненої моделі, необхідно за вибіркою отримати якомога кращі їх оцінки, тобто значення, найближчі до параметрів узагальненої моделі. З цією метою використовують метод найменших квадратів (МНК).

3.2. Метод найменших квадратів

3.2.1. Основні припущення

Застосування методу найменших квадратів до загальної лінійної багатофакторної моделі (3.1) передбачає наявність таких передумов:

1) кожне значення випадкової складової рівняння u_i , $i = 1, 2, \dots, n$, є випадковою величиною і математичне сподівання залишків u_i дорівнює нулю:

$$M(u) = 0;$$

2) компоненти вектора залишків некорельовані (лінійно незалежні) між собою і мають сталу дисперсію:

$$M(u^T u) = \sigma^2 E;$$

3) пояснюючі змінні (регресори, фактори моделі) некорельовані із залишками;

4) пояснюючі змінні некорельовані між собою.

Порушення першої передумови означає, що існує систематичний вплив на залежну змінну, який не враховано в моделі. Таку ситуацію можна трактувати як помилку специфікації, однак наявність вільного члена моделі дає змогу скоригувати модель так, щоб забезпечити виконання першої передумови.

Друга передумова означає, що залишки моделі є помилками вимірювання. Якщо між компонентами вектора залишків існує кореляційна залежність, таке явище називається автокореляцією. Наявність автокореляції в моделі свідчить про існування кореляції між послідовними значеннями деякої незалежної змінної або про неврахований суттєвий фактор, що впливає на залежну змінну і не може бути усунений за рахунок вільного члена моделі. Загальний вплив пояснюючих змінних, не врахованих у моделі, може виявитися також у тому, що дисперсія залишків для окремих груп спостережень змінюватиметься. Таке явище називається гетероскедастичністю. У будь-якому разі порушення другої передумови впливає на методи оцінювання параметрів моделі.

Наявність залежності між залишками та незалежними змінними найчастіше пов'язана з тим, що в моделі присутні лагові (затримані в часі) змінні або вона будується на базі одночасних структурних

рівнянь. Для оцінювання параметрів і в цьому разі застосовують інші методи.

Залежність між незалежними змінними може значною мірою впливати на якість оцінок, отриманих за МНК. Якщо між незалежними змінними моделі існують тісні лінійні зв'язки, це явище називають мультиколінеарністю. Моделі, у яких спостерігається мультиколінеарність, стають надзвичайно чутливими до конкретного набору даних, до специфікації моделі й мають значні відхилення від дійсних значень параметрів узагальненої моделі.

Крім розглянутих чотирьох передумов важливе значення має припущення про нормальний розподіл залишків моделі. Це припущення забезпечує нормальний розподіл коефіцієнтів регресії й дає змогу використовувати відомі критерії для перевірки статистичних гіпотез відносно отриманих оцінок, а також визначати їх довірчі інтервали.

3.2.2. МНК-оцінки параметрів лінійної регресії та їх основні властивості

З теорії ймовірностей відомо (доведено в теоремі Гаусса — Маркова), що коли виконуються перелічені передумови, то отримані за допомогою МНК оцінки параметрів регресійного рівняння є незміщеними, обгрунтованими, ефективними та інваріантними.

Наявність таких властивостей оцінок гарантує, що останні не мають систематичної похибки (незміщеність), надійність їх підвищується зі збільшенням обсягу вибірки (обгрунтованість), вони є найкращими серед інших оцінок параметрів, лінійних відносно ендогенної змінної (ефективність). Крім того, оцінка перетворених параметрів (оцінка функції від параметра) може бути отримана в результаті аналогічного перетворення оцінки параметра (інваріантність).

Зокрема, якщо порушується друга передумова МНК (за наявності автокореляції чи гетероскедастичності), то отримані за цим методом оцінки втрачають властивість ефективності, хоча залишаються незміщеними та обгрунтованими. Якщо порушується четверта передумова, тобто між змінними існують мультиколінеарні зв'язки, це призводить до зміщення МНК-оцінок. Застосування моделей, що мають зміщені чи неефективні оцінки, втрачає сенс.

3.2.3. Оцінювання за методом найменших квадратів та інтерпретація результатів

Нехай відомо n спостережень незалежних змінних x_1, x_2, \dots, x_m і n спостережень залежної змінної y . Необхідно за МНК оцінити параметри $a_0, a_1, a_2, \dots, a_m$ лінійної моделі (3.1).

Якщо виконуються зазначені раніше передумови, то оцінки параметрів можна отримати за таким алгоритмом.

1. Незалежні змінні записати у вигляді матриці

$$X = \{x_0, x_1, x_2, \dots, x_m\},$$

де x_0 – вектор, складений з n одиниць; x_1, x_2, \dots, x_m – вектори спостережень незалежних змінних.

2. Обчислити матрицю $X^T X$ і вектор $X^T y$, де X^T – транспонована матриця X , y – вектор спостережень залежної змінної.

Зауваження. Транспонована матриця утворюється з вихідної перетворенням рядків у стовпці.

3. Обчислити обернену матрицю $(X^T X)^{-1}$.

Зауваження. Матриця A^{-1} називається оберненою до A , якщо виконується співвідношення $A^{-1} A = A A^{-1} = E$, де E – одинична матриця.

4. Обчислити параметри моделі за формулою

$$a = (X^T X)^{-1} (X^T y), \quad (3.2)$$

де a – вектор параметрів, $a = (a_0, a_1, a_2, \dots, a_m)^T$.

Зауваження. Для визначення оцінок параметрів можна скористатися будь-яким методом розв'язання системи лінійних рівнянь відносно вектора невідомих змінних:

$$(X^T X)a = X^T y.$$

Приклад [3]. Підприємство, що складається з багатьох філій, досліджує залежність свого річного товарообігу y (млн у. о.) від торгової площі своїх філій x_1 (тис. кв. м) і середньоденної інтенсивності потоку покупців (тис. чол./день). Просторові дані за філіями наведено в табл. 3.1.

Для відображення залежності між цими показниками обирають лінійну регресійну модель

$$y = a_0 + a_1x_1 + a_2x_2 + u.$$

Таблиця 3.1

Номер філії	Значення y (y_i)	Значення x_1 (x_{1i})	Значення x_2 (x_{2i})
1	2,93	0,31	10,24
2	5,27	0,98	7,15
3	6,85	1,21	10,81
4	7,01	1,29	9,89
5	7,02	1,12	13,72
6	8,35	1,49	13,92
7	4,33	0,78	8,54
8	5,77	0,94	12,36
9	7,68	1,29	12,27
10	3,16	0,48	11,01
11	1,52	0,24	8,25
12	3,15	0,55	9,31

У результаті оцінювання параметрів за методом найменших квадратів отримано такі оцінки: $\hat{a}_0 = -0,832$; $\hat{a}_1 = 4,743$; $\hat{a}_2 = 0,175$.

Оцінки інтерпретуються таким чином. Коефіцієнт $a_1 = 4,743$ означає, що за інших незмінних умов змінна x_1 збільшиться (зменшиться) на одиницю, залежна змінна y збільшиться (зменшиться) відповідно на 4,743 одиниці. Зокрема, у наведеному прикладі збільшення (зменшення) торгової площі на 1 тис. кв. м збільшить (зменшить) річний товарообіг на 4,743 млн у. о. Аналогічно, збільшення (зменшення) середньоденної інтенсивності покупців на 1 тис. чол./день збільшить (зменшить) річний товарообіг на 0,175 млн у. о.

3.3. Верифікація моделі

3.3.1. Показники якості моделі

У класичному регресійному аналізі вважається, що функція регресії відома до оцінювання параметрів, тобто регресійна модель специфікована правильно. Однак в емпіричних економічних і соціаль-

них дослідженнях не завжди відомо, скільки факторів має бути введено в модель і яка форма залежності краще описує реальні зв'язки. Щоб забезпечити найбільш адекватне відтворення досліджуваного явища чи процесу, необхідно вибрати регресійну функцію серед багатьох варіантів, використовуючи спеціальні критерії якості моделі.

Для перевірки коректності побудови моделі визначають насамперед:

- стандартну похибку рівняння;
- коефіцієнт детермінації;
- коефіцієнт множинної кореляції;
- стандартну похибку параметрів.

Зауважимо, що зазначені показники отримують на підставі конкретних статистичних даних, тобто кожна з цих характеристик є вибірковою характеристикою і тому має бути перевірена на значущість за допомогою спеціальних статистичних критеріїв.

Стандартна похибка рівняння (точкова оцінка емпіричної дисперсії залишків) характеризує абсолютну величину розкиду випадкової складової рівняння і обчислюється за формулою

$$S_u^2 = \frac{1}{n} \sum_{i=1}^n u_i^2.$$

Поправка на число ступенів свободи дає незміщену оцінку дисперсії залишків:

$$\hat{\sigma}_u^2 = \frac{1}{n - m - 1} \sum_{i=1}^n u_i^2.$$

Зрозуміло, що перевага віддається моделям, у яких стандартна похибка рівняння менша порівняно з іншими моделями. Однак така оцінка якості має суттєвий недолік: через те що для неї не визначено верхню межу, порівняння різних моделей за цим критерієм досить проблематичне.

Коефіцієнт детермінації R^2 показує, яка частина руху залежної змінної описується даним регресійним рівнянням, і обчислюється за формулою

$$R^2 = 1 - \frac{S_u^2}{S_y^2},$$

де $S_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$; \bar{y} – середнє значення залежної змінної,

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i.$$

На значення коефіцієнта детермінації впливає кількість факторів, що враховано в моделі. Уведення в модель кожної нової змінної збільшує значення коефіцієнта детермінації. Тому щоб запобігти невиправданому розширенню моделі й мати змогу порівнювати моделі з різною кількістю факторів, вводять спеціальний оцінений коефіцієнт детермінації

$$\bar{R}^2 = 1 - \frac{\sigma_u^2}{\sigma_y^2},$$

де σ_u^2 – незміщена оцінка дисперсії залишків; σ_y^2 – незміщена оцінка дисперсії залежної змінної, $\sigma_y^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$.

Неважко помітити, що обидва коефіцієнти пов'язані такою залежністю:

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

Обчислений у такий спосіб коефіцієнт детермінації називається скоригованим за Тейлом і позначається \bar{R}_T^2 . Крім того застосовують також коригування за Амемією, яке виконується за формулою

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n+m}{n-m-1}.$$

Обчислений у такий спосіб коефіцієнт детермінації називається скоригованим за Амемією і позначається \bar{R}_A^2 .

Обидва коефіцієнти \bar{R}_T^2 і \bar{R}_A^2 враховують той факт, що введення в модель кожного нового регресора зменшує число ступенів свободи. А для застосування статистичних критеріїв перевірки якості отриманих результатів ступенів свободи бажано мати якомога більше.

Очевидно, для кожного \bar{R}_T^2 і \bar{R}_A^2 виконується нерівність $\bar{R}^2 \leq R^2$, тобто зі збільшенням кількості факторів моделі оцінені коефіцієнти

детермінації зростають повільніше, ніж R^2 . Крім того, якщо $R^2 = 1$, то і $\bar{R}^2 = 1$. Якщо R^2 прямує до нуля, оцінені коефіцієнти стають від'ємними. Така властивість скоригованих коефіцієнтів детермінації дає змогу більш об'єктивно оцінювати якість моделей з різною кількістю факторів, причому в разі застосування коефіцієнта \bar{R}_A^2 (скоригованого за Амемією) перевага однозначно віддається рівнянню з меншою кількістю регресорів.

Зауваження. Коефіцієнт детермінації має ще два рівноцінних означення. За першим, коефіцієнт детермінації R^2 дорівнює квадрату емпіричного коефіцієнта кореляції між двома рядами спостережень (теоретичними значеннями регресанда y_i та його розрахунковими значеннями \hat{y}_i , $i = 1, 2, \dots, n$) і обчислюється за формулою

$$R^2 = \frac{(\sum (y_i - \bar{y})(\hat{y}_i - \bar{y}))^2}{\sum (y_i - \bar{y})^2 \sum (\hat{y}_i - \bar{y})^2}.$$

За другим, коефіцієнт детермінації R^2 дорівнює відношенню суми квадратів відхилень розрахункових значень регресанда від його середнього значення до суми квадратів відхилень спостережених значень регресанда від того самого середнього значення:

$$R^2 = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}.$$

В обох випадках сума \sum обчислюється за всіма спостереженнями $i = 1, 2, \dots, n$.

Коефіцієнт множинної кореляції R (\bar{R}) визначає міру зв'язку залежної змінної з усіма незалежними факторами і є коренем квадратним з відповідного коефіцієнта детермінації: $R = \sqrt{R^2}$ ($\bar{R} = \sqrt{\bar{R}^2}$).

Стандартна похибка рівняння, коефіцієнт детермінації та множинної кореляції є характеристиками, за якими перевіряється правильність вибору незалежних змінних моделей. При порівнянні регресійних рівнянь з різною кількістю незалежних змінних вирішальними критеріями є стандартна похибка рівняння (найменша) та коефіцієнт детермінації (якогома ближчий до одиниці і з більшим числом ступенів свободи).

3.3.2. Перевірка значущості та довірчі інтервали

Розглянуті показники якості моделі побудовані за даними спостережень, тобто є деякими вибірковими характеристиками генеральної сукупності. З математичної статистики відомо, що будь-яка статистика (функція від елементів вибірки) має бути перевірена на значущість. Іншими словами, за допомогою спеціальних критеріїв необхідно встановити, чи зумовлено значення цієї функції лише похибками вимірювання, чи вона відображає якусь суттєву (значущу) інформацію. Непереверений статистичний результат є лише деякою гіпотезою, яка може бути прийнята чи відхилена.

Нагадаємо, що перевірка гіпотез у загальному випадку виконується в такому порядку: для кожної задачі добирається деяка випадкова величина, що має відомий чи близький до відомого закон розподілу. Функція від елементів вибірки є конкретною реалізацією цієї випадкової величини.

Зауваження. У задачах регресійного аналізу важливе значення має припущення про нормальний розподіл випадкових величин, що задіяні в даній моделі. Певні перетворення нормально розподілених величин забезпечують їх розподіл за законом Стюдента чи за законом Фішера: на підставі першого з них визначаються довірчі інтервали, а другий дає змогу оцінювати відношення двох випадкових величин.

Стосовно кожного статистичного результату висувається так звана нульова гіпотеза (про рівність нулю деякої випадкової величини) і альтернативна до неї гіпотеза (про її суттєву відмінність від нуля). У нульовій гіпотезі формулюють результат, який бажано відхилити, а в альтернативній, яка інакше називається експериментальною, — той, що його необхідно підтвердити.

Зауваження. Рівність двох величин у загальному випадку може розглядатися як рівність нулю їх різниці.

За заданим *рівнем значущості* множина допустимих значень розбивається на дві неперетинні множини: одна містить значення випадкової величини, імовірність досягнення яких перевищує заданий рівень значущості, а інша — критична область — визначає ті значення, що досягаються рідко (імовірність потрапити до такої області нижча від заданого рівня), і розташована вона, як правило, на “хвостах розподілу”.

Залежно від альтернативної гіпотези критична область може складатися з одного чи двох проміжків на числовій осі. Це буде один проміжок

(правий чи лівий “хвіст” розподілу), якщо зазначається напрямок нерівності (більше або менше деякої величини), і два проміжки (обидва “хвости” розподілу), якщо встановлюється нерівність (не дорівнює певній величині).

За даними спостережень обчислюється значення відповідної статистики — функції від елементів вибірки. Якщо ця величина потрапляє до критичної області, це означає, що сталася практично неможлива подія, тобто подія, що має дуже малу ймовірність, а отже, від нульової гіпотези слід відмовитися і віддати перевагу альтернативній. Якщо обчислене значення статистики не потрапило до критичної області, роблять висновок, що дана вибірка не суперечить нульовій гіпотезі, тобто неправильною є експериментальна гіпотеза.

При перевірці гіпотез може бути допущена помилка, наприклад може бути відхилена нульова гіпотеза, хоча насправді вона правильна (помилка першого роду), або ж, навпаки, нульова гіпотеза може бути прийнята, хоча вона неправильна (помилка другого роду). На це слід зважати при формулюванні статистичного висновку.

Якщо значення R^2 “близьке” до одиниці, вважається, що регресійне рівняння досить правильно відбиває наявний зв’язок між залежною та незалежними змінними моделі. Якщо значення R^2 “близьке” до нуля, регресійна модель неправильна. Постає питання, як визначити цю “близькість”? Для цього необхідно застосувати відповідний статистичний критерій, який дасть змогу встановити, чи суттєво відрізняється R^2 від нуля, чи ця відмінність пов’язана з особливостями конкретних даних, тобто зумовлена лише похибками вимірювань.

Для перевірки статистичної значущості коефіцієнта детермінації R^2 висувається нульова гіпотеза $H_0 : R^2 = 0$. Це означає, що досліджуване рівняння не пояснює змінювання регресанда під впливом відповідних регресорів. У такому разі всі коефіцієнти при незалежних змінних мають дорівнювати нулю. При цьому нульову гіпотезу можна подати у вигляді

$$H_0 : a_1 = a_2 = \dots = a_n = 0.$$

Альтернативною до неї є H_A : значення хоча б одного параметра моделі відмінне від нуля, тобто хоча б один із факторів впливає на змінювання залежної змінної.

Для перевірки цих гіпотез застосовують F -критерій Фішера з m і $n - m - 1$ ступенями свободи. За отриманими в моделі значеннями

коефіцієнта детермінації R^2 обчислюють експериментальне значення F -статистики:

$$F_{\text{експ}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m},$$

яке порівнюють з табличним значенням розподілу Фішера при заданому рівні значущості α (як правило, $\alpha = 0,05$ або $\alpha = 0,01$). Якщо $F_{\text{табл}} < F_{\text{експ}}$, нульова гіпотеза відхиляється, тобто існує такий коефіцієнт у регресійному рівнянні, який суттєво відрізняється від нуля, а відповідний фактор впливає на досліджувану змінну. Відхилення нуль-гіпотези свідчить про адекватність побудованої моделі. У протилежному випадку модель вважається неадекватною.

Коефіцієнт кореляції, як вибіркова характеристика, перевіряється на значущість за допомогою t -критерію Стьюдента. Фактичне значення t -статистики обчислюється за формулою

$$t_{\text{експ}} = \frac{R\sqrt{n - m - 1}}{\sqrt{1 - R^2}}$$

і порівнюється з табличним значенням t -розподілу з $n - m - 1$ ступенями свободи та при заданому рівні значущості $\alpha/2$ (такий рівень зумовлений тим, що критична область складається з двох проміжків). Якщо абсолютна величина експериментального значення t -статистики перевищує табличне, тобто

$$|t| > t_{\text{табл}},$$

можна зробити висновок, що коефіцієнт кореляції достовірний (значущий), а зв'язок між залежною змінною та всіма незалежними факторами суттєвий.

Окрім загальних показників адекватності моделі існують також оцінки, що дають змогу встановити якість окремих частин рівняння, зокрема одного чи кількох коефіцієнтів регресії. Як і в попередніх випадках, рішення відносно якості коефіцієнтів приймають на основі відповідних статистичних критеріїв.

На підставі одного з найважливіших припущень МНК — припущення про нормальний розподіл випадкової складової рівняння з нульовим математичним сподіванням і сталою дисперсією — дове-

дено, що кожний параметр лінійної регресії також має нормальний розподіл. Причому математичне сподівання параметра дорівнює значенню параметра узагальненої регресії, а дисперсія — незміщений дисперсії випадкової складової рівняння, помножений на відповідний діагональний елемент оберненої матриці $(X^T X)^{-1}$.

Статистичну значущість кожного параметра моделі можна перевірити за допомогою t -критерію. При цьому нульова гіпотеза має вигляд

$$H_0 : a_j = 0,$$

альтернативна

$$H_A : a_j \neq 0.$$

Експериментальне значення t -статистики для кожного параметра моделі обчислюється за формулою

$$t_j = \frac{a_j}{\sqrt{\sigma_u^2 c_{jj}}} = \frac{a_j}{S_{a_j}},$$

де c_{jj} — діагональний елемент матриці $(X^T X)^{-1}$; S_{a_j} — стандартизована похибка оцінки параметра моделі, $S_{a_j} = \sigma_u \sqrt{c_{jj}}$.

Експериментальне значення t_j -критерію порівнюється з табличним значенням $t_{\text{табл}}$ з $n - m - 1$ ступенями свободи при заданому рівні значущості $\alpha/2$ (критична область розбивається на два фрагменти, межі яких задаються квантилем $\alpha/2$). Якщо значення t_j -статистики потрапляє до критичної області (за абсолютним значенням перевищує $t_{\text{табл}}$), приймається альтернативна гіпотеза про значущість відповідного параметра. Інакше робиться висновок про статистичну незначущість параметра a_j , а це означає, що відповідна незалежна змінна не впливає суттєво на змінування регресанда.

Зауваження. Оскільки t_j -статистика є відношенням відповідного параметра моделі до його стандартної похибки (середньоквадратичного відхилення), то на практиці частіше застосовують грубішу оцінку, а саме допускають, щоб стандартні похибки становили 45–50 % значення параметра, аби стверджувати про його статистичну значущість.

Довірчі інтервали для кожного окремого параметра a_j обчислюються на основі його стандартної похибки та критерію Стьюдента:

$$(a_j - t_{\text{табл}} \sqrt{\sigma_u^2 c_{jj}}; \quad a_j + t_{\text{табл}} \sqrt{\sigma_u^2 c_{jj}}).$$

Табличне значення $t_{\text{табл}}$, як і раніше, має $n - m - 1$ ступенів свободи і рівень значущості $\alpha/2$ ($t_{\text{табл}} = t_{\alpha/2}(n - m - 1)$).

Обчислені значення t_j -статистик застосовують також для розрахунку часткових коефіцієнтів детермінації ΔR_j^2 , які визначають граничний внесок j -го регресора в загальний коефіцієнт детермінації. Коефіцієнт ΔR_j^2 показує, на яку величину зменшиться коефіцієнт детермінації R^2 , якщо j -й регресор (і лише він!) буде вилучений з групи регресорів. Формула для розрахунку часткового коефіцієнта детермінації має вигляд

$$\Delta R_j^2 = \frac{(1 - R^2) t_j^2}{n - m},$$

де R^2 — коефіцієнт детермінації, обчислений для моделі з m регресорами; t_j^2 — квадрат обчисленого значення t -статистики для j -го регресійного коефіцієнта; n — кількість спостережень, m — кількість регресорів.

Зауваження. Часткові коефіцієнти детермінації $\Delta \bar{R}_T^2$ і $\Delta \bar{R}_A^2$, обчислені за відповідними значеннями \bar{R}_T^2 та \bar{R}_A^2 , можуть бути як додатними, так і від'ємними, що дає змогу більш об'єктивно оцінювати моделі з різною кількістю регресорів.

3.4. Прогнозування за лінійною моделлю

Якщо побудована модель адекватна за F -критерієм, то її застосовують для прогнозування залежної змінної.

Про прогнозування регресанда говорять тоді, коли в часових рядах прогнозний період настає пізніше, ніж базовий. Якщо регресія побудована за просторовими даними, прогноз стосується тих елементів генеральної сукупності, що перебувають за межами застосованої вибірки.

Якість прогнозу тим краща, чим повніше виконуються передумови моделі в прогнозний часовий період, надійніше (вірогідніше) оцінено параметри моделі й більш точно визначено прогнознi значення регресорів.

Значення \hat{y}_p для майбутнього періоду чи додаткового елемента обчислюють за формулою (3.1) за відомим вектором оцінених параметрів $\hat{a} = (\hat{a}_0, \hat{a}_1, \hat{a}_2, \dots, \hat{a}_m)$ і за вектором значень незалежних змінних $x_p = (1, x_{1p}, x_{2p}, \dots, x_{mp})$, що не належать до базового періоду. Розрізняють прогноз середній (оцінку математичного сподівання регресанда) та індивідуальний (оцінку певної реалізації регресанда y_p , що відповідає моменту p). Перша з них базується на передумові МНК про нульове математичне сподівання випадкової складової рівняння регресії, а друга застосовує оцінене значення \hat{y}_p . Оцінену дисперсію прогнозу обчислюють відповідно за формулами

$$\hat{\sigma}_e^2 = \hat{\sigma}_u^2 x_p^T (X^T X)^{-1} x_p;$$

$$\hat{\sigma}_{e(i)}^2 = \hat{\sigma}_u^2 + \hat{\sigma}_u^2 x_p^T (X^T X)^{-1} x_p.$$

Зрозуміло, що здебільшого реальне значення показника y_t не збігатиметься зі значенням його математичного сподівання, але якщо розглядати велику кількість вибірок, на підставі яких визначатиметься прогноз, то можна гарантувати, що приблизно $(1 - \alpha) \cdot 100\%$ результатів потраплять відповідно до інтервалів

$$(\hat{y}_p - t_{\alpha/2} \sqrt{\hat{\sigma}_e^2}; \hat{y}_p + t_{\alpha/2} \sqrt{\hat{\sigma}_e^2});$$

$$(\hat{y}_p - t_{\alpha/2} \sqrt{\hat{\sigma}_{e(i)}^2}; \hat{y}_p + t_{\alpha/2} \sqrt{\hat{\sigma}_{e(i)}^2}),$$

де $t_{\alpha/2}$ — табличне значення критерію Стюдента з $n - m - 1$ ступенями свободи та при заданому рівні значущості $\alpha/2$. (Значення $\alpha/2$ вибирають, як і раніше, через двосторонні критичні межі.)

Зауваження. Очевидно, з віддаленням від середнього значення вибірки спостережень похибка прогнозу зростатиме, що призведе до збільшення довірчого інтервалу для індивідуального значення залежної змінної.

3.5. Методи побудови багатofакторної регресійної моделі

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

Для вибору компромісного рішення не існує єдиної процедури.

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

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

2. Метод виключень економніший щодо обчислень і базується на дослідженні часткових F -критеріїв, які дають змогу встановлювати статистичну значущість співвідношення між залишками моделі з найбільшою кількістю факторів і залишками моделі з одним вилученим фактором. Якщо для деякого вилученого фактора таке співвідношення не є значущим (приймається нульова гіпотеза), то він до моделі не повертається. Таке дослідження проводиться також для рівняння з меншою кількістю факторів, але з більшим числом ступенів свободи.

3. Покроковий регресійний метод діє у зворотному порядку порівняно з попереднім методом, тобто до моделі послідовно включаються фактори, що мають найбільший коефіцієнт кореляції із залежною змінною. Модель аналізується за значеннями коефіцієнта детермінації та частковими F -критеріями. Фактори, що не задоволь-

няють критерії, з моделі вилучаються. Процес припиняється, якщо жоден з факторів рівняння вилучити не вдається, а новий претендент на включення не відповідає частковому F -критерію. На практиці цей метод найпоширеніший.

3.6. Етапи дослідження загальної лінійної моделі множинної регресії

Розглядається багатofакторна лінійна регресійна модель

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m,$$

що описує залежність між результативною змінною y та деякими впливовими факторами x_1, x_2, \dots, x_m . Інформація про значення y, x_1, x_2, \dots, x_m міститься у відповідних статистичних даних — n спостереженнях (вимірюваннях) кожного показника.

Для дослідження зазначеної моделі слід виконати такі кроки.

1. За даними спостережень оцінити параметри a_1, a_2, \dots, a_m .
2. Для перевірки адекватності отриманої моделі обчислити:
 - а) залишки моделі — розбіжності між спостереженими та розрахунковими значеннями залежної змінної $u_i = y_i - \hat{y}_i, i = 1, 2, \dots, n$;
 - б) відносну похибку залишків та її середнє значення;
 - в) залишкову дисперсію;
 - г) коефіцієнт детермінації;
 - д) вибірковий коефіцієнт множинної кореляції.
3. Перевірити статистичну значущість отриманих результатів:
 - а) перевірити адекватність моделі загалом: за допомогою F -критерію Фішера перевірити гіпотезу

$$H_0 : a_1 = a_2 = \dots = a_m = 0$$

проти альтернативної H_A : існує хоча б один коефіцієнт $a_j \neq 0$;

б) перевірити значущість коефіцієнта множинної кореляції, тобто розглянути гіпотезу $H_0 : R = 0$;

в) перевірити істотність кожного коефіцієнта регресії: за допомогою t -критерію Стьюдента перевірити гіпотезу

$$H_0 : a_j = 0 \text{ для всіх } j = 1, 2, \dots, m$$

проти відповідних альтернативних гіпотез

$$H_A : a_j \neq 0 \text{ для всіх } j = 1, 2, \dots, m;$$

г) оцінити вплив кожного регресора на якість моделі, тобто обчислити часткові коефіцієнти детермінації ΔR_j^2 , скоригувати їх за Тейлом і за Амемією та дати їх відповідну інтерпретацію;

д) оцінити вплив окремих груп регресорів на змінювання регресанда, застосувавши F -критерій Фішера.

4. Обчислити та інтерпретувати коефіцієнти еластичності.
5. Визначити довірчі інтервали регресії при рівні значущості α .
6. Побудувати довірчі інтервали для параметрів регресії.
7. Обчислити прогнози значення y_p за значеннями $x_{1p}, x_{2p}, \dots, x_{mp}$, що перебувають за межами базового періоду, і знайти межі довірчих інтервалів індивідуальних прогнозованих значень і межі довірчих інтервалів середнього прогнозу.

Приклад параметризації та дослідження багатофакторної регресійної моделі

Розглянемо задачу дослідження впливу на економічний показник y трьох факторів x_1, x_2, x_3 , а саме досліджуватимемо залежність прибутку підприємства $y(i)$ від інвестицій $x_1(i)$, витрат на рекламу $x_2(i)$ та заробітну плату $x_3(i)$.

Вихідні дані в умовних одиницях

Номер спостереження	$y(i)$	$x_1(i)$	$x_2(i)$	$x_3(i)$
1	15,70	17,37	5,28	1,42
2	17,34	18,24	6,47	1,58
3	21,57	22,47	6,98	1,98
4	33,50	18,47	7,05	2,04
5	32,30	16,82	7,94	2,38
6	37,90	17,60	8,12	3,48
7	40,78	17,12	8,69	3,07
8	48,02	19,81	9,31	3,84
9	43,30	18,67	10,45	4,28
10	49,57	20,83	10,47	4,67
11	52,14	22,84	13,48	5,98
12	55,17	28,85	15,78	6,51
13	59,18	29,61	17,65	7,82
14	62,22	35,67	18,47	8,58
15	77,58	47,87	19,64	9,47

Припустимо, що між економічним показником y і факторами x_1, x_2, x_3 існує лінійний зв'язок.

Запишемо рівняння регресії у вигляді

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + u, \quad (3.3)$$

$$\hat{y} = \hat{a}_0 + \hat{a}_1x_1 + \hat{a}_2x_2 + \hat{a}_3x_3, \quad (3.4)$$

де y, \hat{y} – відповідно фактичні та розрахункові значення прибутку; x_1, x_2, x_3 – відповідно інвестиції, витрати на рекламу та заробітну плату; a_0, a_1, a_2, a_3 та $\hat{a}_0, \hat{a}_1, \hat{a}_2, \hat{a}_3$ – відповідно параметри моделі, які потрібно оцінити, та їх оцінки; u – стохастична складова.

1. Знайдемо МНК-оцінки параметрів моделі (3.3). Для цього складемо вектор-стовпець Y і матрицю X :

$$Y = \begin{pmatrix} 15,7 \\ 17,34 \\ 21,57 \\ 33,5 \\ 32,3 \\ 37,9 \\ 40,78 \\ 48,02 \\ 43,3 \\ 49,57 \\ 52,14 \\ 55,17 \\ 59,18 \\ 62,22 \\ 77,58 \end{pmatrix}; \quad X = \begin{pmatrix} 1 & 17,37 & 5,28 & 1,42 \\ 1 & 18,24 & 6,47 & 1,58 \\ 1 & 22,47 & 6,98 & 1,98 \\ 1 & 18,47 & 7,05 & 2,04 \\ 1 & 16,82 & 7,94 & 2,38 \\ 1 & 17,6 & 8,12 & 3,48 \\ 1 & 17,12 & 8,69 & 3,07 \\ 1 & 19,81 & 9,31 & 3,84 \\ 1 & 18,67 & 10,45 & 4,28 \\ 1 & 20,83 & 10,47 & 4,67 \\ 1 & 22,84 & 13,48 & 5,98 \\ 1 & 28,85 & 15,78 & 6,51 \\ 1 & 29,61 & 17,65 & 7,82 \\ 1 & 35,67 & 18,47 & 8,58 \\ 1 & 47,87 & 19,64 & 9,47 \end{pmatrix}.$$

Обчислимо оцінки регресійних коефіцієнтів за формулою

$$\hat{a} = (X'X)^{-1} X'Y,$$

де X' – транспонована матриця X ,

$$X' = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ 17,37 & 18,24 & 22,47 & \dots & 35,67 & 47,87 \\ 5,28 & 6,47 & 6,98 & \dots & 18,47 & 19,64 \\ 1,42 & 1,58 & 1,98 & \dots & 8,58 & 9,47 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 15 & 352,24 & 165,78 & 67,1 \\ 352,24 & 9335,74 & 4404,383 & 1858,071 \\ 165,78 & 4404,38 & 2147,268 & 914,9516 \\ 67,1 & 1858,07 & 914,9516 & 397,2576 \end{pmatrix};$$

$$(X'X)^{-1} = \begin{pmatrix} 2,14866 & -0,0276 & -0,51745 & 0,958316 \\ -0,0276 & 0,00428 & -0,0056 & -0,00245 \\ -0,5174 & -0,0056 & 0,18797 & -0,31932 \\ 0,95831 & -0,0024 & -0,31932 & 0,58755 \end{pmatrix};$$

$$X'Y = \begin{pmatrix} 646,27 \\ 16861,1 \\ 8209,78 \\ 3498,18 \end{pmatrix}; \quad \hat{a} = \begin{pmatrix} 26,10789 \\ -0,2518 \\ -2,72767 \\ 11,85602 \end{pmatrix}.$$

Отже, функція регресії з урахуванням знайдених оцінок коефіцієнтів моделі набуває вигляду

$$\hat{y} = 26,10789 - 0,2518x_1 - 2,72767x_2 + 11,85602x_3. \quad (3.5)$$

2. Для перевірки адекватності отриманої моделі обчислимо:

а) її залишки $u_i = y_i - \hat{y}_i$, $i = 1, 2, \dots, n$, де y_i – задані спостереження, а \hat{y}_i визначені за формулою (3.5) при заданих спостереженнях факторів x_1, x_2, x_3 .

Зауваження. Обчислення значень \hat{y}_i можна виконати у матричному вигляді за формулою $\hat{Y} = Xa$, де \hat{Y} – вектор значень \hat{y}_i , $i = 1, 2, \dots, n$.

$$\hat{Y} = \begin{pmatrix} 24,1675 \\ 22,5995 \\ 24,8857 \\ 26,4133 \\ 28,4322 \\ 40,7864 \\ 34,4915 \\ 41,2522 \\ 43,6463 \\ 47,6718 \\ 54,4867 \\ 52,9835 \\ 63,2227 \\ 68,4707 \\ 72,7592 \end{pmatrix};$$

б) відносну похибку розрахункових значень регресії:

$$\delta_i = \frac{\hat{u}_i}{y_i} \cdot 100\%,$$

$$\delta_i = \begin{pmatrix} -0,53934 \\ -0,30332 \\ -0,15372 \\ 0,21154 \\ 0,11974 \\ -0,07616 \\ 0,15420 \\ 0,14093 \\ -0,008 \\ 0,03829 \\ -0,04501 \\ 0,03963 \\ -0,06831 \\ -0,10046 \\ 0,06213 \end{pmatrix};$$

середнє значення відносної похибки:

$$\bar{\delta} = \frac{\sum_{i=1}^n \delta_i}{n};$$

$$\bar{\delta} = -0,52783;$$

в) середньоквадратичну похибку дисперсії залишків:

$$\hat{S}_u = \sqrt{\hat{\sigma}_u^2} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - m - 1}} = \sqrt{\frac{\sum_{i=1}^n u_i^2}{n - m - 1}} = \sqrt{\frac{u^T u}{n - m - 1}}$$

(чим менша стандартна похибка S , тим краще функція регресії відповідає дослідним даним);

$$\hat{S}_u = 5,7357;$$

г) коефіцієнт детермінації, тобто перевіримо загальний вплив незалежних змінних на залежну змінну:

$$R^2 = 1 - \frac{\sum_{i=1}^n u_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

$$R^2 = 1 - \frac{Y'Y - B'X'Y}{Y'Y - n\bar{y}^2} = \frac{B'X'Y - n\bar{y}^2}{Y'Y - n\bar{y}^2};$$

$$R^2 = 0,91436.$$

Висновок: оскільки коефіцієнт детермінації наближається до одиниці, варіація залежної змінної Y значною мірою визначається варіацією незалежних змінних;

д) вибірковий коефіцієнт множинної кореляції:

$$R = \sqrt{R^2};$$

$$R = 0,956222.$$

Коефіцієнт кореляції досить великий, тому існує тісний лінійний зв'язок усіх незалежних факторів x_1, x_2, x_3 із залежною змінною y .

3. Перевіримо статистичну значущість отриманих результатів:

а) обчислимо F -статистику за формулою (спрощений варіант для перевірки нульової гіпотези: $H_0 : a_1 = a_2 = \dots = a_m = 0$):

$$F_{\text{експ}} = \frac{R^2}{1 - R^2} \frac{n - m - 1}{m};$$

$$F_{\text{експ}} = 39,14827.$$

Знайдемо табличне значення F -статистики $F(m, n - m - 1, \alpha)$ (дод. 5):

$$F(3; 11; 0,05) = 3,59.$$

Порівняємо його з обчисленою F -статистикою.

Оскільки $F_{\text{експ}} > F(3; 11; 0,05)$, нульова гіпотеза відхиляється, тобто коефіцієнти регресії є значущими;

б) обчислимо t -статистику:

$$t = \frac{R\sqrt{n-m-1}}{\sqrt{1-R^2}};$$

$$t = 37,03215.$$

Знайдемо відповідне табличне значення t -розподілу з $(n-m-1) = 11$ ступенями свободи і рівнем значущості $\alpha = 0,05$ (дод. 4): $t_{\text{табл}}(\alpha/2, n-m-1)$;

$$t_{\text{табл}}(0,025; 11) = 2,593097.$$

Оскільки $|t| > t_{\text{табл}}(0,025; 11)$, можна зробити висновок про достовірність коефіцієнта кореляції, який характеризує тісноту зв'язку між залежною та незалежними змінними моделі.

Для вибраного рівня значущості $\alpha = 0,05$ і відповідного ступеня свободи $k = n - m - 1 = 11$ запишемо довірчі межі для множинного коефіцієнта кореляції R :

$$(R - \Delta R; R + \Delta R),$$

$$\text{де } \Delta R = t_{\alpha/2, k} \frac{1-R}{\sqrt{n}};$$

$$\Delta R = 2,593(1 - 0,956222) / \sqrt{15} = 0,029311;$$

$$(R - \Delta R; R + \Delta R) = (0,926911; 0,985533);$$

в) перевіримо значущість окремих коефіцієнтів регресії. Визначимо t -статистику за формулою

$$t_j = \frac{a_j}{\sqrt{\sigma_u^2 c_{jj}}} = \frac{a_j}{S_{a_j}} \quad (j = 0, 1, \dots, m), \quad m = 3,$$

де c_{jj} – діагональний елемент матриці $(X'X)^{-1}$; S_{a_j} – стандартизована похибка оцінки параметра моделі;

$$t_0 = 3,105278; t_1 = -0,67081; t_2 = -1,09688; t_3 = 2,696681.$$

Значення t -критерію порівнюємо з табличним при $k = n - m - 1 = 11$ ступенях свободи і рівні значущості $\alpha = 0,05$: $t_{\text{табл.}}(0,025; 11) = 2,593097$.

Оскільки $|t_0| > t_{\alpha/2, k}$, $|t_1| < t_{\alpha/2, k}$, $|t_2| < t_{\alpha/2, k}$, $|t_3| > t_{\alpha/2, k}$, відповідно оцінки \hat{a}_0, \hat{a}_3 є значущими, а оцінки \hat{a}_1, \hat{a}_2 не є значущими.

Обчислимо відношення

$$\delta_{a_j} = \frac{S_{a_j}}{a_j} \cdot 100\%;$$

$$\delta_{a_0} = 32\%; \delta_{a_1} = 94\%; \delta_{a_2} = 91\%; \delta_{a_3} = 37\%$$

(значення δ характеризують той факт, що оцінки a_0, a_3 – незміщені, а оцінки a_1, a_2 – зміщені);

г) знайдемо значення граничного внеску j -го регресора в коефіцієнт детермінації (тобто визначимо, на яку величину зменшиться частковий коефіцієнт детермінації, якщо j -й регресор буде виключений з рівняння):

$$\Delta R_j^2 = \frac{(1 - R^2)t_j^2}{n - m - 1},$$

$$\text{де } t_j = \frac{a_j}{\sqrt{\sigma_u^2 c_{jj}}} = \frac{a_j}{S_{a_j}};$$

$$\Delta R_0^2 = 3,1053; \Delta R_1^2 = -0,6708; \Delta R_2^2 = -1,0969; \Delta R_3^2 = 2,6967;$$

д) визначимо коефіцієнт детермінації, скоригований за Тейлом:

$$\bar{R}_T^2 = 1 - (1 - R^2) \frac{n - 1}{n - m - 1};$$

$$\bar{R}_T^2 = 0,88.$$

Обчислимо коефіцієнт детермінації, скоригований за Амемією:

$$\bar{R}_A^2 = 1 - (1 - R^2) \frac{n + m + 1}{n - m - 1};$$

$$\bar{R}_A^2 = 0,83.$$

Висновок: із виключенням змінної із рівняння втрачається один ступінь свободи, тоді з двох варіантів рівнянь, які мають однакові інші критерії якості, перевага віддається рівнянню з більшим значенням скоригованого коефіцієнта детермінації (при включенні додаткового регресора \bar{R}_T^2 відображує втрату ступеня свободи більш чітко, ніж \bar{R}_A^2 , тобто в цьому разі $\bar{R}_T^2 > \bar{R}_A^2$).

4. Обчислимо коефіцієнти еластичності:

$$\alpha_i = \frac{\partial \hat{y}}{\partial x_i} \cdot \frac{\bar{x}_i}{\bar{y}};$$

$$\alpha_1 = -0,13724; \alpha_2 = -0,6997; \alpha_3 = 1,23097.$$

Коефіцієнт еластичності є показником впливу зміни питомої ваги x_i на y у припущенні, що вплив інших факторів відсутній: у нашому випадку він показує, що *прибуток підприємства зменшиться на 0,14 %, якщо витрати на рекламу зростуть на 1 %; прибуток підприємства збільшиться на 1,24 %, якщо заробітна плата зросте на 1 %.*

Загальна еластичність Y від усіх факторів x_i :

$$\alpha = \sum_{i=1}^m \alpha_i;$$

$$\alpha = 0,394033.$$

Загальна еластичність показує, що *прибуток підприємства збільшиться на 0,39 %, якщо одночасно збільшити на 1 % усі фактори (інвестиції, витрати на рекламу та заробітну плату).*

5. Обчислимо довірчі інтервали для математичного сподівання \hat{y} і для кожного спостереження $X_i = (x_{1(i)}, x_{2(i)}, x_{3(i)})$ (будемо називати їх довірчими зонами):

$$\left(\hat{y}_i - t_{\alpha/2, k} \hat{S}_u \sqrt{X_i' (X'X)^{-1} X_i}; \quad \hat{y}_i + t_{\alpha/2, k} \hat{S}_u \sqrt{X_i' (X'X)^{-1} X_i} \right),$$

де \hat{S}_u — незміщена оцінка дисперсії залишків: $\hat{S}_u = 5,7357$;

$t_{\text{табл}}(\alpha/2, k)$ — відповідне табличне значення t -розподілу з $k = n - m - 1 = 11$ ступенями свободи і рівнем значущості $\alpha = 0,05$:

$$t_{\text{табл}}(0,025; 11) = 2,593097.$$

Виконавши необхідні розрахунки, отримаємо довірчі зони регресії:

(23,430; 24,904)

(22,594; 22,604)

(24,730; 25,041)

(26,253; 26,573)

(28,241; 28,623)

(40,457; 41,115)

(34,374; 34,608)

(41,109; 41,395)

(43,501; 43,790)

(47,430; 47,913)

(54,290; 54,683)

(52,642; 53,325)

(62,887; 63,558)

(68,229; 68,712)

(71,961; 73,556)

6. Побудуємо довірчі інтервали для параметрів регресії. Довірчі інтервали для параметрів a обчислюються так:

$$\left(\hat{a}_j - t_{\alpha/2, k} \sqrt{\sigma_u^2 c_{jj}}; \quad \hat{a}_j + t_{\alpha/2, k} \sqrt{\sigma_u^2 c_{jj}} \right),$$

де c_{jj} — діагональний елемент матриці $(X'X)^{-1}$; $\sigma_u^2 = 32,89835$;

$$t_{\alpha/2, k} = t_{\text{табл}}(0,025; 11) = 2,593097.$$

Виконавши необхідні розрахунки, отримаємо

$$a_0 \in (4,31; 47,91);$$

$$a_1 \in (-1,23; 0,72);$$

$$a_2 \in (-9,17; 3,72);$$

$$a_3 \in (0,46; 23,26).$$

7. Обчислимо прогнозні значення і знайдемо межі довірчих інтервалів індивідуальних прогнозних значень і межі довірчих інтервалів для математичного сподівання (точковий та інтервальні прогнози):

а) для розрахунку прогнозних значень $y_{pi} = Y_{\text{пр}}$ у рівняння (3.5)

$$\hat{y} = 26,10786 - 0,2518x_1 - 2,72767x_2 + 11,85602x_3$$

підставимо прогнозні значення факторів $x_{1\text{пр}} = 48,82$, $x_{2\text{пр}} = 20,04$, $x_{3\text{пр}} = 10,25$, що лежать за межами базового періоду (точковий прогноз):

$$y_{\text{пр}} = 80,68;$$

б) знайдемо межі інтервального прогнозу індивідуального значення (для $k = n - m - 1 = 11$ ступенів свободи та вибраного рівня значущості $\alpha = 0,05$) за формулою

$$\hat{Y}_{\text{пр}} - t_{\alpha/2} \sigma_u \sqrt{1 + X'_{\text{пр}} (X'X)^{-1} X_{\text{пр}}} \leq \hat{Y}_{\text{пр}} \leq \hat{Y}_{\text{пр}} + t_{\alpha/2} \sigma_u \sqrt{1 + X'_{\text{пр}} (X'X)^{-1} X_{\text{пр}}};$$

$$\hat{Y}_{\text{пр}} = 80,68; \quad \sigma_u = 5,7357; \quad t_{\alpha/2} = t_{\text{табл}}(0,025; 11) = 2,5931;$$

$$X_{\text{пр}} = (48,82; 20,04; 10,25);$$

(58,72; 102,64) – інтервальний прогноз індивідуального значення;

в) знайдемо межі довірчого інтервалу для математичного сподівання значення $y_{\text{пр}}$ за формулою

$$\hat{Y}_{\text{пр}} - t_{\alpha/2} \sigma_u \sqrt{X'_{\text{пр}} (X'X)^{-1} X_{\text{пр}}} \leq M(Y_{\text{пр}}) \leq \hat{Y}_{\text{пр}} + t_{\alpha/2} \sigma_u \sqrt{X'_{\text{пр}} (X'X)^{-1} X_{\text{пр}}};$$

$$\hat{Y}_{\text{пр}} = 80,68; \quad \sigma_u = 5,7357; \quad t_{\alpha/2} = t_{\text{табл}}(0,025; 11) = 2,5931;$$

(64,52; 96,83) – довірчий інтервал для математичного сподівання прогнозного значення.

Контрольні запитання

1. Навести приклади застосування багатofакторного регресійного аналізу.
2. Дати визначення лінійної багатofакторної регресійної моделі. При яких основних припущеннях вона досліджується?
3. Прокоментувати основні етапи побудови лінійної багатofакторної регресійної моделі.
4. Як розраховуються невідомі параметри лінійної моделі багатofакторної регресії за методом МНК?
5. Які основні властивості методу найменших квадратів у випадку багатofакторної лінійної регресії?
6. Як визначаються коефіцієнти множинної кореляції та детермінації в багатofакторній регресійній моделі?
7. Як визначається оцінений коефіцієнт детермінації? Навести його основні властивості.
8. Як побудувати довірчі інтервали параметрів лінійної регресії?
9. У чому відмінність між точковими та інтервальними прогнозами?
10. Якбудеться інтервальний прогноз значень залежної змінної в моделі лінійної регресії?

Вправи та завдання

1. На основі десяти статистичних даних:
 - 1) визначити параметри лінійної моделі залежності витрат на споживання (C) від рівня доходів (D), заощаджень (S) і заробітної плати (L);
 - 2) оцінити коефіцієнт детермінації;
 - 3) оцінити коефіцієнт множинної кореляції;
 - 4) перевірити їх статистичну значущість.

i	$C(i)$	$D(i)$	$S(i)$	$L(i)$
1	45,64	4,75	1,79	58,45
2	44,72	6,28	3,11	53,3
3	52,12	6,87	2,71	56,25
4	50,96	7,85	4,07	58,2
5	50,88	9,26	4,21	45,7
6	58,56	10,39	4,67	58,05
7	56,92	9,42	6,3	65,6
8	58,2	11,01	6	60,85
9	60,36	12,22	4,82	60,55
10	62,76	12,82	4,65	63,50

2. За статистичними показниками Y , K і L за дев'ять років проаналізувати класичну модель виробничої функції Кобба – Дугласа, що описує залежність між продуктивністю праці $y = Y/L$ та фондоозброєністю $x = K/L$ з урахуванням впливу технічного прогресу у виробництві регіону, оцінивши:

- 1) параметри нелінійної моделі;
- 2) коефіцієнт детермінації;
- 3) коефіцієнт множинної кореляції.

i	$Y(i)$	$K(i)$	$L(i)$
1	65,04	4,03	7,45
2	54,27	5,25	8,68
3	78,22	7,57	9,55
4	82,06	7,99	10,67
5	79,14	8,91	11,68
6	90,48	10,67	13,31
7	85,69	11,51	14,27
8	76,26	10,23	13,01
9	82,05	10,84	15,05

Розділ 4. Мультиколінеарність

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

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

Означення 4.1. *Суть мультиколінеарності полягає в тому, що в багатофакторній регресійній моделі дві або більше незалежних змінних пов'язані між собою лінійною залежністю або, іншими словами, мають високий ступінь кореляції:*

$$(r_{x_i x_j} \rightarrow 1, i \neq j).$$

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

Передусім потрібно зрозуміти природу мультиколінеарності.

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

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

Практичні наслідки мультиколінеарності:

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

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

Для ілюстрації розглянемо двофакторну регресійну модель

$$y = a_0 + a_1 x_1 + a_2 x_2 + u$$

та її вибірковий аналог

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2.$$

Дисперсія оцінок параметрів \hat{a}_1 і \hat{a}_2 має вигляд

$$D(\hat{a}_0) = \frac{\sigma_\xi^2}{(1-r^2) \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2}, \quad (4.1)$$

$$D(\hat{a}_2) = \frac{\sigma_\xi^2}{(1-r^2) \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2}, \quad (4.2)$$

$$\text{cov}(\hat{a}_1, \hat{a}_2) = \frac{-r\sigma_\xi^2}{(1-r^2) \sqrt{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2}}, \quad (4.3)$$

де r – коефіцієнт кореляції між x_1 і x_2 .

З (4.1), (4.2) випливає, що якщо r зростає, то $D(\hat{a}_1)$, $D(\hat{a}_2)$ також зростають.

З (4.3) випливає, що якщо r збільшується, $\text{cov}(\hat{a}_1, \hat{a}_2)$ зростає за абсолютною величиною. Причому при наближенні до граничного значення це збільшення має експоненціальний характер.

- збільшення довірчого інтервалу (оскільки збільшується середній квадрат відхилення параметрів);
- незначущість t -статистик.

(Оскільки значення t -статистики Стьюдента $t = \frac{\hat{a}_1}{\sigma_{\hat{a}_1}}$, то у випадку мультиколінеарності $\sigma_{\hat{a}_1} \rightarrow \infty$, а отже, $t \rightarrow 0$).

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

4.2. Тестування наявності мультиколінеарності

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

- велике значення R^2 і незначущість t -статистики.

Наявність цих двох факторів одночасно є “класичною” ознакою мультиколінеарності.

З одного боку, незначущість t -статистики Стьюдента означає, що один або більше оцінених параметрів статистично незначуще відрізняються від нуля. З іншого боку, якщо значення R^2 велике, ми приймаємо з великою ймовірністю F -критерій Фішера, який відкидає нульову гіпотезу ($H_0 : a_1 = a_2 = \dots = a_m = 0$). Суперечність свідчить про наявність мультиколінеарності;

- велике значення парних коефіцієнтів кореляції.

Якщо значення хоча б одного коефіцієнта кореляції $r_{x_i x_j} > 0,8$, $i \neq j$, то мультиколінеарність є серйозною проблемою.

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

Для визначення мультиколінеарності здебільшого застосовують такі тести:

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

- *характеристичні значення та умовний індекс*.

Розглянемо їх більш детально.

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

Нехай $R_{x_i, x_1, x_2, \dots, x_m}^2$ — коефіцієнт детермінації в регресії, яка пов'язує фактор x_i з іншими факторами. Тоді *F-тест* виконується так:

1) для кожного коефіцієнта детермінації розраховуємо F_i -відношення:

$$F_i = \frac{(R_{x_i, x_1, x_2, \dots, x_m}^2) / (m - 1)}{(1 - R_{x_i, x_1, x_2, \dots, x_m}^2) / (n - m)}, \quad (4.4)$$

де n — кількість спостережень; m — кількість факторів.

F-тест перевіряє гіпотезу $H_0: R_{x_i, x_1, \dots, x_m}^2 = 0$ проти гіпотези $H_1:$

$$R_{x_i, x_1, \dots, x_m}^2 \neq 0;$$

2) $F_{кр}$ знаходимо за таблицею *F-розподілу* Фішера з $(m-1)$ і $(n-m)$ ступенями свободи і заданим рівнем значущості;

3) якщо $F_i > F_{кр}$, то гіпотезу H_0 відкидаємо (x_i — мультиколінеарний фактор), якщо $F_i < F_{кр}$, то гіпотезу H_0 приймаємо (фактор x_i не є мультиколінеарним).

Тест, що застосовує *характеристичні значення* (власні числа матриці спостережень) та *умовний індекс R* (що обчислюється як відношення максимального власного числа матриці до її мінімального власного чис-

ла), використовується в сучасних статистичних пакетах. Ми не розглядатимемо його детально, бо це потребує застосування апарату теорії матриць.

Зазначимо лише, що за цим тестом розраховується не тільки умовне число R , а й умовний індекс $CI = \sqrt{R}$. Якщо $100 \leq R \leq 1000$, мультиколінеарність *помірна*, якщо $R > 1000$ — *висока*. Аналогічно, якщо $10 \leq CI \leq 30$, мультиколінеарність *помірна*, якщо $CI > 0$ — *висока*.

Ми розглянули лише основні методи тестування мультиколінеарності. Жоден з них не є універсальним. Усі вони мають один спільний недолік: жоден із них не проводить чіткої межі між тим, що треба вважати “суттєвою” мультиколінеарністю, яку необхідно враховувати, і тим, коли нею можна знехтувати.

4.3. Алгоритм Фаррара — Глобера

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

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

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

Складемо **покроковий алгоритм Фаррара — Глобера**.

1-й крок:

нормалізувати змінні x_1, x_2, \dots, x_m економетричної моделі, обчисливши

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

де n — кількість спостережень ($i = 1, 2, \dots, n$); m — кількість незалежних змінних ($j = 1, m$); \bar{x}_j — середня арифметична j -ї незалежної змінної; $\sigma_{x_j}^2$ — дисперсія j -ї незалежної змінної.

2-й крок:

на основі матриці X^* , елементами якої є нормалізовані незалежні змінні x_{ij}^* , обчислити кореляційну матрицю (матрицю моментів нормалізованої системи нормальних рівнянь):

$$R = X^{*tr} X^* = \begin{pmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{pmatrix},$$

де X^{*tr} — транспонована матриця X^* (елементи матриці R характеризують щільність зв'язку однієї незалежної змінної з іншою);

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

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

3-й крок:

визначити $|R|$ — визначник кореляційної матриці R ;

обчислити критерій χ^2 :

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

порівняти значення χ^2 з табличним при $\frac{1}{2} m(m-1)$ ступенях свободи і рівні значущості α (якщо $\chi^2 > \chi_{\text{табл}}^2$, то в масиві незалежних змінних існує мультиколінеарність).

4-й крок:

визначити матрицю похибок:

$$C = R^{-1} = (X^{*tr} X^*)^{-1} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mm} \end{pmatrix}.$$

5-й крок:

розрахувати F -критерії:

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

де c_{kk} — діагональні елементи матриці C ;

значення критеріїв F_k порівняти з табличним при $(n - m)$ і $(m - 1)$ ступенях свободи й рівні значущості α (якщо $F_k > F_{\text{табл}}$, то відповідна k -та незалежна змінна мультиколінеарна з іншими);

розрахувати коефіцієнти детермінації для кожної змінної:

$$R_k^2 = 1 - \frac{1}{c_{kk}}.$$

6-й крок:

знайти часткові коефіцієнти кореляції, які характеризують щільність зв'язку між двома змінними за умови, що інші змінні $x_{l1}, x_{l2}, \dots, x_{lm}$ не впливають на цей зв'язок (існування парної мультиколінеарності):

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

де c_{kj} — елементи матриці C , що розміщені в k -му рядку та j -му стовпці, $k = 1, 2, \dots, m$; $j = 1, 2, \dots, m$; c_{kk} і c_{jj} — діагональні елементи матриці C .

Однак якщо порівняти конкретні числові значення часткових і парних коефіцієнтів, то можна побачити, що перші значно менші,

ніж останні. Тому на основі знання парних коефіцієнтів кореляції висновок про мультиколінеарність робити неможливо. Для цього необхідно виконати 7-й крок.

7-й крок:

розрахувати t -критерії:

$$t_{kj} = |r_{kj}| \cdot \frac{\sqrt{n-m}}{\sqrt{1-r_{kj}^2}};$$

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

Висновки:

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

2. Якщо $F_k > F_{\text{табл}}$, то x_k залежить від усіх інших незалежних змінних і треба вирішити питання про її виключення з переліку змінних.

3. Якщо $t_{kj} > t_{\text{табл}}$, то x_k і x_j щільно пов'язані між собою.

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

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

Приклад дослідження наявності мультиколінеарності на основі алгоритму Фаррара — Глобера

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

№ п/п	$y(i)$	$x_1(i)$	$x_2(i)$	$x_3(i)$
1	25,74	4,69	11,97	29,23
2	25,34	5,64	13,43	29,35
3	31,26	6,26	12,92	33,40
4	33,50	6,99	14,74	30,97
5	32,30	6,36	14,64	32,92
6	38,90	7,60	17,10	37,27
7	41,58	7,12	15,63	30,97
8	48,02	6,81	15,35	33,58
9	43,30	8,67	15,85	35,62
10	51,78	7,83	18,05	34,99
11	52,14	7,84	17,24	39,34
12	54,94	8,85	20,52	41,50
13	59,18	9,61	19,18	45,58
14	62,22	10,67	19,03	41,08
15	63,62	11,04	21,45	40,54
16	65,01	11,85	22,25	42,75
17	67,78	12,94	24,75	43,89
18	71,45	14,24	25,03	41,95
19	75,24	15,67	27,87	44,06
20	77,38	16,33	30,486	46,77

Розв'язання.

1-й крок:

нормалізуємо змінні x_1, x_2, x_3 економетричної моделі, обчисливши

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

де $n=20$ – кількість спостережень ($i=1, 2, \dots, n$); $m=3$ – кількість незалежних змінних ($j=1, m$); \bar{x}_j – середня арифметична j -ї незалежної змінної:

$$\bar{x}_1 = 9,3505; \quad \bar{x}_2 = 18,874; \quad \bar{x}_3 = 37,788;$$

$\sigma_{x_j}^2$ – дисперсія j -ї незалежної змінної:

$$\sigma_{x_1}^2 = 11,35297; \quad \sigma_{x_2}^2 = 26,06648; \quad \sigma_{x_3}^2 = 31,86169.$$

2-й крок:

на основі нової матриці X^* , елементами якої є нормалізовані не-

залежні змінні x_{ij}^* ,

$$X^* = \begin{pmatrix} 0,309287 & -0,302374 & 0,339018 \\ 0,246242 & -0,238430 & 0,334264 \\ 0,205096 & -0,260766 & 0,173826 \\ 0,156651 & -0,181056 & 0,270089 \\ 0,198460 & -0,185436 & 0,192841 \\ 0,116169 & -0,077695 & 0,020520 \\ 0,148024 & -0,142077 & 0,270089 \\ 0,168596 & -0,154340 & 0,166696 \\ 0,045160 & -0,132441 & 0,085883 \\ -0,100905 & -0,036088 & 0,110840 \\ -0,100242 & -0,071564 & 0,061481 \\ -0,033215 & 0,072089 & 0,147047 \\ 0,017221 & 0,013401 & 0,308673 \\ 0,087566 & 0,006832 & 0,130409 \\ 0,112121 & 0,112820 & 0,109018 \\ 0,165875 & 0,147858 & 0,196565 \\ 0,238212 & 0,257350 & 0,241725 \\ 0,324485 & 0,269613 & 0,164874 \\ 0,419385 & 0,393997 & 0,248460 \end{pmatrix}$$

обчислимо кореляційну матрицю (матрицю моментів нормалізованої системи нормальних рівнянь):

$$R = X^{*tr} X^* = \begin{pmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{pmatrix},$$

де X^{*tr} – транспонована матриця X^* ; елементи матриці R характеризують щільність зв'язку однієї незалежної змінної з іншою ($r_{ij} = r_{x_i x_j}$ – парні коефіцієнти кореляції);

$$R = \begin{pmatrix} 0,95 & 0,92883 & 0,819534 \\ 0,92883 & 0,95 & 0,831277 \\ 0,81953 & 0,83127 & 0,95 \end{pmatrix}.$$

3-й крок:

визначимо $|R|$ – визначник кореляційної матриці R :

$$|R| = 0,00881;$$

обчислимо значення критерію χ^2 :

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

$$\chi^2 = 81,23005;$$

порівняємо значення χ^2 з табличним при $\frac{1}{2}m(m-1) = 3$ ступенях свободи й рівні значущості $\alpha = 0,05$ (дод. 3):

$$\chi_{\text{табл}}^2 = 7,814724.$$

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

4-й крок:

визначимо матрицю похибок:

$$C = R^{-1} = (X^{*tr} X^*)^{-1} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1m} \\ c_{21} & c_{22} & \dots & c_{2m} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mm} \end{pmatrix};$$

$$C = \begin{pmatrix} 24,00377 & -22,82912 & -0,7311322 \\ -22,82912 & 26,20416 & -3,235457 \\ 0,731132 & -3,235457 & 4,5144749 \end{pmatrix}.$$

5-й крок:

розрахуємо F -критерії:

$$F_k = \frac{(c_{kk} - 1)(n - m)}{(m - 1)}, \quad k = 1, 2, 3,$$

де c_{kk} — діагональні елементи матриці C ;

$$F_1 = 195,53205,$$

$$F_2 = 214,2354,$$

$$F_3 = 29,87303;$$

значення критеріїв F_k порівняємо з табличним при $(n - m) = 17$ і $(m - 1) = 2$ ступенях свободи і рівні значущості $\alpha = 0,05$ (дод. 5):

$$F_{\text{табл}} = 19,43703.$$

Оскільки $F_1 > F_{\text{табл}}$, $F_2 > F_{\text{табл}}$, $F_3 > F_{\text{табл}}$, робимо висновок, що *перша, друга й третя незалежні змінні мультиколінеарні з іншими*; визначимо коефіцієнти детермінації для кожної змінної:

$$R_k^2 = 1 - \frac{1}{c_{kk}};$$

$$R_1^2 = 0,958339; \quad R_2^2 = 0,961838; \quad R_3^2 = 0,778490.$$

6-й крок:

знайдемо часткові коефіцієнти кореляції, які характеризують щільність зв'язку між двома змінними за умови, що інші змінні $x_{j1}, x_{j2}, \dots, x_{jm}$ не впливають на цей зв'язок (існування парної мультиколінеарності):

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

де c_{kj} — елементи матриці C , що розміщені в k -му рядку та j -му стовпці, $k=1, 2, \dots, m$, $j=1, 2, \dots, m$; c_{kk} і c_{jj} — діагональні елементи матриці C ;

$$r_{12}=0,910257, \quad r_{13}=0,070234, \quad r_{23}=0,297472.$$

Однак якщо порівняти абсолютні значення часткових парних коефіцієнтів, то можна побачити, що перші значно менші, ніж останні. Тому на основі знання парних коефіцієнтів кореляції висновок про мультиколінеарність робити неможливо. Для цього необхідно ще виконати 7-й крок.

7-й крок:

розрахуємо t -критерії:

$$t_{kj} = |r_{kj}| \frac{\sqrt{n-m}}{\sqrt{1-r_{kj}^2}};$$

$$t_{12}=9,064506, \quad t_{13}=0,290302, \quad t_{23}=1,284666;$$

значення критеріїв t_{kj} порівняємо з табличними при $(n-m)=17$ ступенях свободи й рівні значущості $\alpha=0,05$ (дод. 4):

$$t_{\text{табл}} = 2,109818.$$

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

Якщо F -критерій перевищує табличне значення, а це означає що k -та змінна залежить від інших змінних у масиві, необхідно вирішувати питання про її виключення з переліку змінних.

Якщо t_{kj} -критерій перевищує табличне значення, то ця пара змінних (k і j) тісно взаємопов'язана. Звідси, аналізуючи рівень обох видів критеріїв F і t , можна зробити обґрунтований висновок про те, яку зі змінних необхідно виключити із дослідження чи замінити її іншою. Але заміна масиву незалежних змінних завжди має узгоджуватися з економічною доцільністю, що зумовлена метою дослідження.

4.4. Засоби усунення мультиколінеарності. Метод головних компонентів

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

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

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

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

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

Алгоритм методу головних компонентів

Цей алгоритм включає дев'ять кроків.

1-й крок:

нормалізувати змінні x_1, x_2, \dots, x_m регресійної моделі, обчисливши

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sigma_{xj}},$$

де n — кількість спостережень ($i = 1, n$);

m — кількість пояснюючих змінних у моделі ($j = 1, m$);

\bar{x}_j — середня арифметична j -ї незалежної змінної;

σ_{xj} — середньоквадратичне відхилення j -ї незалежної змінної.

2-й крок:

побудувати нову матрицю X^* , елементами якої є нормалізовані незалежні змінні.

3-й крок:

обчислити кореляційну матрицю (матрицю моментів нормалізованої системи нормальних рівнянь) за формулою

$$r = \frac{1}{n} (X^{*tr} X^*),$$

де X^{*tr} — транспонована матриця X^* :

$$R = X^{*tr} X^* = \begin{pmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{pmatrix}$$

(недіагональні елементи матриці R характеризують щільність зв'язку однієї незалежної змінної з іншою ($r_{ij} = r_{x_i x_j}$), тобто є парними коефіцієнтами кореляції).

4-й крок:

знайти характеристичні числа матриці r , тобто визначити корені $\lambda_1, \lambda_2, \dots, \lambda_m$ рівняння m -го порядку:

$$|r - \lambda E| = 0,$$

де E – одинична матриця розмірності $m \times m$; $\lambda_j, j = 1, 2, \dots, m$ – характеристичні числа матриці r .

5-й крок:

ранжувати власні значення $\lambda_i, i = 1, 2, \dots, m$, за абсолютним рівнем весу кожного головного компонента в загальну дисперсію.

6-й крок:

розв'язати систему рівнянь

$$|r - \lambda E| \vec{a} = 0$$

і обчислити власні вектори $a_i, i = 1, 2, \dots, m$, за умови, що вони відповідають таким співвідношенням:

$$a'_i a_j = \begin{cases} 1, & \text{якщо } i = j, \\ 0, & \text{якщо } i \neq j. \end{cases}$$

7-й крок:

знайти головні компоненти векторів $z_i = \lambda a_i, i = 1, 2, \dots, m$, які задовольняють умови

$$\sum_{j=1}^n z_{ij} = 0,$$

$$\frac{1}{n} z'_i z_i = \lambda_i, \quad i = 1, 2, \dots, m,$$

$$z'_i z_j = 0, \quad i \neq j, \quad j = 1, 2, \dots, m.$$

8-й крок:

визначити параметри моделі $\hat{Y} = Z\hat{P}$:

$$\hat{P} = Z^{-1}Y.$$

9-й крок:

знайти параметри моделі $\hat{Y} = XA$:

$$A = \bar{a}\hat{P}.$$

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

Контрольні запитання

1. Що означає мультиколінеарність змінних?
2. До яких наслідків призводить мультиколінеарність?
3. Як впливає наявність мультиколінеарності на оцінку параметрів моделі?
4. За допомогою яких методів визначають наявність мультиколінеарності?
5. Які статистичні критерії включає алгоритм Фаррара – Глобера?
6. Охарактеризуйте алгоритм Фаррара – Глобера.
7. Які висновки можна зробити, дослідивши мультиколінеарність за алгоритмом Фаррара – Глобера?
8. Що характеризують елементи кореляційної матриці R ?
9. Як усунути мультиколінеарність? Наведіть методи її усунення.
10. Яким методом можуть бути оцінені параметри моделі з мультиколінеарними змінними?

11. Як обчислити головні компоненти?
12. Як оцінити параметри моделі на основі головних компонентів?
13. Які умови мають задовольняти головні компоненти?
14. Коли доцільно застосовувати метод головних компонентів?

Вправи та завдання

1. На базі $n=11$ статистичних даних певного регіону дослідити мультиколінеарність між факторами.

i	$C(i)$	$D(i)$	$S(i)$	$L(i)$
1	5,25	9,11	7,05	16,05
2	11,24	13,57	8,68	18,68
3	16,27	14,01	9,57	20,06
4	18,75	17,29	10,10	29,67
5	21,78	19,58	11,55	31,55
6	24,58	21,07	13,31	34,01
7	27,09	22,47	15,37	35,34
8	31,76	24,68	17,01	36,01
9	35,94	25,75	19,67	38,54
10	38,57	27,05	21,92	41,92
11	41,47	30,87	25,08	43,27

2. Запропонувати способи усунення мультиколінеарності.



Розділ 5. ГЕТЕРОСКЕДАСТИЧНІСТЬ

5.1. Виявлення гетероскедастичності та її природа

Розглянемо класичну лінійну багатofакторну модель

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m + u. \quad (5.1)$$

Як завжди,

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad a = \begin{pmatrix} a_0 \\ \vdots \\ a_m \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix},$$
$$X = \{1, x_{ij}\}, \quad i = \overline{1, n}, \quad j = \overline{1, m}. \quad (5.2)$$

Для застосування МНК при оцінюванні параметрів моделі раніше було сформульовано основні припущення, які на практиці можуть порушуватись.

У попередньому розділі розглядався особливий випадок багатofакторного регресійного аналізу, пов'язаний з проблемою мультиколінеарності. Тепер розглянемо інший особливий випадок, що стосується сталості дисперсії кожної випадкової величини u_i (гомоскедастичність залишків).

Означення 5.1. Якщо дисперсія залишків стала для кожного спостереження, то це явище називається гомоскедастичністю:

$$Du_i = M(u_i - Mu_i)^2 = \sigma_{u_i}^2 = \text{const}, \quad i = \overline{1, n}. \quad (5.3)$$

Якщо це припущення не задовольняється в якомусь окремому випадку, то маємо гетероскедастичність (помилки u_i некорельовані, але мають несталу дисперсію).

Означення 5.2. Якщо дисперсія залишків змінюється для кожного спостереження або групи спостережень, то це явище називається гетероскедастичністю:

$$D u_i = M (u_i - M u_i)^2 = \sigma_{u_i}^2 \neq \text{const}, \quad i = \overline{1, n}. \quad (5.4)$$

Розглянемо питання про доцільність припущення (5.3) і про те, що відбувається, якщо це припущення не задовольняється.

Насамперед зауважимо, що сутність припущення про гомоскедастичність полягає в тому, що варіація кожної випадкової складової u_i навколо її математичного сподівання не залежить від значення факторів x :

$$\sigma_{u_i}^2 \neq f(x_1, x_2, \dots, x_{p_i}).$$

Форма гетероскедастичності залежить від знаків і значень коефіцієнтів у залежності

$$\sigma_{u_i}^2 = f(x_1, x_2, \dots, x_{p_i}).$$

Оскільки u_i — не спостережувана випадкова величина, ми не знаємо справжньої форми гетероскедастичності.

У прикладних дослідженнях, як правило, використовують зручне припущення, а саме в разі простої лінійної регресії гетероскедастичність має форму

$$\sigma_{u_i}^2 = k^2 x^2$$

($k = \text{const}$, яку потрібно оцінити).

Наслідки порушення припущення про гомоскедастичність:

1) неможливо знайти середньоквадратичне відхилення параметрів $\sigma_{a_i}^2$ регресії, а отже, неможливо оцінити значущість параметрів;

2) неможливо побудувати довірчий інтервал для прогнозних значень $y_{\text{пр}}$;

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

Зазначимо, що якщо незважаючи на гетероскедастичність ми використовуватимемо звичайні процедури перевірки гіпотез, то вис-

новки можуть бути неправильними. Зрозуміло, гетероскедастичність є суттєвою проблемою, а тому потрібно вміти з'ясувати її наявність.

5.2. Тестування наявності гетероскедастичності

Як і в разі мультиколінеарності, єдиних правил виявлення гетероскедастичності немає, а є різноманітні тести (критерії): критерій μ , параметричний та непараметричний тести Гольдфельда — Квандта, тест Глейсера, тест рангової кореляції Спірмана та ін. Розглянемо лише деякі з них.

Зауважимо, що інколи в ході проведення економетричних досліджень гетероскедастичність вгадується інтуїтивно або висувається як абсолютне припущення:

$$\sigma_{u_i}^2 = f(x_1, x_2, \dots, x_{p_i}).$$

Наприклад, вивчаючи бюджет сім'ї, можна помітити, що дисперсія залишків зростає відповідно до зростання доходу. Отже, перший крок до виявлення гетероскедастичності – глибокий **аналіз змісту** досліджуваної **проблеми**.

Крім того, існує графічний метод тестування наявності гетероскедастичності, що ґрунтується на встановленні наявності систематичного зв'язку квадратів залишків регресійної моделі, побудованої на основі припущення про відсутність гетероскедастичності (**графічний аналіз**).

5.2.1. Параметричний тест Гольдфельда — Квандта

Зауваження. 1. Цей тест застосовується до великих вибірок.
2. Тест припускає нормальний розподіл і незалежність випадкових величин u_i .

1-й крок:

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

2-й крок:

відкинути s спостережень, які розміщені всередині векторів вихідних даних, де $s = \frac{4n}{15}$, n – кількість елементів вектора x_i .

3-й крок:

побудувати дві моделі на основі звичайного МНК за двома створеними сукупностями спостережень обсягом $\frac{n-c}{2}$ за умови, що

$$\frac{n-c}{2} \geq m, \text{ де } m \text{ — кількість змінних.}$$

4-й крок:

знайти суму квадратів залишків S_1 і S_2 за першою і другою моделями:

$$S_1 = u_1^{tr} u_1, \quad S_2 = u_2^{tr} u_2,$$

де u_1 і u_2 — залишки відповідно за першою і другою моделями.

5-й крок:

розрахувати критерій $F^* = \frac{S_2}{S_1}$, який у разі виконання гіпотези

про гомоскедастичність відповідатиме F -розподілу з $\gamma_1 = \frac{(n-c-2m)}{2}$,

$$\gamma_2 = \frac{(n-c-2m)}{2} \text{ ступенями свободи;}$$

значення критерію F^* порівняти з табличним значенням F -критерію при вибраному рівні значущості α і відповідних ступенях свободи; якщо $F^* \leq F_{\text{табл}}$, то гетероскедастичність відсутня.

Зауваження: чим більше значення F^* , тим більша гетероскедастичність залишків.

5.2.2. Непараметричний тест Гольдфельда — Квандта

Цей тест базується на встановленні кількості піків значень залишків після впорядкування (ранжування) спостережень за x_{ij} . Якщо для всіх значень змінної x_{ij} залишки розподіляються приблизно однаково, то дисперсія їх однорідна і гетероскедастичність відсутня. Якщо вона змінюється, то гетероскедастичність присутня.

Зазначимо, що цей тест не цілком надійний для перевірки на гетероскедастичність. Однак він дуже простий і часто використовується для першої оцінки наявності гетероскедастичності множини спостережень.

5.2.3. Тест Глейсера

Перевірка на гетероскедастичність базується на побудові регресійної функції, що характеризує залежність величини залишків за модулем від пояснюючої змінної x_j , яка може зумовити зміну дисперсії залишків.

Аналітична форма регресійних функцій може мати вигляд

$$\begin{aligned} |u| &= a_0 + a_1 x_j, \\ |u| &= a_0 + a_1 x_j^{-1}, \\ |u| &= a_0 + a_1 x_j^{1/2} \text{ і т. ін.} \end{aligned}$$

Рішення про відсутність гетероскедастичності залишків приймається на основі значущості коефіцієнтів a_0 і a_1 . Перевага цього методу полягає в можливості розрізнити випадок чистої ($a_0 = 0$, $a_1 \neq 0$) і змішаної ($a_0 \neq 0$, $a_1 \neq 0$) гетероскедастичності. Залежно від цього потрібно використовувати різні матриці S .

Оскільки явище гетероскедастичності пов'язане з тим, що змінюються дисперсії залишків, а коваріація між ними відсутня, то матриця S у співвідношенні $M(uu') = \sigma_u^2 S$ має бути додатно визначеною й діагональною.

Приклад. Перевірити гіпотезу про відсутність гетероскедастичності для побудови моделі, яка характеризує залежність заощаджень від доходів населення. Статистичні дані наведено в таблиці.

Рік	Заощадження	Дохід
1	1,36	14,87
2	1,2	14,4
3	1,7	13,8
4	1,84	15,6
5	2,1	15,94
6	1,12	16,9
7	1,89	17,7
8	2,3	18,67
9	2,5	18,04
10	1,17	19,5
11	1,9	21,4
12	1,95	22,7
13	2,87	25,7
14	2,6	27,18
15	1,75	28,9
16	1,96	29,45
17	1,4	30,07
18	2,99	30,2

Розв'язання. Ідентифікуємо змінні: y – заощадження, x – дохід. Специфікуємо модель у вигляді

$$y = a_0 + a_1x + u,$$

$$\hat{y} = \hat{a}_0 + \hat{a}_1x,$$

де u – стохастична складова моделі.

Для перевірки гіпотези про відсутність гетероскедастичності залишків моделі застосуємо параметричний тест Гольдфельда – Квандта.

1-й крок:

спостереження впорядкуємо за зростанням за величиною доходу (вектор x), який може спричинити зміну дисперсії залишків.

Рік	Заощадження	Дохід
1	1,7	13,8
2	1,2	14,4
3	1,36	14,87
4	1,84	15,6
5	2,1	15,94
6	1,12	16,9
7	1,89	17,7
8	2,5	18,04
9	2,3	18,67
10	1,17	19,5
11	1,9	21,4
12	1,95	22,7
13	2,87	25,7
14	2,6	27,18
15	1,75	28,9
16	1,96	29,45
17	1,4	30,07
18	2,99	30,2

2-й крок:

відкинемо c спостережень усередині вектора вихідних даних, де

$$c = \frac{4n}{15}, n - \text{кількість елементів вектора } x.$$

Отже,

$$c = \frac{4 \cdot 18}{15} \approx 4.$$

Отримаємо дві сукупності спостережень обсягом $\frac{18-4}{2} = 7$.

Перша сукупність спостережень:

1	1,7	13,8
2	1,2	14,4
3	1,36	14,87
4	1,84	15,6
5	2,1	15,94
6	1,12	16,9
7	1,89	17,7

Друга сукупність спостережень:

1	1,95	22,7
2	2,87	25,7
3	2,6	27,18
4	1,75	28,9
5	1,96	29,45
6	1,4	30,07
7	2,99	30,2

3-й крок:

побудуємо дві моделі на основі звичайного МНК за двома створеними сукупностями спостережень:

$$\hat{y}_1 = 0,7411 + 0,05514x,$$

$$\hat{y}_2 = 2,9371 - 0,02595x.$$

4-й крок:

знайдемо суму квадратів залишків S_1 і S_2 за першою і другою моделями:

$$S_1 = u_1^{tr} u_1 = 0,8149, \quad S_2 = u_2^{tr} u_2 = 2,1628,$$

де u_1 і u_2 — залишки відповідно за першою і другою моделями.

5-й крок:

розрахуємо критерій $F^* = \frac{S_2}{S_1} = \frac{2,1628}{0,8149} \approx 2,654$, який у разі вико-

нання гіпотези про гомоскедастичність відповідатиме F -розподілу

з $\gamma_1 = \frac{(n - c - 2m_1)}{2} = 5$, $\gamma_2 = \frac{(n - c - 2m_1)}{2} = 5$ ступенями свободи;
 $m_1 = m + 1$;

значення критерію F^* порівняємо з табличним значенням F -критерію при рівні значущості $\alpha = 0,05$ і відповідних ступенях свободи:

$$F_{\text{табл}} = F(0,05; 5) = 5,05.$$

Оскільки $F^* \leq F_{\text{табл}}$, то гетероскедастичність відсутня. Отже, МНК-оцінки параметрів регресійної моделі можуть застосовуватися для подальших досліджень.

5.3. Трансформування початкової моделі

Розглянемо питання усунення гетероскедастичності трансформуванням початкової моделі.

Припустимо, що за статистичними даними побудовано початкову регресійну модель

$$y = f(x, u)$$

і на базі будь-якого тесту встановлено наявність гетероскедастичності:

$$D_{u_i} = \sigma_{u_i}^2 \neq \text{const.}$$

Для усунення гетероскедастичності початкову модель змінюють (трансформують) так, щоб помилки мали сталу дисперсію:

$$D_{u_i} = \sigma_{u_i}^2 = \text{const.}$$

Трансформація моделі зводиться до зміни початкової форми моделі методом, який залежить від специфічної форми гетероскедастичності, тобто від форми залежності між дисперсіями залишків і значеннями незалежних змінних:

$$\sigma_{u_i}^2 = \varphi(x_i). \quad (5.5)$$

Розглянемо можливі випадки трансформації моделі на прикладі простої лінійної регресії.

Нехай початкова модель

$$y = a_0 + a_1x + u, \quad (5.6)$$

де компоненти випадкового вектора u гетероскедастичні, але відповідають іншим класичним припущенням лінійної регресії.

Розглянемо такі випадки.

Випадок 1. Припустимо, що гетероскедастичність має форму

$$\sigma_{u_i}^2 = Mu_i^2 = k^2 x^2, \quad (5.7)$$

де $k = \text{const}$ (тобто дисперсія залишків зростає пропорційно до x^2).

Із припущення (5.7) випливає

$$k^2 = \sigma_{u_i}^2 / x^2.$$

Це означає, що трансформація моделі полягає в діленні початкової моделі на $\sqrt{x^2} = x$.

Отже, трансформована модель має вигляд

$$\frac{y_i}{x_i} = \frac{a_0}{x_i} + a_1 + \frac{u_i}{x_i}. \quad (5.8)$$

Значимо, що параметр при змінній $1/x_i$ у трансформованій моделі є перетином (вільним членом) початкової моделі, тоді як перетин трансформованої моделі є нахилом початкової.

Розглянемо

$$M\left(\frac{u_i}{x_i}\right)^2 = \frac{1}{x_i^2} Mu_i^2 = \frac{1}{x_i^2} k^2 x_i^2 = k^2.$$

Отже, нова випадкова величина моделі (5.8) має скінченну сталу дисперсію k^2 . Таким чином, модель (5.8) має гомоскедастичну випадкову змінну, що означає правомірність застосування класичного МНК для розрахунку невідомих параметрів трансформованої моделі (5.8).

Випадок 2. Припустимо, що гетероскедастичність має форму

$$\sigma_{u_i}^2 = Mu_i^2 = k^2 x, \quad (5.9)$$

де $k = \text{const}$ (тобто дисперсія залишків зростає пропорційно до x).

Із припущення (5.8) випливає

$$k^2 = \sigma_{u_i}^2 / x.$$

Це означає, що допустима трансформація полягає в діленні початкової моделі на \sqrt{x} .

Отже, трансформована модель має вигляд

$$\frac{y}{\sqrt{x}} = \frac{a_0}{\sqrt{x}} + a_1 \frac{x}{\sqrt{x}} + \frac{u}{\sqrt{x}}. \quad (5.10)$$

Розглянемо

$$M\left(\frac{u_i}{\sqrt{x}}\right)^2 = \frac{1}{x_i} M(u_i)^2 = \frac{1}{x_i} k^2 x_i = k^2.$$

Отже, для трансформованої моделі випадкова величина $\frac{u}{\sqrt{x}}$ гомоскедастична зі сталою дисперсією k^2 . Це означає, що, виконавши зазначене вище перетворення, ми виключили гетероскедастичність.

Випадок 3. Припустимо, що гетероскедастичність має форму

$$\sigma_{u_i}^2 = Mu_i^2 = k^2(b_0 + b_1 x_i)^2 \quad (5.11)$$

(дисперсія зростає пропорційно до квадрата лінійної функції від x).

Із припущення (5.11) випливає

$$k^2 = \sigma_{u_i}^2 / (b_0 + b_1 x_i)^2. \quad (5.12)$$

Допустима трансформація полягає в діленні початкової моделі на

$$\sqrt{(b_0 + b_1 x_i)^2} = (b_0 + b_1 x_i).$$

Отже, трансформована модель має вигляд

$$\frac{y_i}{b_0 + b_1 x_i} = a_0 \frac{1}{b_0 + b_1 x_i} + a_1 \frac{x_i}{b_0 + b_1 x_i} + \frac{u_i}{b_0 + b_1 x_i}. \quad (5.13)$$

Розглянемо

$$M\left(\frac{u_i}{b_0 + b_1 x_i}\right)^2 = \frac{1}{(b_0 + b_1 x_i)^2} Mu_i^2 = \frac{k^2 (b_0 + b_1 x_i)^2}{(b_0 + b_1 x_i)^2} = k^2.$$

Отже, нова випадкова величина $\frac{u_i}{b_0 + b_1 x_i}$ є гомоскедастичною із сталою дисперсією k^2 .

Загальний випадок. Припустимо, що гетероскедастичність має форму

$$\sigma_{u_i}^2 = M u_i^2 = k^2 \varphi(x_i), \quad (5.14)$$

де $k = \text{const}$, $\varphi(x_i)$ – функція від x_i .

Трансформація початкової моделі здійснюється діленням її на $\sqrt{\varphi(x_i)}$.

Зазначимо, що така трансформація еквівалентна застосуванню зваженого методу найменших квадратів (ЗМНК), який є особливим випадком узагальненого методу найменших квадратів (УМНК). Суть ЗМНК полягає в мінімізації зваженої суми квадратичних відхилень:

$$\sum_{i=1}^n \frac{u_i^2}{\sigma_{u_i}^2} = \sum_{i=1}^n \frac{1}{\sigma_{u_i}^2} (y_i - a_0 - a_1 x_i)^2 \rightarrow \min. \quad (5.15)$$

Зазначимо також, що ЗМНК, застосований до початкової моделі, дає такі самі результати, що й МНК, застосований до трансформованої моделі.

Твердження. Оцінки трансформованої моделі мають меншу дисперсію (ефективніші), ніж оцінки, отримані із застосуванням МНК до початкової моделі.

Нарешті, потрібно пам'ятати, що гетероскедастичність може існувати за рахунок неврахованих факторів (поганої специфікації моделі). У цьому разі можливим рішенням є включення неврахованих факторів у модель. Слепе застосування трансформації (без аналізу причин гетероскедастичності) зробить гомоскедастичною випадкову змінну, однак оцінки параметрів залишаться неправильними через неврахування важливих факторів.

5.4. Оцінювання параметрів багатofакторної регресійної моделі на основі узагальненого методу найменших квадратів

Розглянемо детальніше загальний випадок оцінювання параметрів моделі з гетероскедастичними залишками.

Запишемо узагальнену багатofакторну регресійну модель у матричному вигляді

$$y = Xa + u, \quad (5.16)$$

де y – вектор-стовпець залежної змінної розмірності $(n \times 1)$;

X – матриця незалежних змінних розмірності $(n \times (m + 1))$;

a – вектор-стовпець невідомих параметрів розмірності $((m + 1) \times 1)$;

u – вектор-стовпець випадкових помилок розмірності $(n \times 1)$.

Нехай виконуються всі припущення класичної лінійної багатofакторної моделі, за винятком припущення про гомоскедастичність похибок. Якщо до моделі (5.16) застосувати звичайний МНК, отримана оцінка параметрів буде незміщеною, обґрунтованою, однак не ефективною (не має найменшої дисперсії серед незміщених оцінок).

За наявності гетероскедастичності для оцінювання параметрів моделі доцільно застосувати узагальнений метод найменших квадратів (метод Ейткена), вектор оцінювання якого має вигляд

$$a = (X'S^{-1}X)^{-1}X'S^{-1}Y. \quad (5.17)$$

Вектор a містить незміщену лінійну оцінку параметрів моделі, яка має найменшу дисперсію і матрицю коваріацій:

$$\sigma^2(a) = \sigma_u^2(X'S^{-1}X)^{-1}.$$

Зауваження. Для отримання УМНК-оцінок необхідно знати коваріаційну матрицю S вектора похибок, яка на практиці дуже рідко відома. Тому природно спершу оцінити матрицю S , а потім застосувати її оцінку у формулі (5.17). Цей підхід є суть узагальненого методу найменших квадратів.

Визначення матриці S . Оскільки явище гетероскедастичності пов'язане лише з тим, що змінюються дисперсії залишків, а кова-

ріація між ними відсутня, то матриця S має бути діагональною, а саме

$$S = \begin{pmatrix} \frac{1}{\lambda_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\lambda_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\lambda_n} \end{pmatrix}.$$

Зазначимо, що матриця S залежить від специфічної форми гетероскедастичності й може бути розрахована виходячи з припущень про залежність похибок від однієї із незалежних змінних (випадки 1–3).

У матриці S значення λ_i , $i = 1, 2, \dots, n$, можна обчислити, користуючись гіпотезами:

1) $M(uu') = \sigma_u^2 x_{ij}$, тобто дисперсія залишків пропорційна до змін пояснюючої змінної x_{ij} ;

2) $M(uu') = \sigma_u^2 x_{ij}^2$, тобто зміна дисперсії пропорційна до зміни квадрата пояснюючої змінної x_{ij}^2 ;

3) $M(uu') = \sigma_u^2 \{|u|\}^2$, тобто дисперсія залишків пропорційна до зміни квадрата залишків за модулем.

Для першої гіпотези $\lambda_i = \frac{1}{x_{ij}}$.

Для другої гіпотези $\lambda_i = \frac{1}{x_{ij}^2}$.

Для третьої гіпотези $\lambda_i = \{|u|\}^2$, або $\lambda_i = (a_0 - a_1 x_{ij})^2$, або $\lambda_i = (a_0 - a_1 x_{ij}^{-1})^2$.

Оскільки матриця S симетрична і додатно визначена, то при $S = PP^{-1}$ матриця P має вигляд

$$P = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\sqrt{\lambda_n}} \end{pmatrix}, \quad P^{-1} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sqrt{\lambda_n} \end{pmatrix}.$$

Зауваження. Коефіцієнт детермінації не може бути задовільною мірою якості моделі в разі застосування УМНК (на відміну від класичної моделі). У загальному випадку значення коефіцієнта детермінації навіть не повинно перебувати в інтервалі $[0,1]$, а додавання чи вилучення незалежної змінної (фактора) не обов'язково зумовлює його збільшення або зменшення.

Основні висновки щодо наявності гетероскедастичності в регресійній моделі

1. Якщо виявлено гетероскедастичність, а дисперсії невідомі, необхідно трансформувати початкову модель з метою усунення гетероскедастичності.
2. Якщо $\sigma_{u_i}^2$ відомі (що, взагалі, рідкість), то невідомі параметри регресійної моделі розраховуються за МНК.
3. Якщо $\sigma_{u_i}^2$ невідомі, але відомий вигляд залежності між $\sigma_{u_i}^2$ та однією із незалежних змінних x_i , то параметри регресійної моделі розраховуються за УМНК.
4. Важливим є припущення про нормальний закон розподілу випадкової змінної u . Якщо це припущення порушується (або, як часто буває на практиці, ігнорується), то оцінки параметрів залишаються найкращими, однак ми не можемо визначити їх статистичну значущість (надійність) за допомогою класичних тестів значущості (t , F тощо), оскільки ці тести базуються на нормальному законі розподілу.

Контрольні запитання

1. Яке явище називається гомоскедастичністю?
2. Яке явище називається гетероскедастичністю?

3. У чому полягає суть гетероскедастичності?
4. Яку форму звичайно має гетероскедастичність?
5. До яких наслідків призводить порушення припущення про гомоскедастичність?
6. Як встановити наявність гетероскедастичності?
7. Назвіть методи визначення гетероскедастичності.
8. За яких умов застосовується параметричний тест Гольдфельда – Квандта?
9. У чому суть непараметричного тесту?
10. На чому базується тест Глейсера?
11. У чому суть трансформації моделі?
12. Наведіть форми гетероскедастичності в найпоширеніших випадках трансформації.
13. Які властивості мають оцінки параметрів трансформованої моделі?
14. За рахунок чого може існувати гетероскедастичність?
15. У чому суть узагальненого методу найменших квадратів?

МАУП

Розділ 6. АВТОКОРЕЛЯЦІЯ

6.1. Природа автокореляції та її наслідки

Розглянемо класичну лінійну багатофакторну модель

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m + u \quad (6.1)$$

або в матричному вигляді

$$Y = Xa + u, \quad (6.2)$$

де y – вектор-стовпець залежної змінної розмірності $(n \times 1)$;

X – матриця незалежних змінних розмірності $(n \times (m + 1))$;

a – вектор-стовпець невідомих параметрів розмірності $((m + 1) \times 1)$;

u – вектор-стовпець випадкових помилок розмірності $(n \times 1)$;

$$\text{cov}(u_i, u_j) = 0, \quad i \neq j.$$

Одним із припущень класичного регресійного аналізу є припущення про незалежність випадкових величин u_i , $i = 1, \dots, n$, тобто якщо це припущення порушується (незважаючи на те, що дисперсія залишків є сталою – наявна гомоскедастичність), то ми маємо справу з явищем, яке називається **автокореляцією** залишків.

Важливо зрозуміти, що спричинює автокореляцію, які її практичні та теоретичні наслідки, чи є ефективні методи тестування наявності автокореляції, чи змінюються методи знаходження невідомих параметрів моделі в умовах автокореляції.

Автокореляція залишків виникає найчастіше тоді, коли економетрична модель будується на основі часових рядів. Якщо існує кореляція між послідовними значеннями деякої незалежної змінної, то спостерігатиметься й кореляція послідовних значень залишків, так звані лагові затримки (запізнювання) в економічних процесах.

Автокореляція може виникати через інерційність і циклічність багатьох економічних процесів. Провокувати автокореляцію також може неправильно специфікована функціональна залежність у регресійних моделях.

Припустимо, модель (6.1) має автокорельовані залишки, тобто випадкові величини u_i залежні між собою: $M(u_i u_j) \neq 0, i \neq j$.

Отже, як і у випадку гетероскедастичності, дисперсія залишків

$$M(uu') \neq \sigma_u^2 I. \quad (6.3)$$

Але при гетероскедастичності змінюються дисперсії залишків за відсутності їх коваріації, а при автокореляції існує коваріація залишків за незмінної дисперсії.

Значимо, що за наявності автокореляції залишків, як і за наявності гетероскедастичності, дисперсія залишків має вигляд

$$M(uu') = \sigma_u^2 \Omega, \quad (6.4)$$

однак матриця Ω матиме тут інший вигляд:

$$\Omega = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \dots & 1 \end{pmatrix}, \quad (6.5)$$

де параметр ρ характеризує коваріацію кожного наступного значення залишків із попереднім.

Так, якщо для залишків записати авторегресійну модель першого порядку

$$u_t = \rho u_{t-1} + v_t, \quad (6.6)$$

то ρ характеризує силу зв'язку величин залишків у період t з величинами залишків у період $t-1$.

Якщо проігнорувати матрицю Ω при визначенні дисперсії залишків і для оцінювання параметрів моделі застосувати МНК, то можливі такі **наслідки**:

1. Оцінки параметрів моделі можуть бути незміщеними, але неефективними, тобто вибірковій дисперсії вектора оцінок a можуть бути невиправдано великими.

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

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

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

6.2. Тестування наявності автокореляції

Тестування наявності автокореляції, як правило, здійснюється за d -тестом Дарбіна – Уотсона, хоча існують й інші не менш відомі тести: критерій фон Неймана, нециклічний коефіцієнт автокореляції, циклічний коефіцієнт автокореляції.

6.2.1. Критерій Дарбіна — Уотсона

(складається з кількох етапів і включає зони невизначеності)

Крок 1. Розраховується значення d -статистики за формулою

$$DW = d = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2}. \quad (6.7)$$

Зауваження. Доведено, що значення d -статистики Дарбіна – Уотсона перебуває в межах $0 \leq DW \leq 4$.

Крок 2. Задаємо рівень значущості α . За таблицею Дарбіна – Уотсона при заданому рівні значущості α , кількості факторів m і кількості спостережень n знаходимо два значення DW_1 і DW_2 :

- Якщо $0 < DW < DW_1$, то наявна додатна автокореляція.
 - Якщо $DW_1 \leq DW \leq DW_2$ або $4 - DW_2 \leq DW \leq 4 - DW_1$, ми не можемо зробити висновки ані про наявність, ані про відсутність автокореляції (DW потрапляє в зону невизначеності).
 - Якщо $4 - DW_1 < DW < 4$, маємо від'ємну автокореляцію.
 - Якщо $DW_2 < DW < 4 - DW_2$, то автокореляція відсутня.
- Графічне зображення розподілу ілюструє рис. 6.1.



Рис. 6.1. Зони автокореляційного зв'язку за критерієм Дарбіна – Уотсона

6.2.2. Критерій фон Неймана

Розраховується

$$Q = Q_{\text{факт}} = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2} \frac{n}{n-1}. \quad (6.8)$$

Звідси $Q = DW \frac{n}{n-1}$. Отже, при $n \rightarrow \infty$ $Q = DW$.

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

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

6.2.3. Коефіцієнти автокореляції та їх застосування

Окрім статистик Дарбіна – Уотсона та Неймана, для перевірки автокореляції застосовують також нециклічний коефіцієнт автокорре-

ляції r^* , який відображає ступінь взаємозв'язку рядів u_1, u_2, \dots, u_{n-1} і обчислюється за формулою

$$r^* = \frac{\sum_{t=2}^n u_t u_{t-1} - \frac{1}{n-1} \left(\sum_{t=2}^n u_t \right) \left(\sum_{t=2}^n u_{t-1} \right)}{\left[\sum_{t=2}^n u_t^2 - \frac{1}{n-1} \left(\sum_{t=2}^n u_t \right)^2 \right]^{1/2} \left[\sum_{t=2}^n u_{t-1}^2 - \frac{1}{n-1} \left(\sum_{t=2}^n u_{t-1} \right)^2 \right]^{1/2}}.$$

Коефіцієнт r^* може набувати значень в інтервалі $(-1,1)$. Його від'ємні значення свідчать про від'ємну автокореляцію залишків, а додатні – про додатну автокореляцію. Значення, що лежать в деякій критичній області поблизу нуля, підтверджують нульову гіпотезу про відсутність автокореляції в залишках. Оскільки ймовірнісний розподіл r^* встановити важко, то на практиці замість r^* обчислюють *циклічний* коефіцієнт автокореляції r^0 . Загалом, якщо часовий ряд має циклічний характер, тобто припускається, що після значення u_τ загальний характер зміни членів ряду повторюється, то автокореляцію визначають за допомогою коефіцієнта r^0 , запровадженого Андерсоном.

У цьому разі автокореляція визначається між послідовностями, зрутими на період τ :

$$u_1, u_2, \dots, u_\tau \text{ і } u_{\tau+1}, u_{\tau+2}, \dots, u_{2\tau}$$

$$u_1, u_2, \dots, u_n \text{ і } u_{\tau+1}, \dots, u_n, u_1, u_2, \dots, u_\tau.$$

Якщо період $\tau = 1$, то маємо коефіцієнт циклічної автокореляції першого порядку, який відбиває інтенсивність взаємозв'язку між послідовностями

$$u_1, u_2, \dots, u_{n-1}, u_n$$

та

$$u_2, u_3, \dots, u_n, u_1.$$

Для досить довгих рядів вплив циклічних членів стає незначним, тому ймовірнісний розподіл коефіцієнта r^* наближається до ймовірнісного розподілу коефіцієнта циклічної автокореляції r^0 , який обчислюється за формулою

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

Якщо останній член ряду дорівнює першому, тобто $u_1 = u_n$, то нециклічний коефіцієнт автокореляції дорівнює циклічному. Очевидно, якщо залишки не містять тренда, то припущення про рівність $u_1 = u_n$ недалеко від дійсності й циклічний коефіцієнт автокореляції близький до нециклічного. Крім того, припускаючи, що середня залишків дорівнює нулю, тобто $\bar{u} = 0$, а отже,

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

отримуємо приблизну формулу для обчислення циклічного коефіцієнта автокореляції:

$$r = \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2},$$

причому $r \in (-1,1)$. Значення r використовується при оцінюванні параметрів моделі.

6.3. Параметризація моделі з автокорельованими залишками

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

- 1) Ейткена (УМНК);
- 2) перетворення вихідної інформації;
- 3) Кочрена – Оркатта;
- 4) Дарбіна.

Перші два методи доцільно застосовувати тоді, коли залишки описуються авторегресійною моделлю першого порядку (6.6):

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

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

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + v_t,$$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + v_t.$$

6.3.1. Метод Ейткена

Як зазначалося, оператор оцінювання УМНК можна записати так:

$$\hat{a} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y \quad (6.9)$$

або

$$\hat{a} = (X' S^{-1} X)^{-1} X' S^{-1} Y, \quad (6.10)$$

де Ω^{-1} — матриця, обернена до дисперсійно-коваріаційної матриці залишків Ω ; S^{-1} — матриця, обернена до матриці $S = \sigma_u^2 \Omega$.

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

$$\Omega^{-1} = \frac{1}{1 - \rho^2} \cdot \begin{pmatrix} 1 & -\rho & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 + \rho^2 & -\rho & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}. \quad (6.11)$$

На практиці для розрахунку ρ використовується співвідношення

$$\rho \approx r \approx \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2} \quad (6.12)$$

або

$$\rho \approx r' \approx \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2} \frac{n}{n-1}. \quad (6.13)$$

6.3.2. Метод Кочрена – Оркатта

Зауваження. Метод Кочрена – Оркатта є ітераційним методом наближеного пошуку оцінок параметрів моделі з автокорельованими залишками, який базується на МНК.

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

$$\sum_{t=2}^n \varepsilon_t^2 = \sum_{t=2}^n [(y_t - \rho y_{t-1}) - a_0(1 - \rho) - a_1(x_t - \rho x_{t-1})]^2, \quad (6.14)$$

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

Крок 2. Покласти $\hat{a}_0 = \hat{a}_0^{(1)}$ і $\hat{a}_1 = \hat{a}_1^{(1)}$; підставивши їх у рівняння (6.14), обчислити $\rho = r_1$.

Крок 3. Підставити в рівняння (6.14) значення $\rho = r_2$, знайти $\hat{a}_0^{(2)}$ і $\hat{a}_1^{(2)}$.

Крок 4. Використати $\hat{a}_0 = \hat{a}_0^{(2)}$ і $\hat{a}_1 = \hat{a}_1^{(2)}$ для мінімізації суми квадратів залишків (6.14) за невідомим параметром $\rho = r_3$. Процедуру повторювати доти, доки наступні значення параметрів \hat{a}_0 , \hat{a}_1 і ρ не відрізняться менш як на задану величину.

Значимо, що наведений метод завжди забезпечує:

- знаходження глобального оптимуму;
- порівняно добру збіжність.

6.4. Приклад оцінювання параметрів моделі з автокорельованими залишками

На основі двох взаємопов'язаних часових рядів про роздрібний товарообіг і доходи населення побудувати модель, що характеризує залежність роздрібногo товарообігу від доходу.

Вихідні дані наведено в таблиці.

Рік	Товарообіг	Дохід
1	25,5	27,6
2	26,4	27,4
3	27,9	28,7
4	28,1	29,5
5	28,8	30,9
6	29,3	31,4
7	29,8	31,8
8	30,7	32,2
9	31,5	33,6
10	32,4	34,7

Розв'язання.

1. Ідентифікуємо змінні моделі: y_t – роздрібний товарообіг у період t , залежна змінна; x_t – дохід у період t , пояснююча змінна;

$$y_t = f(x_t, u_t),$$

де u_t – стохастична складова (залишки).

2. Специфікуємо модель у лінійній формі:

$$y_t = a_0 + a_1 x_t + u_t;$$

$$\hat{y}_t = a_0 + a_1 x_t;$$

$$u_t = y_t - \hat{y}_t.$$

3. Визначимо \hat{a}_0 , \hat{a}_1 на основі МНК, припустивши, що залишки некорельовані:

$$\hat{a} = (X^T X)^{-1} X^T y.$$

$$X = \begin{pmatrix} 1 & 27,6 \\ 1 & 27,4 \\ 1 & 28,7 \\ 1 & 29,5 \\ 1 & 30,9 \\ 1 & 31,4 \\ 1 & 31,8 \\ 1 & 32,2 \\ 1 & 33,6 \\ 1 & 34,7 \end{pmatrix}; \quad X^T X = \begin{pmatrix} 10,0 & 307,8 \\ 307,8 & 9528,36 \end{pmatrix};$$

$$(X^T X)^{-1} = \begin{pmatrix} 17,55538 & -0,5671 \\ -0,5671 & 0,018424 \end{pmatrix}; \quad X^T y = \begin{pmatrix} 290,4 \\ 8985,64 \end{pmatrix};$$

$$\hat{a} = \begin{pmatrix} 2,313641 \\ 0,868303 \end{pmatrix}.$$

Отже, модель має вигляд

$$\hat{y}_t = 2,313641 + 0,868303 x_t.$$

4. Знайдемо оцінені значення \hat{y}_t на основі отриманої моделі та визначимо залишки u_t .

Рік	y_t	\hat{y}_t	u_t	u_t^2	u_{t-1}	$(u_t - u_{t-1})^2$	$u_t u_{t-1}$
1	25,5	26,2787973	-0,7788	0,606525			
2	26,4	26,1051367	0,294863	0,086944	-0,7788	1,152747	-0,22964
3	26,9	27,2339303	-0,33393	0,111509	0,294863	0,395381	-0,09846
4	27,1	27,9285725	-0,82857	0,686532	-0,33393	0,244671	0,276685
5	27,8	29,1441963	-1,3442	1,806864	-0,82857	0,265868	1,113764
6	28,3	29,5783477	-1,27835	1,634173	-1,3442	0,004336	1,71835
7	29,4	29,9256688	-0,52567	0,276328	-1,27835	0,566526	0,671988
8	30,7	30,2729899	0,42701	0,182338	-0,52567	0,907597	-0,22447
9	31,5	31,4886138	0,011386	0,00013	0,42701	0,172743	0,004862
10	33,4	32,4437468	0,956253	0,91442	0,011386	0,892774	0,010888
Сума	287	290,4	-3,4	6,305763	-4,35625	4,602643	3,243969

5. Обчислимо оцінку d -статистики Дарбіна – Уотсона:

$$DW = \frac{\sum_{t=2}^{10} (u_t - u_{t-1})^2}{\sum_{t=1}^{10} u_t^2} = \frac{4,602643}{6,305763} \approx 0,7299.$$

Задамо $\alpha = 0,05$ і при $n = 10$ та $m = 1$ знайдемо за таблицею d -статистики Дарбіна – Уотсона (дод. 4) критичні значення критерію:

$DW_1 = 0,604$ – нижня межа;

$DW_2 = 1,001$ – верхня межа.

Оскільки $DW_1 < DW < DW_2$, то з похибкою щонайбільше у 5 % випадків можна стверджувати, що автокореляція залишків u_t невизначена.

Завдання для самостійної роботи: перевірити автокореляцію залишків моделі на основі критерію фон Неймана.

Щоб оцінити параметри моделі з автокорельованими залишками, використовуємо УМНК. Оператор оцінювання

$$\hat{a} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \quad \text{або} \quad \hat{a} = (X'S^{-1}X)^{-1}X'S^{-1}Y,$$

де Ω – дисперсійно-коваріаційна матриця залишків, яка має вигляд

$$\Omega = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^9 \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^8 \\ \rho^2 & \rho & 1 & \rho & \dots & \rho^7 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^9 & \rho^8 & \rho^7 & \rho^6 & \dots & 1 \end{pmatrix}; \quad S = \sigma_u^2 \Omega.$$

Щоб сформувати Ω або S , необхідно знати величину ρ , яка характеризує взаємозв'язок між послідовними членами ряду залишків.

Припустимо, залишки описуються автокореляційною моделлю першого порядку (6.6):

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

де

$$\rho \approx r' \approx \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2} \cdot \frac{n}{n-1} + \frac{m+1}{n} = \frac{3,243969}{6,305763} \cdot \frac{10}{9} + \frac{2}{10} \approx 0,77.$$

Отже,

$$\Omega = \begin{pmatrix} 1 & 0,77 & 0,5929 & 0,4565 & \dots & 0,0951 \\ 0,77 & 1 & 0,77 & 0,5929 & \dots & 0,1235 \\ 0,5929 & 0,77 & 1 & 0,77 & \dots & 0,1604 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0,0951 & 0,1235 & 0,1604 & 0,2084 & \dots & 1 \end{pmatrix};$$

$$\Omega^{-1} = \begin{pmatrix} 2,4564 & -1,8914 & \dots & 5,6E-17 \\ -1,8914 & 3,9128 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ -4,4657E-17 & -3,65E-17 & \dots & 2,4564 \end{pmatrix}.$$

6. Розрахуємо:

$$X_t^T \Omega^{-1} = \begin{pmatrix} 0,565 & 0,1041 & \dots & 0,5649 \\ 15,9715 & -0,1246 & \dots & 21,685 \end{pmatrix};$$

$$X_t^T \Omega^{-1} X_t = \begin{pmatrix} 2,171 & 67,163 \\ 67,157 & 2110,708 \end{pmatrix}; \quad X_t^T \Omega^{-1} y_t = \begin{pmatrix} 63,007 \\ 1976,647 \end{pmatrix};$$

$$(X_t^T \Omega^{-1} X_t)^{-1} = \begin{pmatrix} 29,1312 & -0,9269 \\ -0,9269 & 0,0299 \end{pmatrix}; \quad \hat{a} = \begin{pmatrix} 3,20236 \\ 0,834594 \end{pmatrix}.$$

Отже, $\hat{y}_t = 3,20236 + 0,834594x_t$.

7. Знайдемо оцінені значення \hat{y}_t на основі побудованої моделі та визначимо її залишки v_i

Рік	y_t	\hat{y}_t	v_t	v_t^2	v_{t-1}	$(v_t - v_{t-1})^2$	$v_t v_{t-1}$
1	25,5	26,23716	-0,73716	0,54341	—	—	—
2	26,4	26,07024	0,329755	0,108739	-0,73716	1,13831587	-0,24308
3	27,9	27,15522	0,744783	0,554701	0,329755	0,17224772	0,245596
4	28,1	27,82289	0,277107	0,076788	0,744783	0,21872034	0,206385
5	28,8	28,99132	-0,19132	0,036605	0,277107	0,2194286	-0,05302
6	29,3	29,40862	-0,10862	0,011799	-0,19132	0,00683976	0,020782
7	29,8	29,74246	0,05754	0,003311	-0,10862	0,0276099	-0,00625
8	30,7	30,0763	0,623702	0,389005	0,05754	0,32053971	0,035888
9	31,5	31,24473	0,25527	0,065163	0,623702	0,13574218	0,159213
10	32,4	32,16278	0,237217	0,056272	0,25527	0,00032594	0,060554
Сума	290,4	288,9117	1,488264	1,845793	1,251048	2,23977001	0,426067

8. Розрахуємо d -статистику Дарбіна – Уотсона:

$$DW = \frac{\sum_{t=2}^{10} (v_t - v_{t-1})^2}{\sum_{t=1}^{10} v_t^2} = \frac{2,23977001}{1,845793} \approx 1,2134.$$

Порівняємо d -статистику Дарбіна – Уотсона з критичними значеннями при $\alpha=0,05$, $n=10$ та $m=1$.

Оскільки $DW_2 < DW < 4-DW_2$, робимо висновок, що ми усунули автокореляцію залишків. А це в свою чергу означає, що дотримується гіпотеза про те, що залишки описуються авторегресійною схемою першого порядку. Якщо залишки описуються авторегресійною схемою вищого порядку, доцільно оцінити параметри моделі методом Кочрена – Оркатта або Дарбіна.

9. Прогноз: визначимо прогнозний рівень товарообігу, якщо дохід становитиме $x_{n+1}=35$. Співвідношення, що визначає прогнозний рівень залежної змінної, має вигляд: $\hat{y}_{n+1} = x_{n+1}a$.

Отже,

$$\hat{y}_{n+1} = 3,20236 + 0,834594x_t = 3,20236 + 0,834594 \cdot 55 = 49,105085.$$

Це означає, що прогнозний рівень роздрібного товарообігу на $(n+1)$ -й рік становить 49,105085.

Контрольні запитання

1. Що означає автокореляція залишків?
2. Які причини виникнення автокореляції?
3. Чим відрізняється УМНК при автокореляції?
4. До яких наслідків призводить автокореляція залишків?
5. Як впливає автокореляція на оцінку параметрів моделі?
6. Наведіть тести наявності автокореляції залишків.
7. Дайте стислу характеристику алгоритму методу Дарбіна – Уотсона.
8. Зобразіть графічно зони автокореляційного зв'язку за критерієм Дарбіна – Уотсона.
9. Охарактеризуйте алгоритм критерію фон Неймана.

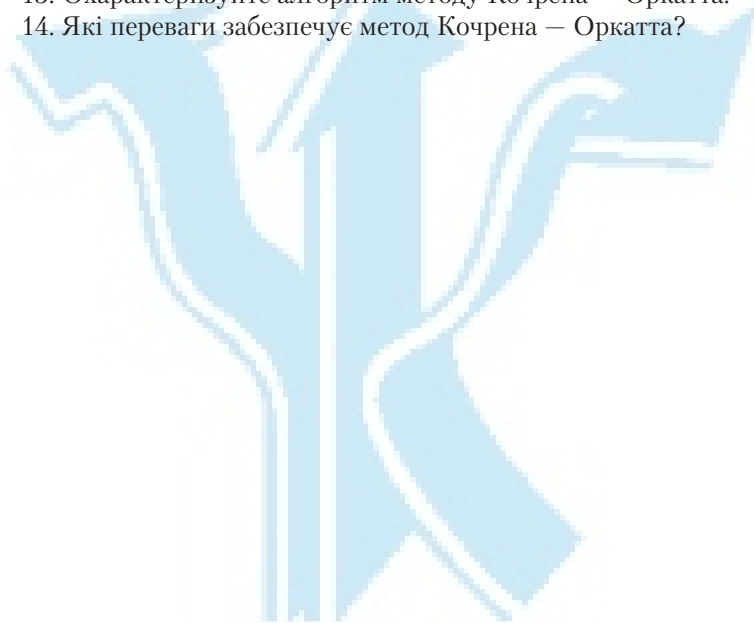
10. На основі яких методів можна оцінити параметри моделі з автокорельованими залишками?

11. У яких випадках для оцінки параметрів економетричної моделі з автокорельованими залишками доцільно застосовувати ітераційні методи Кочрена – Оркатта або Дарбіна?

12. Охарактеризуйте метод Ейткена оцінки параметрів економетричної моделі з автокорельованими залишками.

13. Охарактеризуйте алгоритм методу Кочрена – Оркатта.

14. Які переваги забезпечує метод Кочрена – Оркатта?



МАУП

Розділ 7. МОДЕЛІ РОЗПОДІЛЕНОГО ЛАГА

7.1. Поняття лага та лагових моделей в економіці

Для багатьох економічних процесів типово, що ефект від впливу деякого фактора на показник, який характеризує процес, виявляється не одразу, а поступово, через деякий час або протягом деякого часу. Таке явище називається запізнюванням (затримкою), а проміжок часу, у який спостерігається це запізнювання, — часовим лагом, або просто лагом.

Наприклад, функція споживання після різкої зміни доходу також змінюється, але не пропорційно до доходу. Зокрема, зі збільшенням доходу витрати значно зростають (задовольняються нагальні потреби), а потім можуть зменшуватися за рахунок збільшення інвестицій (плануються великі витрати і зростають нагромадження, а споживання зменшується). Вкладання коштів у наукові дослідження не одразу впливає на зростання продуктивності праці — має пройти певний час від виникнення наукової ідеї до її впровадження у виробництво. Капітальне будівництво також не дає миттєвих прибутків і т. ін.

Означення 7.1. Моделі, у яких досліджуваний показник у момент часу t визначається не лише поточними, а й попередніми значеннями незалежних змінних, називаються дистрибутивно-лаговими:

$$y_t = a + a_0x_t + a_1x_{t-1} + a_2x_{t-2} + \dots + u_t.$$

Означення 7.2. Моделі, у яких досліджуваний показник у момент часу t визначається своїми попередніми значеннями, називаються авторегресійними або динамічними моделями. Наприклад,

$$y_t = a + a_0x_t + b_1y_{t-1} + b_2y_{t-2} + \dots + u_t.$$

Змінні $x_{t-\tau}$ і $y_{t-\tau}$ у період $t - \tau$ називаються лаговими змінними, а величина τ — періодом зсуву (лагом).

Якщо в економетричній моделі незалежні змінні використовують за кілька попередніх періодів, то такі моделі називають моделями з кінцевим лагом (скінченними моделями). Якщо вплив незалежної змінної не обмежується певним періодом, розглядають нескінченні лагові моделі.

Звичайно, нескінченна лагова модель більш загальна, однак практичне застосування такої моделі досить проблематичне через велику кількість факторів, складність внутрішньої структури та обмеженість часових рядів — інформаційної бази моделей.

Коефіцієнти a_j , $j = 0, 1, 2, \dots$, називаються коефіцієнтами лага, а послідовність $\{a_j, j = 0, 1, 2, \dots\}$ — структурою лага.

Коефіцієнт a_0 при незалежній змінній x_t , що відбиває її вплив на залежну змінну в поточний період, називається короткостроковим, або впливовим, мультиплікатором. Часткові суми коефіцієнтів $(a_0 + a_1)$, $(a_0 + a_1 + a_2)$, \dots , що відображають зміну y_t в другий, третій і наступні періоди, називаються проміжними інтервалами. Загальна сума лагових коефіцієнтів для всієї моделі називається довгостроковим, або загальним, дистрибутивно-лаговим мультиплікатором.

Остання сума для скінченних моделей, очевидно, є скінченим числом. Для нескінченної моделі лагові коефіцієнти за певних умов також можуть утворити скінченну суму. Якщо кожен із коефіцієнтів розділити на їх суму, отримаємо відповідно нормовані коефіцієнти лага та нормовану структуру лага.

Усі нормовані коефіцієнти менші від одиниці, а їх сума дорівнює одиниці.

Дистрибутивно-лагові моделі, які ще називають моделями розподіленого лага, задовільно описують економічні процеси лише в стабільних (незмінних) умовах. Необхідність враховувати ще й поточні умови функціонування вимагає застосування узагальнених моделей.

Означення 7.3. Якщо економетрична модель містить не лише лагові змінні, а й змінні, що безпосередньо впливають на досліджуваний показник (тобто містить й поточні умови функціонування), то така модель називається узагальненою моделлю розподіленого лага:

$$y_t = \sum_{\tau \geq 0} a_\tau x_{t-\tau} + \sum_{i=1}^m b_i x_{t,i} + u_t.$$

Оцінювання параметрів таких рівнянь ускладнюється обмеженнями, що накладаються на коефіцієнти при лагових змінних.

Перш ніж будувати економетричну модель з лаговими змінними, необхідно обґрунтувати величину лага. Для цього застосовують взаємну кореляційну функцію — послідовність коефіцієнтів кореляції, які визначають ступінь зв'язку кожного елемента вектора залежної змінної з елементом вектора незалежної змінної, зсунутими один відносно одного на часовий лаг τ :

$$r_\tau = \frac{(n-\tau) \sum_{t=1}^{n-\tau} y_t x_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} x_{t+\tau}}{\sqrt{\left[(n-\tau) \sum_{t=1}^{n-\tau} y_t^2 - \left(\sum_{t=1}^{n-\tau} y_t \right)^2 \right] \left[(n-\tau) \sum_{t=1}^{n-\tau} x_{t+\tau}^2 - \left(\sum_{t=1}^{n-\tau} x_{t+\tau} \right)^2 \right]}}. \quad (7.1)$$

Серед отриманих коефіцієнтів кореляції вибирають найбільший за модулем, а відповідне значення часового зсуву вважають лагом. Якщо таких коефіцієнтів кілька, застосовують модель розподіленого лага.

Приклад [17]. На основі взаємопов'язаних часових рядів, які характеризують чисту продукцію та капітальні вкладення Республіки Сирії за 20 років, побудувати взаємну кореляційну функцію, використавши дані табл. 7.1.

У результаті розрахунків, виконаних за формулою (7.1) для різних значень τ , отримано значення коефіцієнтів кореляції, що наведені в табл. 7.2.

З табл. 7.2 видно, що найбільше значення коефіцієнта кореляції $r_\tau = 0,92$. Він відповідає трьом значенням $\tau = \{3, 4, 5\}$. Це означає, що найбільшого впливу капітальних вкладень на обсяг чистої продукції слід очікувати впродовж третього, четвертого та п'ятого років.

Таблиця 7.1

Рік	Капітальні вкладення, млн сирійських лір	Чиста продукція, млн сирійських лір	Рік	Капітальні вкладення, млн сирійських лір	Чиста продукція, млн сирійських лір
1	3857	24334	11	17006	72165
2	4686	28678	12	17352	78743
3	5515	33021	13	17838	80381
4	5209	32432	14	18878	82204
5	7522	40325	15	19090	77833
6	10390	49334	16	20016	81413
7	13678	54717	17	17736	77484
8	15976	53818	18	11951	75443
9	13880	55968	19	11469	85038
10	13949	61517	20	9068	75809

Таблиця 7.2

τ	1	2	3	4	5	6	7	8	9	10
r_τ	0,89	0,86	0,89	0,92	0,92	0,85	0,75	0,64	0,4	0,55

Динамічна модель розподіленого лага в такому разі має вигляд

$$y_t = a_0 x_t + a_1 x_{t-3} + a_2 x_{t-4} + a_3 x_{t-5} + u_t,$$

де y_t — чиста продукція в період t ; $a_j, j = 0, 1, 2, 3$ — вагові коефіцієнти лагових змінних; $x_{t-\tau} (\tau = 3, 4, 5)$ — капітальні вкладення в період $t - \tau$.

7.2. Оцінювання параметрів дистрибутивно-лагових моделей

Для оцінювання параметрів дистрибутивно-лагових моделей звичайно застосовують два можливих підходи: послідовне оцінювання і апріорне оцінювання.

Ідея першого підходу полягає в тому, щоб поступово досліджувати вплив запізнених змінних на залежну змінну. Другий підхід базується на припущенні, що параметри моделі мають певну закономірність, тобто пов'язані між собою деякими співвідношеннями.

Послідовне оцінювання параметрів виконується так: спочатку будують регресію залежної та незалежної змінних в один і той самий момент часу, потім до моделі додають ще одну змінну — незалежну змінну в попередній момент часу, тобто розглядають залежність показника від двох змінних. Далі в регресію вводиться ще одна змінна — у момент часу, зсунутий на два попередніх проміжки, і т. д. Кожна з моделей досліджується на адекватність і значущість її параметрів. Процедура закінчується, коли параметри при лагових змінних починають бути статистично незначущими та (або) коефіцієнт хоча б однієї змінної змінює свій знак.

Такий метод хоч і повний, однак має певні недоліки. По-перше, те, що невідомою є максимальна тривалість лага, а це не дає змоги передбачити, скільки змінних увійде в модель. По-друге, між послідовними значеннями змінних здебільшого спостерігається висока кореляція, що породжує проблему мультиколінеарності в моделі. Крім того, через зменшення ступенів свободи в таких моделях оцінки стають дещо непевними, що також знижує їх якість.

Наявність мультиколінеарності між лаговими змінними ускладнює побудову моделі. Щоб усунути мультиколінеарність, на коефіцієнти при лагових змінних накладають додаткові обмеження (апріорне оцінювання), а саме вибирають їх так, щоб вплив лага на досліджуваний показник був “односпрямований” (тобто коефіцієнти були б однакового знака) і зменшувався з кожним наступним кроком у минуле. Такі припущення реалізують, як правило, у моделях, де параметри змінюються в геометричній прогресії. Крім того, нескінченна сума членів спадної геометричної прогресії є скінченною величиною, що дає змогу узагальнити модель з кінцевим лагом і застосовувати однакові методи оцінювання параметрів.

Однак і в цьому разі залишається велика кількість оцінюваних параметрів.

Уведення в модель лагової залежної змінної y_{t-1} (затримка на один період), відоме як перетворення Койка, значно спрощує модель:

$$y_t = w(1 - \lambda)x_t + \lambda y_{t-1} + (u_t - \lambda u_{t-1}), \quad (7.2)$$

де $w = \sum_{j=1}^{\infty} a_j$ (скінченне число), $0 \leq \lambda \leq 1$.

Така модель містить не лише поточні, а й попередні значення залежної змінної, тобто є авторегресійною.

Заміна незалежних лагових змінних x_{t-1}, x_{t-2}, \dots однією залежною змінною y_{t-1} зменшує кількість оцінюваних параметрів і усуває проблему мультиколінеарності, однак призводить до нових труднощів. Наявність у моделі лагової залежної змінної потребує перевірки передумови про незалежність змінних і залишків при застосуванні звичайного МНК. Крім того, залишки моделі $v_t = u_t + \lambda u_{t-1}$ часто виявляються серійно корельованими, а тому при дослідженні їх на автокореляцію необхідно використати спеціальні тести.

Отримана алгебраїчним способом модель Койка позбавлена теоретичного обґрунтування і фактично є послідовною моделлю.

З певних економічних міркувань можна отримати моделі, що зовні нагадують модель Койка, але з іншою інтерпретацією коефіцієнтів лагових змінних. Такими моделями є модель адаптивних сподівань

$$y_t = \alpha(1 - \lambda) + \beta(1 - \lambda)x_t + \lambda y_{t-1} + u_t(1 - \lambda u_{t-1}) \quad (7.3)$$

та модель часткового коригування

$$y_t = \alpha\gamma + \beta\gamma x_t + (1 - \gamma)y_{t-1} + u_t, \quad 0 \leq \gamma \leq 1. \quad (7.4)$$

Ці моделі відрізняються від моделі Койка наявністю вільного члена, але при цьому реалізують різні ідеї щодо економічної діяльності. У першій моделі відображено думку про те, що люди навчаються з попереднього досвіду, причому нещодавній досвід має більший вплив, аніж попередній; друга базується на тому, що через інертність економічної системи зміна одного економічного показника не одразу впливає на зміну іншого і відповідний рівень залежної змінної досягається через певний час.

7.3. Оцінювання параметрів авторегресійних моделей

Три авторегресійних моделі – Койка (7.2), адаптивних сподівань (7.3) і часткового коригування (7.4) – можна подати в загальній формі:

$$y_t = a_0 + a_1 x_t + a_2 y_{t-1} + v_t. \quad (7.5)$$

Наявність лагових залежних змінних у динамічних моделях створює певні проблеми при оцінюванні параметрів: серед пояснюючих

змінних є стохастичні (залежні лагові змінні), а також існує проблема серійної кореляції залишків моделі та лагових змінних. Залежно від гіпотез щодо залишків таких моделей використовують відповідні методи оцінювання.

Гіпотеза 1. Залишки є нормально розподіленими випадковими величинами з нульовим математичним сподіванням та сталою дисперсією.

Гіпотеза 2. Залишки описуються авторегресійною схемою першого порядку:

$$v_t = u_t - \lambda u_{t-1}, \quad 0 \leq \lambda \leq 1.$$

Гіпотеза 3. Залишки автокорельовані та описуються авторегресійною схемою першого порядку:

$$v_t = u_t + \lambda u_{t-1},$$

крім того, $v_t = \rho v_{t-1} + \epsilon_t$.

Перша гіпотеза виконується лише для моделі часткового коригування (7.4); саме для неї можливе застосування звичайного МНК. Однак залежність залишків від лагової змінної y_{t-1} у цій моделі призводить до зміщення оцінок параметрів. Та хоча оцінки параметрів будуть завищеними, вони матимуть найменшу середньоквадратичну похибку. І після визначення величини зміщення МНК-оцінки будуть найприйнятнішими.

Якщо залишки моделі визначаються через автокорельовані випадкові величини, то МНК-оцінки параметрів моделі також матимуть зміщення, до того ж зміщення матиме також критерій Дарбіна – Уотсона. Тому для перевірки автокореляції залишків застосовують узагальнений критерій Дарбіна – Уотсона. Оцінювання параметрів таких моделей виконують узагальненим методом найменших квадратів (методом Ейткена), в операторі оцінювання якого

$$\hat{a} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

коригуюча матриця має вигляд

$$V = \sigma_u^2 \begin{pmatrix} 1 + \lambda^2 & -\lambda & 0 & \dots & 0 \\ -\lambda & 1 + \lambda^2 & -\lambda & \dots & 0 \\ 0 & -\lambda & 1 + \lambda^2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 + \lambda^2 \end{pmatrix}.$$

Якщо лагову модель можна подати у вигляді

$$y_t - \lambda y_{t-1} = a_0 + a_1 x_t + v_t,$$

то до перетворених у такий спосіб даних залежної змінної застосовують звичайний МНК. Причому параметр λ вибирають з інтервалу $0 \leq \lambda \leq 1$ так, щоб мінімізувати суму квадратів залишків $u'V^{-1}u$.

Якщо відносно залишків моделі приймається третя гіпотеза, то параметри оцінюють за допомогою таких методів:

- 1) класичного МНК після попереднього перетворення вхідних даних;
- 2) методу Ейткена (узагальненого МНК);
- 3) ітераційного методу;
- 4) методу інструментальних змінних;
- 5) алгоритму Уолліса.

Контрольні запитання

1. Що в економіці називається лагом?
2. Що є причиною лага?
3. Які регресійні моделі називають авторегресійними та дистрибутивно-лаговими?
4. Якою є природа авторегресійних моделей?
5. Які проблеми виникають при оцінюванні параметрів дистрибутивно-лагових і авторегресійних моделей?
6. Які гіпотези висуваються щодо залишків лагових моделей?
7. Якими методами оцінюються параметри дистрибутивно-лагових і авторегресійних моделей?

Розділ 8. СИСТЕМИ ОДНОЧАСНИХ РІВНЯНЬ

8.1. Поняття про системи одночасних рівнянь

Багато економічних взаємозв'язків допускають моделювання одним рівнянням. Однак деякі економічні процеси моделюються не одним, а кількома рівняннями. Співвідношення між економічними показниками можуть мати стохастичний і детермінований характер. Стохастичні зв'язки між змінними описуються регресійними рівняннями, а детерміновані визначаються тотожностями й не містять невідомих параметрів.

У системах рівнянь через наявність прямих і зворотних зв'язків залежна змінна одного рівняння може бути незалежною змінною в інших рівняннях. Змінні, що стоять у лівій частині рівнянь, називаються ендогенними, причому їх кількість не перевищує загальної кількості всіх рівнянь. Інші змінні, що входять до моделі, називаються екзогенними.

Наприклад, повна кейнсіанська модель доходу складається з двох співвідношень:

$$C_t = a_0 + a_1 Y_t + u_t, \quad (8.1)$$

$$Y_t = C_t + Z_t, \quad (8.2)$$

де C_t – витрати на споживання; Y_t – дохід; a_0, a_1 – невідомі параметри; u_t – залишки моделі; Z_t – неспоживчі витрати (інвестиції).

Перше співвідношення – це регресійна функція споживання, а друге – тотожність доходу. Величина доходу Y_t для першого рівняння є незалежною змінною, для другого – залежною, а величина C_t –

навпаки: у першому рівнянні вона є залежною змінною, у другому – незалежною. Для системи загалом змінні Y_t і C_t є ендогенними, а змінна Z_t – екзогенною.

Означення 8.1. Для систем одночасних рівнянь усі змінні, що можуть бути визначені із системи рівнянь, називаються ендогенними, причому їх кількість не перевищує загальної кількості рівнянь.

Означення 8.2. Для систем одночасних рівнянь усі змінні, які задаються за межами моделі або є заздалегідь відомими, називаються відповідно екзогенними або предетермінованими.

У розглянутій кейнсіанській моделі доходу величини C_t і Y_t є ендогенними змінними, що визначаються всередині моделі. Змінна Z_t задається (визначається) поза моделлю, отже, вона є екзогенною.

Із першого співвідношення цієї моделі видно, що змінна C_t залежить від доходу Y_t і від залишків u_t , а з другого співвідношення очевидна залежність доходу Y_t від споживчих C_t і неспоживчих витрат Z_t . Незважко помітити, що обидві змінні C_t і Y_t можуть бути виражені через Z_t і залишки u_t .

8.2. Приклади систем одночасних рівнянь

1. Модель “попит – пропозиція”.

Одна з найпростіших систем одночасних рівнянь, що використовується при моделюванні попиту та пропозиції в ринковій економіці, має вигляд

$$\begin{array}{l} \text{Функція попиту} \\ \text{Функція пропозиції} \\ \text{Функція рівноваги} \end{array} \quad \left\{ \begin{array}{l} q_t^D = \alpha_0 + \alpha_1 p_t + u_{t1}, \quad \alpha_1 < 0, \\ q_t^S = \beta_0 + \beta_1 p_t + u_{t2}, \quad \beta_1 < 0, \\ q_t^D = q_t^S. \end{array} \right.$$

Припускається, що обсяг попиту q^D і обсяг пропозиції q^S певного товару в момент часу t є лінійними регресійними функціями від ціни цього товару p_t у цей самий момент часу. Останнє співвідношення в цій моделі – функція рівноваги – є тотожністю.

Наявність випадкових відхилень u_{t1} і u_{t2} у даній моделі пов'язана передусім з відсутністю ряду важливих пояснюючих змінних (прибутку споживачів, цін на супутні товари, цін на ресурси, податків тощо).

Зміна одного з цих факторів може відбитися на моделі. Наприклад, зростання прибутку споживачів може зсунути лінію попиту вгору (рис. 8.1). Це призведе до зміни рівноважної ціни та рівноважної кількості.

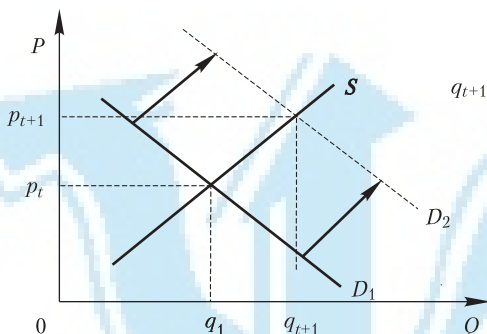


Рис. 8.1

Модель “попит — пропозиція” можна вдосконалити. Наприклад, якщо до функції попиту додати прибуток споживачів y_t , дістанемо систему

Функція попиту	$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + u_{t1}, & \alpha < 0, \\ q_t^S = \beta_0 + \beta_1 p_t + u_{t2}, & \beta_1 < 0, \\ q_t^D = q_t^S. \end{cases}$
Функція пропозиції	
Функція рівноваги	

2. Модель рівноваги на ринку товарів (модель IS).

Однією з можливих нестохастичних форм моделі IS (рівноваги на ринку товарів) є така модель:

Функція споживання	$c_t = \beta_0 + \beta_1 y_t,$	(8.3)
--------------------	--------------------------------	-------

Функція податків	$\tau_t = \alpha_0 + \alpha_1 y_t,$	(8.4)
------------------	-------------------------------------	-------

Функція інвестицій	$i_t = \gamma_0 + \gamma_1 r_t,$	(8.5)
--------------------	----------------------------------	-------

Визначення	$y(d)_t = y_t + \tau_t,$	(8.6)
------------	--------------------------	-------

Державні витрати	$g_t = \bar{g},$	(8.7)
------------------	------------------	-------

Макроекономічна тотожність	$y_t = c_t + i_t + g_t,$	(8.8)
----------------------------	--------------------------	-------

де $c_t, y_t, \tau_t, i_t, g_t, r_t, y(d)_t$ — відповідно значення в момент часу t споживання (c_t), національного доходу (y_t), обсягу податків (τ_t), бажаного обсягу чистих інвестицій (i_t), процентної ставки (r_t), розміщеного прибутку ($y(d)_t$), державних витрат (g_t), у даному разі $g_t = \bar{g}_t = \text{const}$.

Щоб отримати в явному вигляді співвідношення між процентною ставкою й рівнем прибутку, при якому ринок товарів перебуває у стані рівноваги, необхідно у (8.3) підставити (8.4) і (8.6). Підставивши отримане співвідношення, а також (8.5) і (8.7) у (8.8), дістанемо

$$y_t = \pi_0 + \pi_1 r_t, \quad (8.9)$$

$$\text{де } \pi_0 = \frac{\beta_0 + \alpha_0 \beta_1 + \gamma_0 + \bar{g}}{1 - \beta_1(1 - \alpha_1)}; \quad \pi_1 = \frac{1}{1 - \beta_1(1 - \alpha_1)}.$$

3. Модель рівноваги на ринку грошей (модель *LM*).

Рівновага на ринку грошей задається таким співвідношенням між процентною ставкою та рівнем доходу, при якому попит на гроші дорівнює їх пропозиції. Наведемо одну із нестохастичних форм такої моделі:

$$\text{Функція попиту на гроші} \quad \left\{ \begin{array}{l} M_t^D = a + by_t - cr_t, \end{array} \right. \quad (8.10)$$

$$\text{Функція пропозиції грошей} \quad \left\{ \begin{array}{l} M_t^S = M, \end{array} \right. \quad (8.11)$$

$$\text{Умова рівноваги} \quad \left\{ \begin{array}{l} M_t^D = M_t^S. \end{array} \right. \quad (8.12)$$

Співвідношення (8.10) можна записати у вигляді

$$y_t = \lambda_0 + \lambda_1 M + \lambda_2 r_t, \quad (8.13)$$

$$\text{де } \lambda_0 = -a/b, \quad \lambda_1 = 1/b, \quad \lambda_2 = c/b.$$

Співвідношення (8.11) відоме як рівняння *LM*. Спільну модель *IS-LM* зображено на рис. 8.2.

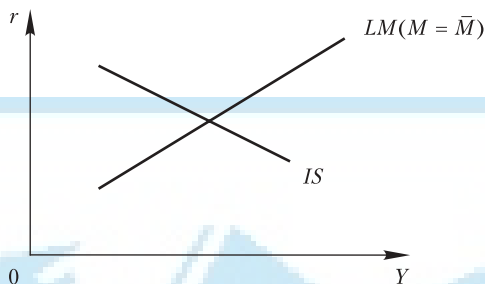


Рис. 8.2

Точка перетину ліній IS і LM визначає співвідношення між процентною ставкою й рівнем доходу, при якому обидва ринки перебувають у стані рівноваги. Ця точка визначається як розв'язок системи рівнянь (8.1)–(8.6) і (8.8)–(8.10).

8.3. Структурна та зведена (прогнозна) форми системи рівнянь

1. Структурна форма економетричної моделі.

Структурна форма економетричної моделі описує одно- та багатоваріантні стохастичні причинні співвідношення між економічними величинами в їх безпосередньому вигляді. Вона містить усю суттєву інформацію про залежності між економічними явищами та процесами. Кожне співвідношення такої системи (рівняння чи тотожність) має певну економічну інтерпретацію. Структурні рівняння системи описують окремо економічні явища з урахуванням економічних, технологічних, демографічних, соціологічних та інших факторів, що спричиняють зміну залежних змінних. Характерною особливістю структурних рівнянь є їх певна автономність щодо визначених змінних, оскільки зміна останніх в одному структурному рівнянні не обов'язково зумовлює зміну залежних змінних в інших рівняннях.

Для адекватного відображення реальної дійсності та повного охоплення економічних показників одночасними співвідношеннями в системах застосовують також тотожності — детерміновані залежності економічних величин. Тотожності не містять випадкових складових, а параметри їх заздалегідь відомі (найчастіше вони дорівнюють одиниці), тому вони не підлягають оцінюванню. Отже, справедливим буде таке означення.

Означення 8.3. Економетрична модель, що відображає структуру зв'язків між змінними, називається структурною формою моделі.

У загальному випадку структурна форма моделі має вигляд

$$Ay_t = Bx_t + u_t, \quad (8.14)$$

де y_t — вектор залежних (ендогенних) змінних; x_t — вектор незалежних (екзогенних) змінних; u_t — вектор залишків, $t = 1, 2, \dots, T$.

2. Повна економетрична модель.

Економетрична модель називається *повною*, якщо:

- а) вона охоплює змінні, що суттєво впливають на *спільно залежні* змінні, а вектор залишків має випадковий характер;
- б) містить стільки рівнянь, скільки в ній є спільно залежних змінних, тобто кожна залежна змінна пояснюється окремим рівнянням;
- в) система рівнянь має однозначний розв'язок відносно спільно залежних змінних, тобто матриця A в моделі (8.14) не вироджена (має відмінний від нуля визначник):

$$\det A \neq 0.$$

Повна модель застосовується у випадках, коли необхідно кількісно описати економічне явище чи процес або спрогнозувати їх розвиток.

3. Зведена форма економетричної моделі.

Якщо економетрична модель застосовується не для аналізу системи, а для передбачення чи оцінювання параметрів, структурна форма моделі неприйнятна. Алгебраїчними перетвореннями систему структурних рівнянь зводять до форми, у якій кожне рівняння містить лише одну ендогенну змінну, яка є функцією від екзогенних змінних. Така форма рівнянь називається *зведеною*.

Зведену форму рівнянь можна назвати *скороченою*. Це пов'язано з тим, що при певних перетвореннях багато окремих економічних залежностей можуть бути виключені з розгляду, а отже, загальна кількість рівнянь може скоротитися.

Внаслідок таких перетворень зведена форма рівнянь, на відміну від структурної, не має ні безпосередньої, ні будь-якої економічної інтерпретації. Рівняння у зведеній формі дають змогу передбачити, як зміниться значення ендогенної змінної, якщо змінюватимуться значення екзогенних змінних, однак на підставі цих рівнянь неможливо пояснити, як і чому це відбувається. Саме через це зведену форму рівнянь називають також *прогнозною*.

Отже, коли виникає питання про консультації чи практичні поради, системи рівнянь у зведеній формі особливо корисні, оскільки дають змогу формальну модель звести до мінімальної кількості співвідношень. Звичайно, зведена модель матиме цінність, якщо правильною є початкова структурна модель.

Зокрема, якщо економетрична модель повна, то її залежні змінні можна представити в явному вигляді як функції від спільно незалежних змінних, розв'язавши її відносно вектора залежних змінних y_t . Це можливо, оскільки за означенням матриця A такої моделі є не виродженою; після множення системи (8.14) на її обернену матрицю отримаємо

$$y_t = A^{-1}Bx_t + A^{-1}u_t. \quad (8.15)$$

При таких перетвореннях параметри зведеної форми стають функціями від параметрів вихідних структурних рівнянь і залишки такої моделі, очевидно, є лінійною комбінацією залишків структурної моделі.

Увівши позначення $v_t = A^{-1}u_t$, $R = A^{-1}B$, отримаємо спрощений вигляд моделі:

$$y_t = Rx_t + v_t. \quad (8.16)$$

У такій системі кожна залежна змінна визначається через незалежні змінні моделі, тобто система (8.16) є зведеною формою економетричної моделі.

8.4. Поняття ідентифікації (ототожнення) системи рівнянь

Маючи дві форми системи одночасних рівнянь, необхідно визначити, яка з них краще підходить для оцінювання параметрів моделі. Передусім необхідно дослідити можливості застосування звичайного МНК до окремих рівнянь системи.

Як зазначалося (тема 3), для отримання незміщених і обґрунтованих оцінок параметрів регресійного рівняння за звичайним МНК необхідно виконати ряд передумов: залишки моделі мають бути випадковими величинами з нульовим математичним сподіванням, зі сталими дисперсіями, некорельованими між собою та незалежними відносно ендогенних змінних моделі.

З'ясуємо, як виконуються ці передумови для кейнсіанської моделі (8.1)–(8.2).

Нехай залишки моделі u_t є випадковими, з нульовим математичним сподіванням, некорельовані між собою, мають однакові дисперсії для всіх спостережень, тобто задовольняють перші дві передумови застосування МНК. Перевіримо передумову відносно незалежності ендогенних змінних і залишків моделі, тобто переконаємося, що $\text{cov}(Y_t, u_t) = 0$ для будь-яких відхилень.

Підставивши значення C_t з першого рівняння (8.1) моделі в друге (8.2), отримуємо співвідношення

$$Y_t = a_0 + a_1 Y_t + Z_t + u_t,$$

розв'язавши яке відносно Y_t , матимемо

$$Y_t = \frac{a_0}{1 - a_1} + \frac{1}{1 - a_1} Z_t + \frac{u_t}{1 - a_1}. \quad (8.17)$$

Значимо, що коефіцієнт $\frac{1}{1 - a_1}$ в останньому співвідношенні є грошовим мультиплікатором, що визначає, на яку величину зростає сукупний прибуток зі збільшенням обсягу інвестицій на одиницю.

Наявність коефіцієнта $\frac{1}{1 - a_1}$ при u_t свідчить про залежність між змінною Y_t і залишками моделі. Дійсно, з (8.17) маємо

$$M(Y_t) = \frac{a_0}{1 - a_1} + \frac{1}{1 - a_1} Z_t. \quad (8.18)$$

В останньому співвідношенні враховано те, що $M(u_t) = 0$, а також те, що змінна Z_t є екзогенною (незалежною) для даної моделі. Тоді різниця між (8.17) і (8.18) становить

$$Y_t - M(Y_t) = \frac{u_t}{1 - a_1}.$$

Отже,

$$\begin{aligned} \text{cov}(Y_t, u_t) &= M((Y_t - M(Y_t))(u_t - M(u_t))) = \\ &= M\left(\frac{u_t}{1 - a_1} u_t\right) = \frac{1}{1 - a_1} M(u_t^2) = \frac{\sigma^2}{1 - a_1} > 0. \end{aligned}$$

Тут ми скористалися твердженням економічної теорії про те, що гранична схильність до споживання a_1 перебуває в межах $0 < a_1 < 1$.

Отже, залишки моделі корелюють із залежною змінною, тому застосування звичайного МНК дасть зміщені та необґрунтовані оцінки параметрів моделі. В останньому можна перекоонатися, проаналізувавши оцінку \hat{a}_1 параметра a_1 рівняння (8.1), отриману за МНК.

Щоб забезпечити необхідну якість оцінок параметрів (незміщеність, ефективність і обґрунтованість), намагаються на підставі оцінених параметрів скороченої (зведеної) форми системи рівнянь отримати оцінки параметрів структурної форми. Однак тут виникає проблема однозначних залежностей між параметрами: при поверненні від скороченої форми моделі до структурної (обернені перетворення) можна отримати єдине значення шуканого параметра чи кілька різних значень або взагалі не мати змоги отримати жодного.

Щоб передбачити можливі варіанти розв'язання задачі оцінювання параметрів системи одночасних рівнянь, необхідно попередньо дослідити модель, а саме перевірити ідентифікованість системи. Під проблемою ідентифікації розуміють можливість чисельної оцінки параметрів структурних рівнянь за оцінками коефіцієнтів зведених рівнянь.

Означення 8.4. Економетрична модель, задана системою одночасних рівнянь, називається точно (строго) ідентифікованою (ототоженою), якщо однозначно можна отримати оцінки її параметрів на основі оцінених параметрів зведеної моделі.

Означення 8.5. Надідентифікованою (переототоженою) називається така модель, що для деяких її параметрів можна отримати кілька кількісних значень на підставі параметрів зведеної форми.

Крім того, модель може бути неідентифікованою (неототоженою). Це трапляється в тому разі, якщо кількість невідомих параметрів набагато перевищує кількість рівнянь, через які їх треба оцінити.

Отже, перехід від структурної до зведеної форми системи рівнянь хоча й дає змогу усунути проблему корельованості пояснюючої змінної та випадкового відхилення, однак призводить до іншої не менш серйозної проблеми — *проблеми ідентифікованості*.

Щоб зрозуміти проблему ідентифікованості, необхідно усвідомити суть принципів розбіжностей між структурними та зведеними

рівняннями. Наприклад, у моделі “попит – пропозиція”

$$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + u_{1t}, & \alpha_1 < 0, \\ q_t^S = \beta_0 + \beta_1 p_t + u_{2t}, & \beta_1 < 0, \\ q_t^D = q_t^S \end{cases}$$

оцінки коефіцієнтів поведінкових рівнянь визначають функції попиту та пропозиції. Оцінюючи коефіцієнти зведених рівнянь, ми визначаємо точку перетину кривих попиту та пропозиції, тобто рівноважну ціну товару та його рівноважну кількість. Очевидно, обчисливши ці значення, неможливо відновити функції попиту та пропозиції, тому що через одну точку на площині можна провести нескінченно багато ліній.

Побудуємо зведені рівняння для цієї моделі. Використавши умову рівноваги, отримаємо

$$\alpha_0 + \alpha_1 p_t + u_{1t} = \beta_0 + \beta_1 p_t + u_{2t}.$$

Останнє рівняння, розв’язане відносно p_t , має вигляд

$$p_t = \lambda_0 + u_t, \quad (8.19)$$

де $\lambda_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$, $u_t = \frac{u_{2t} - u_{1t}}{\alpha_1 - \beta_1}$ – випадкова складова.

Підставляючи знайдене значення p_t у початкові рівняння, отримуємо

$$q_t = \lambda_1 + v_t, \quad (8.20)$$

де $\lambda_1 = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1}$, $v_t = \frac{\alpha_1 u_{2t} - \alpha_1 u_{1t}}{\alpha_1 - \beta_1}$ – випадковий член.

Рівняння (8.19) і (8.20) утворюють систему зведених рівнянь. Однак система структурних рівнянь має чотири невідомі коефіцієнти: $\alpha_0, \alpha_1, \beta_0, \beta_1$. Із курсу алгебри відомо, що для однозначного визначення k невідомих необхідно мати щонайменше k (незалежних) рівнянь. Отже, ми не зможемо однозначно визначити чотири коефіцієнти, маючи лише систему з двох рівнянь:

$$\begin{cases} \lambda_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}, \\ \lambda_1 = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1}. \end{cases}$$

Неважко помітити, що, відкинувши випадкові залишки у зведених рівняннях (8.19) і (8.20), можна встановити значення $p_t = \lambda_0$ та $q_t = \lambda_1$, яке фактично визначає точку перетину кривих попиту та пропозиції (точку ринкової рівноваги). Але через одну точку можна провести як завгодно багато ліній (рис. 8.3, а). Тому для визначення конкретних прямих необхідна додаткова інформація, яку можна отримати за рахунок екзогенних змінних, що входять до структурних рівнянь.

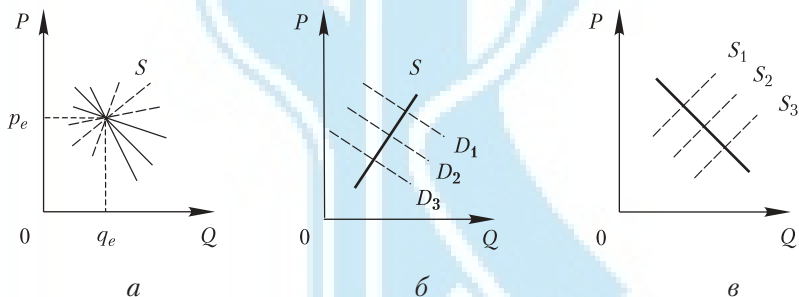


Рис. 8.3

Наприклад, нехай до функції попиту додано ще одну пояснюючу (екзогенну) змінну y_t — прибуток споживачів. Тоді модель “попит — пропозиція” матиме вигляд:

$$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + u_{1t} & (\alpha_1 < 0, \alpha_2 > 0), \\ q_t^S = \beta_0 + \beta_1 p_t + u_{2t}. \end{cases} \quad (8.21)$$

Таке доповнення до моделі дає деяку додаткову інформацію про поведінку споживача. Згідно з економічною теорією, для нормальних товарів $\alpha_2 > 0$.

Прирівнявши обсяг попиту і обсяг пропозиції, матимемо

$$\alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \varepsilon_{1t} = \beta_0 + \beta_1 p_t + \varepsilon_{2t}$$

або

$$p_t = \lambda_0 + \lambda_1 y_t + v_t, \quad (8.22)$$

$$\text{де } \lambda_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}, \quad \lambda_1 = -\frac{\alpha_2}{\alpha_1 - \beta_1}, \quad v_t = \frac{\varepsilon_{2t} - \varepsilon_{1t}}{\alpha_1 - \beta_1}.$$

Прирівнявши ціну попиту та ціну пропозиції в точці рівноваги, отримаємо

$$q_t = \lambda_2 + \lambda_3 y_t + w_t, \quad (8.23)$$

$$\text{де } \lambda_2 = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1}, \quad \lambda_3 = -\frac{\alpha_2 \beta_1}{\alpha_1 - \beta_1}, \quad w_t = \frac{\alpha_1 \varepsilon_{2t} - \beta_1 \varepsilon_{1t}}{\alpha_1 - \beta_1}.$$

Рівняння (8.22) і (8.23) є зведеними. Застосувавши МНК, неважко знайти оцінки їх параметрів λ_0 , λ_1 , λ_2 , λ_3 . Однак цього недостатньо для того, щоб оцінити п'ять параметрів α_0 , α_1 , α_2 , β_0 , β_1 початкової системи структурних рівнянь. Ми можемо визначити параметри β_0 і β_1 функції пропозиції системи (8.21):

$$\begin{cases} \beta_1 = \frac{\lambda_3}{\lambda_1}, \end{cases} \quad (8.24)$$

$$\begin{cases} \beta_0 = \lambda_2 - \beta_1 \lambda_0. \end{cases} \quad (8.25)$$

Але α_0 , α_1 , α_2 визначити однозначно не можна. Отже, потрібно деяке довизначення. Зауважимо, що введенням пояснючої змінної у функцію попиту (перше рівняння системи (8.21) ми визначили функцію пропозиції (друге рівняння цієї самої моделі) (рис. 8.3).

Якщо у функцію пропозиції ввести пояснючу змінну (наприклад, заздалегідь визначену змінну p_{t-1}), виключивши при цьому з функції попиту змінну, що визначає прибуток, можна отримати конкретну функцію попиту при невизначеній функції пропозиції (рис. 8.3). Цей висновок обґрунтовується за аналогією з попередньо описаною схемою та рекомендується як вправа для самостійної роботи.

Зазначимо, що якщо в кожне зі структурних рівнянь моделі “попит – пропозиція” поряд із ціною товару буде введено по одній пояснюючій (екзогенно визначеній) змінній (наприклад, y_t у функцію попиту й p_{t-1} у функцію пропозиції), то коефіцієнти структурних рівнянь можуть бути оцінені однозначно. У цьому разі модель буде однозначно визначеною, тобто ідентифікованою.

Розглянемо модель “попит – пропозиція” з кількістю екзогенних змінних, що перевищує кількість структурних рівнянь:

$$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \alpha_3 s_t + \varepsilon_{1t}, \\ q_t^S = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + \varepsilon_{2t}, \end{cases}$$

де змінна s_t – обсяг заощаджень до моменту часу t .

З умови ринкової рівноваги нескладно отримати такі зведені рівняння:

$$\begin{cases} p_t = \lambda_0 + \lambda_1 y_t + \lambda_2 s_t + \lambda_3 p_{t-1} + v_t, \\ q_t = \lambda_4 + \lambda_5 y_t + \lambda_6 s_t + \lambda_7 p_{t-1} + w_t, \end{cases}$$

де

$$\begin{cases} \lambda_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}; & \lambda_1 = -\frac{\alpha_2}{\alpha_1 - \beta_1}; & \lambda_2 = -\frac{\alpha_3}{\alpha_1 - \beta_1}; \\ \lambda_3 = \frac{\beta_2}{\alpha_1 - \beta_1}; & \lambda_4 = \frac{\alpha_1 \beta_0 - \alpha_0 \beta_1}{\alpha_1 - \beta_1}; \\ \lambda_5 = -\frac{\alpha_2 \beta_1}{\alpha_1 - \beta_1}; & \lambda_6 = -\frac{\alpha_3 \beta_1}{\alpha_1 - \beta_1}; & \lambda_7 = -\frac{\alpha_1 \beta_2}{\alpha_1 - \beta_1}; \end{cases} \quad (8.26)$$

$$v_t = \frac{\alpha_1 \varepsilon_{2t} - \beta_1 \varepsilon_{1t}}{\alpha_1 - \beta_1}; \quad w_t = \frac{\varepsilon_{2t} - \varepsilon_{1t}}{\alpha_1 - \beta_1}.$$

Для оцінки семи структурних коефіцієнтів $\alpha_0, \alpha_1, \alpha_2, \alpha_3, \beta_0, \beta_1, \beta_2$ у цьому разі отримано вісім рівнянь (8.26). Як наслідок, однозначне визначення структурних коефіцієнтів неможливе через суперечливість співвідношень. Наприклад, з (8.26) випливає неможливість визначення β_1 . Дійсно, $\beta_1 = \lambda_6 / \lambda_2$ і $\beta_1 = \lambda_5 / \lambda_1$. Але це

можливо лише за умови $\lambda_6/\lambda_2 = \lambda_5/\lambda_1$, що нереально, оскільки коефіцієнт β_1 , який міститься в усіх рівняннях для оцінки зведених коефіцієнтів, також недосконалий. У цьому разі маємо ситуацію невизначеності або надідентифікованості, тобто “занадто багато” інформації (обмежень) для визначення лінії доходу. Через суперечливість інформації неможливо отримати шуканий розв’язок.

У ситуації неідентифікованості “занадто мало” інформації, а тому існує кілька різних ліній, що задовольняють обмеження моделі.

Необхідні й достатні умови ідентифікованості

Щоб швидше формально визначити ідентифікованість структурних рівнянь, застосовують такі необхідні й достатні умови. Нехай система одночасних рівнянь містить N рівнянь відносно N ендогенних змінних, а також M екзогенних або заздалегідь визначених змінних. Крім того, для деякого рівняння кількість ендогенних і екзогенних змінних у перевірці на ідентифікованість дорівнює відповідно n і t . Змінні, що не входять у дане рівняння, але входять в інші рівняння системи, назвемо *виключеними змінними* (з даного рівняння). Їх кількість дорівнює $N-n$ для ендогенних і $M-t$ для екзогенних змінних.

Перша необхідна умова. Рівняння ідентифіковане, якщо воно виключає принаймні $N-1$ змінну (ендогенну чи екзогенну), що присутня в моделі:

$$(N - n) + (M - t) \geq N - 1.$$

Друга необхідна умова. Рівняння ідентифіковане, якщо кількість виключених з нього екзогенних змінних не менше кількості ендогенних змінних у цьому рівнянні, зменшеної на одиницю: $M - t \geq n - 1$.

Знаки рівності в обох необхідних умовах відповідають точній ідентифікованості рівняння.

Наведемо приклади використання зазначених умов для визначення ідентифікованості структурних рівнянь.

У простій моделі “попит – пропозиція”

$$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \varepsilon_{1t}, \\ q_t^S = \beta_0 + \beta_1 p_t + \varepsilon_{2t} \end{cases}$$

$N = 2, M = 0$. Для кожного з рівнянь $n = 2, m = 0$. Отже, перша необхідна умова, а саме $(N - n) + (M - m) \geq N - 1$, не виконується для обох рівнянь, тому що в цьому разі $(N - n) + (M - m) = 0 < N - 1 = 1$. Це означає, що вони обидва неідентифіковані.

2. У моделі (8.21) до функції попиту додано екзогенну змінну y_t (прибуток споживачів):

$$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \varepsilon_{1t} & (\alpha_1 < 0, \alpha_2 > 0), \\ q_t^S = \beta_0 + \beta_1 p_t + \varepsilon_{2t} & (\beta_1 > 0), \end{cases}$$

$N = 2, M = 1$. Для кожного з рівнянь $n = 2$. Для першого рівняння $m = 1$, для другого $m = 0$. Тоді для першого рівняння $(N - n) + (M - m) = 0 < 1 = N - 1$. А це означає, що перша необхідна умова не виконується і дане рівняння неідентифіковане. Для другого рівняння цієї системи $(N - n) + (M - m) = 1 = N - 1$, тобто дане рівняння точно ідентифіковане. Отже, функція пропозиції може бути визначена однозначно.

3. У моделі

$$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \varepsilon_{1t}, \\ q_t^S = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + \varepsilon_{2t} \end{cases}$$

$N = 2, M = 2$. Для кожного рівняння $n = 2, m = 1$. У цьому разі для кожного з рівнянь виконується умова $(N - n) + (M - m) = 1 = N - 1$. Отже, обидва рівняння цієї системи точно ідентифіковані.

3. У моделі “попит – пропозиція”, де враховано три екзогенні змінні:

$$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \alpha_3 s_t + \varepsilon_{1t}, \\ q_t^S = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + \varepsilon_{2t}, \end{cases}$$

$N = 2, M = 3$. Для кожного рівняння системи $n = 2$. Кількість виключених змінних у першому рівнянні $m = 2$. Тоді перше рівняння точно ідентифіковане, тому що для нього $(N - n) + (M - m) = 1 = N - 1$. Для другого рівняння $m = 1$. Отже, для нього $(N - n) + (M - m) = 2 > 1 = N - 1$. Це рівняння є перевизначеним.

Для однозначної оцінки коефіцієнтів функції пропозиції в цьому разі необхідно використовувати інші спеціальні методи оцінювання параметрів.

Необхідна і достатня умова ідентифікованості

У моделі, що містить N рівнянь відносно N ендогенних змінних, умова ідентифікованості виконується тоді і тільки тоді, коли ранг матриці, складеної з виключених з даних рівнянь змінних, але таких, що містяться в інших рівняннях системи, дорівнює $N - 1$.

8.5. Методи оцінювання параметрів систем рівнянь

Як зазначалося, застосування звичайного МНК до рівнянь структурної форми системи рівнянь призводить до отримання зміщених оцінок параметрів через корельованість (залежність) змінних і залишків моделі, що є порушенням однієї з передумов застосування МНК. Перехід від структурної форми моделі до скороченої є одним із способів, що усуває проблему корельованості, однак породжує іншу, а саме проблему ідентифікованості окремих рівнянь системи, а також системи загалом.

Залежно від розв'язання цієї проблеми, тобто після перевірки умови ідентифікованості кожного рівняння системи, застосовують такі методи:

1) якщо кожне рівняння системи точно ідентифіковане, то параметри зведеної моделі оцінюють непрямим методом найменших квадратів (НМНК); ідея методу полягає в тому, щоб від структурної форми перейти до зведеної, звичайним МНК оцінити параметри останньої й оберненим перетворенням отримати оцінки параметрів структурної форми;

2) усунути кореляцію між змінними та залишками моделі можна також за допомогою методу інструментальних змінних, ідея якого полягає в тому, щоб змінні, що корелюють із залишками, замінити іншими – інструментальними, які тісно пов'язані з незалежними змінними моделі, але зовсім не пов'язані з її залишками;

3) якщо рівняння структурної форми моделі надідентифіковані, то параметри моделі оцінюють за допомогою двокрокового методу найменших квадратів (2МНК), що передбачає виконання двох етапів:

а) перший – ендогенні змінні “звільняють” від стохастичних залишків;

б) другий – оцінені рівняння підставляють у структурну систему рівнянь, до яких потім застосовують звичайний МНК;

4) трикроковий метод найменших квадратів для одночасного оцінювання всіх рівнянь системи за певних обставин ефективніший порівняно з непрямым і двокроковим МНК.

8.5.1. Непрямий метод найменших квадратів оцінювання параметрів точно ідентифікованих систем

Оскільки на основі звичайного МНК неможливо отримати якісні оцінки параметрів системи одночасних рівнянь, варто скористатися іншими методами оцінювання параметрів. Одним із них є непрямий метод найменших квадратів, що ґрунтується на використанні зведених рівнянь.

Для ілюстрації НМНК розглянемо кейнсіанську модель формування прибутків:

$$C_t = a_0 + a_1 Y_t + u_t,$$

$$Y_t = C_t + Z_t.$$

У зведеній формі ця модель має вигляд

$$Y_t = \frac{a_0}{1 - a_1} + \frac{1}{1 - a_1} Z_t + \frac{u_t}{1 - a_1}, \quad (8.27)$$

$$C_t = \frac{a_0}{1 - a_1} + \frac{a_1}{1 - a_1} Z_t + \frac{u_t}{1 - a_1}. \quad (8.28)$$

Позначимо

$$\frac{a_0}{1 - a} = \lambda_{10}, \quad \frac{1}{1 - a_1} = \lambda_{11}, \quad \frac{a_0}{1 - a_1} = \lambda_{20},$$

$$\frac{a_1}{1 - a_1} = \lambda_{21}, \quad \frac{u_t}{1 - a_1} = v_t \sim N\left(0, \frac{\sigma^2}{1 - a_1}\right).$$

Тоді замість останніх співвідношень отримаємо

$$\begin{cases} Y_t = \lambda_{10} + \lambda_{11}Z_t + v_t, \\ C_t = \lambda_{20} + \lambda_{21}Z_t + v_t. \end{cases}$$

Через те, що обсяг інвестицій Z_t є екзогенною змінною моделі, ця змінна не корелює з випадковим залишком v_t у зведеній формі системи рівнянь (8.27), (8.28), а отже, і з залишками v_t останньої системи. Це означає, що для випадкового члена v_t виконуються передумови МНК. Тому оцінки $\hat{\lambda}_{10}, \hat{\lambda}_{11}, \hat{\lambda}_{20}, \hat{\lambda}_{21}$ параметрів $\lambda_{10}, \lambda_{11}, \lambda_{20}, \lambda_{21}$, отримані за МНК, будуть незміщеними, обґрунтованими і ефективними. Знаючи ці оцінки, нескладно визначити оцінки \hat{a}_0 і \hat{a}_1 коефіцієнтів a_0 і a_1 рівняння початкової структурної системи:

$$\hat{a}_1 = \frac{\hat{\lambda}_{21}}{\hat{\lambda}_{11}}; \quad \hat{a}_0 = \frac{\hat{\lambda}_{20}}{\hat{\lambda}_{11}}.$$

Визначення оцінок за зазначеною схемою називається непрямим методом найменших квадратів. Зміст такої назви очевидний: перш ніж застосувати звичайний МНК, початкову систему перетворили до зведеної форми, за МНК-оцінками якої визначили оцінки початкової системи.

Оцінки параметрів a_0 та a_1 , отримані за НМНК, є обґрунтованими, а тому при великих вибірках існує висока ймовірність, що вони наближатимуться до істинних значень параметрів.

Зазначимо, що в цьому разі оцінки визначаються точно, а регресійне рівняння в розглянутій моделі доходу є ідентифікованим (однозначно визначеним).

Отже, при непрямому МНК виконуються такі кроки:

- 1) перевіряється умова ідентифікованості для кожного рівняння структурної форми. Якщо всі рівняння точно ідентифіковані, то виконується наступний крок, інакше застосовується інший метод;
- 2) вихідна структурна система рівнянь перетворюється до зведеної форми;
- 3) оцінюються за МНК параметри рівнянь у зведеній формі;
- 4) на основі оцінок, знайдених для зведеної форми, обчислюються параметри структурних рівнянь за допомогою обернених перетворень.

Приклад. Розглядається модель “попит – пропозиція”:

Пропозиція	}	$q_t = \beta_0 + \beta_1 p_t + \varepsilon_{1t},$
Попит		

де q_t, p_t — ендогенні змінні (кількість товару і ціна в році t); $\varepsilon_{1t}, \varepsilon_{2t}$ — випадкові відхилення; y_t — екзогенна змінна (прибуток споживачів).

На підставі наступних статистичних даних необхідно оцінити коефіцієнти функції пропозиції, використовуючи МНК і НМНК. Порівняти результати.

p_t	q_t	y_t	p_t^2	y_t^2	$p_t q_t$	$p_t y_t$	$q_t y_t$	
1	8	2	1	4	8	2	16	
2	10	4	4	16	20	8	40	
3	7	3	9	9	21	9	21	
4	5	5	16	25	20	20	25	
5	1	2	25	4	5	10	2	
15	31	16	55	58	74	49	104	Сума
3	6,2	3,2	11	11,6	14,8	9,8	20,8	Середнє

Побудуємо зведені рівняння зазначеної системи, віднявши від функції пропозиції функцію попиту. Отримаємо

$$(\beta_0 - \alpha_0) + (\beta_1 - \alpha_1)p_t - \alpha_2 y_t + (\varepsilon_{1t} - \varepsilon_{2t}) = 0,$$

звідки

$$p_t = \pi_{10} + \pi_{11}y_t + v_{1t},$$

$$q_t = \pi_{20} + \pi_{21}y_t + v_{2t},$$

де
$$\pi_{10} = \frac{\alpha_0 - \beta_0}{\beta_1 - \alpha_1}, \quad \pi_{11} = \frac{\alpha_2}{\beta_1 - \alpha_1}, \quad v_{1t} = \frac{\varepsilon_{2t} - \varepsilon_{1t}}{\beta_1 - \alpha_1};$$

$$\pi_{20} = \beta_0 + \beta_1 \pi_{10}, \quad \pi_{21} = \beta_1 \pi_{11}, \quad v_{2t} = \beta_1 v_{1t} + \varepsilon_{1t}.$$

Неважко помітити, що функція пропозиції точно ідентифікована. Оцінки b_1 і b_0 параметрів β_1 і β_0 можуть бути визначені на основі оцінок коефіцієнтів таких рівнянь:

$$\beta_1 = \frac{\pi_{21}}{\pi_{11}}, \quad \beta_0 = \pi_{20} - \beta_1 \pi_{10} \Rightarrow \beta_1 = \frac{\bar{\pi}_{21}}{\bar{\pi}_{11}}, \quad b_0 = \bar{\pi}_{20} - b_1 \bar{\pi}_{10}.$$

За наявними статистичними даними оцінимо коефіцієнти зведених рівнянь:

$$\bar{\pi}_{11} = \frac{\overline{yp} - \bar{y} \cdot \bar{p}}{y^2 - \bar{y}^2} = \frac{0,2}{1,36} = 0,1471,$$

$$\bar{\pi}_{10} = \bar{p} - \bar{\pi}_{11}\bar{y} = 3 - 0,147 \cdot 3,2 = 2,5293,$$

$$\bar{\pi}_{21} = \frac{\overline{yq} - \bar{y} \cdot \bar{q}}{y^2 - \bar{y}^2} = \frac{0,96}{1,36} = 0,9411,$$

$$\bar{\pi}_{20} = \bar{q} - \bar{\pi}_{21}\bar{y} = 3,9411.$$

Оцінки коефіцієнтів функції пропозиції за МНК будуть такі:

$$b_1 = 0,7059 / 0,1471 = -4,7988, \quad b_0 = 3,9411 - 4,7988 \cdot 2,5293 = -8,1965.$$

Отже, функція пропозиції має вигляд

$$q_t = -8,1965 + 4,7988 p_t.$$

Водночас розраховані безпосередньо за МНК оцінки даного рівняння становитимуть:

$$b_1 = \frac{\overline{pq} - \bar{p} \cdot \bar{q}}{p^2 - \bar{p}^2} = \frac{-3,8}{2} = -1,9, \quad b_0 = \bar{q} - b_1\bar{p} = 11,9,$$

тобто функція пропозиції має вигляд

$$\bar{q}_t = 11,9 - 1,9 p_t.$$

За отриманими результатами можна зробити висновок про те, що застосування МНК у невідповідних ситуаціях може істотно спотворити картину залежності.

8.5.2. Метод інструментальних змінних

Ще одним способом усунення корелювання пояснюючої змінної з випадковим відхиленням є метод інструментальних змінних.

Сутність цього методу полягає в заміні змінної, що корелює із залишками, інструментальною змінною (ІЗ), яка повинна мати такі властивості:

- корелювати (бажано значною мірою) із заміненою пояснюючою змінною;
- не корелювати з випадковим відхиленням.

Опишемо схему використання ІЗ на прикладі парної регресії, у якій

$$Y = \beta_0 + \beta_1 X + \varepsilon.$$

Змінну X замінюють змінною Z такою, що $\text{cov}(Z; X) \neq 0$ і $\text{cov}(Z, \varepsilon) = 0$. Принципи використання ІЗ передбачають виконання таких умов:

$$M(\varepsilon_t) = 0, \text{cov}(z_t, \varepsilon_t).$$

Відповідні вибіркові оцінки даних умов такі:

$$\begin{cases} \frac{1}{T} \sum e_t = 0, \\ \frac{1}{T} \sum z_t e_t = 0. \end{cases}$$

У розгорненому вигляді остання система має вигляд

$$\begin{cases} \sum (y_t - b_0^{\text{із}} - b_1^{\text{із}} x_t) = 0, \\ \sum z_t (y_t - b_0^{\text{із}} - b_1^{\text{із}} x_t) = 0, \end{cases}$$

звідки

$$\begin{cases} b_1^{\text{із}} = \frac{\sum (z_t - \bar{z})(y_t - \bar{y})}{\sum (z_t - \bar{z})(x_t - \bar{x})}, \\ b_0^{\text{із}} = \bar{y} - b_1^{\text{із}} \bar{x}. \end{cases}$$

Нехай зі збільшенням обсягу вибірки $D(X)$ прямує до деякої скінченної межі σ_x^2 , а коваріація $\text{cov}(Z, X)$ — до скінченної межі $\sigma_{zx} \neq 0$.

Покажемо, що в цьому разі b_1 прямує до істинного значення β_1 .
З останньої системи маємо

$$\begin{aligned} b_1^{i3} &= \frac{\text{cov}(Z, Y)}{\text{cov}(Z, X)} = \frac{\text{cov}(Z, \beta_0 + \beta_1 X + \varepsilon)}{\text{cov}(Z, X)} = \\ &= \frac{\text{cov}(Z, \beta_0)}{\text{cov}(Z, X)} + \frac{\text{cov}(Z, \beta_1 X)}{\text{cov}(Z, X)} + \frac{\text{cov}(Z, \varepsilon)}{\text{cov}(Z, X)} = \\ &= \beta_1 + \frac{\text{cov}(Z, \varepsilon)}{\text{cov}(Z, X)} \xrightarrow{n \rightarrow \infty} \beta_1 + \frac{0}{\sigma_{z\varepsilon}} = \beta_1. \end{aligned}$$

Тут ми скористалися такими співвідношеннями: $\text{cov}(Z, \beta_0) = 0$, тому що $\beta_0 = \text{const}$; $\text{cov}(Z, \beta_1 X) = \beta_1 \text{cov}(Z, X)$. При великих обсягах вибірки розподіл b_1^{i3} прямує до нормального:

$$b_1^{i3} \sim N(\beta_1, S^2(b_1^{i3})),$$

де

$$S^2(b_1^{i3}) = \frac{S^2 \sum (z_t - \bar{z})}{[\sum (z_t - \bar{z})(x_t - \bar{x})]^2}; \quad S^2 = \frac{1}{T} \sum e_t^2;$$

$$e_t = y_t - b_0^{i3} - b_1^{i3} x_t.$$

8.5.3. Двокроковий метод найменших квадратів оцінювання параметрів надідентифікованих систем

Опис цього методу супроводимо прикладом його використання для моделі рівноваги на ринках товарів і грошей ($IS-LM$) для закритої економіки при фіксованій податковій ставці t :

$$\begin{cases} Y = \alpha_0 + \alpha_1 r + \alpha_2 G + \alpha_3 t + \varepsilon_1 & (\alpha_1 < 0), \\ Y = \beta_0 + \beta_1 r + \beta_2 M + \varepsilon_2 & (\beta_1 > 0). \end{cases} \quad (8.29)$$

Перше рівняння системи є перевизначеним (щодо змінної r). Щоб оцінити його коефіцієнти, рекомендується скористатися *двокроковим методом найменших квадратів (2МНК)*, суть якого полягає у використанні як ІЗ оцінки перевизначеної змінної, отриманої на базі екзогенних (чи заздалегідь визначених) змінних моделі.

Крок 1.

У першому рівнянні цієї системи перевизначеною змінною є процентна ставка r . Її можна оцінити, спираючись лише на екзогенні змінні (наприклад, віднімаючи від другого співвідношення перше):

$$r = \lambda_0 + \lambda_1 M + \lambda_2 G + \lambda_3 t + v. \quad (8.30)$$

(Як вправу пропонується знайти коефіцієнти $\lambda_0, \lambda_1, \lambda_2, \lambda_3$ і v .)

Застосовуючи для (8.30) МНК, отримуємо оцінку \hat{r} змінної r :

$$\hat{r} = \hat{\lambda}_0 + \hat{\lambda}_1 M + \hat{\lambda}_2 G + \hat{\lambda}_3 t, \quad (8.31)$$

де \hat{r} — умовна середня при фіксованих значеннях M, G, t .

Крок 2.

Підставляючи оцінку (8.31) у друге рівняння системи (8.29), маємо

$$Y = \beta_0 + \beta_1 \hat{r} + \beta_2 M + \varepsilon_2. \quad (8.32)$$

Ця заміна дає змогу розв'язати таку істотну проблему перевизначених моделей, як корельованість пояснюючої змінної з випадковим членом (нагадаємо, що така корельованість призводить до отримання зміщених і необгрунтованих оцінок). Дійсно, оцінка \hat{r} виражається лише через екзогенні змінні, а отже, не корелює з випадковим відхиленням. Фактично її можна розглядати як нову екзогенну змінну.

Замінивши в моделі (8.29) друге рівняння на (8.32), отримаємо систему, яку можна оцінити за допомогою МНК.

Якщо модель містить більш як одну перевизначену змінну, на першому етапі необхідно оцінити всі такі змінні.

2МНК має певні властивості, що зумовлюють його широке практичне застосування.

1. У даному методі перший етап (етап побудови зведених рівнянь) виконується для частини перевизначених рівнянь, не за-

чіпаючи інші рівняння моделі. Це дає змогу мінімізувати обсяг обчислень.

2. За наявності перевизначених рівнянь 2МНК на відміну від МНК визначає єдині оцінки параметрів моделі.

3. Застосовуючи даний метод, достатньо використовувати лише екзогенні й визначені змінні моделі.

4. Застосування 2МНК ефективно лише в тому разі, якщо коефіцієнт детермінації R^2 для зведених рівнянь, побудованих на першому етапі, буде досить великий. При цьому ІЗ (у нашому прикладі це змінна \hat{r}) незначною мірою корелює з випадковим відхиленням і наближається до істинного значення (\hat{r}) заміненої змінної. При невеликому значенні R^2 використання 2МНК малопродуктивне, тому що в цьому разі ІЗ мало відповідає істинному значенню заміненої змінної.

Зазначимо, що за допомогою методу інструментальних змінних як складової 2МНК можна отримувати обґрунтовані оцінки й оцінки стандартних відхилень для вибірок великих обсягів. Однак для малих вибірок висновки будуть не настільки конкретними.

8.5.4. Трикроковий метод найменших квадратів

Розглянуті методи дають змогу оцінювати параметри окремих рівнянь системи. Кожен з них має переваги та недоліки, однак їх об'єднує спільна риса — значний обсяг розрахунків при роботі із системами великої розмірності, тобто із такими, що містять велику кількість рівнянь, а отже, велику кількість змінних і параметрів. Скорочення обсягу розрахунків стає особливо актуальним у процесі вивчення швидкоплинних процесів, а також у тому разі, якщо змінюється пріоритетність окремих незалежних змінних моделі. У цих випадках краще скористатися методом, що одночасно оцінює параметри всіх рівнянь системи, зокрема трикроковим методом найменших квадратів (ЗМНК).

Особливістю ЗМНК є те, що при оцінюванні параметрів системи загалом слід зважати на залежності між окремими рівняннями. Ці залежності виявляються в тому, що залишки окремих рівнянь корелюють між собою, тобто загальна матриця коваріацій системи є недіагональною. У такій ситуації найкращим методом оцінювання є узагальнений метод найменших квадратів (метод Ейткена). Однак у цьому разі необхідно знати перше наближення матриці коваріацій. Для рівнянь множинної

регресії з автокорельованими залишками цю матрицю отримують на підставі залишків моделі, параметри якої оцінено за звичайним МНК, і вже після обчислення коефіцієнта кореляції коригують загальний оператор оцінювання параметрів рівняння.

Для систем рівнянь, особливо в разі надідентифікованості окремих рівнянь, краще початкове наближення матриці коваріацій визначають за залишками, які отримано в результаті оцінювання параметрів рівнянь за двокроковим МНК. Отже, саме поєднання 2МНК і методу Ейткена дало назву цьому методу.

Для практичного застосування 3МНК потрібно виконати такі вимоги [17]:

1) усі тотожності, які входять до системи рівнянь, виключають з розгляду, тому що вони не містять невідомих параметрів і не параметризуються;

2) кожне неідентифіковане рівняння також виключають із системи, оскільки оцінити їх параметри в принципі неможливо;

3) точно ідентифіковані та надідентифіковані рівняння поділяють на дві різні групи і 3МНК застосовують до кожної з них окремо;

4) якщо група надідентифікованих рівнянь складається лише з одного рівняння, то 3МНК перетворюється на 2МНК;

5) кореляція залишків окремих рівнянь системи призводить до того, що загальна матриця коваріацій системи є недіагональною, однак водночас не між усіма рівняннями системи існує залежність, тому матриця коваріації часто буває блочно-діагональною, тоді оцінювання параметрів на основі 3МНК виконують окремо для кожної групи рівнянь, що відповідають одному блоку.

8.5.5. МНК для рекурсивних моделей

Одним із випадків успішного застосування МНК для оцінювання структурних коефіцієнтів моделі є *рекурсивні (трикутні) моделі*, у яких ендогенні змінні послідовно (рекурсивно) пов'язані одна з одною. Перша ендогенна змінна Y_1 залежить лише від екзогенних змінних X_i , $i=1, 2, \dots, m$, і випадкового відхилення ϵ_1 . Друга ендогенна змінна Y_2 визначається лише значеннями екзогенних змінних X_i , $i=1, 2, \dots, m$, випадковим відхиленням ϵ_2 , а також ендогенною змінною Y_1 . Третя ендогенна змінна Y_3 залежить від тих самих змінних, що і Y_2 , випадкового відхилення ϵ_3 , а також від попередніх ендогенних змінних (Y_1 , Y_2) і т. д.

У цих моделях структурні рівняння оцінюються поетапно ($Y_1 \rightarrow Y_2 \rightarrow Y_3 \rightarrow \dots \rightarrow Y_N$). Застосовуючи МНК для таких моделей, можна отримати незміщені та обґрунтовані оцінки.

Однак моделі даного типу трапляються досить рідко. У загальному випадку для оцінки структурних коефіцієнтів спочатку необхідно перетворити вихідні рівняння до зведеної форми, а потім застосувати звичайний МНК.

8.6. Прогноз і загальні довірчі інтервали

Точковий прогноз залежних змінних визначається на підставі зведеної (прогнозої) форми економетричної моделі, заданої системою одночасних рівнянь.

Визначення довірчих інтервалів для цього прогнозу залежить від способу, яким було отримано зведену форму моделі.

У загальному випадку довірчі інтервали для кожної ендогенної змінної задаються співвідношенням

$$(\hat{Y}_i - \sqrt{t_{\alpha/2}^2 S_{ii}^2}, \hat{Y}_i + \sqrt{t_{\alpha/2}^2 S_{ii}^2}), \quad i = 1, 2, \dots, r,$$

де $t_{\alpha/2}^2 = F_{\alpha}$ – табличне значення критерію Фішера з довірчим рівнем α при $(r, n - k - r + 1)$ ступенях свободи, де r – кількість рівнянь системи, n – загальна кількість спостережень, k – кількість екзогенних змінних i -го рівняння системи; S_{ii}^2 – дисперсія залишків i -го рівняння моделі.

Довірчі інтервали для всіх ендогенних змінних визначають за формулами

$$\left(\hat{Y}_i - \sqrt{\frac{[1 + X_j^T (X^T X)^{-1} X_j](n - k)r}{n - k - r + 1} S_{ii}^2}, \right.$$

$$\left. \hat{Y}_i + \sqrt{\frac{[1 + X_j^T (X^T X)^{-1} X_j](n - k)r}{n - k - r + 1} S_{ii}^2} \right),$$

де X_j – матриця спостережень k екзогенних змінних, що ввійшли в i -те рівняння системи, розміром $n \times k$; X – загальна матриця спостережень розміром $n \times r$; r – загальна кількість рівнянь системи.

Контрольні запитання

1. Які основні причини використання систем одночасних рівнянь?
2. У чому полягає основна розбіжність між структурними рівняннями системи і рівняннями у зведеної формі?
3. Чому звичайний МНК практично не використовується для оцінювання систем одночасних рівнянь?
4. У чому полягає суть непрямого методу найменших квадратів?
5. Яка проблема часто виникає в процесі чисельного оцінювання параметрів структурних рівнянь за оцінками коефіцієнтів зведених рівнянь?
6. Назвіть причини неідентифікованості й надідентифікованості систем структурних рівнянь.
7. Наведіть необхідні й достатні умови ідентифікованості систем.
8. Які системи можна оцінювати звичайним МНК?
9. У чому полягає суть двокрокового методу найменших квадратів (2МНК)?
10. Які з наступних тверджень є істинними, помилковими чи невизначеними? Відповідь поясніть.
 - а. Звичайний МНК незастосовний для оцінювання коефіцієнтів структурних рівнянь систем одночасних рівнянь.
 - б. МНК рідко використовується для оцінювання коефіцієнтів структурних рівнянь систем одночасних рівнянь, тому що існують методи отримання більш якісних оцінок.
 - в. Екзогенні та визначені змінні моделі по суті одне й те саме.
 - г. Інструментальні змінні дають змогу розв'язати одну із серйозних проблем систем одночасних рівнянь — проблему корельованості пояснюючої змінної з випадковим відхиленням.
 - д. Проблема неідентифікованості передусім пов'язана з неможливістю отримання оцінок коефіцієнтів структурних рівнянь.
 - е. Не існує єдиного критерію для оцінювання загальної якості всієї системи одночасних рівнянь.
 - є. Якщо рівняння точно ідентифіковані, то оцінки, отримані за методом інструментальних змінних, і оцінки, отримані за 2МНК, будуть ідентичні.
 - ж. Оцінки, отримані за 2МНК, мають *бажані* властивості лише при великих вибірках.
 - з. Для точно ідентифікованих систем 2МНК не використовується.

11. Нехай макроекономічна модель закритої економіки має такий спрощений вигляд:

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t + \varepsilon_t, \\ i_t = \gamma_0 + \gamma_1 r_t + v_t, \\ y_t = c_t + i_t + g_t, \end{cases}$$

де c_t — обсяг споживання в році t ; y_t — ВВП у році t ; i_t — обсяг інвестицій у році t ; r_t — процентна ставка в році t ; g_t — обсяг державних витрат в році t .

а. Які із зазначених змінних даної моделі є екзогенними, а які — ендогенними?

б. Чи є модель точно ідентифікованою?

в. Як можна оцінити параметри моделі?

12. Поясніть на прикладі системи трьох рівнянь, що відкидання від кожного рівняння системи по одній змінній не може гарантувати ідентифікованості кожного з розглянутих рівнянь.

Вправи та завдання

1. Розглядається модель

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t + \varepsilon_t, \\ i_t = \gamma_0 + \gamma_1 y_t + \gamma_2 g_{t-1} + v_t, \\ y_t = c_t + i_t + g_t, \end{cases}$$

де c_t — обсяг споживання; i_t — обсяг інвестицій; y_t — прибуток; g_t — обсяг державних витрат.

а. Подайте дану систему у зведеній формі.

б. Що можна сказати відносно ідентифікованості функції споживання та функції інвестицій?

в. Що можна було б сказати щодо оцінки граничної схильності до споживання, якби вона була визначена звичайним МНК на основі рівняння $c_t = \beta_0 + \beta_1 y_t + \varepsilon_t$?

2. Розглядається модель

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t + \varepsilon_t, \\ i_t = \gamma_0 + \gamma_1 y_t + \gamma_2 y_{t-1} + v_t, \\ y_t = c_t + i_t + g_t, \end{cases}$$

де c_t – обсяг споживання; i_t – обсяг інвестицій; y_t – прибуток; g_t – обсяг державних витрат.

а. Подайте дану систему у зведений формі.

б. Визначте, які зі структурних рівнянь ідентифіковані.

в. Який метод можна використовувати для оцінювання параметрів розглянутої моделі?

3. Розглядається модель “попит – пропозиція”

$$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \alpha_3 p_{t-1} + \varepsilon_t, & \sigma(\varepsilon_i, \varepsilon_j) = 0 \\ q_t^S = \beta_0 + \beta_1 p_t + v_t, & \sigma(v_i, v_j) = 0 \\ q_t^D = q_t^S, & \text{при } i \neq j. \end{cases}$$

а. Чи будуть ідентифіковані рівняння даної системи?

б. Які оцінки параметрів можна отримати при використанні

МНК?

в. Як можна оцінити рівняння пропозиції за допомогою методу інструментальних змінних?

г. Як можна оцінити рівняння пропозиції за допомогою

2МНК?

д. Як пов’язані між собою оцінки, отримані в пунктах (в) і (г)?

е. Чи можна оцінити рівняння попиту на основі непрямого

МНК?

4. Розглядається модель “попит – пропозиція”

$$\begin{array}{l} \text{Попит} \\ \text{Пропозиція} \end{array} \quad \begin{cases} Q^D = \alpha_0 + \alpha_1 P + \varepsilon, \\ Q^S = \beta_0 + \beta_1 P + \beta_2 W + v, \\ Q^D = Q^S, \end{cases}$$

де Q – кількість товару; P – ціна товару; ε, v – випадкові відхилення, що задовольняють передумови МНК; W – заробітна плата.

Нехай є такі спостереження:

P	10	15	5	8	4
Q	6	6	18	12	8
W	2	6	2	7	4

а. Які зі змінних у даній моделі є екзогенними, а які – ендогенними?

- б. Подайте дану систему у зведеному вигляді.
 в. Визначте за МНК коефіцієнти зведених рівнянь (якщо можливо).

г. Чи збігаються знаки знайдених коефіцієнтів з передбачуваними теоретично?

д. На основі знайдених зведених коефіцієнтів за НМНК визначте структурні коефіцієнти для функції попиту.

е. Чи можна за НМНК оцінити структурні коефіцієнти для функції пропозиції? Якщо так, то як це зробити?

5. Нехай модель “прибуток – споживання” має такий вигляд:

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t + \varepsilon_t, \\ i_t = \gamma_0 + \gamma_1 r_t + v_t, \\ y_t = c_t + i_t + g_t, \end{cases}$$

де c_t – обсяг споживання в році t ; y_t – ВВП у році t ; i_t – обсяг інвестицій у році t ; r_t – процентна ставка в році t ; g_t – обсяг державних витрат у році t .

а. Які із зазначених змінних даної моделі є екзогенними, а які – ендогенними?

б. Поясніть, які знаки коефіцієнтів очікуються з погляду економічної теорії.

в. Наведіть формули розрахунку коефіцієнтів таких рівнянь:

$$\begin{cases} y_t = \pi_{10} + \pi_{11} r_t + \pi_{12} g_t + v_{1t}, \\ c_t = \pi_{20} + \pi_{21} r_t + \pi_{22} g_t + v_{2t}, \\ i_t = \pi_{30} + \pi_{31} r_t + \pi_{32} g_t + v_{3t}. \end{cases}$$

г. Які з параметрів структурних рівнянь ідентифіковані?

д. Визначте на основі НМНК параметри β_0 і β_1 .

6. Нехай модель “прибуток – споживання” має такий вигляд:

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t + \beta_1 c_{t-1} + \varepsilon_t, \\ i_t = \gamma_0 + \gamma_1 r_t + v_t, \\ y_t = c_t + i_t + g_t, \end{cases}$$

де c_t , c_{t-1} – обсяг споживання в роках відповідно t і $t-1$; y_t – ВВП у році t ; i_t – обсяг інвестицій у році t ; r_t – процентна ставка в році t ; g_t – обсяг державних витрат у році t .

а. Які із зазначених змінних даної моделі є екзогенними, ендогенними, а які — визначеними?

б. Поясніть, які знаки коефіцієнтів очікуються з погляду економічної теорії.

в. Визначте ідентифікованість структурних рівнянь на основі необхідних і достатніх умов ідентифікації.

г. Наведіть формули розрахунку коефіцієнтів відповідних зведених рівнянь:

$$\begin{cases} y_t = \pi_{10} + \pi_{11}r_t + \pi_{12}g_t + \pi_{13}c_{t-1} + v_{1t}, \\ c_t = \pi_{20} + \pi_{21}r_t + \pi_{22}g_t + \pi_{23}c_{t-1} + v_{2t}, \\ i_t = \pi_{30} + \pi_{31}r_t + \pi_{32}g_t + \pi_{33}c_{t-1} + v_{3t}. \end{cases}$$

д. Які з параметрів структурних рівнянь ідентифіковані?

е. Опишіть схему використання 2МНК для оцінювання параметрів структурних рівнянь.

7. Розглядається система одночасних рівнянь:

$$\begin{cases} q_t = \beta_0 + \beta_1 p_t + \beta_2 i_t + \varepsilon_t, \\ q_t = \gamma_1 p_t + v_t. \end{cases}$$

а. Які зі змінних у даній моделі є екзогенними, а які — ендогенними?

б. Нехай за статистичними даними отримано такі результати:

$$\sum q_t^2 = 110, \quad \sum p_t^2 = 50, \quad \sum i_t^2 = 100, \quad \sum q_t p_t = 100,$$

$$\sum q_t i_t = 90, \quad \sum p_t i_t = 100.$$

Знайдіть на основі МНК оцінку параметра γ_t .

в. Знайдіть оцінку цього самого параметра на основі НМНК і 2МНК.

г. Порівняйте знайдені оцінки. Яку б з них ви обрали й чому?

8. Розглядається система одночасних рівнянь

$$\begin{cases} y_{1t} = \beta_0 + \beta_1 y_{2t} + \beta_2 x_t + \varepsilon_t, \\ y_{2t} = \gamma_0 + \gamma_1 y_{1t} + v_t. \end{cases}$$

Нехай дана система у зведеному вигляді виражена такими співвідношеннями:

$$\begin{cases} y_{1t} = 2 + 5x_t, \\ y_{2t} = 1 + 10x_t. \end{cases}$$

- а. Оцініть ідентифіковані параметри структурних рівнянь.
 б. Оцініть ідентифіковані параметри структурних рівнянь у припущенні, що $\beta_1 = 0$.
 в. Оцініть ідентифіковані параметри структурних рівнянь у пропозиції, якщо $\beta_0 = 0$.

9. Наведено дані щодо ВВП (Y), споживання (C) та інвестицій (I) для умовної економіки за 20 років:

Y	95,75	98,55	103,55	109,00	108,25
C	60,45	62,45	65,90	68,90	68,45
I	14,30	15,85	17,75	19,70	18,10

Y	107,40	112,70	117,75	123,45	126,55
C	70,00	73,55	76,55	79,70	81,60
I	14,60	17,35	20,00	22,15	22,30

Y	125,85	128,10	125,35	130,25	138,30
C	81,55	82,55	83,45	87,35	91,55
I	19,80	21,00	18,00	20,00	25,25

Y	142,65	146,80	151,30	157,40	161,25
C	95,50	99,00	101,75	105,40	107,45
I	24,85	24,50	25,00	25,80	26,15

- а. У припущенні, що споживання залежить лінійно від прибутку за схемою найпростішої кейнсіанської моделі формування прибутків $c_t = \beta_0 + \beta_1 y_t + \varepsilon_t$ (див. с. 126), оцініть за МНК параметри β_0 і β_1 функції споживання.
 б. Оцініть ті самі параметри на основі НМНК.
 в. Порівняйте отримані результати. Зробіть висновки щодо якості оцінок.

10. Розглядається кейнсіанська модель

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t + \beta_2 T_t + \varepsilon_t, \\ i_t = \alpha_0 + \alpha_1 y_{t-1} + v_t, \\ T_t = \gamma_0 + \gamma_1 y_t + w_t, \\ y_t = c_t + i_t + g_t, \end{cases}$$

де y_t – величина прибутку; c_t – споживання; i_t – інвестиції; T_t – розмір податків; ε, v, w – випадкові члени.

а. Які змінні в даній моделі є ендогенними, які – екзогенними, а які – визначеними?

б. На підставі необхідних і достатніх умов ідентифікованості визначте, які з рівнянь ідентифіковані.

в. Чи буде ідентифікована система загалом?

г. Що зміниться, якщо до функції інвестицій додати екзогенну змінну r_t – процентну ставку в році t ?

11. Розглядається модель

$$\begin{cases} r_t = \beta_0 + \beta_1 y_t + \beta_2 m_t + \varepsilon_t, \\ y_t = \alpha_0 + \alpha_1 r_t + v_t, \end{cases}$$

де r_t – процентна ставка в році t ; y_t – ВВП у році t ; m_t – грошова маса в році t .

а. Чи можна ідентифікувати рівняння розглянутої моделі?

б. Який метод визначення оцінок параметрів доцільний для розглянутої моделі?

в. На підставі наведених далі статистичних даних оцініть параметри ідентифікованих рівнянь. Чи збігаються знаки знайдених оцінок з передбачуваними теоретично?

r_t	6,55	4,55	4,45	7,00	7,50
y_t	95,75	98,5	103,55	109,00	108,25
m_t	58,30	60,00	60,55	64,50	65,00

r_t	8,75	9,70	10,00	11,50	7,75
y_t	107,40	112,70	117,75	23,45	126,55
m_t	63,45	67,60	70,50	74,00	76,50

r_t	6,00	6,10	5,90	9,80	8,00
y_t	125,85	128,10	125,35	130,25	138,30
m_t	75,00	77,25	74,00	78,45	85,50

r_t	7,50	7,00	6,50	7,40	5,50
y_t	142,65	146,80	151,30	157,40	161,25
m_t	87,00	88,00	90,50	94,40	96,50

12. Розглядається макроекономічна модель

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t + \beta_2 r_t + \varepsilon_{1t}, \\ r_t = \alpha_0 + \alpha_1 i_t + \alpha_2 m_t + \varepsilon_{2t}, \\ i_t = \gamma_0 + \gamma_1 r_t + \gamma_2 (y_t - y_{t-1}) + \varepsilon_{3t}, \\ y_t = c_t + i_t + g_t, \end{cases}$$

де c_t — обсяг споживання в році t ; y_t — ВВП у році t ; r_t — процентна ставка в році t ; i_t — обсяг інвестицій у році t ; m_t — грошова маса M_t в році t ; g_t — обсяг державних витрат у році t .

а. Які з розглянутих рівнянь ідентифіковані?

б. Яким методом можуть бути оцінені параметри ідентифікованих рівнянь?

13. Розглядається модель, що складається з двох рівнянь:

$$\begin{cases} y_t = \alpha_0 + \alpha_1 x_t + \varepsilon_{1t}, \\ z_t = \beta_0 + \beta_1 y_t + \varepsilon_{2t}, \end{cases} \quad \sigma(\varepsilon_1, \varepsilon_2) = 0.$$

а. Яким методом можна оцінити розглянуту систему? Чи може бути для цього використаний звичайний МНК?

б. Чи будуть оцінки, отримані за МНК, збігатися з оцінками 2МНК?

в. Як можна оцінити дану систему, якщо в її друге рівняння як пояснюючу змінну буде введено змінну X ?

14. Розглядається модель попиту та пропозиції для грошей:

$$\begin{cases} m_t^D = \alpha_0 + \alpha_1 y_t + \alpha_2 r_{t-1} + \alpha_3 p_{t-1} + \varepsilon_{1t}, \\ m_t^S = \beta_0 + \beta_1 y_t + \varepsilon_{2t}, \end{cases}$$

де m_t^D – обсяг попиту на гроші; m_t^S – обсяг пропозиції грошей у році t ; y_t – прибуток; r_{t-1} – процентна ставка; p_{t-1} – індекс цін у році $(t-1)$.

- а. Чи будуть ідентифіковані обидві функції?
- б. Яким методом можуть бути знайдені оцінки ідентифікованих параметрів?
- в. Що відбудеться з ідентифікованою системою, якщо до функції пропозиції додати як пояснюючі змінні y_{t-1} і m_{t-1} ?
- г. Який метод визначення оцінок доцільний при виконанні попереднього пункту?

МАУП

Розділ 9. МЕТОДИ ДОСЛІДЖЕННЯ ЯКІСНИХ ЕКОНОМІЧНИХ ПОКАЗНИКІВ

9.1. Якісні економічні показники

Звичайно незалежні змінні в регресійних моделях мають “неперервні” області змінювання (національний дохід, обсяги виробництва, розмір заробітної плати тощо), тобто є метрично (кількісно) вимірними величинами. У реальних ситуаціях економічні явища більш різноманітні. На залежну змінну крім кількісних факторів впливають і якісні: якість продукції, рівень професійної підготовки працівників, їхня стать, страйки, зміни в економічній політиці тощо. Часто змінні, що відображають якісні характеристики об’єкта, набувають лише двох значень: 1 — якщо певна ознака присутня; 0 — якщо вона відсутня. Такі змінні називають бінарними, дихотомними або *dummy*-змінними.

У перекладі з англійської мови *dummy variables* означає “фіктивні змінні”, хоча насправді їх “фіктивність” полягає лише в тому, що вони кількісно описують деяку якісну ознаку.

Дихотомні змінні використовують у регресійних моделях поряд з кількісними змінними або утворюють регресійні моделі, у яких всі фактори є якісними (бінарними) змінними.

Поєднання в моделі кількісних та якісних факторів значно розширює можливості регресійного аналізу, а отже, можливості прогнозування та підготовки прийняття рішень.

Наприклад, при дослідженні заробітної плати може виникнути питання залежності її від рівня освіти, від статі працівника тощо. За певними якісними ознаками, звичайно, дані можна поділити за категоріями й вивчати кожну залежність окремо, а вже потім шукати відмінності між ними. Але введення додаткової бінарної змінної дає

зможу оцінювати одне рівняння, у якому різні класи спостережень розділяються за допомогою цієї змінної.

Приклад 9.1. Нехай регресійна модель залежності заробітної плати y від деяких кількісних факторів x_1, x_2, \dots, x_m має вигляд

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m + u.$$

Тоді для вивчення впливу вищої освіти на рівень оплати праці вводять нову змінну d , яка може набувати двох значень: $d = 1$, якщо робітник має вищу освіту, та $d = 0$, якщо не має. Модель, що враховує цей фактор, матиме вигляд

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m + \delta d + u,$$

тобто за наявності вищої освіти заробітна плата в середньому становить

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m + \delta,$$

а за її відсутності

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m.$$

У такому разі коефіцієнт δ має відображати зміни в зарплаті при переході робітників з однієї категорії (без вищої освіти) в іншу (з вищою освітою).

Параметри такої моделі оцінюються за допомогою методу найменших квадратів, а значущість параметра δ , встановлена в процесі перевірки нульової гіпотези $H_0 : \delta = 0$, означає наявність суттєвих відмінностей у заробітній платі робітників двох зазначених категорій.

Якщо якісна ознака має не два, а більше значень, то використовують кілька бінарних змінних. Причому їх кількість на одиницю менша, ніж кількість розглянутих категорій. Це пов'язано з тим, що сума бінарних змінних, які відповідають різним категоріям, завжди дорівнюватиме одиниці для всіх спостережень (тобто кожне спостереження, напевно, потрапляє до якоїсь однієї категорії). А таке співвідношення означає наявність мультиколінеарності між незалежними змінними і унеможливорює оцінювання параметрів моделі за методом найменших квадратів.

9.2. Регресійні моделі з бінарними незалежними змінними

Однією із сфер застосування бінарних змінних є аналіз сезонних коливань. За допомогою цих змінних можна усунути сезонні коливання з метою визначення головних тенденцій розвитку певного економічного процесу.

Приклад 9.2. Нехай y – обсяг споживання певного продукту, який залежить від пори року. Для виявлення сезонності можна ввести бінарні змінні d_1, d_2, d_3 :

$d_1 = 1$, якщо місяць року зимовий, $d_1 = 0$ – в інших випадках;

$d_2 = 1$, якщо місяць року весняний, $d_2 = 0$ – в інших випадках;

$d_3 = 1$, якщо місяць року літній, $d_3 = 0$ – в інших випадках.

На базі відповідних статистичних даних методом найменших квадратів можна оцінити параметри a_0, a_1, a_2, a_3 лінійного регресійного рівняння

$$y = a_0 + a_1 d_1 + a_2 d_2 + a_3 d_3 + u.$$

Отримані результати мають такий зміст: коефіцієнт a_0 визначає середньомісячний обсяг споживання досліджуваного продукту; суми коефіцієнтів $a_0 + a_1, a_0 + a_2, a_0 + a_3$ – обсяг споживання відповідно взимку, навесні та влітку. Отже, параметри a_1, a_2, a_3 вказують на сезонні відхилення в обсягах споживання продукту відносно осінніх місяців. Перевірка статистичної значущості кожного з коефіцієнтів регресії виконується за допомогою традиційного t -тесту. Прийняття гіпотези про рівність нулю кожного з параметрів означає несуттєву різницю між споживанням в осінній період і споживанням в інший сезон. Комплексна гіпотеза $a_1 = a_2 = a_3 = 0$ перевіряється за допомогою F -тесту. Зокрема, якщо приймається припущення $a_1 = a_2$, то це означає, що споживання взимку та весною не відрізняються між собою і т. ін.

Приклад 9.3. Розглянемо ще один приклад застосування фіктивних змінних.

Нехай y – середньомісячний обсяг споживання деякого індивіда, а I – його середньомісячний дохід. У лінійній регресійній моделі залежності споживання від доходу

$$y = a_0 + a_1I + u$$

коефіцієнт a_1 називається “схильністю до споживання”. Щоб визначити вплив сезону на схильність до споживання, як і в попередньому прикладі, застосовують бінарні змінні d_1, d_2, d_3 , а модель при цьому набирає вигляду

$$y = a_0 + a_1d_1 + a_2d_2 + a_3d_3 + a_4d_1I + a_5d_2I + a_6d_3I + a_7I + u.$$

Коефіцієнти цієї моделі $a_7, a_4 + a_7, a_5 + a_7, a_6 + a_7$ визначають схильність до споживання відповідно восени, зимою, весною та влітку. Як і в попередній моделі перевіряються гіпотези про відсутність сезонних впливів на схильність до споживання.

Крім того, бінарні змінні використовують також при дослідженні моделей, які описують структурні зміни в економіці. Розглянемо такий приклад.

Приклад 9.4. Нехай досліджується залежність обсягу випущеної підприємством продукції y від обсягу його основного фонду x . Припускається, що після досягнення основним фондом підприємства розміру \bar{x} відбувається певна структурна перебудова підприємства. Залежність випуску продукції від основного фонду в результаті перебудови змінюється, але загалом залишається неперервною. У такому разі функція залежності матиме кусково-лінійний графік, який відображає така регресійна модель:

$$y = a_0 + a_1x + a_2(x - \bar{x})d + u,$$

де бінарна змінна $d = 0$, якщо $x \leq \bar{x}$, і $d = 1$, якщо $x > \bar{x}$.

Якщо в результаті тестування значущості параметрів моделі приймається нульова гіпотеза $H_0 : a_2 = 0$, то це означає, що структурна зміна на підприємстві не відбулася.

Зауважимо, що спосіб уведення в модель бінарних змінних залежить від апріорної інформації щодо впливу якісних факторів на залежну змінну і від гіпотез, які необхідно перевірити на підставі цієї інформації. У свою чергу цей самий спосіб визначає, як будуть інтерпретовані отримані оцінки параметрів моделі.

9.3. Регресійні моделі з бінарними залежними змінними

Бінарними (дихотомними) можуть бути не лише незалежні, а й залежні змінні. Такі дані отримують, як правило, під час опитування населення, перепису тощо. Дані опитувань зазвичай якісні, тобто відтворюють певний якісний стан досліджуваного об'єкта. Залежна змінна при цьому набуває двох значень: $y_i = 1$, якщо i -й елемент об'єкта переходить у певний стан чи має певну властивість (ознаку), $y_i = 0$ – в інших випадках. Наприклад, $y_i = 1$, якщо покупець (i -й респондент) купив певний товар, $y_i = 0$, якщо не купив; безробітний знайшов ($y_i = 1$) чи не знайшов ($y_i = 0$) робоче місце; сім'я купила ($y_i = 1$) чи не купила ($y_i = 0$) власну квартиру і т. ін. Фактори, що впливають на той чи інший стан об'єкта, звичайно можуть бути кількісними. Змінювання залежної змінної в цьому разі можна інтерпретувати як імовірність певної події. Наприклад, купівля деякого товару залежить від рівня доходу певної особи чи сім'ї, але якщо особа чи сім'я цей товар має, то навряд чи найближчим часом буде здійснено ще таку саму покупку.

Діаграма розсіювання залежності цих двох показників така: незалежна змінна (дохід) набуває певних значень на числовій осі x , а дані спостереження залежної змінної y розміщені лише на двох паралельних прямих $y = 0$ і $y = 1$. Застосування класичної регресійної залежності в таких випадках не дає бажаних результатів: на кінцях проміжку спостережень регресійна пряма значно відхиляється від точок спостереження. Зокрема, на початковому етапі вона набуватиме від'ємних значень, а на кінцевому – значень, більших від одиниці (рис. 9.1). Якщо залежна змінна інтерпретується як імовірність купівлі, такі результати взагалі абсурдні. У таких випадках доцільніше припустити, що залежність між розглянутими показниками нелінійна. Дійсно, для сімей (осіб) з низьким рівнем доходу приріст Δx мало змінює ймовірність додаткових витрат, а при значному підвищенні рівня доходу той самий приріст Δx значно збільшує ймовірність нових придбань. Якщо сім'я вже має досить високий рівень доходу і забезпечила себе необхідними товарами, марно сподіватися на нові покупки.

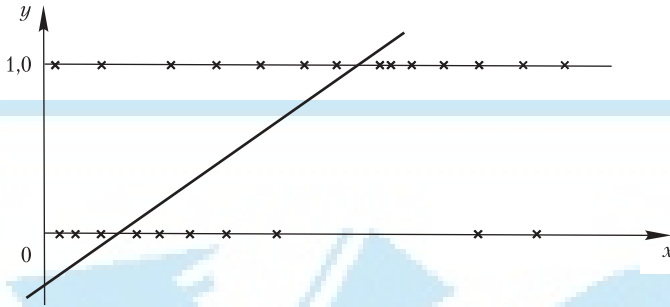


Рис. 9.1

Логічно припустити, що регресійна функція, як і функція розподілу випадкової величини, має S-подібну траєкторію розвитку (рис. 9.2). Практикою перевірено, що функції розподілу доходів можуть бути підпорядковані нормальному чи логістичному закону розподілу.

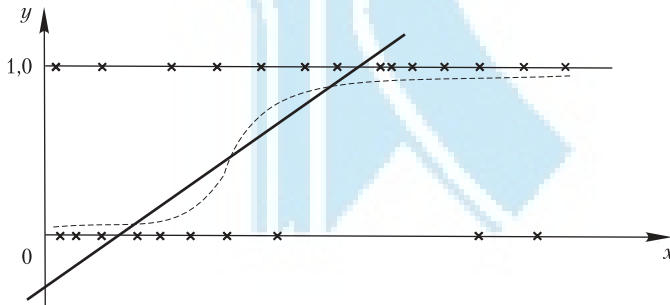


Рис. 9.2

Означення 9.1. Регресійна модель з бінарною (дихотомною) залежною змінною, що має нормальний розподіл, називається *пробіт-моделлю*.

Означення 9.2. Регресійна модель, у якій залежна змінна підпорядкована логістичному закону розподілу, називається *логіт-моделлю*.

Вивчення взаємозв'язку регресії з бінарною залежною змінною дає підставу для вибору доцільної форми регресійного співвідношення,

відмінної від звичайної лінійної регресії, чим розширює можливості моделювання та прогнозування специфічних залежностей між економічними показниками (кількісними та якісними).

Прогнози ймовірностей за перетвореними моделями регресії (зокрема, за логіт- і пробіт-моделями) застосовуються в багатьох галузях людської діяльності, в економічних і соціальних дослідженнях. Аналогічні підходи можуть застосовуватись і для інших якісних змінних та узагальнених моделей регресії.

Контрольні запитання

1. Які особливості економічного явища можна дослідити, використовуючи бінарні змінні?
2. Наведіть приклади використання бінарних незалежних змінних.
3. Наведіть приклади використання бінарних залежних змінних.
4. Як називаються моделі, що містять бінарні залежні змінні?
5. Якими є відмінності між моделями, що містять бінарні залежні змінні?

МАУП

ТЕСТОВІ ЗАВДАННЯ З ЕКОНОМЕТРІЇ

Варіант 1

1. Дайте визначення економетрії:
 - 1) наука, що вивчає вимірність зв'язків у відповідному економічному аналізі;
 - 2) наука, що застосовує математичні та математико-статистичні методи в економіці;
 - 3) наука, що вивчає методи оцінювання параметрів моделей, які характеризують кількісні взаємозв'язки між економічними показниками.
2. Структуру економетричної моделі визначають:
 - 1) незалежні змінні;
 - 2) залежні змінні;
 - 3) параметри.
3. Виробнича функція — це:
 - 1) діяльність деякого підприємства, спрямована на виробництво певного виду продукції;
 - 2) система взаємозв'язків, які встановлюються між виробничими одиницями (підрозділами підприємства) у процесі їх функціонування;
 - 3) залежність між обсягом виробленої продукції та спожитими для цього певними ресурсами.
4. Економетрична модель є:
 - 1) стохастичною;
 - 2) детермінованою;
 - 3) структурною.
5. Показник, що характеризує величину розкиду випадкової складової рівняння регресії, називається:
 - 1) коефіцієнтом кореляції;
 - 2) стандартною похибкою параметра;
 - 3) стандартною похибкою рівняння.

6. Коефіцієнт детермінації визначається за формулою:

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

$$3) R^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2}.$$

7. Критерій Фішера застосовується для перевірки значущості:
- 1) оцінок параметрів моделі;
 - 2) економетричної моделі;
 - 3) коефіцієнта множинної кореляції.
8. Якщо існує взаємозалежність послідовних членів часового чи просторового ряду, то маємо явище:
- 1) гетероскедастичності;
 - 2) автокореляції;
 - 3) мультиколінеарності.
9. Вимірювання зв'язку між економічними показниками з урахуванням часових зсувів виконується на основі:
- 1) динамічних моделей;
 - 2) моделей розподіленого лага;
 - 3) систем структурних рівнянь.
10. Для оцінювання параметрів систем одночасних рівнянь застосовується:
- 1) зважений метод найменших квадратів;
 - 2) узагальнений метод найменших квадратів;
 - 3) непрямий метод найменших квадратів.

Варіант 2

1. Зазначте найсуттєвішу задачу економетричного дослідження:
- 1) побудова економетричних моделей;
 - 2) оцінка та перевірка економетричних моделей;
 - 3) прогнозування економічних процесів на основі економетричних моделей.

2. При побудові економетричної моделі необхідно:
1) розглянути всі без винятку елементи, що впливають на результат процесу;

2) виключати ті елементи, що видаються нетиповими стосовно даної проблеми;

3) використовувати всю можливу інформацію, що стосується даного дослідження.

3. Для оцінювання параметрів економетричної моделі застосовують:

1) закон нормального розподілу Гаусса;

2) критерій Стюдента;

3) метод найменших квадратів.

4. Показник, що визначає міру зв'язку залежної змінної з усіма незалежними змінними, називається:

1) коефіцієнтом кореляції;

2) стандартною похибкою рівняння;

3) коефіцієнтом детермінації.

5. Коефіцієнт детермінації обчислюється за формулою:

$$1) R^2 = 1 - \frac{\sigma_u^2}{\sigma_y^2}; \quad 2) R^2 = \frac{\sigma_u^2}{\sigma_y^2}; \quad 3) R^2 = 1 - \frac{\sigma_y^2}{\sigma_u^2}.$$

6. Критерій Фішера застосовують для перевірки значущості:

1) оцінок параметрів моделі;

2) економетричної моделі;

3) коефіцієнта кореляції.

7. Якщо дисперсія залишків змінюється для кожного спостереження чи групи спостережень, то маємо явище:

1) автокореляції;

2) гетероскедастичності;

3) мультиколінеарності.

8. У разі мультиколінеарності при визначенні залежності між парами незалежних змінних застосовується:

1) F -критерій Фішера;

2) t -критерій Стюдента;

3) критерій χ^2 .

9. Для обґрунтування величини лага в дистрибутивно-лагових моделях застосовують:

1) кореляційну матрицю;

- 2) коваріаційну матрицю;
 - 3) взаємну кореляційну функцію.
10. Системи одночасних рівнянь можуть містити:

- 1) регресійні та функціональні рівняння;
- 2) регресійні рівняння та тотожності;
- 3) функціональні рівняння та тотожності.

Варіант 3

1. Регресійні рівняння описують:
 - 1) структурний зв'язок між показниками економічних процесів;
 - 2) функціональний зв'язок між суб'єктами економічної діяльності;
 - 3) кореляційний зв'язок між економічними показниками.
2. Екзогенні змінні:
 - 1) визначаються як розв'язок системи рівнянь;
 - 2) залишаються незмінними протягом усього періоду спостережень;
 - 3) задаються за межами економетричної моделі.
3. Однією з передумов застосування методу найменших квадратів є:
 - 1) математичне сподівання залишків моделі є сталою величиною;
 - 2) сума залишків моделі відмінна від нуля;
 - 3) математичне сподівання залишків моделі дорівнює нулю.
4. Показник, що визначає, яка частина руху залежної змінної описується даним регресійним рівнянням, називається:
 - 1) коефіцієнтом детермінації;
 - 2) коефіцієнтом кореляції;
 - 3) оцінкою узгодженості (конкордації).
5. Коефіцієнт детермінації обчислюється за формулою:

$$1) R^2 = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \text{var}(y)}};$$

$$2) R^2 = \frac{\text{cov}^2(x, y)}{\sqrt{\text{var}(x) \text{var}(y)}};$$

$$3) R^2 = \frac{\text{cov}(x, y)}{\text{var}(x) \text{var}(y)}.$$

6. Якщо дисперсія залишків стала для кожного спостереження, то маємо явище:

- 1) автокореляції;
- 2) мультиколінеарності;
- 3) гомоскедастичності.

7. Наявність мультиколінеарності перевіряється за допомогою:

- 1) μ -критерію;
- 2) алгоритму Фаррара – Глобера;
- 3) методу Дарбіна.

8. У разі мультиколінеарності для виявлення залежної змінної від сукупності інших незалежних змінних застосовують:

- 1) F -критерій Фішера;
- 2) t -критерій Стьюдента;
- 3) критерій Пірсона.

9. Моделі, у яких на залежні змінні впливають значення незалежних змінних у попередні періоди, називаються:

- 1) динамічними;
- 2) дистрибутивно-лаговими;
- 3) структурними.

10. Для оцінювання параметрів систем одночасних рівнянь застосовують:

- 1) узагальнений метод найменших квадратів;
- 2) двокроковий метод найменших квадратів;
- 3) метод головних компонентів.

Варіант 4

1. Найпоширенішими функціями в економетричному моделюванні є:

- 1) показникові;
- 2) лінійні;
- 3) логарифмічні.

2. Економетрична модель, яка кількісно описує зв'язок основних результативних показників виробничо-господарської діяльності з факторами, що визначають ці показники, називається:

- 1) показниковою функцією;
- 2) виробничою функцією;
- 3) моделлю розподіленого лага.

3. Оцінка параметра називається ефективною, якщо:

- 1) задовольняє закон великих чисел;
- 2) має найменшу дисперсію;
- 3) математичне сподівання її дорівнює значенню параметра.

4. Дисперсійно-коваріаційна матриця визначається на підставі:
- 1) матриці спостережень незалежних змінних;
 - 2) матриці нормалізованих змінних моделі;
 - 3) системи нормальних рівнянь.
5. Зв'язок між економічними характеристиками випуску продукції та спожитими для цього ресурсами визначається економіко-математичним співвідношенням:

$$1) Y = \alpha(1+r)^x; \quad 2) Y = \alpha F^\alpha L^\beta; \quad 3) Y = e^{\alpha\beta x + \gamma}.$$

6. Довірчі інтервали функції регресії визначаються за допомогою:
- 1) t -критерію Стьюдента та залишкової дисперсії;
 - 2) стандартної похибки рівняння;
 - 3) стандартної похибки параметрів.
7. Критерій Дарбіна – Уотсона застосовується для виявлення:
- 1) автокореляції;
 - 2) гомоскедастичності;
 - 3) мультиколінеарності.
8. За наявності гетероскедастичності параметри моделі оцінюються за:
- 1) умовним методом найменших квадратів;
 - 2) узагальненим методом найменших квадратів;
 - 3) двокроковим методом найменших квадратів.
9. Моделі, що містять лагові значення залежної змінної, називаються:
- 1) моделями розподіленого лага;
 - 2) авторегресійними моделями;
 - 3) моделями сезонних коливань.
10. Система одночасних рівнянь називається ідентифікованою, якщо:
- 1) параметри структурної форми моделі однозначно визначаються через параметри зведеної форми;
 - 2) кількість змінних, виключених з кожного рівняння системи, дорівнює кількості рівнянь моделі;
 - 3) параметри зведеної форми моделі визначаються через параметри структурної форми.

Варіант 5

1. Залежність між величинами x та y називається статистичною, якщо:

- 1) кожному значенню x відповідає лише одне значення y , яке обчислюється за відомою формулою;
- 2) змінювання однієї величини зумовлює змінювання розподілу іншої;
- 3) змінювання величини x зумовлює змінювання y за заданим законом.

2. Кореляційна матриця:

- 1) є матрицею парних коефіцієнтів кореляції;
- 2) характеризує щільність зв'язку всіх незалежних змінних із залежною змінною;
- 3) описує кореляційні зв'язки між незалежними змінними моделі.

3. Оцінка параметра називається обґрунтованою, якщо:

- 1) задовольняє закон великих чисел;
- 2) отримана за методом найменших квадратів;
- 3) має найменшу дисперсію.

4. Точковий прогноз – це:

- 1) побудова регресійної залежності за заданими точками;
- 2) значення залежної змінної, обчислене за моделлю при заданому значенні пояснюючих змінних;
- 3) визначення крайніх точок довірчого інтервалу для прогнозного значення залежної змінної.

5. Статистична значущість параметрів моделі визначається за допомогою:

- 1) стандартної похибки рівняння;
- 2) t -критерію Стьюдента;
- 3) F -критерію Фішера.

6. Якщо виникає явище гетероскедастичності, то оцінки параметрів моделі, отримані за 1МНК, будуть:

- 1) необґрунтованими;
- 2) зміщеними;
- 3) неефективними.

7. За наявності автокореляції для оцінювання параметрів застосовується:

- 1) непрямий метод найменших квадратів;

- 2) метод Фаррара – Глобера;
 - 3) узагальнений метод найменших квадратів.
8. Для виявлення незалежної змінної, що залежить від усіх інших незалежних змінних, у разі мультиколінеарності застосовується:

- 1) F -критерій;
- 2) t -критерій;
- 3) критерій χ^2 .

9. Умова ідентифікованості рівняння структурної форми має вигляд:

- 1) $k_s - 1 < m - m_s$;
- 2) $k_s - 1 = m - m_s$;
- 3) $k_s - 1 \leq m - m_s$,

де k_s , m_s – відповідно кількість залежних і незалежних змінних, що входять до s -го рівняння структурної форми, m – загальна кількість екзогенних змінних моделі.

10. Для оцінювання параметрів рекурсивних систем рівнянь застосовується:

- 1) метод найменших квадратів;
- 2) двокроковий метод найменших квадратів;
- 3) метод головних компонентів.

Варіант 6

1. Економетрична модель – це:
 - 1) рівняння чи система рівнянь, що описують строгі функціональні залежності між економічними показниками;
 - 2) функція чи система функцій, що описує кореляційно-регресійний зв'язок між економічними показниками;
 - 3) система рівнянь і тотожностей, що описує існуючі зв'язки між показниками економічних процесів.

2. Застосування методу найменших квадратів можливе, якщо незалежні змінні:

- 1) містять стохастичну складову;
- 2) не пов'язані із залишками;
- 3) мають сталу дисперсію.

3. Оцінка параметра називається незміщеною, якщо:
 - 1) задовольняє закон великих чисел;
 - 2) математичне сподівання її не залежить від вибірки і близьке до значення параметра;
 - 3) має найменшу дисперсію.

4. Дисперсія прогнозу для середнього значення залежної змінної обчислюється за формулою:

$$1) \sigma^2 = \sigma_u^2 X_0' (X'X)^{-1} X_0;$$

$$2) \sigma^2 = \sigma_u \sqrt{X_0' (X'X)^{-1} X_0};$$

$$3) \sigma^2 = \sqrt{\sigma_u^2 (1 + X_0' (X'X)^{-1} X_0)}.$$

5. Явище мультиколінеарності виникає, якщо:

- 1) існує лінійний зв'язок між незалежними змінними;
- 2) дисперсія залишків стала для кожного спостереження;
- 3) існує лінійна залежність між послідовними членами ряду залишків.

6. Критерій фон Неймана застосовується для виявлення:

- 1) автокореляції;
- 2) гетероскедастичності;
- 3) мультиколінеарності.

7. Кореляційна матриця обчислюється на підставі:

- 1) матриці спостережень незалежних змінних;
- 2) матриці нормалізованих змінних;
- 3) системи нормальних рівнянь.

8. За наявності гетероскедастичності для оцінювання параметрів застосовують метод:

- 1) Дарбіна;
- 2) Ейткена;
- 3) Фаррара — Глобера.

9. Побудова моделей розподіленого лага ускладнена через наявність:

- 1) автокореляції;
- 2) мультиколінеарності;
- 3) гетероскедастичності.

10. Система рівнянь, розв'язана відносно ендогенних змінних, називається:

- 1) структурною формою моделі;
- 2) нормальною системою рівнянь;
- 3) зведеною формою моделі.

Варіант 7

1. Виробнича функція – це:
 - 1) діяльність деякого підприємства, спрямована на виробництво певного виду продукції;
 - 2) система взаємозв'язків, що встановлюються між виробничими одиницями (підрозділами підприємства) у процесі їх функціонування;
 - 3) залежність між обсягом виробленої продукції та спожитими для цього певними ресурсами.
2. Незалежні змінні моделі:
 - 1) визначаються як розв'язок рівняння чи системи рівнянь;
 - 2) задаються за межами економетричної моделі;
 - 3) залишаються незмінними протягом усього періоду спостереження.
3. Застосування методу найменших квадратів можливе, якщо незалежні змінні моделі утворюють:
 - 1) систему нормальних рівнянь;
 - 2) лінійно незалежну систему векторів;
 - 3) однорідну систему рівнянь.
4. Коефіцієнт детермінації:
 - 1) характеризує абсолютну величину розкиду випадкової складової рівняння;
 - 2) показує, яка частина руху залежної змінної описується даним регресійним рівнянням;
 - 3) визначає міру зв'язку залежної змінної з усіма незалежними факторами.
5. При визначенні загальної мультиколінеарності масиву незалежних змінних застосовують:
 - 1) F -критерій Фішера;
 - 2) t -критерій Стьюдента;
 - 3) критерій Пірсона χ^2 .
6. Явище гетероскедастичності виникає, якщо:
 - 1) існує взаємозалежність послідовних членів часового ряду;
 - 2) дисперсія залишків змінюється для кожного спостереження чи групи спостережень;
 - 3) існує лінійний зв'язок між незалежними змінними моделі.

7. Стандартна похибка рівняння обчислюється за формулою:

$$1) S_u^2 = \frac{1}{n} \sum_{i=1}^n u_i^2; \quad 2) S_u^2 = \frac{1}{n-m-1} \sum_{i=1}^n u_i^2; \quad 3) S_u = \sqrt{R^2}.$$

8. Критерій Глейсера застосовується для виявлення:

- 1) автокореляції;
- 2) гетероскедастичності;
- 3) мультиколінеарності.

9. Якщо економетрична модель крім лагових змінних містить змінні, що характеризують поточні умови функціонування економічної системи, то маємо:

- 1) модель адаптивних сподівань;
- 2) узагальнену модель розподіленого лага;
- 3) модель часткового коригування.

10. Якщо рівняння структурної форми моделі надіентифіковані, то для оцінювання параметрів рівнянь застосовують:

- 1) непрямий МНК;
- 2) двокроковий МНК;
- 3) трикроковий МНК.

Варіант 8

1. Залежні змінні моделі:

- 1) визначаються як розв'язок рівняння чи системи рівнянь;
- 2) змінюють свої значення залежно від періоду спостереження;
- 3) задаються за межами економетричної моделі.

2. Застосування методу найменших квадратів можливе, якщо виконуються передумови про відсутність:

- 1) автокореляції змінних;
- 2) автокореляції залишків;
- 3) гомоскедастичності.

3. Стандартна похибка рівняння:

- 1) показує, яка частина руху залежної змінної описується даним регресійним рівнянням;
- 2) характеризує величину розкиду випадкової складової рівняння;
- 3) визначає міру зв'язку залежної змінної з усіма незалежними факторами.

4. При визначенні залежності між парами незалежних змінних у разі мультиколінеарності застосовують:

- 1) F -критерій Фішера;
- 2) t -критерій Стьюдента;
- 3) критерій Пірсона χ^2 .

5. Явище автокореляції виникає, якщо:

- 1) дисперсія залишків змінюється для кожного спостереження;
- 2) існує нелінійний зв'язок між незалежними змінними;
- 3) існує взаємозалежність послідовних елементів ряду залишків.

6. Метод Фаррара – Глобера застосовується для виявлення:

- 1) автокореляції;
- 2) гомоскедастичності;
- 3) мультиколінеарності.

7. За наявності гетероскедастичності для оцінювання параметрів застосовують:

- 1) непрямий метод найменших квадратів;
- 2) умовний метод найменших квадратів;
- 3) узагальнений метод найменших квадратів.

8. Стандартна похибка інтервального прогнозу обчислюється за формулою:

$$1) S_{\text{пр}} = \sigma_u^2 X_0' (X'X)^{-1} X_0;$$

$$2) S_{\text{пр}} = \sigma_u \sqrt{X_0' (X'X)^{-1} X_0};$$

$$3) S_{\text{пр}} = \sqrt{\sigma_u^2 (1 + X_0' (X'X)^{-1} X_0)}.$$

9. Для оцінювання параметрів моделей розподіленого лага застосовують:

- 1) непрямий МНК;
- 2) трикроковий МНК;
- 3) ітераційний метод.

10. Система одночасних рівнянь називається надідентифікованою, якщо виконується умова:

$$1) k_s - 1 < m - m_s; \quad 2) k_s - 1 = m - m_s; \quad 3) k_s - 1 \leq m - m_s,$$

де k_s , m_s — відповідно кількість залежних і незалежних змінних, що входять до s -го рівняння структурної форми, m — загальна кількість екзогенних змінних моделі.

Варіант 9

1. Економіко-математичне співвідношення, що визначає в аналітичній формі зв'язок між економічними характеристиками випуску продукції та використаними для цього ресурсами, називається:

- 1) структурною формою моделі;
- 2) виробничою функцією;
- 3) моделлю “витрати – випуск”.

2. Параметри моделі:

- 1) задаються за межами економетричної моделі;
- 2) визначаються на основі статистичних даних;
- 3) змінюються протягом усього періоду спостережень.

3. Коефіцієнт множинної кореляції:

- 1) показує, яка частина руху залежної змінної описується даним регресійним рівнянням;
- 2) визначає міру зв'язку залежної змінної з усіма незалежними факторами;
- 3) визначає внесок кожної незалежної змінної в дисперсію результативної змінної.

4. Для перевірки значущості моделі застосовують:

- 1) F -критерій Фішера;
- 2) t -критерій Стьюдента;
- 3) критерій Пірсона.

5. Точковий прогноз – це:

- 1) побудова регресійної залежності за заданими точками;
- 2) значення залежної змінної, обчислене за моделлю при заданому значенні пояснюючих змінних;
- 3) визначення крайніх точок довірчого інтервалу для прогнозного значення залежної змінної.

6. Явище гомоскедастичності має місце, якщо:

- 1) існує нелінійний зв'язок між незалежними змінними;
- 2) дисперсія залишків стала для кожного спостереження;
- 3) існує взаємозалежність послідовних членів часового ряду.

7. За наявності автокореляції параметри моделі оцінюються за методом:

- 1) найменших квадратів;
- 2) Ейткена;
- 3) Фаррара – Глобера.

8. Моделі, що містять лагові значення залежної змінної, називаються:

- 1) авторегресійними моделями;
- 2) моделями адаптивних сподівань;
- 3) моделями сезонних коливань.

9. Для обґрунтування величини лага в дистрибутивно-лагових моделях застосовують:

- 1) кореляційну матрицю;
- 2) коваріаційну матрицю;
- 3) взаємну кореляційну функцію.

10. Непрямий метод найменших квадратів застосовують при оцінюванні:

- 1) лагових коефіцієнтів багатофакторної моделі;
- 2) параметрів систем одночасних рівнянь;
- 3) параметрів моделей, що містять автокореляцію.

Варіант 10

1. Дайте визначення економетрії:

- 1) наука, що вивчає вимірність зв'язків у відповідному економічному аналізі;
- 2) наука, що застосовує математичні та математико-статистичні методи в економіці;
- 3) наука, що вивчає методи оцінювання параметрів моделей, які характеризують кількісні взаємозв'язки між економічними показниками.

2. При побудові економетричної моделі необхідно:

- 1) розглядати всі без винятку елементи, що впливають на результат процесу;
- 2) виключати елементи, що видаються нетиповими стосовно даної проблеми;
- 3) використовувати всю можливу інформацію, що стосується даного дослідження.

3. Для оцінювання параметрів економетричної моделі застосовують:

- 1) закон нормального розподілу Гаусса;
- 2) метод найменших квадратів;
- 3) метод виключення Жордана — Гаусса.

4. Стандартні похибки параметрів:

- 1) визначають внесок кожної незалежної змінної в дисперсію результативного фактора;

- 2) показують статистичну значущість параметрів;
- 3) визначають міру зв'язку кожного незалежного фактора із залежною змінною.

5. Критерій Фішера застосовують для перевірки значущості:

- 1) оцінок параметрів моделі;
- 2) економетричної моделі;
- 3) коефіцієнта множинної кореляції.

6. Дисперсія прогнозу обчислюється за формулою:

$$1) \sigma^2 = \sigma_u^2 X_0' (X'X)^{-1} X_0;$$

$$2) \sigma^2 = \sigma_u \sqrt{X_0' (X'X)^{-1} X_0};$$

$$3) \sigma^2 = \sqrt{\sigma_u^2 (1 + X_0' (X'X)^{-1} X_0)}.$$

7. Якщо існує лінійний зв'язок між незалежними змінними, то це явище називається:

- 1) автокореляцією;
- 2) гомоскедастичністю;
- 3) мультиколінеарністю.

8. За наявності мультиколінеарності для оцінювання параметрів моделі застосовують:

- 1) узагальнений МНК;
- 2) метод головних компонентів;
- 3) непрямий метод найменших квадратів.

9. Ітераційний метод застосовують при оцінюванні:

- 1) параметрів багатофакторних моделей;
- 2) параметрів моделей, у яких спостерігається явище мультиколінеарності;
- 3) лагових коефіцієнтів багатофакторних динамічних моделей.

10. Якщо рівняння структурної форми моделі надіентифіковані, то для оцінювання параметрів рівнянь застосовують:

- 1) непрямий МНК;
- 2) двокроковий МНК;
- 3) узагальнений МНК.

ДОДАТКИ

РОБОТА З ТАБЛИЦЯМИ СТАНДАРТИЗОВАНОГО НОРМАЛЬНОГО РОЗПОДІЛУ

При проведенні статистичного аналізу дуже часто використовується таблиця значень функції Лапласа

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_0^u e^{-\frac{t^2}{2}} dt = P(0 \leq U < u) = F(u) - 0,5,$$

що визначає ймовірність потрапляння випадкової величини (ВВ) U в інтервал $[0, u)$.

У лівому стовпці табл. 1 наведено значення U з точністю до десятих, у верхньому рядку — соті частки U (значення U визначаються з точністю до сотих). Значення $\Phi(u)$ визначається на перетині відповідних даному значенню u рядка і стовпця ($\Phi(u)$ визначається з точністю до четвертого знака після коми). Наприклад, $\Phi(0,17) = 0,0675$, тобто $P(0 \leq U < 0,17) = 0,0675$.

Таблиця 1

U	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,0000	0,0040	0,0080	0,0120	0,0160	0,0199	0,0239	0,0279	0,319	0,0359
0,1	0,0398	0,0438	0,0478	0,0517	0,0557	0,0596	0,0636	0,0675	0,0714	0,0753
...
3,0	0,4987	0,4987	0,4987	0,4988	0,4988	0,4989	0,4989	0,4989	0,4990	0,4990
...
5,0	0,49999997									

Суть функції Лапласа $\Phi(u)$ і її зв'язок з функцією розподілу $F(u)$ стандартизованої нормальної ВВ ілюструє рис. 1.

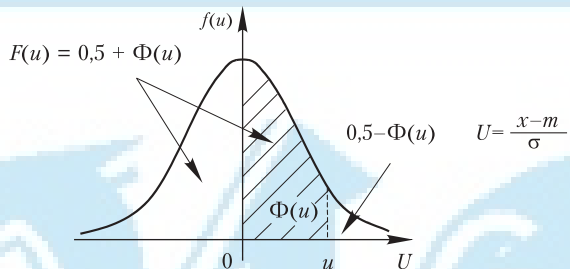


Рис. 1

Зазначимо, що кожне значення $F(u)$ перевищує відповідне значення $\Phi(u)$ на 0,5. Тому таблиці для $F(u)$ мають аналогічний вигляд.

РОБОТА З ТАБЛИЦЯМИ t -РОЗПОДІЛУ СТЬЮДЕНТА

Таблиця критичних точок розподілу Стьюдента має такий вигляд:

Таблиця 2

k	α						
	0,40	0,25	0,10	0,05	0,025	0,01	...
1	0,325	1,000	3,078	6,314	12,706	31,821	...
...
10	0,260	0,700	1,372	1,812	2,228	2,764	...
...
30	0,256	0,683	1,310	1,697	2,042	2,457	...
...
∞	0,253	0,674	1,282	1,645	1,960	2,326	...

У табл. 2 у першому стовпці наведено числа ступенів свободи k , у верхньому рядку – ймовірності (рівні значущості) α . Критична точка $t_{\alpha, k}$ визначається перетином стовпця із заданою ймовірністю α і рядка, що відповідає числу ступенів свободи k . Наприклад, $t_{0,05; 10} = 1,812$. Іншими словами, $P(t_{10} > 1,812) = 0,05$.

Іноді таблиці розподілу Стьюдента наводяться для двох критичних точок $t_{\alpha, k}^*$, що визначаються з умови $P(|t| > t_{\alpha, k}^*) = \alpha$ (рис. 2).

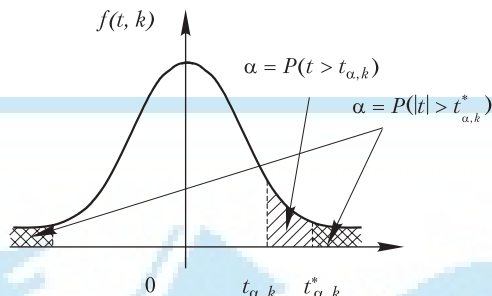


Рис. 2

РОБОТА З ТАБЛИЦЯМИ χ^2 -РОЗПОДІЛУ

Нижче наведено таблицю критичних точок χ^2 -розподілу.

Таблиця 3

k	α								
	...	0,975	0,950	0,900	...	0,100	0,050	0,025	...
1	...	10^{-5}	$4 \cdot 10^{-4}$	0,016	...	2,71	3,84	5,02	...
...
10	...	3,25	3,94	4,87	...	15,99	18,31	20,48	...
...
30	...	16,79	18,49	20,60	...	40,26	43,77	46,98	...
...

У табл. 3 у лівому стовпці наведено різні числа ступенів свободи k . У верхньому рядку зазначено ймовірності (рівні значущості) α потрапляння величини, що розглядається, у “правий хвіст” розподілу χ^2 (рис. 3, а).

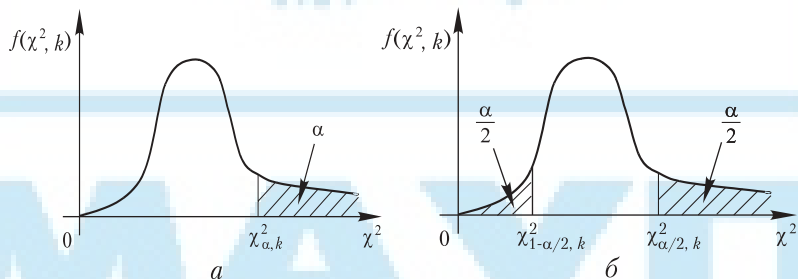


Рис. 3

Критична точка $\chi^2_{\alpha, k}$ розміщена на перетині стовпця із заданою ймовірністю α і рядка, що відповідає числу ступенів свободи k . На-

приклад, $\chi^2_{0,025; 10} = 20,48$. Іншими словами, $P(\chi^2_{10} > 20,48) = 0,025$. За-значимо, що часто таблиці χ^2 -розподілу наводяться для двох критич-них точок $\chi^2_{1-\alpha/2, k}$ і $\chi^2_{\alpha/2, k}$. У цьому разі припускається, що ймовірності потрапляння ВВ, що розглядається, в обидва “хвости” розподілу однакові й дорівнюють половині рівня значущості α , тобто $\alpha/2$ (рис. 3, б).

РОБОТА З ТАБЛИЦЯМИ F-РОЗПОДІЛУ ФІШЕРА

Таблиці критичних точок розподілу Фішера звичайно наводять-ся для різних значень імовірності (рівня значущості) α потраплян-ня у “хвіст” розподілу (рис. 4). Наприклад, для $\alpha = 0,05$ таблиця має такий вигляд:

Таблиця 4

k_2	k_1						
	1	...	10	...	100	...	∞
1	161	...	242	...	253	...	254
...
10	4,96	...	2,98	...	2,59	...	2,54
...
100	3,94	...	1,92	...	1,39	...	1,28
...
∞	3,84	...	1,83	...	1,24	...	1,00

На перетині стовпця і рядка (табл. 4), що відповідають числам ступенів свободи k_1 і k_2 , розміщена критична точка $F(\alpha, k_1, k_2)$.

Наприклад, $F_{0,05; 1; 10} = 4,96$ ($P(F_{1,10} > 4,96) = 0,05$).

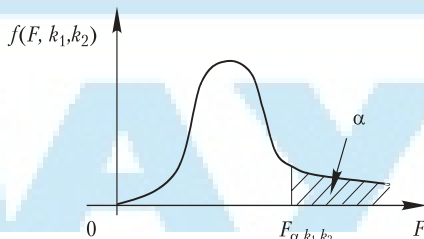


Рис. 4

Таблиця стандартизованого нормального розподілу
 $A(z)$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5703	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9533
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9700	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987									

Джерело: **Дугерти К.** Введение в эконометрику: Пер. с англ. – М.: ИНФРА-М, 1997. – С. 367.

$A(z)$ – це інтеграл щільності ймовірності стандартизованого нормального розподілу від $-\infty$ до z (площа під кривою зліва від z), який є ймовірність того, що величина нормально розподіленої випадкової змінної не перевищує середнього значення більш як на z стандартних відхилень.

Таблиця t -розподілу Стьюдента
(критичні значення $t(\alpha, k)$)

Тест	Рівень значущості α							
Двосторонній	50 %	20 %	10 %	5 %	2 %	1 %	0,2 %	0,1 %
Односторонній	25 %	10 %	5 %	2,5 %	1 %	0,5 %	0,1 %	0,05 %
k								
1	1,000	3,078	6,314	12,706	31,821	63,657	318,31	636,62
2	0,861	1,886	2,920	4,303	6,965	9,925	22,327	31,598
3	0,765	1,638	2,353	3,182	4,541	5,841	10,214	12,924
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307	3,551
60	0,679	1,296	1,671	2,000	2,390	2,660	3,232	3,460
120	0,677	1,289	1,658	1,980	2,358	2,617	3,160	3,373
∞	0,674	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Джерела: **Доугерти К.** Введение в эконометрику: Пер. с англ. — М.: ИНФРА-М, 1997. — С. 368.

Лук'яненко І. Г., Краснікова Л. І. Економетрика. — К.: Т-во "Знання", КОО, 1998. — С. 484.

Таблиця $F(\alpha, k_1, k_2)$ розподілу Фішера
для рівня значущості $\alpha = 0,05$ (5 %)

k_2	k_1												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	161	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	1,09
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 $F(\alpha, k_1, k_2)$ розподілу Фішера
для рівня значущості $\alpha = 0,05$ (5 %)

k_2	k_1												
	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	2,00	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	1,08	2,06	2,05	1,96	1,91	1,88	1,82	1,79	1,76
24	2,13	2,11	1,09	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	1,50	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	1,57	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

Таблиця $F(\alpha, k_1, k_2)$ розподілу Фішера
для рівня значущості $\alpha = 0,01$ (1 %)

k_2	k_1												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	4052	4999	5403	5625	5764	5859	5928	5981	6023	6056	6083	6106	6126
2	98,5	99,0	99,2	99,3	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,4
3	34,1	30,8	29,4	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	27,1	27,0
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,5	14,4	14,3
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,96	9,89	9,82
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,66
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47	6,41
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67	5,61
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11	5,05
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71	4,65
11	9,64	7,20	6,21	5,67	5,31	5,07	4,88	4,74	4,63	4,54	4,46	4,39	4,34
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16	4,10
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,90
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,75
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,61
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55	3,50
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46	3,40
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37	3,32
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30	3,24
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23	3,18
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,24	3,17	3,12
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12	3,07
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07	3,02
24	7,82	5,64	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03	2,98
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	3,06	2,99	2,94
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02	2,96	2,90
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,99	2,93	2,87
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,96	2,90	2,84
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,93	2,87	2,81
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91	2,84	2,79
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66	2,61
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,62	2,56	2,51
60	7,07	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50	2,44
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,51	2,45	2,40
80	6,96	4,88	4,04	3,56	3,25	3,04	2,87	2,74	2,64	2,55	2,48	2,42	2,36
90	6,92	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52	2,45	2,39	2,33
100	6,88	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,43	2,37	2,31
120	6,84	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,40	2,34	2,28
140	6,81	4,76	3,92	3,46	3,15	2,93	2,77	2,64	2,54	2,45	2,38	2,31	2,26
160	6,79	4,74	3,90	3,44	3,13	2,92	2,75	2,62	2,52	2,43	2,36	2,30	2,24
180	6,77	4,72	3,89	3,43	3,12	2,90	2,74	2,61	2,51	2,42	2,35	2,28	2,23
200	6,75	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,34	2,27	2,22
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,25	2,18	2,13

Таблиця $F(\alpha, k_1, k_2)$ розподілу Фішера
для рівня значущості $\alpha = 0,01$ (1 %)

k_2	k_1												
	14	15	16	17	18	19	20	30	40	50	100	200	∞
1	6143	6157	6169	6182	6192	6201	6209	6261	6287	6303	6335	6350	6366
2	99,4	99,4	99,4	99,4	99,4	99,4	99,5	99,5	99,5	99,5	99,5	99,5	99,5
3	26,9	26,9	26,8	26,8	26,8	26,7	26,7	26,5	26,4	26,4	26,2	26,2	26,1
4	14,2	14,2	14,2	14,1	14,1	14,0	14,0	13,8	13,7	13,7	13,6	13,5	13,5
5	9,77	9,72	9,68	9,64	9,61	9,58	9,55	9,38	9,29	9,24	9,13	9,08	9,02
6	7,60	7,56	7,52	7,48	7,45	7,42	7,40	7,23	7,14	7,09	6,99	6,93	6,88
7	6,36	6,31	6,28	6,24	6,21	6,18	6,16	5,99	5,91	5,86	5,75	5,70	5,65
8	5,56	5,52	5,48	5,44	5,41	5,38	5,36	5,20	5,12	5,07	4,96	4,91	4,86
9	5,01	4,96	4,92	4,89	4,86	4,83	4,81	4,65	4,57	4,52	4,41	4,36	4,31
10	4,60	4,56	4,52	4,49	4,46	4,43	4,41	4,25	4,17	4,12	4,01	3,96	3,91
11	4,29	4,25	4,21	4,18	4,15	4,12	4,10	3,94	3,86	3,81	3,70	3,65	3,60
12	4,05	4,01	3,97	3,94	3,91	3,88	3,86	3,70	3,62	3,57	3,47	3,41	3,36
13	3,86	3,82	3,78	3,74	3,72	3,69	3,66	3,51	3,42	3,37	3,27	3,22	3,17
14	3,70	3,66	3,62	3,59	3,56	3,53	3,51	3,35	3,27	3,22	3,11	3,06	3,00
15	3,56	3,52	3,49	3,45	3,42	3,40	3,37	3,21	3,13	3,08	2,98	2,92	2,87
16	3,45	3,41	3,37	3,34	3,31	3,28	3,26	3,10	3,02	2,97	2,86	2,81	2,75
17	3,35	3,31	3,27	3,24	3,21	3,19	3,16	3,00	2,92	2,87	2,76	2,71	2,65
18	3,27	3,23	3,19	3,16	3,13	3,10	3,08	2,92	2,84	2,78	2,68	2,62	2,57
19	3,19	3,15	3,12	3,08	3,05	3,03	3,00	2,84	2,76	2,71	2,60	2,55	2,49
20	3,13	3,09	3,05	3,02	2,99	2,96	2,94	2,78	2,69	2,64	2,54	2,48	2,42
21	3,07	3,03	2,99	2,96	2,92	2,90	2,88	2,72	2,64	2,58	2,48	2,42	2,36
22	3,02	2,98	2,94	2,91	2,88	2,85	2,83	2,67	2,58	2,53	2,42	2,36	2,31
23	2,97	2,93	2,89	2,86	2,83	2,80	2,78	2,62	2,54	2,48	2,37	2,32	2,26
24	2,93	2,89	2,85	2,82	2,79	2,76	2,74	2,58	2,49	2,44	2,33	2,27	2,21
25	2,89	2,85	2,81	2,78	2,75	2,72	2,70	2,54	2,45	2,40	2,29	2,23	2,17
26	2,86	2,81	2,73	2,75	2,72	2,69	2,66	2,60	2,42	2,36	2,25	2,19	2,13
27	2,82	2,78	2,75	2,71	2,68	2,66	2,63	2,47	2,38	2,33	2,22	2,16	2,10
28	2,79	2,75	2,72	2,68	2,65	2,63	2,60	2,44	2,35	2,30	2,19	2,13	2,06
29	2,77	2,73	2,69	2,66	2,63	2,60	2,57	2,41	2,33	2,27	2,16	2,10	2,03
30	2,74	2,70	2,66	2,63	2,60	2,57	2,55	2,39	2,30	2,25	2,13	2,07	2,01
40	2,56	2,52	2,48	2,45	2,42	2,39	2,37	2,20	2,11	2,06	2,02	1,96	1,80
50	2,46	2,42	2,38	2,35	2,32	2,29	2,27	2,10	2,01	1,95	1,94	1,87	1,68
60	2,39	2,35	2,31	2,28	2,25	2,22	2,20	2,03	1,94	1,88	1,75	1,68	1,60
70	2,35	2,31	2,27	2,23	2,20	2,18	2,15	1,98	1,89	1,83	1,70	1,62	1,54
80	2,31	2,27	2,23	2,20	2,17	2,14	2,12	1,94	1,85	1,79	1,65	1,58	1,49
90	2,29	2,24	2,21	2,17	2,14	2,11	2,09	1,92	1,82	1,76	1,62	1,55	1,46
100	2,27	2,22	2,19	2,15	2,12	2,08	2,07	1,89	1,80	1,74	1,60	1,52	1,43
120	2,23	2,19	2,15	2,12	2,09	2,06	2,03	1,86	1,76	1,70	1,56	1,48	1,38
140	2,21	2,17	2,13	2,10	2,07	2,04	2,01	1,84	1,74	1,67	1,53	1,45	1,35
160	2,20	2,15	2,11	2,08	2,05	2,02	1,99	1,82	1,72	1,66	1,51	1,42	1,32
180	2,18	2,14	2,10	2,07	2,04	2,01	1,98	1,81	1,71	1,64	1,49	1,41	1,30
200	2,17	2,13	2,09	2,06	2,03	2,00	1,97	1,79	1,69	1,63	1,48	1,39	1,28
∞	2,08	2,04	2,00	1,97	1,93	1,90	1,88	1,70	1,59	1,52	1,36	1,25	1,00

Джерело. Бронштейн И. Н., Семенов К. А. Справочник по математике: для инженеров и учащихся ВТУЗов. — М.: Наука, 1981. — С. 82–87.

d -статистика Дарбіна – Уотсона: d_n і d_0 при рівні значущості
 $\alpha = 0,05$ (5 %); n – кількість спостережень

n	Кількість факторів									
	$m = 1$		$m = 2$		$m = 3$		$m = 4$		$m = 5$	
n	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0
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	1,96	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	2,06
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	1,96
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
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,14	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,53	1,35	1,59	1,30	1,65	1,24	1,72	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,20	1,79
39	1,44	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,43	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	1,79	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
180	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	1,90	1,73	1,81	1,72	1,82

d -статистика Дарбіна – Уотсона: d_n і d_0 при рівні значущості $\alpha = 0,05$ (5 %); n – кількість спостережень

n	Кількість факторів									
	$m = 6$		$m = 7$		$m = 8$		$m = 9$		$m = 10$	
	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0
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	1,55	1,88	0,47	2,06	0,40	2,24	0,33	2,43	0,27	2,63
22	1,57	1,85	0,51	2,02	0,44	2,19	0,37	2,37	0,30	2,55
23	1,62	1,82	0,55	1,98	0,47	2,14	0,40	2,31	0,34	2,48
24	1,65	1,79	0,58	1,94	0,51	2,10	0,44	2,26	0,38	2,42
25	1,68	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
31	0,83	1,70	0,77	1,80	0,71	1,91	0,65	2,02	0,59	2,13
32	0,86	1,69	0,79	1,78	0,73	1,89	0,67	2,00	0,62	2,10
33	0,88	1,68	0,82	1,78	0,76	1,87	0,70	1,98	0,64	2,00
34	0,90	1,68	0,84	1,77	0,78	1,86	0,72	1,96	0,67	2,06
35	0,91	1,67	0,86	1,76	0,80	1,85	0,74	1,94	0,69	2,04
36	0,93	1,67	0,88	1,75	0,82	1,84	0,77	1,93	0,71	2,01
37	0,95	1,66	0,90	1,74	0,84	1,83	0,79	1,91	0,73	2,00
38	0,97	1,66	0,91	1,74	0,86	1,82	0,81	1,90	0,75	1,99
39	0,98	1,66	0,93	1,73	0,88	1,81	0,83	1,89	0,77	1,98
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	1,96
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	1,73	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	1,72	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

d -статистика Дарбіна – Уотсона: d_n і d_0 при рівні значущості $\alpha = 0,05$ (5 %); n – кількість спостережень

n	Кількість факторів									
	$m = 11$		$m = 12$		$m = 13$		$m = 14$		$m = 15$	
	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0
16	0,06	3,45								
17	0,08	3,29	0,05	3,51						
18	0,11	3,15	0,08	3,36	0,05	3,56				
19	0,15	3,02	0,10	3,23	0,07	3,42	0,04	3,60		
20	0,18	2,91	0,13	3,11	0,09	3,30	0,61	3,47	0,04	3,64
21	0,21	2,82	0,16	3,00	0,12	3,19	0,08	3,36	0,06	3,52
22	0,25	2,73	0,19	2,91	0,15	3,08	0,11	3,25	0,08	3,41
23	0,28	2,65	0,23	2,82	0,18	2,99	0,14	3,16	0,10	3,31
24	0,32	2,58	0,26	2,74	0,21	2,91	0,17	3,07	0,13	3,22
25	0,35	2,52	0,29	2,67	0,24	2,83	0,19	2,98	0,15	3,13
26	0,38	2,46	0,32	2,61	0,27	2,76	0,22	2,91	0,18	3,05
27	0,41	2,41	0,36	2,55	0,30	2,69	0,25	2,84	0,21	2,98
28	0,44	2,36	0,39	2,50	0,33	2,64	0,28	2,77	0,24	2,91
29	0,48	2,32	0,42	2,45	0,36	2,58	0,31	2,71	0,27	2,84
30	0,50	2,28	0,45	2,41	0,39	2,53	0,34	2,66	0,29	2,79
31	0,53	2,25	0,48	2,37	0,42	2,49	0,37	2,61	0,32	2,73
32	0,56	2,22	0,50	2,33	0,45	2,45	0,40	2,56	0,35	2,68
33	0,59	2,19	0,53	2,30	0,48	2,41	0,43	2,52	0,38	2,63
34	0,61	2,16	0,56	2,27	0,50	2,37	0,45	2,48	0,40	2,59
35	0,63	2,14	0,58	2,24	0,53	2,34	0,48	2,44	0,43	2,55
36	0,66	2,11	0,61	2,21	0,55	2,31	0,50	2,41	0,46	2,51
37	0,68	2,09	0,63	2,19	0,58	2,28	0,53	2,38	0,48	2,48
38	0,70	2,07	0,65	2,16	0,60	2,26	0,55	2,35	0,50	2,45
39	0,72	2,06	0,67	2,14	0,62	2,23	0,58	2,32	0,53	2,41
40	0,74	2,04	0,67	2,12	0,65	2,21	0,60	2,30	0,55	2,39
45	0,84	1,97	0,79	2,04	0,74	2,12	0,70	2,19	0,66	2,27
50	0,94	1,93	0,87	1,99	0,83	2,05	0,87	2,12	0,75	2,18
55	0,98	1,89	0,94	1,95	0,90	2,00	0,86	2,06	0,83	2,12
60	1,04	1,86	1,00	1,95	0,97	1,96	0,93	2,02	0,89	2,07
65	1,09	1,85	1,05	1,89	1,02	1,93	0,99	1,98	0,95	2,03
70	1,13	1,83	1,10	1,87	1,07	1,91	1,04	1,95	1,01	2,00
75	1,17	1,82	1,14	1,86	1,11	1,89	1,08	1,93	1,05	1,97
80	1,21	1,81	1,18	1,84	1,15	1,88	1,12	1,91	1,09	1,95
85	1,23	1,80	1,21	1,83	1,18	1,87	1,16	1,90	1,13	1,93
90	1,26	1,80	1,24	1,83	1,22	1,86	1,19	1,89	1,17	1,92
95	1,29	1,79	1,27	1,82	1,24	1,85	1,22	1,88	1,20	1,91
100	1,31	1,79	1,29	1,82	1,27	1,84	1,25	1,87	1,23	1,90
150	1,47	1,78	1,46	1,80	1,44	1,81	1,43	1,83	1,41	1,85
200	1,56	1,79	1,55	1,80	1,54	1,81	1,53	1,82	1,52	1,84

d -статистика Дарбіна – Уотсона: d_n і d_0 при рівні значущості $\alpha = 0,05$ (5 %); n – кількість спостережень

n	Кількість факторів									
	$m = 16$		$m = 17$		$m = 18$		$m = 19$		$m = 20$	
	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0
21	0,04	3,67								
22	0,05	3,56	0,03	3,70						
23	0,07	3,46	0,05	3,60	0,03	3,73				
24	0,09	3,36	0,07	3,50	0,04	3,63	0,03	3,75		
25	0,12	3,27	0,09	3,41	0,06	3,54	0,04	3,66	0,03	3,77
26	0,14	3,19	0,11	3,33	0,08	3,45	0,06	3,57	0,04	3,68
27	0,17	3,11	0,13	3,25	0,10	3,37	0,07	3,49	0,05	3,60
28	0,19	3,04	0,16	3,17	0,12	3,30	0,93	3,41	0,07	3,52
29	0,22	2,97	0,18	3,10	0,15	3,22	0,11	3,34	0,09	3,45
30	0,25	2,91	0,21	3,03	0,17	3,15	0,14	3,27	0,11	3,38
31	0,28	2,85	0,23	2,97	0,20	3,09	0,16	3,20	0,13	3,31
32	0,30	2,80	0,26	2,91	0,22	3,03	0,18	3,14	0,15	3,25
33	0,33	2,75	0,29	2,86	0,25	2,97	0,21	3,08	0,17	3,14
34	0,36	2,70	0,31	2,81	0,27	2,92	0,23	3,02	0,20	3,13
35	0,38	2,66	0,34	2,76	0,30	2,87	0,26	2,97	0,22	3,07
36	0,41	2,61	0,36	2,72	0,32	2,82	0,28	2,92	0,24	3,02
37	0,43	2,58	0,39	2,68	0,35	2,77	0,31	2,87	0,27	2,97
38	0,46	2,54	0,41	2,64	0,37	2,73	0,33	2,83	0,29	2,92
39	0,48	2,51	0,44	2,60	0,40	2,69	0,35	2,79	0,32	2,88
40	0,51	2,48	0,46	2,57	0,42	2,66	0,38	2,75	0,34	2,84
45	0,61	2,35	0,57	2,42	0,53	2,50	0,49	2,58	0,45	2,66
50	0,71	2,25	0,67	2,32	0,63	2,39	0,59	2,46	0,55	2,53
55	0,79	2,18	0,75	2,24	0,71	2,30	0,67	2,36	0,64	2,42
60	0,86	2,12	0,82	2,17	0,79	2,23	0,75	2,28	0,72	2,34
65	0,92	2,08	0,89	2,12	0,85	2,17	0,82	2,22	0,79	2,27
70	0,97	2,04	0,94	2,06	0,91	2,13	0,88	2,17	0,85	2,22
75	1,02	2,01	0,99	2,05	0,96	2,09	0,93	2,13	0,91	2,17
80	1,07	1,96	1,04	2,02	1,01	2,06	0,99	2,10	0,96	2,14
85	1,11	1,97	1,08	2,00	1,05	2,03	1,03	2,07	1,01	2,10
90	1,14	1,95	1,12	1,48	1,09	2,01	1,07	2,04	1,04	2,08
95	1,17	1,93	1,15	1,96	1,13	1,99	1,10	2,02	1,08	2,05
100	1,20	1,92	1,18	1,95	1,16	1,98	1,14	2,01	1,11	2,03
150	1,40	1,86	1,39	1,86	1,37	1,90	1,36	1,91	1,35	1,93
200	1,51	1,85	1,50	1,86	1,48	1,87	1,47	1,88	1,46	1,90

d -статистика Дарбіна – Уотсона: d_n і d_0 при рівні значущості $\alpha = 0,01$ (1 %); n – кількість спостережень

n	Кількість факторів									
	$m = 1$		$m = 2$		$m = 3$		$m = 4$		$m = 5$	
	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0
6	0,39	1,14								
7	0,44	1,04	0,29	1,68						
8	0,50	1,00	0,35	1,49	0,23	2,10				
9	0,55	1,00	0,41	1,39	0,28	1,88	0,18	2,43		
10	0,60	1,00	0,47	1,33	0,34	1,73	0,23	2,19	0,15	2,69
11	0,65	1,01	0,52	1,29	0,40	1,64	0,29	2,03	0,19	2,45
12	0,70	1,02	0,57	1,27	0,45	1,58	0,34	1,91	0,24	2,28
13	0,74	1,04	0,62	1,26	0,50	1,53	0,39	1,83	0,29	2,15
14	0,78	1,05	0,66	1,25	0,55	1,49	0,46	1,76	0,34	2,05
15	0,81	1,07	0,70	1,25	0,59	1,46	0,49	1,70	0,39	1,97
16	0,84	1,09	0,74	1,25	0,63	1,45	0,53	1,66	0,44	1,90
17	0,87	1,11	0,77	1,26	0,67	1,43	0,57	1,63	0,48	1,85
18	0,90	1,12	0,81	1,26	0,71	1,42	0,61	1,60	0,52	1,80
19	0,93	1,13	0,84	1,27	0,74	1,42	0,65	1,58	0,56	1,77
20	0,95	1,15	0,87	1,27	0,77	1,41	0,69	1,57	0,60	1,74
21	0,98	1,16	0,89	1,28	0,80	1,41	0,72	1,55	0,63	1,71
22	1,00	1,17	0,92	1,28	0,83	1,41	0,75	1,54	0,67	1,69
23	1,02	1,19	0,94	1,29	0,86	1,41	0,78	1,53	0,70	1,67
24	1,04	1,20	0,96	1,30	0,88	1,41	0,81	1,53	0,73	1,66
25	1,06	1,21	0,98	1,31	0,91	1,41	0,83	1,52	0,76	1,65
26	1,07	1,22	1,00	1,31	0,93	1,41	0,86	1,52	0,78	1,64
27	1,09	1,23	1,02	1,32	0,95	1,41	0,88	1,52	0,81	1,63
28	1,10	1,24	1,04	1,33	0,97	1,42	0,90	1,51	0,83	1,62
29	1,12	1,25	1,05	1,33	0,99	1,42	0,92	1,51	0,86	1,61
30	1,13	1,26	1,07	1,34	1,01	1,42	0,94	1,51	0,88	1,61
31	1,15	1,27	1,09	1,35	1,02	1,43	0,96	1,51	0,90	1,60
32	1,16	1,28	1,10	1,35	1,04	1,43	0,98	1,51	0,92	1,60
33	1,17	1,29	1,11	1,36	1,06	1,43	1,00	1,51	0,94	1,59
34	1,18	1,30	1,13	1,36	1,07	1,44	1,01	1,51	0,95	1,59
35	1,20	1,31	1,14	1,37	1,09	1,44	1,03	1,51	0,97	1,59
36	1,21	1,32	1,15	1,38	1,10	1,44	1,04	1,51	0,99	1,59
37	1,22	1,32	1,17	1,38	1,11	1,45	1,06	1,51	1,00	1,59
38	1,23	1,33	1,18	1,39	1,13	1,45	1,07	1,52	1,02	1,59
39	1,24	1,34	1,19	1,39	1,14	1,45	1,09	1,52	1,03	1,58
40	1,21	1,34	1,20	1,40	1,15	1,46	1,10	1,52	1,04	1,59
45	1,29	1,38	1,25	1,42	1,20	1,47	1,16	1,53	1,11	1,59
50	1,32	1,40	1,29	1,45	1,25	1,49	1,21	1,54	1,16	1,59
55	1,36	1,43	1,32	1,47	1,28	1,51	1,25	1,55	1,21	1,59
60	1,38	1,45	1,35	1,48	1,32	1,52	1,28	1,56	1,25	1,60
65	1,41	1,47	1,38	1,50	1,35	1,53	1,32	1,57	1,26	1,60
70	1,43	1,49	1,40	1,52	1,37	1,55	1,34	1,58	1,31	1,61
75	1,45	1,50	1,42	1,53	1,40	1,56	1,37	1,59	1,34	1,62
80	1,47	1,52	1,44	1,54	1,42	1,57	1,39	1,60	1,36	1,62
85	1,48	1,53	1,46	1,55	1,44	1,58	1,41	1,60	1,39	1,63
90	1,50	1,54	1,47	1,56	1,45	1,59	1,43	1,61	1,41	1,64
95	1,51	1,55	1,49	1,57	1,47	1,60	1,45	1,62	1,43	1,64
100	1,52	1,56	1,50	1,58	1,48	1,60	1,46	1,63	1,44	1,65
150	1,61	1,64	1,60	1,65	1,58	1,67	1,57	1,68	1,56	1,69
200	1,66	1,68	1,65	1,69	1,64	1,70	1,63	1,72	1,62	1,73

***d*-статистика Дарбіна – Уотсона: d_n і d_0 при рівні значущості $\alpha = 0,01$ (1 %); n – кількість спостережень**

<i>n</i>	Кількість факторів									
	<i>m</i> = 6		<i>m</i> = 7		<i>m</i> = 8		<i>m</i> = 9		<i>m</i> = 10	
	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0
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	1,55	1,88	0,47	2,06	0,40	2,24	0,33	2,43	0,27	2,63
22	1,57	1,85	0,51	2,02	0,44	2,19	0,37	2,37	0,30	2,55
23	1,62	1,82	0,55	1,98	0,47	2,14	0,40	2,31	0,34	2,48
24	1,65	1,79	0,58	1,94	0,51	2,10	0,44	2,26	0,38	2,42
25	1,68	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
31	0,83	1,70	0,77	1,80	0,71	1,91	0,65	2,02	0,59	2,13
32	0,86	1,69	0,79	1,78	0,73	1,89	0,67	2,00	0,62	2,10
33	0,88	1,68	0,82	1,78	0,76	1,87	0,70	1,98	0,64	2,00
34	0,90	1,68	0,84	1,77	0,78	1,86	0,72	1,96	0,67	2,06
35	0,91	1,67	0,86	1,76	0,80	1,85	0,74	1,94	0,69	2,04
36	0,93	1,67	0,88	1,75	0,82	1,84	0,77	1,93	0,71	2,01
37	0,95	1,66	0,90	1,74	0,84	1,83	0,79	1,91	0,73	2,00
38	0,97	1,66	0,91	1,74	0,86	1,82	0,81	1,90	0,75	1,99
39	0,98	1,66	0,93	1,73	0,88	1,81	0,83	1,89	0,77	1,98
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	1,96
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	1,73	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	1,72	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

d -статистика Дарбіна – Уотсона: d_n і d_0 при рівні значущості $\alpha = 0,01$ (1 %); n – кількість спостережень

n	Кількість факторів									
	$m = 11$		$m = 12$		$m = 13$		$m = 14$		$m = 15$	
	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0
16	0,06	3,45								
17	0,08	3,29	0,05	3,51						
18	0,11	3,15	0,08	3,36	0,05	3,56				
19	0,15	3,02	0,10	3,23	0,07	3,42	0,04	3,60		
20	0,18	2,91	0,13	3,11	0,09	3,30	0,61	3,47	0,04	3,54
21	0,21	2,82	0,16	3,00	0,12	3,19	0,08	3,36	0,06	3,52
22	0,25	2,73	0,19	2,91	0,15	3,08	0,11	3,25	0,08	3,41
23	0,28	2,65	0,23	2,82	0,18	2,99	0,14	3,16	0,10	3,31
24	0,32	2,58	0,26	2,74	0,21	2,91	0,17	3,07	0,13	3,22
25	0,35	2,52	0,29	2,67	0,24	2,83	0,19	2,98	0,15	3,13
26	0,38	2,46	0,32	2,61	0,27	2,76	0,22	2,91	0,18	3,05
27	0,41	2,41	0,36	2,55	0,30	2,69	0,25	2,84	0,21	2,98
28	0,44	2,36	0,39	2,50	0,33	2,64	0,28	2,77	0,24	2,91
29	0,48	2,32	0,42	2,45	0,36	2,58	0,31	2,71	0,27	2,84
30	0,50	2,28	0,45	2,41	0,39	2,53	0,34	2,66	0,29	2,79
31	0,53	2,25	0,48	2,37	0,42	2,49	0,37	2,61	0,32	2,73
32	0,56	2,22	0,50	2,33	0,45	2,45	0,40	2,56	0,35	2,68
33	0,59	2,19	0,53	2,30	0,48	2,41	0,43	2,52	0,38	2,63
34	0,61	2,16	0,56	2,27	0,50	2,37	0,45	2,48	0,40	2,59
35	0,63	2,14	0,58	2,24	0,53	2,34	0,48	2,44	0,43	2,55
36	0,66	2,11	0,61	2,21	0,55	2,31	0,50	2,41	0,46	2,51
37	0,68	2,09	0,63	2,19	0,58	2,28	0,53	2,38	0,48	2,48
38	0,70	2,07	0,65	2,16	0,60	2,26	0,55	2,35	0,50	2,45
39	0,72	2,06	0,67	2,14	0,62	2,23	0,58	2,32	0,53	2,41
40	0,74	2,04	0,67	2,12	0,65	2,21	0,60	2,30	0,55	2,39
45	0,84	1,97	0,79	2,04	0,74	2,12	0,70	2,19	0,66	2,27
50	0,94	1,93	0,87	1,99	0,83	2,05	0,87	2,12	0,75	2,18
55	0,98	1,89	0,94	1,95	0,90	2,00	0,86	2,06	0,83	2,12
60	1,04	1,86	1,00	1,95	0,97	1,96	0,93	2,02	0,89	2,07
65	1,09	1,85	1,05	1,89	1,02	1,93	0,99	1,98	0,95	2,03
70	1,13	1,83	1,10	1,87	1,07	1,91	1,04	1,95	1,01	2,00
75	1,17	1,82	1,14	1,86	1,11	1,89	1,08	1,93	1,05	1,97
80	1,21	1,81	1,18	1,84	1,15	1,88	1,12	1,91	1,09	1,95
85	1,23	1,80	1,21	1,83	1,18	1,87	1,16	1,90	1,13	1,93
90	1,26	1,80	1,24	1,83	1,22	1,86	1,19	1,89	1,17	1,92
95	1,29	1,79	1,27	1,82	1,24	1,85	1,22	1,88	1,20	1,91
100	1,31	1,79	1,29	1,82	1,27	1,84	1,25	1,87	1,23	1,90
150	1,47	1,78	1,46	1,80	1,44	1,81	1,43	1,83	1,41	1,85
200	1,56	1,79	1,55	1,80	1,54	1,81	1,53	1,82	1,52	1,84

d -статистика Дарбіна – Уотсона: d_n і d_0 при рівні значущості $\alpha = 0,01$ (1 %); n – кількість спостережень

n	Кількість факторів									
	$m = 16$		$m = 17$		$m = 18$		$m = 19$		$m = 20$	
	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0
21	0,04	3,67								
22	0,05	3,56	0,03	3,70						
23	0,07	3,46	0,05	3,60	0,03	3,73				
24	0,09	3,36	0,07	3,50	0,04	3,63	0,03	3,75		
25	0,12	3,27	0,09	3,41	0,06	3,54	0,04	3,66	0,03	3,77
26	0,14	3,19	0,11	3,33	0,08	3,45	0,06	3,57	0,04	3,68
27	0,17	3,11	0,13	3,25	0,10	3,37	0,07	3,49	0,05	3,60
28	0,19	3,04	0,16	3,17	0,12	3,30	0,09	3,41	0,07	3,52
29	0,22	2,97	0,18	3,10	0,15	3,22	0,11	3,34	0,09	3,45
30	0,25	2,91	0,21	3,03	0,17	3,15	0,14	3,27	0,11	3,38
31	0,28	2,85	0,23	2,97	0,20	3,09	0,16	3,20	0,13	3,31
32	0,30	2,80	0,26	2,91	0,22	3,03	0,18	3,14	0,15	3,25
33	0,33	2,75	0,29	2,86	0,25	2,97	0,21	3,08	0,17	3,14
34	0,36	2,70	0,31	2,81	0,27	2,92	0,23	3,02	0,20	3,13
35	0,38	2,66	0,34	2,76	0,30	2,87	0,26	2,97	0,22	3,07
36	0,41	2,61	0,36	2,72	0,32	2,82	0,28	2,92	0,24	3,02
37	0,43	2,58	0,39	2,68	0,35	2,77	0,31	2,87	0,27	2,97
38	0,46	2,54	0,41	2,64	0,37	2,73	0,33	2,83	0,29	2,92
39	0,48	2,51	0,44	2,60	0,40	2,69	0,35	2,79	0,32	2,88
40	0,51	2,48	0,46	2,57	0,42	2,66	0,38	2,75	0,34	2,84
45	0,61	2,35	0,57	2,42	0,53	2,50	0,49	2,58	0,45	2,66
50	0,71	2,25	0,67	2,32	0,63	2,39	0,59	2,46	0,55	2,53
55	0,79	2,18	0,75	2,24	0,71	2,30	0,67	2,36	0,64	2,42
60	0,86	2,12	0,82	2,17	0,79	2,23	0,75	2,28	0,72	2,34
65	0,92	2,08	0,89	2,12	0,85	2,17	0,82	2,22	0,79	2,27
70	0,97	2,04	0,94	2,06	0,91	2,13	0,88	2,17	0,85	2,22
75	1,02	2,01	0,99	2,05	0,96	2,09	0,93	2,13	0,91	2,17
80	1,07	1,96	1,04	2,02	1,01	2,06	0,99	2,10	0,96	2,14
85	1,11	1,97	1,08	2,00	1,05	2,03	1,03	2,07	1,01	2,10
90	1,14	1,95	1,12	1,48	1,09	2,01	1,07	2,04	1,04	2,08
95	1,17	1,93	1,15	1,96	1,13	1,99	1,10	2,02	1,08	2,05
100	1,20	1,92	1,18	1,95	1,16	1,98	1,14	2,01	1,11	2,03
200	1,51	1,85	1,50	1,86	1,48	1,87	1,47	1,88	1,46	1,90

Джерело. Толбатов Ю. А. Економетрика. — К.: Четверта хвиля, 1997. — С. 293–300.

ТАБЛИЦЯ КРИТИЧНИХ ЗНАЧЕНЬ
ДЛЯ ВІДНОШЕННЯ ФОН НЕЙМАНА

$$p\left(\frac{\delta^2}{s^2} < k\right) = \int_0^k \omega\left(\frac{\delta^2}{s^2}\right) d\left(\frac{\delta^2}{s^2}\right)$$

k	n								
	4	5	6	7	8	9	10	11	12
0,25				0,00001	0,00001	0,00001	0,00001		0,00001
0,30				0,00007	0,00007	0,00005	0,00004	0,00002	0,00003
0,35			0,00006	0,00027	0,00021	0,00014	0,00009	0,00005	0,00007
0,40			0,00047	0,00065	0,00047	0,00031	0,00019	0,00012	0,00016
0,45			0,00126	0,00126	0,00088	0,00059	0,00038	0,00025	0,00031
0,50		0,00038	0,00246	0,00214	0,00150	0,00103	0,00069	0,00046	0,00055
0,55		0,00223	0,00409	0,00333	0,00237	0,00168	0,00116	0,00080	0,00094
0,60		0,00493	0,00615	0,00486	0,00355	0,00259	0,00185	0,00132	0,00152
0,65		0,00830	0,00865	0,00678	0,00511	0,00382	0,00282	0,00208	0,00235
0,70		0,01225	0,01161	0,00913	0,00710	0,00544	0,00414	0,00313	0,00351
0,75		0,01673	0,01505	0,01197	0,00958	0,00753	0,00587	0,00455	0,00508
0,80	0,00356	0,02171	0,01900	0,01534	0,01263	0,01015	0,00809	0,00642	0,00714
0,85	0,01302	0,02717	0,02348	0,01932	0,01631	0,01338	0,01089	0,00883	0,00980
0,90	0,02257	0,03310	0,02851	0,02403	0,02068	0,01729	0,01436	0,01188	0,01316
0,95	0,03223	0,03949	0,03412	0,02957	0,02579	0,02196	0,01858	0,01565	0,01733
1,00	0,04199	0,04634	0,04035	0,03598	0,03171	0,02745	0,02363	0,02025	0,02241
1,05	0,05186	0,05364	0,04728	0,04325	0,03849	0,03384	0,02959	0,02578	0,02852
1,10	0,06184	0,06140	0,05500	0,05137	0,04618	0,04120	0,03655	0,03232	0,03577
1,15	0,07194	0,06963	0,06361	0,06036	0,05482	0,04957	0,04458	0,03997	0,04425
1,20			0,07323	0,07020	0,06445	0,05901	0,05375	0,04882	0,05407
1,25						0,06956	0,06412	0,05894	0,06531
1,30								0,07040	

k	n						
	15	20	25	30	40	50	60
0,35	0,00001						
0,40	0,00002						
0,45	0,00004						
0,50	0,00009	0,00001					
0,55	0,00018	0,00002					
0,60	0,00033	0,00005	0,00001				
0,65	0,00059	0,00012	0,00002				
0,70	0,00100	0,00024	0,00005	0,00001			
0,75	0,00161	0,00044	0,00011	0,00003			
0,80	0,00250	0,00076	0,00023	0,00007	0,00001		
0,85	0,00375	0,00127	0,00044	0,00015	0,00002		
0,90	0,00547	0,00206	0,00079	0,00030	0,00004	0,00001	
0,95	0,00778	0,00323	0,00135	0,00057	0,00010	0,00002	
1,00	0,01079	0,00489	0,00222	0,00102	0,00022	0,00005	0,00001
1,05	0,01465	0,00720	0,00355	0,00176	0,00044	0,00012	0,00003
1,10	0,01950	0,01033	0,00550	0,00294	0,00085	0,00026	0,00008
1,15	0,02550	0,01448	0,00826	0,00474	0,00158	0,00054	0,00019
1,20	0,03280	0,01986	0,01208	0,00738	0,00280	0,00108	0,00043
1,25	0,04155	0,02670	0,01723	0,01117	0,00476	0,00206	0,00092
1,30	0,05189	0,03524	0,02402	0,01644	0,00780	0,00376	0,00185
1,35	0,06396	0,04571	0,03276	0,02357	0,01235	0,00656	0,00355
1,40	0,07787	0,05834	0,04379	0,03298	0,01892	0,01098	0,00649
1,45		0,07333	0,05743	0,04511	0,02810	0,01769	0,01133
1,50			0,07398	0,06038	0,04055	0,02750	0,01893
1,55				0,07920	0,05696	0,04131	0,03034
1,60					0,07797	0,06006	0,04675
1,65						0,08465	0,06942
1,70							0,09949

Значення k , для яких $P\left(\frac{\delta^2}{s^2} < k\right) = 0$.

n	4	5	6	7	8	9	10	11	12	15
k	0,7811	0,4775	0,3215	0,2311	0,1740	0,1357	0,1088	0,0891	0,0743	0,0468

n	20	25	30	40	50	60
k	0,0259	0,0164	0,0113	0,0063	0,0040	0,0028

Джерело. Наконечний С. І., Терещенко Т. О., Романюк Т. П. Економетрія: Навч. посіб. — К.: КНЕУ, 1997. — С. 345–346.

СПИСОК ВИКОРИСТАНОЇ ТА РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

Основна

1. Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрики: Учебник для вузов. — М.: ЮНИТИ, 1998. — 1022 с.
2. Бородич С. А. Эконометрика: Учеб. пособие. — Минск: Новое знание, 2001. — 408 с.
3. Грубер Й. Економетрія: Вступ до множинної регресії та економетрії: У 2 т. — К.: Нічлава, 1998–1999.
4. Джонстон Дж. Эконометрические методы. — М.: Статистика, 1980. — 444 с.
5. Доугерти К. Введение в эконометрику: Пер. с англ. — М.: ИНФРА-М, 1997. — 402 с.
6. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. — М.: Финансы и статистика, 1986. — Т. 1 — 365 с.; Т. 2 — 379 с.
7. Емельянов А. С. Эконометрия и прогнозирование. — М.: Экономика, 1985. — С. 82–89.
8. Єлейко В. Основи економетрії. — Львів: “Марка Лтд”, 1995. — 191с.
9. Замков О. О., Толстопятенко А. В., Черемных Ю. Н. Математические методы в экономике: Учебник / Под общ. ред. А. В. Сидоровича. — 3-е изд., перераб. — М.: Дело и Сервис, 2001. — 368 с. — (Сер. “Учебники МГУ им. М. В. Ломоносова”).
10. Кейн Э. Экономическая статистика и эконометрия. Введение в количественный экономический анализ. — М.: Статистика, 1977. — 254 с.
11. Корольов О. А. Економетрія: Навч. посіб. — К.: Європейський ун-т, 2002. — 660 с.
12. Ланге О. Введение в эконометрию. — М.: Прогресс, 1964. — 360 с.
13. Леоненко М. М., Мішура Ю. С., Пархоменко В. М., Ядренко М. Й. Теоретико-ймовірнісні та статистичні методи в економетриці та фінансовій математиці. — К.: Інформтехніка, 1995. — 380 с.
14. Лук'яненко І. Г., Краснікова Л. І. Економетрика: Підручник. — К.: Т-во “Знання”, КОО, 1998. — 494 с.
15. Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика: Навч. курс. — М.: Дело, 1997. — 248 с.
16. Маленко Э. Статистические методы эконометрии. — М.: Статистика, 1975. — 423 с.
17. Наконечний С. І., Терещенко Т. О., Романюк Т. П. Економетрія: Навч. посіб. — К.: КНЕУ, 1997. — 352 с.
18. Тинтнер Г. Введение в эконометрию. — М.: Статистика, 1965. — 368 с.

19. *Толбатов Ю. А.* Эконометрика: Підруч. для студ. екон. спец. вищ. навч. закл. — К.: Четверта хвиля, 1997. — 320 с.
20. *Фишер Ф.* Проблема идентификации в эконометрии. — М.: Статистика, 1978. — 224 с.
21. *Хейс Д.* Причинный анализ в статистических исследованиях. — М.: Финансы и статистика, 1981. — 224 с.

Додаткова

22. *Венецкий И. Г., Венецкая В. И.* Основные математико-статистические понятия и формулы в экономическом анализе: Справочник. — 2-е изд., перераб. и доп. — М.: Статистика, 1979. — 448 с.
23. *Вини Р., Холден К.* Введение в прикладной эконометрический анализ. — М.: Финансы и статистика, 1981. — 294 с.
24. *Геец В. М.* Отраслевое прогнозирование: методологический и организационный аспекты. — К.: Наук. думка, 1990. — 120 с.
25. *Гранберг А. Г.* Динамические модели народного хозяйства. — М.: Экономика, 1985. — 204 с.
26. *Гранберг А. Г.* Статистическое моделирование и прогнозирование. — М.: Финансы и статистика, 1990. — 378 с.
27. *Дадаия В. С.* Моделирование глобальных экономических процессов. — М.: Экономика, 1984. — 278 с.
28. *Демиденко Е. З.* Линейная и нелинейная регрессии. — М.: Финансы и статистика, 1981. — 302 с.
29. *Дружинин В. В., Конторов Д. С.* Проблемы системологии. Проблемы теории сложных систем. — М.: Радио и связь, 1986. — 296 с.
30. *Дружинин В. В., Конторов Д. С.* Системотехника. — М.: Радио и связь, 1985. — 200 с.
31. *Дюран Б., Одел П.* Кластерный анализ. — М.: Статистика, 1977. — 128 с.
32. *Емельянов А. С.* Общественное производство: Динамика, тенденции, модели. — К.: Наук. думка, 1980. — С. 347–409.
33. *Емельянов А. С., Кузьменко В. П.* Многорегиональная эконометрическая модель УКР-3: Плановое управление экономикой развитого социализма: В 5 т. — К.: Наук. думка, 1985. — Т. 1. Народнохозяйственные процессы, их планирование и прогнозирование. — С. 285–289.
34. *Иванова В. М.* Экономическая теория. Основы бизнеса / Ред. совет: А. Д. Смирнов, В. Ф. Максимова и др. — М.: СОМИНТЭК, 1991. — Ч. IV. Эконометрика.
35. *Ивахненко А. Г.* Долгосрочное прогнозирование и управление сложными системами. — К.: Тэхника, 1975. — 312 с.
36. *Ивахненко А. Г.* Алгоритмы метода группового учета аргументов (МГУА) при непрерывных и бинарных признаках. — К.: ИК НАНУ, 1992. — 52 с.

37. *Ивахненко А. Г., Юрачковский Ю. П.* Моделирование сложных систем по экспериментальным данным. — М.: Радио и связь, 1987. — 116 с.
38. *Колек Ю., Шуял И.* Эконометрические модели в социалистических странах: Пер. со словац. — М.: Экономика, 1978. — 152 с.
39. *Королев О. А.* Проблемы конструирования и использования макроэкономических эконометрических моделей переходной экономики: на примере Украины. — К.: ТОВ “Міжнар. фін. агенція”, 1997. — 224 с.
40. *Корольов О. А.* Эконометрія в задачах, ситуаціях та проблемах для студентів спеціальності «Фінанси і кредит»: Конспект лекцій і практикум: У 3 ч. — К.: КДТЕУ, 1996–1997.
41. *Корольов О. А., Рязанцева В. В.* Практикум з економітриї. Навч. посіб. — К.: Київ. нац. торг.-екон. ун-т, 2000. — 249 с.
42. *Култич О. І.* Економітрия: Навч. посіб. — Хмельницький: Поділля, 1997. — 120 с.
43. *Мартынюс С.* Методологические проблемы построения и применения эконометрических моделей. — Вильнюс: Макскас, 1979. — 170 с.
44. *Михалевич М. В., Чижевская А. Ю.* Динамические макромоделли нестабильных процессов при переходе к рыночной экономике // Кибернетика и системный анализ. — 1993. — № 4. — С. 81–88.
45. *Михалевич В. С., Михалевич М. В.* Динамические макромоделли процессов ценообразования в переходной экономике // Кибернетика и системный анализ. — 1995. — № 3. — С. 116–129.
46. *Новицкий П. В., Зограф И. А.* Оценка погрешностей результатов измерений. — Л.: Энергоатомиздат, 1991. — 304 с.
47. *Петров А. А., Поспелов И. Г.* Системный анализ развивающейся экономики: Многосекторная модель и учет природных ресурсов / Изв. АН СССР. Техническая кибернетика. — 1979. — № 4. — С. 11–23.
48. *Пуарье Д.* Эконометрия структурных изменений (с применением сплайн-функций). — М.: Финансы и статистика, 1981. — 184 с.
49. *Сегал В. В.* Анализ и синтез сложных систем: Опыт системного исследования. — К.: ЦЭМИ “Тридента”, 1994. — С. 59–153.
50. *Устойчивые* статистические методы оценки данных: Пер. с англ. / Под ред. Р. Л. Лонера, Г. Н. Уилкинсона. — М.: Машиностроение, 1984. — 232 с.
51. *Фомин Б. С.* Эконометрические теории и модели международных экономических отношений. — М.: Мысль, 1970. — 268 с.

Зміст

<i>Вступ</i>	3
Розділ 1. Математичне моделювання як метод наукового пізнання економічних явищ і процесів	11
1.1. Загальні принципи моделювання в економіці	11
1.2. Кореляційно-регресійний аналіз в економіці	14
1.3. Економетрична модель та її елементи	19
1.4. Статистична база економетричних досліджень	21
1.5. Особливості математичного моделювання економічних систем	23
Розділ 2. Моделі парної регресії та їх дослідження	25
2.1. Приклади парних зв'язків в економіці	25
2.2. Лінійна модель з двома змінними	26
2.3. Метод найменших квадратів	29
Розділ 3. Загальна лінійна економетрична модель	39
3.1. Багатофакторні економетричні моделі та їх специфікація.	39
3.2. Метод найменших квадратів	44
3.3. Верифікація моделі	47
3.4. Прогнозування за лінійною моделлю	55
3.5. Методи побудови багатофакторної регресійної моделі	57
3.6. Етапи дослідження загальної лінійної моделі множинної регресії	58
Розділ 4. Мультиколінеарність	72
4.1. Поняття про мультиколінеарність та її вплив на оцінку параметрів моделі	72
4.2. Тестування наявності мультиколінеарності	74
4.3. Алгоритм Фаррара — Глобера	76
4.4. Засоби усунення мультиколінеарності. Метод головних компонентів	84
Розділ 5. Гетероскедастичність	89
5.1. Виявлення гетероскедастичності та її природа	89
5.2. Тестування наявності гетероскедастичності	91
5.3. Трансформування початкової моделі	96
5.4. Оцінювання параметрів багатофакторної регресійної моделі на основі узагальненого методу найменших квадратів	100

<i>Розділ 6. Автокореляція</i>	104
6.1. Природа автокореляції та її наслідки	104
6.2. Тестування наявності автокореляції	106
6.3. Параметризація моделі з автокорельованими залишками	109
6.4. Приклад оцінювання параметрів моделі з автокорельованими залишками	111
<i>Розділ 7. Моделі розподіленого лага</i>	118
7.1. Поняття лага та лагових моделей в економіці	118
7.2. Оцінювання параметрів дистрибутивно-лагових моделей	121
7.3. Оцінювання параметрів авторегресійних моделей	123
<i>Розділ 8. Системи одночасних рівнянь</i>	126
8.1. Поняття про системи одночасних рівнянь	126
8.2. Приклади систем одночасних рівнянь	127
8.3. Структурна та зведена (прогнозна) форми системи рівнянь	130
8.4. Поняття ідентифікації (ототожнення) системи рівнянь	132
8.5. Методи оцінювання параметрів систем рівнянь	141
8.6. Прогноз і загальні довірчі інтервали	151
<i>Розділ 9. Методи дослідження якісних економічних показників</i>	161
9.1. Якісні економічні показники	161
9.2. Регресійні моделі з бінарними незалежними змінними	163
9.3. Регресійні моделі з бінарними залежними змінними	165
<i>Тестові завдання з економетрії</i>	168
<i>Додатки</i>	183
<i>Список використаної та рекомендованої літератури</i>	203



Educational manual “Econometrics” covers main tasks of econometric investigations and methods of their solving, place and significance of this scientific branch, its connection with economics, statistics and mathematics. Least-squares method (LSM) was chosen as the main method of estimating regressive models parameters. The manual contains peculiarities of parameters estimation in case of violating main suppositions of classical LSM use. Alternative methods of estimating models parameters (main components method, generalized LSM etc.) are described here. Modelling of economic processes dynamics is considered by the examples of distributive-lag and autoregressive models. Systems of simultaneous equations were used to model economic processes with direct and feedback coupling. Means of investigation and use of models with qualitative independent variables are described in the manual.

For students of economic specialties who studied higher mathematics, algebra, probability theory and mathematical statistics.

Навчальне видання
Лещинський Олег Львович
Рязанцева Валентина Василівна
Юнькова Олена Олександрівна

ЕКОНОМЕТРІЯ
Навчальний посібник
2-ге видання, стереотипне

Educational edition
Leschynsky, Oleg L.
Ryazantseva, Valentyna V.
Yunjkova, Olena O.

ECONOMETRICS
Educational manual
2nd edition, stereotype

Відповідальний редактор *С. Г. Рогузько*
Редактор *Л. В. Логвиненко*
Коректор *Л. Г. Бурлакiна*
Комп'ютерне верстання *М. І. Фадєєва, С. В. Фадєєв*
Оформлення обкладинки *О. О. Стеценко*

Підп. до друку 07.07.08. Формат 60×84/16. Папір офсетний. Друк офсетний.
Ум. друк. арк. 12,10. Обл.-вид. арк. 14,04. Наклад 2500 пр.

Міжрегіональна Академія управління персоналом (МАУП)
03039 Київ-39, вул. Фрометівська, 2, МАУП

ДП «Видавничий дім «Персонал»
03039 Київ-39, просп. Червонозоряний, 119, літ. XX

*Свідоцтво про внесення до Державного реєстру
суб'єктів видавничої справи ДК № 8 від 23.02.2000*