

Я. В. Федько

ЕКОНОМЕТРІЯ

Міністерство освіти і науки України
Державний заклад
«Луганський національний університет
імені Тараса Шевченка»

Я. В. Федько

ЕКОНОМЕТРІЯ

*Курс лекцій для студентів
денної та заочної форм навчання
спеціальностей «Прикладна статистика»,
«Менеджмент»*

Луганськ
ДЗ «ЛНУ імені Тараса Шевченка»
2011

УДК 330.43(075.8)

ББК 65в631я73

Ф35

Рецензенти :

Матросова Л. М. – доктор економічних наук, професор кафедри економічної теорії та прикладної статистики Луганського національного університету імені Тараса Шевченка.

Димарський Я. М. – доктор фізико-математичних наук, професор, завідувач кафедри інформатики Луганського державного університету внутрішніх справ імені Е. О. Дідоренка.

Мортіков В. В. – доктор економічних наук, професор кафедри управління персоналом та економічної теорії Східноукраїнського національного університету імені Володимира Даля.

Федько Я. В.

К93 Економетрія : курс лекцій для студ. ден. та заоч. форми навч. спец. «Прикладна статистика», «Менеджмент» / Я. В. Федько ; Держ. закл. «Луган. нац. ун-т імені Тараса Шевченка». – Луганськ : Вид-во ДЗ «ЛНУ імені Тараса Шевченка», 2011. – 150 с.

Курс лекцій охоплює всі змістовні модулі, передбачені навчальною програмою. Містить глосарій, контрольні питання з кожної теми та список літератури, рекомендованої для самостійної роботи студентів.

Цей курс лекцій стане до нагоди також аспірантам, викладачам, слухачам ліцеїв, курсів, коледжів, які цікавляться економетричними дослідженнями.

УДК 330.43(075.8)

ББК 65в631я73

*Рекомендовано до друку навчально-методичною радою
Луганського національного університету імені Тараса Шевченка
(протокол № 5 від 19 січня 2011 року)*

© Федько Я. В., 2011

© ДЗ «ЛНУ імені Тараса Шевченка», 2011

З М І С Т

Передмова	7
------------------------	---

Перший модуль

Тема 1. Економетрія як самостійна наукова

дисципліна	9
1.1. Коротка історична довідка	9
1.2. Економетрія як наукова дисципліна	12
1.3. Об'єкт, предмет, мета і завдання економетрії.....	13
1.4. Місце курсу серед дисциплін фундаментальної підготовки бакалаврів з економічних спеціальностей.....	14
Контрольні питання	15

Тема 2. Математичне моделювання як метод наукового пізнання економічних явищ і процесів.....

2.1. Поняття математичної моделі. Класифікація моделей	16
2.2. Етапи побудови економічної моделі.....	18
2.3. Кореляційно-регресійний аналіз в економіці.....	19
Контрольні питання	26

Тема 3. Основи економетричного моделювання.....

3.1. Економетрична модель, її елементи та приклади.....	26
3.2. Випадкова складова економетричної моделі.....	33
3.3. Оцінювання параметрів моделі методом найменших квадратів.....	38
Контрольні питання	42

Тема 4. Проста лінійна регресія.....

4.1. Проста лінійна регресія.....	42
4.2. Побудова лінійної регресії.....	48
4.3. Коефіцієнт кореляції.....	57
4.4. Статистичні критерії перевірки значущості оцінок параметрів і моделі.....	58
4.5. Нелінійні моделі парної регресії й кореляції.....	61
Контрольні питання.....	62

Тема 5. Багатофакторна регресія	63
5.1. Відбір факторів при побудові рівняння множинної регресії.....	63
5.2. Метод найменших квадратів (МНК) для багатофакторної регресії. Властивості оцінок на основі МНК.....	68
5.3. Перевірка істотності факторів і показники якості регресії.....	69
5.4. Узагальнений метод найменших квадратів (УМНК).....	73
Контрольні питання.....	75
Тема 6. Методи дослідження якісних економічних показників	75
6.1. Якісні економічні показники.....	76
6.2. Регресійні моделі з бінарними незалежними змінними.....	78
6.3. Регресійні моделі з бінарними залежними змінними.....	79
Контрольні питання.....	81
Тема 7. Дисперсійний аналіз економетричної моделі	82
7.1. Поняття дисперсійного аналізу. 1МНК.....	82
7.2. Приклад дисперсійного аналізу економетричної моделі та прогноз.....	87
Контрольні питання.....	90
Другий модуль	
Тема 8. Мультиколінеарність	91
8.1. Визначення мультиколінеарності та її природа.....	91
8.2. Практичні наслідки мультиколінеарності.....	92
8.3. Тестування наявності та рівня мультиколінеарності. Метод Фаррара-Глобера.....	92

8.4. Засоби вилучення мультиколінеарності.....	95
Контрольні питання	95
Тема 9. Гетероскедастичність.....	96
9.1. Визначення гетероскедастичності та її природи.....	96
9.2. Наслідки порушення припущення про гомоскедастичність.	97
9.3. Знаходження гетероскедастичності та її вилучення.....	98
Контрольні питання	103
Тема 10. Автокореляція: поняття, причини виникнення, наслідки.....	104
10.1. Поняття та визначення автокореляції. Причини виникнення автокореляції.....	104
10.2. Методи перевірки наявності автокореляції.....	106
10.3. Методи оцінки параметрів моделі з автокорельованими залишками.....	108
Контрольні питання	112
Тема 11. Метод інструментальних змінних.....	113
11.1. Властивості оцінок моделі при стохастичних змінних.....	113
11.2. Метод інструментальних змінних.....	114
11.3. Визначення інструментальних змінних.....	117
Контрольні питання	120
Тема 12. Моделі розподіленого лагу.....	112
12.1. Поняття лагу і лагових змінних.	121
12.2. Взаємна кореляційна функція	123
12.3. Лаги залежних і незалежних змінних.....	124
12.4. Методи оцінювання та прогноз.....	128
Контрольні питання	129

Тема 13. Системи одночасних структурних рівнянь	130
13.1. Структурна й наведена форми моделі.....	130
13.2. Проблема ідентифікації.....	134
13.3. Методи оцінки параметрів структурної форми моделі.....	138
Контрольні питання	140
Глосарій	141
Список використаної та рекомендованої літератури	147

ПЕРЕДМОВА

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.

Запропонований курс лекцій було розроблено згідно з навчальною програмою дисципліни «Економетрія». Його переваги у тому, що у ньому в рамках робочої програми коротко та лаконічно розглядаються основні модулі, передбачені навчальною програмою та містить весь

теоритичний матеріал необхідний для практичних занять з економетрії. При розробці курсу було ураховано отримані студентами знання, уміння та навички з економічної теорії, вищої математики та статистики, а також кількість годин, які відведено на вивчення курсу з економетрії навчальними планами підготовки бакалаврів з економічних спеціальностей. Курс лекцій містить глосарій, контрольні питання з кожної теми та список літератури, рекомендованої для самостійної підготовки студентів.

У ньому 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. Описано альтернативні методи оцінювання параметрів моделей та 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.

Основними **завданнями**, що мають бути вирішені у процесі вивчення курсу, є формування у студентів певних знань та вмінь:

- розрізняти специфікації економічних моделей;
- знати методи обчислення оцінок параметрів моделі;
- вміти оцінювати якість самої моделі;
- надавати економіко-статистичне тлумачення одержаних результатів;
- визначати наявність мультиколінеарності, автокореляції та гетероскедастичності в моделі та вміти застосовувати методи їх усунення;
- вміти використовувати узагальнений метод найменших квадратів;

- будувати багатофакторні лінійні економетричні моделі динаміки;
- оцінювати параметри системи одночасних рівнянь;
- перевіряти моделі на ідентифікованість;
- використовувати математичні методи дослідження якісних економічних показників.

Перший модуль

Тема 1. Економетрія як самостійна наукова дисципліна.

План

- 1.1. Коротка історична довідка.
- 1.2. Економетрія як наукова дисципліна.
- 1.3. Об'єкт, предмет і завдання економетрії
- 1.4. Місце курсу серед дисциплін фундаментальної підготовки бакалаврів з економічних спеціальностей.

Література [3, 4, 5, 10, 12, 14, 15, 20].

1.1. 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 XIX 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 XX 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 математизації економічної теорії. Хоча їхні праці безпосередньо не

пов'язані з економетричними дослідженнями, усе ж вони значною мірою вплинули 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.

1.2. Економетрія як наукова дисципліна.

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:

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.*

1.3. Об'єкт, предмет 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.

1.4. 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.

Контрольні питання

1. Дайте визначення предмета курсу економетрії.
2. У чому полягає основний зміст проблематики економетрії?
3. Як пов'язані між собою математична економіка, описова статистика та математична статистика?
4. Визначте основне завдання економетричного дослідження.
5. Яка роль цього курсу в підготовці бакалаврів з економічних спеціальностей?

Тема 2. Математичне моделювання як метод наукового пізнання економічних явищ і процесів.

План

- 2.1. Поняття математичної моделі. Класифікація моделей.
 - 2.2. Етапи побудови економічної моделі.
 - 2.3. Кореляційно-регресійний аналіз в економіці.
- Література [1, 5, 8, 13, 22, 24].

2.1. Поняття математичної моделі. Класифікація моделей.

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 факторів 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.

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.

2.2. 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) уводяться 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.

2.3. 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 (опади, 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

$$M(Y|x) = f(x), \quad (2.1)$$

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

$$M(Y | x_1, x_2, \dots, x_m) = P(x_1, x_2, \dots, x_m), \quad (2.2)$$

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

$$Y = f(x) \quad (2.3)$$

має бути доповнена випадковою складовою u , що, власне, і вказує на стохастичну сутність залежності.

Означення 1. Зв'язки між залежною та незалежною (незалежними) змінними, що описуються співвідношеннями (2.3)

$$Y = P(x_1, x_2, \dots, x_m) + u, \dots \dots \dots (2.4)$$

називають *регресійними рівняннями (моделями)*.

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)*.

Будують діаграму розсіювання таким чином: на осі абсцис відзначають значення незалежної змінної (x), на осі ординат – значення залежної змінної (y), результат кожного спостереження (x_i, y_i) деякого економічного процесу відображається точкою на площині. Сукупність цих точок утворює хмарку, яка відображає зв'язок між двома змінними. За шириною розкиду точок можна зробити висновок про тісноту зв'язку сукупності. Якщо точки розміщені близько одна до одної (у вигляді вузької смужки), то можна стверджувати про наявність відносно тісного зв'язку. Якщо точки на діаграмі розкидані широко, то має місце слабкий зв'язок між змінними, або взагалі не існує зв'язку. Наведемо приклад діаграми розсіювання у випадку лінійного зв'язку (рис 2.1.).

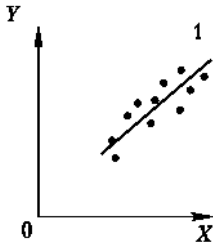
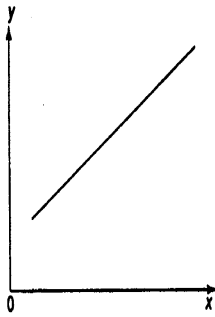
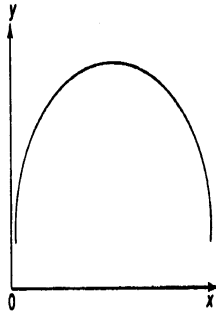


Рис 2.1 Діаграма розсіювання у випадку лінійного зв'язку.

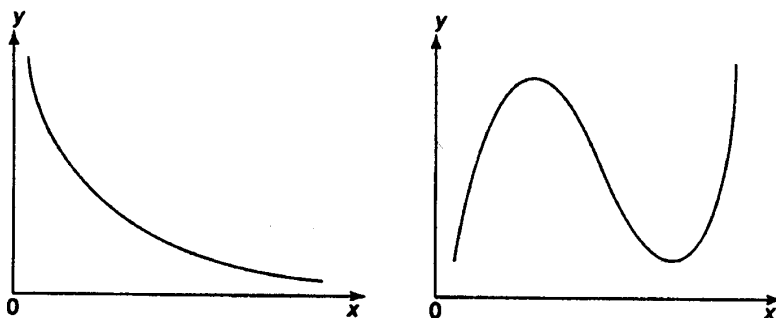
Розглянемо основні типи кривих, які використовуються при кількісній оцінці зв'язків (рис. 2.2.):



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



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



$$\hat{y}_x = a + b/x \qquad \hat{y}_x = a + b \cdot x + c \cdot x^2 + d \cdot x^3$$

Рис. 2.2. Основні типи кривих, які використовуються при кількісній оцінці зв'язків між двома змінними.

Щоб краще вибрати форму зв'язку, необхідно збільшити кількість спостережень — точок кореляційного поля або скористатися 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 :

$$C = c_0 + bY, \quad (2.5)$$

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.

Контрольні питання

1. З яких елементів складається математична модель?

2. Назвіть типи математичних моделей. Чим вони різняться між собою?

3. До якого типу математичних моделей належить економетрична модель?

4. Які особливості має економетрична модель?

5. Як треба розуміти сукупність спостережень та її однорідність?

6. Чим забезпечується порівнянність даних у просторі і часі?

7. У яких випадках використовують кореляційно-регресійний аналіз?

Тема 3. Основи економетричного моделювання

План

3.1. Економетрична модель її елементи.

3.2. Випадкова складова економетричної моделі.

3.3. Оцінювання параметрів моделі методом найменших квадратів.

Література [2, 10, 12, 14, 21, 25, 27].

3.1. Економетрична модель, її елементи та приклади

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. 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

$$Y = f(x_1, x_2, \dots, x_m, u), \dots \dots \dots (3.1)$$

де Y — результат, або залежна змінна, змінювання 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 2. 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 3. 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.

Приклад 1. Виробнича функція Кобба — Дугласа

Виробнича функція — це економетрична модель, яка кількісно описує зв'язок основних результативних показників виробничо-господарської діяльності з факторами, що визначають ці показники. До основних показників можна віднести дохід, прибуток, рентабельність, продуктивність праці, собівартість і т.ін. Перше поняття виробничої функції пов'язане з математичним моделюванням технологічної залежності між обсягом продукції, що випускається, і кількісними характеристиками витрат ресурсів. Звідси і назва функції «виробнича». Уперше така функція була побудована американськими дослідниками Коббом і Дугласом ще в 30-ті роки ХХ ст. за даними про функціонування обробної промисловості США протягом двадцяти років і є класичним прикладом економетричного моделювання.

Функція Кобба — Дугласа (CDPF) належить до найвідоміших виробничих функцій, що набули широкого застосування в економічних дослідженнях, особливо на макрорівні. Класична виробнича функція Кобба — Дугласа має вигляд

$$Y = aF^\alpha L^{1-\alpha}, \quad (3.2)$$

де Y — обсяг продукції; F — основний капітал; L — робоча сила. У цій функції параметри a , α і $1 - \alpha$ є невід’ємними. Таке твердження можна довести, якщо з виробничої функції виключити один з факторів. Для цього, поділивши ліву і праву частини залежності

$$Y = f(F, L) \quad (3.3)$$

на L , дістанемо функцію двох змінні

$$W = f(V) \quad (3.4),$$

де $w = \frac{Y}{L}$ — продуктивність праці; $v = \frac{F}{L}$ — фондоозброєність праці.

Нехай залежність між W і V має вигляд степеневі функції, тобто.. $W = aV^\alpha$.

Підставивши в цю функцію $w = \frac{Y}{L}$ і $v = \frac{F}{L}$, дістанемо: $\frac{Y}{L} = a \left(\frac{F}{L}\right)^\alpha$, або $Y = aF^\alpha L^{1-\alpha}$.

Сума параметрів або степінь однорідності, класичної функції Кобба — Дугласа дорівнює одиниці. А це означає, що при збільшенні обох виробничих ресурсів на одиницю обсяг продукції також збільшиться на одиницю. Отже, ефективність ресурсів у такому разі стала.

Практичні дослідження функції Кобба — Дугласа показали, що припущення про лінійну однорідність на практиці виконується рідко. Тому була запропонована виробнича функція загальнішого вигляду $Y = aF^\alpha L^\beta$.

Сума параметрів ($\alpha + \beta$) на відміну від попереднього випадку може бути як меншою, так і більшою від одиниці.

Якщо $(\alpha + \beta) > 1$, то темпи росту обсягу продукції вищі за темпи росту виробничих ресурсів, а якщо $(\alpha + \beta) < 1$, то, навпаки, темпи росту продукції нижчі за темпи росту ресурсів.

Припустимо, що рівень кожного виробничого ресурсу збільшився на $r\%$, тоді величини їх дорівнюватимуть $F(1 + \frac{r}{100})$ і $L(1 + \frac{r}{100})$.

Обсяг продукції на основі виробничої функції запишеться так:

$$Y = a \left[F \left(1 + \frac{r}{100} \right) \right]^\alpha \left[L \left(1 + \frac{r}{100} \right) \right]^\beta = a F^\alpha L^\beta \left(1 + \frac{r}{100} \right)^\alpha \left(1 + \frac{r}{100} \right)^\beta = Y' \left(1 + \frac{r}{100} \right)^{\alpha + \beta}.$$

Звідси при $\alpha + \beta > 1$ обсяг продукції зростає більш ніж на $r\%$; при $\alpha + \beta < 1$ — менш ніж на $r\%$; при $\alpha + \beta = 1$ продукція збільшиться на $r\%$. Визначивши окремі коефіцієнти еластичності для виробничої функції Кобба — Дугласа, дістанемо:

$$Y_F = \frac{\partial Y}{\partial F} = \alpha \frac{Y'}{F}; \quad Y_L = \frac{\partial Y}{\partial L} = \beta \frac{Y'}{L}.$$

Це означає, що граничний приріст продукції за рахунок приросту кожного ресурсу визначається як добуток коефіцієнта еластичності на середню ефективність ресурсу. Параметр a у функції Кобба — Дугласа залежить од вибраних одиниць вимірювання Y , F , L ; водночас числове значення цього параметра визначається також ефективністю виробничого процесу. У цьому можна переконатись, порівнявши дві виробничі функції, які відрізняються одна від одної лише значенням параметра a .

Для фіксованих значень F і L тій функції, в якій більше числове значення параметра a , відповідає більше значення Y . Отже, і виробничий процес, який описується цією функцією, буде ефективнішим. Другі похідні функції Кобба — Дугласа мають такий вигляд:

$$Y_{FF} = \frac{\alpha(\alpha-1)Y}{F^2}; \quad Y_{LL} = \frac{\beta(\beta-1)Y}{L^2}.$$

Узявши до уваги, що $0 < \alpha < 1$ і $0 < \beta < 1$, $Y_{FF} < 0$ і $Y_{LL} < 0$, дійдемо висновку: при збільшенні ресурсів граничний приріст обсягу продукції зменшуватиметься. Якщо обсяг продукції у функції Кобба — Дугласа вважати сталим (таким, що дорівнює const), то можна обчислити граничні норми заміщення ресурсів:

$$h = \frac{Y_F}{Y_L} = \frac{\alpha \frac{Y}{F}}{\beta \frac{Y}{L}} = \frac{\alpha}{\beta} \frac{L}{F}.$$

Звідси бачимо, що гранична норма заміщення ресурсів у функції Кобба — Дугласа визначається як добуток співвідношень величин ресурсів та їх коефіцієнтів еластичності. Швидкість зміни норми заміщення ресурсів у зв'язку зі зміною величини ресурсів обчислюється так:

$$\frac{\partial h}{\partial F} = \frac{\beta}{\alpha L}; \quad \frac{\partial h}{\partial L} = -\frac{\beta F}{\alpha L^2}.$$

Мірою швидкості зміни h є еластичність заміщення ресурсів F і L , що визначається як відношення зміни величини ресурсів до зміни величини h :

$$h_{F/A} = \frac{L/F \cdot \partial L/F}{\partial h/h} = \frac{h(Lh + F)}{FL \left(h \frac{\partial h}{\partial F} - \frac{\partial h}{\partial L} \right)} = 1.$$

Отже, еластичність заміщення в кожній точці кривої, що характеризує виробничу функцію Кобба — Дугласа, дорівнює одиниці.

Розглянемо тепер поведінку функції при зміні масштабу виробництва. Для цього припустимо, що витрати кожного ресурсу виробництва збільшилися в λ раз, тоді нове значення Y визначатиметься так:

$$Y = a(\lambda F)^\alpha (\lambda L)^\beta = \lambda^{\alpha + \beta} Y.$$

Степінь однорідності цієї функції дорівнює $\alpha + \beta$. Якщо $\alpha + \beta = 1$, то рівень ефективності ресурсів не

залежить від масштабів виробництва. Якщо $\alpha + \beta < 1$, то з розширенням масштабів виробництва середні витрати в розрахунку на одиницю продукції зменшуються, а при $\alpha + \beta > 1$ — збільшуються. Причому ці властивості не залежать від числових значень F і L і зберігають силу в кожній точці виробничої функції.

За припущення, що мета господарської діяльності — максимізація прибутку, можна проілюструвати інші властивості виробничої функції. Запишемо функцію прибутку:

$$\Pi = bY^{r+1} - wL - rF + \lambda [f(F, L) - Y].$$

Підприємець вибирає такі значення Y , L , F , які максимізують прибуток при обмеженнях, що накладаються виробничою функцією. Величини b , w , r — параметри функції прибутку, λ — множник Лагранжа. Якщо виробничий процес у даному співвідношенні описується функцією Кобба — Дугласа, то можна записати умови максимізації прибутку:

$$w = \frac{\lambda \beta Y}{L}; \quad r = \frac{\lambda \alpha Y}{F}; \quad \frac{w}{r} = \frac{\beta F}{\alpha L}, \quad \lambda = (r + 1) P$$

при $r \neq -1$, де $P = bY^r$.

Звідси обсяги ресурсів такі:

$$L = \frac{(r+1)P\beta Y}{w}; \quad F = \frac{(r+1)P\alpha Y}{r}.$$

У такому випадку максимальне значення випуску продукції, якщо $\alpha + \beta \neq 1$, можна записати так:

$$Y = aF^\alpha L^\beta = a \left[\frac{(r+1)P\alpha Y}{r} \right]^\alpha \left[\frac{(r+1)P\beta Y}{w} \right]^\beta.$$

При $r = 1$ згідно із записаними щойно умовами максимізації дістанемо:

$$F = \frac{w\alpha L}{\beta r}; \quad Y = a \left(\frac{w\alpha}{\beta r} \right)^\alpha L^{\alpha+\beta}.$$

Отже, необхідні умови для забезпечення максимізації прибутку дають змогу визначити відповідні витрати

робочої сили і основного капіталу. З розширенням масштабів виробництва ефективність витрат ресурсів падає, що відповідає максимізації прибутку в умовах досконалої конкуренції. Наведений приклад виробничої функції показує, що ця економетрична модель дає змогу досить широко проаналізувати виробничу діяльність, визначити шляхи її вдосконалення з метою підвищення ефективності. Обґрунтованість такого аналізу повністю залежить од вірогідності економетричної моделі, від того, наскільки вона адекватна реальному процесу.

Проблема побудови виробничої функції або інших технологічних взаємозалежностей у виробництві — класична проблема економетрії.

3.2. Випадкова складова економетричної моделі.

Розглянемо економетричну модель з двома змінними у загальному вигляді:

$$Y = f(X) + u, \quad (3.5)$$

де Y — залежна змінна; X — незалежна змінна; u — випадкова складова.

Це означає, що ми ідентифікували змінну X , яка впливає на змінну Y . Назвемо таку економетричну модель простою моделлю.

Як відомо, численні взаємозв'язки між економічними показниками не можна формалізувати лише на базі простої економетричної моделі. Наведені раніше приклади економетричних моделей показують, що вони описують вплив багатьох чинників на економічні процеси та явища. Причому для формалізації цих зв'язків може використовуватись не одне рівняння, а їх система.

На базі простої економетричної моделі розглянемо принципову структуру економетричної моделі та основні методи оцінювання її параметрів. Змістовне тлумачення взаємозв'язку між економічними показниками [модель (3.5)] має підказати його конкретну аналітичну форму. Але

оскільки одні й ті самі економічні умови можуть задовольняти різні функції, то краще звернутися до статистичного аналізу і з його допомогою зробити вибір серед можливих альтернативних варіантів.

Найпростішою є лінійна форма зв'язку між двома змінними:

$$Y = a_0 + a_1X, \quad (3.6)$$

де a_0 і a_1 — невідомі параметри, перший з яких визначає довжину відрізка, утворюваного перетином прямої з віссю ординат, а другий — тангенс кута нахилу цієї прямої до осі абсцис.

Можливі й інші форми залежностей між двома змінними, наприклад:

$$Y = a_0 e^{a_1 X}; \quad Y = a_0 X^{a_1}; \quad Y = a_0 + \frac{a_1}{X}.$$

Останнє з цих співвідношень є лінійним відносно $\frac{1}{X}$,

а перші два можна звести до лінійної форми для перетворених змінних, якщо взяти логарифми від виразів в обох частинах кожного з рівнянь:

$$\ln Y = \ln a_0 + a_1 X; \quad \ln Y = \ln a_0 + a_1 \ln X.$$

Навіть побіжне знайомство з економічними показниками, взаємозв'язок між якими вимірюється, показує, що окремі експериментальні значення залежної змінної не можуть міститися строго на прямій лінії чи на графіку функції іншої форми. Певна частина фактичних спостережень над змінною лежатиме вище або нижче від значень, обчислених згідно з вибраною функцією. Якщо фактичні значення залежної змінної містяться на значній відстані від обчислених з допомогою функції, то можна припустити, що формалізація залежності між економічними показниками на основі функції типу (3.5) чи якоїсь іншої функції не адекватна реальному процесу взаємозв'язків в економіці. Проте поняття «значна

відстань» не є конкретним, а тому не може бути критерієм для оцінювання адекватності моделі.

Щоб розв'язати цю задачу, до економетричної моделі вводять стохастичну складову, яка акумулює в собі всі відхилення фактичних спостережень змінної Y від обчислених згідно з моделлю.

Математичний аналіз цієї складової дасть змогу зробити висновок щодо того, чи можна вважати її стохастичною і чи містить вона систематичну частину відхилень, яка може бути зумовлена наявністю тих чи інших помилок у моделюванні.

Нехай вектор змінної Y описує витрати на споживання, а вектор X — величину доходу сім'ї. Очевидно, що для окремих груп сімей існує певна залежність між споживчими витратами і доходом сім'ї. Проте, як уже зазначалося, на розмір споживчих витрат крім доходу можуть впливати інші фактори, частина яких є випадковими. Ці фактори й зумовлюють відхилення фактичних витрат на споживання від обчислених, наприклад, на основі регресійної функції:

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X. \quad (3.7)$$

Знак « $\hat{}$ » означає оцінку відповідного параметра.

Наблизити обчислені значення до фактичних формально можна введенням до моделі стохастичної складової:

$$Y = a_0 + a_1 X + u \quad (3.8)$$

У моделі (3.8) символом u позначено змінну, яка може набувати додатних та від'ємних значень, оскільки вона вимірює відхилення витрат на споживання кожної окремої сім'ї від обчисленого значення згідно з (3.7).

Зауважимо, що в моделі (3.8) a_0 і a_1 — оцінювані параметри, а в моделі (3.7) \hat{a}_0 і \hat{a}_1 — їх оцінки.

Означення 1. Стохастичну складову u економетричної моделі називають помилкою (залишком, відхиленням).

Введення до моделі (3.8) стохастичної складової має три підстави, кожна з яких не виключає решти двох.

1. Величину витрат на споживання визначає не лише рівень доходів, а й інші об'єктивні чинники, наприклад розмір сім'ї, середній вік і т.ін.;

2. На величину споживання впливають випадкові фактори, наприклад схильність до ощадливості, стриманість чи навпаки — надмірність у витратах і т.ін.;

3. Частина факторів, які впливають на величину споживчих витрат, не оцінюються кількісно, вони не квантифікуються. Крім того, можлива помилка вимірювання змінних.

Отже, замість залежності

$$Y = f(x_1, x_2, x_3 \dots x_m) \quad (3.9),$$

де m — досить велике, розглядається модель з невеликим числом незалежних змінних, причому Y відіграє роль функції від найважливіших $X_j (j = \overline{1, m})$, тоді як чистий сумарний ефект від впливу всіх інших чинників відбиває змінна u . У крайньому разі, якщо лишається одна незалежна змінна, маємо:

$$Y = f(X, u) \quad (3.10)$$

У класичній лінійній економетричній моделі змінна u інтерпретується як випадкова змінна, яка має розподіл з математичним сподіванням, що дорівнює нулю, і сталою дисперсією σ_u^2 . Це дає змогу розглядати змінну u як стохастичне збурення (помилку, відхилення). З огляду на те, що u охоплює вплив багатьох чинників, які можна вважати незалежними, на підставі центральної граничної теореми теорії ймовірностей, доходимо висновку: стохастична

складова економетричної моделі розподілена за нормальним законом.

Щодо нашого прикладу, коли витрати на споживання перебувають у лінійній залежності від доходу сімей, а змінна u є випадковою складовою, можна

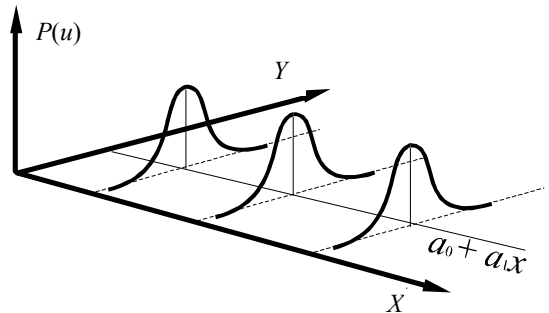


Рис. 3.1. Розділ залишків

графічно зобразити цю залежність за умови, що величина доходів упорядкована від меншого значення до більшого (рис.3.1).

Розподіл імовірностей $P(u)$ групуватиметься при цьому навколо лінії регресії $\hat{a}_0 + \hat{a}_1 X$. Можливо, у цьому прикладі доцільніше було б припускати, що дисперсія відхилення u зростає зі збільшенням доходу X . Цю особливість розглянемо пізніше, бо вона може бути притаманна й іншим економічним залежностям (наприклад, залежності заощаджень від доходу, дивідендів від прибутку і т.ін.).

В економетричній моделі (3.4) параметри \hat{a}_0 , \hat{a}_1 невідомі. На підставі вибірових спостережень X і Y потрібно не лише статистично оцінити ці параметри, а й перевірити виконання щодо них деяких гіпотез:

1) Чи можна вважати споживання пропорційними до доходу ($\hat{a}_0 = 0$)?

2) Чи буде гранична схильність до споживання (\hat{a}_1) більша за половину?

3) Чи виправдана для цієї вибіркової сукупності гіпотеза про сталу дисперсію залишків для всіх значень X ?

Усі наведені щойно запитання є типовими задачами економетричних досліджень, і основна мета економетрії — вивчити класичні методи розв’язування поставлених задач, а також опанувати нові методи розв’язування складніших економічних задач, які максимально наближені до реальних умов.

3.3. Оцінювання параметрів моделі методом найменших квадратів.

Звернемося до прикладу простої економетричної моделі, де потрібно кількісно оцінити зв’язок між витратами на споживання та доходами сім’ї. Щоб оцінити параметри моделі (3.4), необхідно сформулювати вихідну сукупність спостережень, кожна одиниця якої характеризуватиметься витратами на споживання і доходами сімей. Припустимо, що економетрична модель споживання будується для тієї групи людей, в якій зі збільшенням доходів зростають витрати на споживання, тобто модель має вигляд (3.9).

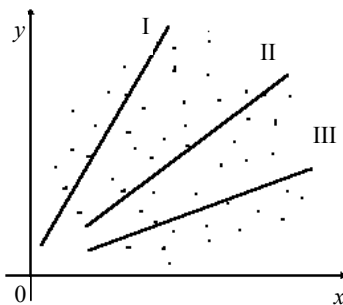


Рис.3. 2. Кореляційне поле

точок (рис.3.2).

На підставі гіпотези про лінійність зв’язку між витратами на споживання і доходом сімей (див. рис.3.2), через кореляційне поле точок можна провести безліч прямих ліній, які різняться між собою параметрами \hat{a}_0 і \hat{a}_1 . Так, якщо витрати на споживання описуватимуться

Зобразимо кожну пару спостережень у системі координат, де величина витрат на споживання відкладається на осі ординат, а доходів — на осі абсцис. У результаті дістанемо кореляційне поле

прямою I, то відхилення їх фактичних значень від розрахункових матимуть переважно знак «мінус». Якщо вони описуватимуться прямою III, то ці відхилення будуть переважно додатними, а якщо прямою II, то кількість від'ємних і додатних відхилень буде приблизно однаковою. Наявність серед відхилень переважно від'ємних чи додатних значень підтверджує, що вони мають не випадковий характер. А це означає: певна пряма лінія не адекватно описує фактичну залежність між витратами на споживання і доходом сімей. Звідси постає задача — застосувати метод найменших квадратів для оцінювання параметрів моделі, щоб відхилення фактичних витрат од розрахункових на основі прямої мали приблизно однакову суму від'ємних і додатних значень, а також були б найменшими. Останнє буде свідчити про те, що розрахункові значення витрат на споживання максимально наближені до фактичних, а це є гарантом вірогідності моделі.

Не доцільно знаходити параметри економетричної моделі, мінімізуючи суму лінійних відхилень фактичних витрат на споживання від розрахункових, бо вона може дорівнювати нулю, якщо сума від'ємних і додатних відхилень буде однаковою. Тому мінімізації підлягає сума квадратів відхилень, і величина її залежатиме безпосередньо від розсіювання точок навколо лінії регресії, а саме:

$$\min \left\{ \sum_{i=1}^n u_i^2 \right\} = f(\hat{a}_0, \hat{a}_1) \dots \dots \dots (3.11)$$

Принцип найменших квадратів полягає в знаходженні таких \hat{a}_0 і \hat{a}_1 , для яких $\sum_{i=1}^n u_i^2$ найменша.

Необхідна умова для цього — перетворення на нуль похідних цієї функції за кожним із параметрів \hat{a}_0 і \hat{a}_1 . Метод, який реалізує принцип найменших квадратів,

називається методом найменших квадратів (МНК).
Оскільки

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

$$\text{то } \begin{cases} \frac{\partial (\sum_{i=1}^n u_i^2)}{\partial a_0} = -2 \sum_{i=1}^n (y_i - \hat{a}_0 - \hat{a}_1 x_i) = 0; \\ \frac{\partial (\sum_{i=1}^n u_i^2)}{\partial a_1} = -2 \sum_{i=1}^n x_i (y_i - \hat{a}_0 - \hat{a}_1 x_i) = 0. \end{cases}$$

Виконавши елементарні перетворення, дістанемо систему нормальних рівнянь

$$\begin{cases} n\hat{a}_0 + \hat{a}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n y_i; \\ \hat{a}_0 \sum_{i=1}^n x_i + \hat{a}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i. \end{cases} \quad (3.12)$$

Підставимо в систему (3.12) значення $\sum_{i=1}^n x_i$, $\sum_{i=1}^n y_i$, $\sum_{i=1}^n x_i^2$, $\sum_{i=1}^n x_i y_i$, які можна дістати на підставі сукупності спостережень, і розв'яжемо її відносно невідомих параметрів \hat{a}_0 і \hat{a}_1 :

$$\hat{a}_1 = \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i - \sum_{i=1}^n x_i y_i}{\left(\sum_{i=1}^n x_i\right)^2 - n \sum_{i=1}^n x_i^2}; \quad (3.13)$$

$$\hat{a}_0 = \frac{\sum_{i=1}^n x_i \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i^2 \sum_{i=1}^n y_i}{\left(\sum_{i=1}^n x_i\right)^2 - n \sum_{i=1}^n x_i^2}. \quad (3.14)$$

Оскільки оцінки найменших квадратів такі, що лінія регресії обов'язково проходить через точку середніх значень (\bar{x}, \bar{y}) , то оцінки параметрів моделі можна знайти дещо інакше.

Поділивши перше рівняння системи (3.12) на n , дістанемо:

$$\bar{Y} = \hat{a}_0 + \hat{a}_1 \bar{X} \quad (3.15)$$

Віднімемо (3.15) від (3.7):

$$\hat{Y} - \bar{Y} = \hat{a}_1 (X - \bar{X}) \quad (3.16)$$

Нехай $Y_i - \bar{Y} = y_i$, $X_i - \bar{X} = x_i$ і $\hat{Y} - \bar{Y} = \hat{y}_i$, тоді

$$\hat{y}_i = \hat{a}_1 x_i \quad (3.17),$$

а відхилення фактичних значень від розрахункових будуть

$$\text{такі: } u_i = y_i - \hat{y}_i = y_i - \hat{a}_1 x_i. \quad (3.18)$$

Сума квадратів залишків при цьому: $\sum_{i=1}^n u_i^2 = \sum_{i=1}^n (y_i - \hat{a}_1 x_i)^2$.

Мінімізація цієї суми за невідомим параметром \hat{a}_1 дає співвідношення:

$$\hat{a}_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

Крім того, можна помітити, що $\frac{\partial^2 (\sum_{i=1}^n u_i^2)}{\partial \hat{a}_1^2} = 2 \sum_{i=1}^n x_i^2$, тобто

друга похідна за параметром \hat{a}_1 від суми квадратів відхилень додатна. Отже, знайдене значення \hat{a}_1 відповідає мінімуму суми квадратів відхилень.

Параметр \hat{a}_0 можна обчислити, використавши співвідношення (3.14):

$$\hat{a}_0 = \bar{Y} - \hat{a}_1 \bar{X}. \quad (3.19)$$

Співвідношення (3.19) можна було б дістати також, записавши друге рівняння системи (3.12) через відхилення кожної змінної від її середнього арифметичного значення, згадавши при цьому, що сума таких відхилень завжди дорівнює нулю.

Контрольні питання

1. Як визначається набір змінних для побудови економетричної моделі?
2. Наведіть приклади економетричних моделей.
3. Дайте тлумачення випадкової складової економетричної моделі.
4. Які методи застосовуються для оцінювання параметрів класичної регресійної моделі?
5. У чому сутність методу найменших квадратів (МНК)?
6. Визначіть дисперсію залишків економетричної моделі.

Тема 4. Проста лінійна регресія

План

- 4.1. Проста лінійна регресія.
 - 4.2. Побудова лінійної регресії.
 - 4.3. Коефіцієнт кореляції
 - 4.4. Статистичні критерії перевірки значущості оцінок параметрів моделі.
 - 4.5. Нелінійні моделі парної регресії й кореляції
- Література [4, 11, 14, 18, 22,].

4.1. Проста лінійна регресія.

Проста (парна) регресія являє собою регресію між двома змінними – Y і X , тобто модель виду:

$$Y = f(X) \dots \dots \dots (4.1)$$

де Y – залежна змінна (результативна ознака);

X – незалежна, або пояснююча, змінна (ознака-фактор).
Знак « \wedge » означає, що між змінними X й Y немає строгої функціональної залежності, тому практично в кожному

окремому випадку величина y складається із двох доданків:

$$y = \hat{y}_x + \varepsilon \quad (4.2),$$

де y – фактичне значення результативної ознаки;

\hat{y}_x – теоретичне значення результативної ознаки, знайдене виходячи з рівняння регресії; ε – випадкова величина, що характеризує відхилення реального значення результативної ознаки від теоретичного, знайденого по рівнянню регресії.

Випадкова величина ε включає вплив не врахованих у моделі факторів, випадкових помилок і особливостей виміру. Її присутність у моделі породжено трьома джерелами: специфікацією моделі, вибіркоvim характером вихідних даних, особливостями виміру змінних.

Від правильно обраної специфікації моделі залежить величина випадкових помилок: вони тим менше, ніж у більшій мірі теоретичні значення результативної ознаки \hat{y}_x , підходять до фактичним даним y . До помилок специфікації ставляться неправильний вибір тої або іншої математичної функції для \hat{y}_x й недооблік у рівнянні регресії якого-небудь істотного фактору, тобто використання парної регресії замість множинної. Поряд з помилками специфікації можуть мати місце помилки вибірки, які мають місце в чинність неоднорідності даних у вихідній статистичній сукупності, що, як правило, буває при вивченні економічних процесів. Якщо сукупність неоднорідна, то рівняння регресії не має практичного змісту. Для одержання гарного результату звичайно виключають із сукупності одиниці з аномальними

значеннями досліджуваних ознак. І в цьому випадку результати регресії являють собою вибіркові характеристики.

Використання тимчасової інформації також являє собою вибірку із усього безлічі хронологічних дат. Змінивши часовий інтервал, можна одержати інші результати регресії.

Найбільшу небезпеку в практичному використанні методів регресії представляють помилки виміру. Якщо помилки специфікації можна зменшити, змінюючи форму моделі (вид математичної формули), а помилки вибірки - збільшуючи обсяг вихідних даних, то помилки виміру практично зводять нанівець всі зусилля по кількісній оцінці зв'язку між ознаками.

Особливо велика роль помилок виміру при дослідженні на макрорівні. Так, у дослідженнях попиту й споживання в якості пояснюючої змінної широко використовується «дохід на душу населення». Разом з тим, статистичний вимір величини доходу сполучено з рядом труднощів і не позбавлено можливих помилок, наприклад, у результаті наявності схованих доходів.

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

У парній регресії вибір виду математичної функції

$\hat{y}_x = f(x)$ може бути здійснений трьома методами:

- 1) графічним;
- 2) аналітичним, тобто виходячи з теорії досліджуваного взаємозв'язку;
- 3) експериментальним.

При вивченні залежності між двома ознаками графічний метод підбора виду рівняння регресії досить наочний. Він заснований на поле кореляції. Значний

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

При обробці інформації на комп'ютері вибір виду рівняння регресії звичайно здійснюється експериментальним методом, тобто шляхом порівняння величини залишкової дисперсії $\sigma_{\text{ост}}^2$, розрахованої при різних моделях.

Якщо рівняння регресії проходить через всі крапки кореляційного поля, що можливо тільки при функціональному зв'язку, коли всі крапки лежать на лінії регресії

$$\hat{y}_x = f(x) \tag{4.4}$$

то фактичні значення результативної ознаки збігаються з теоретичними

$$y = \hat{y}_x \tag{4.5}$$

тобто вони повністю обумовлені впливом фактору x . У цьому випадку залишкова дисперсія $\sigma_{\text{ост}}^2 = 0$.

У практичних дослідженнях, як правило, має місце деяке розсіювання крапок щодо лінії регресії. Воно обумовлено впливом інших, факторів, що ураховуються не в рівнянні регресії. Іншими словами, мають місце відхилення фактичних даних від теоретичних $(y - \hat{y}_x)$.

Величина цих відхилень і лежить в основі розрахунку залишкової дисперсії:

$$\sigma_{\text{ост}}^2 = \frac{1}{n} \sum (y - \hat{y}_x)^2 \tag{4.6}$$

Чим менше величина залишкової дисперсії, тим менше вплив факторів, що враховуються не в рівнянні

регресії, і тем краще рівняння регресії підходить до вихідним даним.

Уважається, що число спостережень повинне в 7-8 разів перевищувати число параметрів, що розраховуються, при X змінній. Це означає, що шукати лінійну регресію, маючи менш 7 спостережень, взагалі не має змісту. Якщо вид функції ускладнюється, то потрібне збільшення обсягу спостережень, тому що кожний параметр при X повинен розраховуватися хоча б по 7 спостереженням. Виходить, якщо ми вибираємо параболу другого ступеня

$$\hat{y}_x = a + b \cdot x + c \cdot x^2 \quad (4.7)$$

то потрібен обсяг інформації вже не менш 14 спостережень.

Розглянемо найпростішу модель парної регресії - лінійну регресію.

Лінійна регресія знаходить широке застосування в економетрії через чітку економічну інтерпретацію її параметрів.

Лінійна регресія зводиться до знаходження рівняння виду:

$$\hat{y}_x = a + b \cdot x \quad (4.8)$$

або

$$y = a + b \cdot x + \varepsilon \quad (4.9)$$

Рівняння виду (4.8) дозволяє за заданим значенням фактору X знаходити теоретичні значення результативної ознаки, підставляючи в нього фактичні значення фактору x .

Побудова лінійної регресії зводиться до оцінки її параметрів - a і b . Класичний підхід до оцінювання параметрів лінійної регресії заснований на методі найменших квадратів (МНК). МНК дозволяє одержати такі оцінки параметрів a і b , при яких сума квадратів

відхилень фактичних значень результативної ознаки y від теоретичних \hat{y}_x мінімальна:

$$\sum_{i=1}^n (y_i - \hat{y}_{x_i})^2 = \sum_{i=1}^n \varepsilon_i^2 \rightarrow \min \dots \dots \dots (4.10)$$

Після нескладних перетворень, отримаємо систему лінійних рівнянь для оцінки параметрів a и b :

$$\begin{cases} a \cdot n + b \cdot \sum x = \sum y; & \dots \dots \dots (4.11) \\ a \cdot \sum x + b \cdot \sum x^2 = \sum x \cdot y. \end{cases}$$

Вирішуючи систему рівнянь (4.11), знайдемо шукані оцінки параметрів a і b . Можна скористатися наступними готовими формулами, які випливають безпосередньо з рішення системи (4.4):

$$a = \bar{y} - b \cdot \bar{x}, b = \frac{\text{cov}(x, y)}{\sigma_x^2},$$

де $\text{cov}(x, y) = \overline{y \cdot x} - \bar{y} \cdot \bar{x}$ – коваріація ознак x і

y , $\sigma_x^2 = \overline{x^2} - \bar{x}^2$ – дисперсія ознаки x ,

$$\bar{x} = \frac{1}{n} \sum x, \bar{y} = \frac{1}{n} \sum y, \overline{y \cdot x} = \frac{1}{n} \sum y \cdot x,$$

Коваріація - числова характеристика спільного розподілу двох випадкових величин, рівна математичному очікуванню добутку відхилень цих випадкових величин від їхніх математичних очікувань. Дисперсія - характеристика випадкової величини, обумовлена як математичне очікування квадрата відхилення випадкової величини від її математичного очікування. Математичне очікування - сума добутків значень випадкової величини на відповідні ймовірності.

Параметр b називається коефіцієнтом регресії. Його величина показує середню зміну результату зі зміною фактору на одну одиницю. Можливість чіткої економічної інтерпретації коефіцієнта регресії зробила лінійне рівняння регресії досить розповсюдженим в економетричних дослідженнях.

Формально a – значення y при $x = 0$. Якщо ознака-фактор x не може мати нульового значення, то вищевказане трактування вільного члена a не має змісту, тобто параметр a може не мати економічного втримування.

4.2. Побудова лінійної регресії.

Для того щоб кількісно описати зв'язок між кількома або багатьма змінними, одна з яких є залежною, інші — незалежними змінними, необхідно розглянути лінійну економетричну модель, яка базується на регресійному аналізі.

У загальному вигляді цю модель можна записати так:

$$Y = f(X_1, X_2, \dots, X_m, u), \dots\dots\dots(4.12)$$

де Y — залежна змінна;

$X_j, (j = \overline{1, m})$ — незалежні змінні;

u — стохастична складова.

Залежна змінна Y називається також пояснюваною, ендогенною змінною, незалежні змінні X_j — пояснюючими, предетермінованими, екзогенними змінними.

Аналітична форма загальної лінійної економетричної моделі:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_m X_m + u \quad (4.13),$$

де $a_j (j = \overline{0, m})$ — параметри моделі.

В матричній формі економетрична модель має такий вигляд:

$$Y = XA + u \quad (4.14),$$

X — матриця незалежних змінних; A — вектор оцінок параметрів моделі; u — вектор залишків.

Щоб оцінити параметри моделі на основі методу 1МНК, необхідно дотримуватися таких передумов (гіпотез):

1) математичне сподівання залишків має дорівнювати нулю, тобто

$$M(u) = 0$$

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

$$M(uu') = \sigma^2 E ;$$

3) незалежні змінні моделі не зв'язані із залишками, тобто

$$M(X'u) = 0$$

4) незалежні змінні моделі створюють лінійно-незалежну систему векторів, тобто

$$(X'_k X_j) = 0, \quad k \neq j; \quad k = \overline{1, m}$$

$$(X'_k X_j) = 1, \quad k = j; \quad j = \overline{1, m}$$

Оператор оцінювання параметрів моделі на основі 1МНК: $A = (X' X)^{-1} X' Y$.

Неважно довести, що оцінки \hat{A} , які можна отримати на основі оператора оцінювання 1МНК, мінімізують суму квадратів залишків u . При цьому значення вектора \hat{A} є розв'язком нормальної системи рівнянь:

$$(X' X) \hat{A} = X' Y.$$

Якщо незалежні змінні в матриці X взяті як відхилення кожного значення від своєї середньої, то матрицю $X'X$ називають матрицею моментів. Числа, що стоять на її головній діагоналі, характеризують величину дисперсій незалежних змінних, інші елементи відповідають взаємним коваріаціям.

Оцінки параметрів загальної економетричної моделі повинні мати такі *властивості*:

- 1) незміщеності;
- 2) обгрунтованості;
- 3) ефективності;
- 4) інваріантності.

Оцінка параметра моделі буде *незміщеною*, коли дотримується рівність:

$$M(\hat{A}) = A.$$

Якщо ця рівність не дотримується, то різниця $M(\hat{A}) - A = Q$ називається зміщенням оцінки.

Оцінки \hat{A} параметрів A називаються *ефективними*, коли вони мають найменшу дисперсію. Якщо функція $g(\hat{A})$ відповідає функції $g(A)$, то оцінки \hat{A} параметрів A є інваріантними.

Приклад 1.

Побудувати економетричну модель, яка характеризує залежність між витратами на харчування, загальними затратами та складом сім'ї на основі даних, наведених у табл. 4.1. Проаналізувати зв'язок, визначений на основі побудованої моделі.

Таблиця 4.1.

№ з / п	Витрати на харчування	Загальні затрати	Склад сім'ї
<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>
1	20	45	1,5
2	32	75	1,6
3	48	125	1,9
4	65	223	1,8
5	45	92	3,4
6	64	146	3,6
7	79	227	3,5
8	104	358	5,5
9	68	135	5,4
10	93	218	5,4
11	117	331	5,3
12	145	490	8,5
13	91	175	8,3
14	131	205	8,1
15	167	468	7,3
16	195	749	8,4

Розв'язання:

1. Ідентифікуємо змінні моделі:

Y — витрати на харчування (залежна змінна);

X_1 — загальні витрати (незалежна змінна);

X_2 — розмір сім'ї (незалежна змінна);

u — залишки (стохастична складова).

Загальний вигляд моделі:

$$Y = f(X_1, X_2, u) .$$

2. Специфікуємо модель, тобто в даному випадку визначимо її аналітичну форму:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + u;$$

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X_1 + \hat{a}_2 X_2 .$$

3. Оцінимо параметри моделі на основі методу ІМНК, попередньо висунувши гіпотезу, що всі чотири передумови для його застосування дотримані.

Оператор оцінювання на основі ІМНК:

$$\hat{A} = (X' X)^{-1} X' Y .$$

У даному операторі матриця X характеризує всі незалежні змінні моделі. Оскільки економетрична модель має вільний член \hat{a}_0 , для якого всі $X_i = 1$, то матрицю X треба доповнити першим стовпцем, в якому всі шістнадцять членів є одиницями. X' — матриця, транспонована до матриці X , а вектор Y — вектор залежної змінної.

$$X = \begin{pmatrix} 1 & 45 & 1,5 \\ 1 & 75 & 1,6 \\ 1 & 125 & 1,9 \\ \\ 1 & 229 & 1,8 \\ 1 & 92 & 3,4 \\ 1 & 146 & 3,6 \\ \\ 1 & 227 & 3,5 \\ 1 & 358 & 5,5 \\ 1 & 135 & 5,4 \\ \\ 1 & 218 & 5,4 \\ 1 & 331 & 5,3 \\ 1 & 490 & 8,5 \\ \\ 1 & 175 & 8,9 \\ 1 & 305 & 8,1 \\ 1 & 468 & 7,3 \\ \\ 1 & 749 & 8,4 \end{pmatrix}$$

$$X'X = \begin{pmatrix} 16 & 4168 & 79,5 \\ 4168 & 1604274 & 25895,6 \\ 79,5 & 25895,6 & 495,7 \end{pmatrix};$$

$$X'Y = \begin{pmatrix} 1464 \\ 512156 \\ 8916,8 \end{pmatrix};$$

$$(X'X)^{-1} = \begin{pmatrix} 0,307813 & -0,000018 & -0,0484215 \\ -0,000018 & 0,000004 & -0,0002049 \\ -0,04842 & -0,0002 & 0,02048829 \end{pmatrix}$$

Підставимо отримані значення оберненої матриці $(X'X)^{-1}$ і добуток матриць $X'Y$ в оператор оцінювання і визначимо оцінки параметрів моделі:

$$\hat{A} = (X'X)^{-1} X'Y = \begin{pmatrix} 9,595983 \\ 0,183683 \\ 6,853759 \end{pmatrix};$$

Таким чином, $\hat{a}_0 = 9,596$; $\hat{a}_1 = 0,1837$; $\hat{a}_2 = 6,854$.

Звідси економетрична модель має вигляд:

$$\hat{Y} = 9,596 + 0,1837 X_1 + 6,854 X_2$$

5. Розрахуємо дисперсії залишків та залежної змінної Y :

$$\sigma_U^2 = \frac{1}{n-m} u'u = \frac{1237,5704}{13} = 95,19772 ;$$

$$\sigma_Y^2 = \frac{1}{n-1} Y'Y = \frac{36518,0}{15} = 2434,533 .$$

Таблиця 4.2.

№ з / п	\hat{Y}	u	u^2	$Y - \bar{Y}$	$(Y - \bar{Y})^2$

1	28,1424	-8,1424	66,2979	-71,5000	5112,2500
2	34,3382	-2,3382	5,4673	-59,5000	3540,2500
3	45,5785	2,4215	5,8637	-43,5000	1892,2500
4	63,9961	1,0039	1,0077	-26,5000	702,2500
5	49,7976	-4,7976	23,0169	-46,5000	2162,2500
6	61,0872	2,9128	8,4843	-27,5000	756,2500
7	75,2802	3,7198	13,8372	-12,5000	156,2500
8	113,0501	-9,0501	81,9051	12,5000	156,2500
9	71,4035	-3,4035	11,5837	-23,5000	552,2500
10	86,6492	6,3508	40,3332	1,5000	2,2500
11	106,7200	10,2800	105,6793	23,5000	650,2500
12	157,8576	-12,8576	165,3171	53,5000	2862,2500
13	98,6267	-7,6267	58,1665	-0,5000	0,2500
14	121,1347	9,8653	97,3237	39,5000	1560,2500
15	145,5920	21,4080	458,3011	75,5000	5700,2500
16	204,7461	-9,7461	94,9859	103,5000	10712,25
Всью- го		0,0000	1237,5704		36518,00

6. Визначимо матрицю коваріацій оцінок параметрів моделі:

$$\text{var}(\hat{A}) = \sigma_v^2 (X'X)^{-1} = \begin{pmatrix} 29,303 & -0,00172 & -4,609614 \\ -0,00172 & 0,0003787 & -0,0195072 \\ -4,609614 & -0,0195072 & 1,95043834 \end{pmatrix}$$

Діагональні елементи цієї матриці характеризують дисперсії оцінок параметрів моделі:

$$\sigma_{\hat{a}_0}^2 = 29,303 ;$$

$$\sigma_{\hat{a}_1}^2 = 0,0003787 ;$$

$$\sigma_{\hat{a}_2}^2 = 1,95043834 .$$

Інші елементи даної матриці визначають рівень коваріації між оцінками параметрів моделі.

7. Знайдемо стандартні помилки оцінок параметрів:

$$S_{\hat{a}_0} = \sqrt{\sigma_{\hat{a}_0}^2} = \sqrt{29,303} = 5,413 ;$$

$$S_{\hat{a}_1} = \sqrt{\sigma_{\hat{a}_1}^2} = \sqrt{0,0003787} = 0,01946 ;$$

$$S_{\hat{a}_2} = \sqrt{\sigma_{\hat{a}_2}^2} = \sqrt{1,95043834} = 1,39658 .$$

Порівняємо стандартні помилки оцінок параметрів моделі з величиною оцінки. Так, співвідношення стандартної помилки й абсолютного значення параметра \hat{a}_0 становить **56%**, параметра \hat{a}_1 — **10,6%**, параметра \hat{a}_2 — **20,4%**. Перше й третє співвідношення свідчать про те, що оцінки параметрів моделі \hat{a}_0 і \hat{a}_2 можуть мати зміщення, а друге співвідношення підтверджує незміщеність оцінки параметра \hat{a}_1 .

8. Дамо змістовне тлумачення параметрів моделі.

Оцінка параметра \hat{a}_1 характеризує граничну зміну величини витрат на харчування залежно від зміни загальних затрат на одиницю. Тобто, якщо загальні затрати сім'ї зростуть на одиницю, то витрати на харчування в них збільшаться на **0,18** одиниці при незмінному складі сім'ї.

Оцінка параметра \hat{a}_2 характеризує граничне зростання витрат на харчування при збільшенні сім'ї на одного члена. Так, якщо склад сім'ї збагатиться ще одним членом, то витрати на харчування зростуть на **6,854** одиниці при незмінній величині доходу.

4.3. Коефіцієнт кореляції

Рівняння регресії завжди доповнюється показником тісноти зв'язку. При використанні лінійної регресії як такий показник виступає *лінійний коефіцієнт кореляції* r_{xy} , якому можна розрахувати по формулі:

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y}. \quad (4.15)$$

Лінійний коефіцієнт кореляції перебуває в межах: $-1 \leq r_{xy} \leq 1$. Чим ближче абсолютне значення r_{xy} до одиниці, тим сильніше лінійний зв'язок між факторами (при $r_{xy} = \pm 1$ маємо строгу функціональну залежність). Але варто мати на увазі, що близькість абсолютної величини лінійного коефіцієнта кореляції до нуля ще не означає відсутності зв'язку між ознаками. При іншій (нелінійній) специфікації моделі зв'язок між ознаками може виявитися досить тісною.

Для оцінки якості підбора лінійної функції розраховується квадрат лінійного коефіцієнта кореляції r_{xy}^2 , називаний *коефіцієнтом детермінації*. Коефіцієнт

детермінації характеризує частку дисперсії результативної ознаки Y , що пояснюється регресією, у загальній дисперсії результативної ознаки:

$$r_{xy}^2 = 1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2}, \quad (4.16)$$

$$\text{де } \sigma_{\text{ост}}^2 = \frac{1}{n} \sum (y - \hat{y}_x)^2, \quad \sigma_y^2 = \frac{1}{n} \sum (y - \bar{y})^2 = \overline{y^2} - \bar{y}^2.$$

Відповідно величина $1 - r_{xy}^2$ характеризує частку дисперсії Y , викликану впливом інших, не врахованих у моделі, факторів.

4.4. Статистичні критерії перевірки значущості оцінок параметрів моделі.

Після того як знайдене рівняння лінійної регресії, проводиться оцінка значимості (адекватності) як рівняння в цілому, так і окремих його параметрів.

Перевірити значимість рівняння регресії - значить установити, чи відповідає математична модель, що виражає залежність між змінними, експериментальним даним і чи досить включених у рівняння пояснюючих змінних (однієї або декількох) для опису залежної змінної.

Оцінка значимості рівняння регресії в цілому провадиться на основі F -критерію Фішера, якому передую дисперсійний аналіз. У математичній статистиці дисперсійний аналіз розглядається як самостійний інструмент статистичного аналізу. В економетрії він застосовується як допоміжні кошти для вивчення якості регресійної моделі.

Відповідно до основної ідеї дисперсійного аналізу, загальна сума квадратів відхилень змінної Y від середнього значення \bar{Y} розкладається на дві частини – «пояснену» і «непояснену»:

$$\sum (y - \bar{y})^2 = \sum (\hat{y}_x - \bar{y})^2 + \sum (y - \hat{y}_x)^2 \quad (4.17),$$

де $\sum (y - \bar{y})^2$ – загальна сума квадратів відхилень;

$\sum (\hat{y}_x - \bar{y})^2$ – сума квадратів відхилень, пояснена регресією (або факторна сума квадратів відхилень);

$\sum (y - \hat{y}_x)^2$ – залишкова сума квадратів відхилень, що характеризує вплив неврахованих у моделі факторів.

Визначення дисперсії на один ступінь волі приводить дисперсії до порівнянного виду. Зіставляючи факторну й залишкову дисперсії розраховуючи на один ступінь волі, одержимо величину F -критерію Фішера:

$$F = \frac{S_{\text{факт}}^2}{S_{\text{ост}}^2}. \quad (4.18)$$

Фактичне значення F -критерію Фішера (4.23) рівняється з табличним значенням $F_{\text{табл}}(\alpha; k_1; k_2)$ при рівні значимості α й ступенях волі $k_1 = m$ й $k_2 = n - m - 1$ (4.24). При цьому, якщо фактичне значення F -критерію більше табличного, то зізнається статистична значимість рівняння в цілому.

Для парної лінійної регресії $m = 1$, тому

$$F = \frac{S_{\text{факт}}^2}{S_{\text{ост}}^2} = \frac{\sum (\hat{y}_x - \bar{y})^2}{\sum (y - \hat{y}_x)^2} \cdot (n - 2). \quad (4.19)$$

Величина F -критерію пов'язана з коефіцієнтом детермінації r_{xy}^2 , і її можна розрахувати по наступній формулі:

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n - 2). \quad (4.20)$$

Для оцінки істотності коефіцієнта регресії його величина рівняється з його стандартною помилкою, тобто визначається фактичне значення t -критерію Стюдента:

$$t_b = \frac{b}{m_b},$$

яке потім рівняється з табличним значенням при

певному рівні значимості α й числі ступенів волі $(n - 2)$. Довірчий інтервал для коефіцієнта регресії

визначається як $b \pm t_{\text{табл}} \cdot m_b$. Оскільки знак коефіцієнта регресії вказує на ріст результативної ознаки y при збільшенні ознаки-фактору x ($b > 0$), зменшення результативної ознаки при збільшенні ознаки-фактору ($b < 0$) або його незалежність від незалежної змінної ($b = 0$), -це границі довірчого інтервалу для коефіцієнта регресії не повинні містити суперечливих результатів, наприклад, $-1,5 \leq b \leq 0,8$. Такого роду запис укажує, що шире значення коефіцієнта регресії одночасно містить позитивні й негативні величини й навіть нуль, чого не може бути.

Значимість лінійного коефіцієнта кореляції перевіряється на основі величини помилки коефіцієнта

$$\text{кореляції } m_r: m_r = \sqrt{\frac{1 - r^2}{n - 2}}.$$

Фактичне значення t -критерію Стьюдента визначається як $t_r = \frac{r}{m_r}$. Існує зв'язок між t -критерієм

Стьюдента й F -критерієм Фішера: $t_b = t_r = \sqrt{F}$

4.5. Нелінійні моделі парної регресії й кореляції.

Якщо між економічними явищами існують нелінійні співвідношення, то вони виражаються за допомогою відповідних нелінійних функцій. Розрізняють два класи нелінійних регресій:

1. Регресії, нелінійні щодо включених в аналіз пояснюючих змінних, але лінійні по оцінюваних параметрах, наприклад

– поліноми різних ступенів – $\hat{y}_x = a + b \cdot x + c \cdot x^2$,

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

– рівностороння гіпербола – $\hat{y}_x = a + b/x$;

– напівлогарифмічна функція – $\hat{y}_x = a + b \cdot \ln x$.

2. Регресії, нелінійні по оцінюваних параметрах, наприклад

– статична – $\hat{y}_x = a \cdot x^b$;

– показникова – $\hat{y}_x = a \cdot b^x$;

– експонентна – $\hat{y}_x = e^{a+b \cdot x}$.

Регресії нелінійні по включеним змінним приводяться до лінійного виду простою заміною змінних, а подальша оцінка параметрів провадиться за допомогою методу найменших квадратів. Розглянемо деякі функції.

Парабола другого ступеня $\hat{y}_x = a + b \cdot x + c \cdot x^2$ приводиться до лінійного виду за допомогою заміни: $x = x_1$, $x^2 = x_2$.

У результаті приходимо до двофакторному рівняння $\hat{y}_x = a + b \cdot x_1 + c \cdot x_2$. Парабола другого ступеня звичайно застосовується у випадках, коли для певного інтервалу значень фактору міняється характер зв'язку розглянутих ознак: пряма зв'язок міняється на зворотну або зворотна на пряму.

Рівностороння гіпербола $\hat{y}_x = a + b/x$ може бути використана для характеристики зв'язку питомих видатків сировини, матеріалів, палива від обсягу випускається продукції, що, часу обігу товарів від величини товарообігу, відсотка приросту заробітної плати від рівня безробіття (наприклад, крива А. В. Філіпса), видатків на непродовольчі товари від доходів або загальної суми видатків (наприклад, криві Енгеля) і в інших випадках. Гіпербола приводиться до лінійного рівняння простою заміною: $z = 1/x$.

Аналогічно приводяться до лінійного виду залежності $\hat{y}_x = a + b \cdot \ln x$ (4.44), $\hat{y}_x = a + b \cdot \sqrt{x}$ та інші.

Контрольні питання

1. Назвіть основні складові елементи моделі парної лінійної регресії?
2. Що таке випадковий член? Назвіть основні причини існування випадкового члена?
3. Як визначається залишки в моделі парної регресії?
4. Назвіть основні напрямки інтерпретації рівняння регресії?
5. За допомогою якого критерія здійснюється перевірка моделі на адекватність?

Тема 5. Багатофакторна регресія

План

5.1. Відбір факторів при побудові рівняння множинної регресії.

5.2. Метод найменших квадратів (МНК) для багатофакторної регресії. Властивості оцінок на основі МНК.

5.3. Перевірка істотності факторів і показники якості регресії.

5.4. Узагальнений метод найменших квадратів (УМНК).

Література [4, 5, 18, 22, 25].

5.1. Відбір факторів при побудові рівняння множинної регресії.

Парна регресія може дати гарний результат при моделюванні, якщо впливом інших факторів, що впливають на об'єкт дослідження, можна зневажити. Якщо ж цим впливом зневажити не можна, то в цьому випадку варто спробувати виявити вплив інших факторів, увівши їх у модель, тобто побудувати рівняння множинної регресії

$$y = f(x_1, x_2, \dots, x_m), \quad (5.1)$$

де y – залежна змінна (результативна ознака),

x_i – незалежні, або пояснюючі, змінні (ознаки-фактори).

Множинна регресія широко використовується в рішенні проблем попиту, прибутковості акцій, при вивченні функції витрат виробництва, у макроекономічних розрахунках і цілому ряді інших питань економетрії. У цей час множинна регресія - один з найпоширеніших методів в економетрії. Основна мета множинної регресії - побудувати модель із більшим числом факторів,

визначивши при цьому вплив кожного з них окремо, а також сукупний їхній вплив на моделюємий показник.

Побудова рівняння множинної регресії починається з рішення питання про специфікацію моделі. Він містить у собі два кола питань: відбір факторів і вибір виду рівняння регресії.

Включення в рівняння множинної регресії того або іншого набору факторів зв'язано насамперед з поданням дослідника про природу взаємозв'язку моделюємого показника з іншими економічними явищами. Фактори, що включаються в множинну регресію, повинні відповідати наступним вимогам.

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

2. Фактори не повинні бути інтеркоррельовані й тим більше перебувати в точному функціональному зв'язку.

Включення в модель факторів з високої інтеркорреляцією, може привести до небажаних наслідків - система нормальних рівнянь може виявитися погано обумовленою й спричинити нестійкість і ненадійність оцінок коефіцієнтів регресії.

Якщо між факторами існує висока кореляція, то не можна визначити їхній ізольований вплив на результативний показник і параметри рівняння регресії виявляються не інтерпретованими і включаються в множинну регресію, повинні пояснити варіацію незалежної змінної. Якщо будується модель із набором m факторів, то для неї розраховується показник детермінації

R^2 , що фіксує частку поясненої варіації результативної ознаки за рахунок розглянутих у регресії m факторів.

Вплив інших, не врахованих у моделі факторів, оцінюється як $1 - R^2$ з відповідною залишковою дисперсією S^2 .

При додатковому включенні в регресію $m + 1$ фактору коефіцієнт детермінації повинен зростати, а залишкова дисперсія зменшуватися:

$$R_{m+1}^2 \geq R_m^2 \quad \text{и.} \quad S_{m+1}^2 \leq S_m^2$$

Якщо ж цього не відбуваються й дані показники практично не відрізняються друг від друга, то фактор, що X_{m+1} включається в аналіз, не поліпшує модель і практично є зайвим чинником.

Насичення моделі зайвими факторами не тільки не знижує величину залишкової дисперсії й не збільшує показник детермінації, але й приводить до статистичної незначимості параметрів регресії за критерієм Стьюдента.

Таким чином, хоча теоретично регресійна модель дозволяє врахувати будь-яке число факторів, практично в цьому немає необхідності. Відбір факторів провадиться на основі якісного теоретико-економічного аналізу. Однак теоретичний аналіз часто не дозволяє однозначно відповісти на запитання про кількісний взаємозв'язок розглянутих ознак і доцільності включення фактору в модель. Тому відбір факторів звичайно здійснюється у дві стадії: на першій підбираються фактори виходячи із сутності проблеми; на другий - на основі матриці показників кореляції визначають статистики для параметрів регресії.

Коефіцієнти інтеркореляції (тобто кореляції між пояснюючими змінними) дозволяють виключати з моделі дублюючі фактори. Уважається, що дві змінні явно колінеарні, тобто перебувають між собою в лінійній залежності, якщо $r_{x_i x_j} \geq 0,7$. Якщо фактори явно колінеарні, то вони дублюють один одного й один з них

рекомендується виключити з регресії. Перевага при цьому віддається не фактору, більш тісно пов'язаному з результатом, а тому фактору, що при досить тісному зв'язку з результатом має найменшу тісноту зв'язку з іншими факторами. У цій вимозі проявляється специфіка множинної регресії як методу дослідження комплексного впливу факторів в умовах їхньої незалежності друг від друга.

По величині парних коефіцієнтів кореляції виявляється лише явна колінеарність факторів. Найбільші труднощі у використанні апарата множинної регресії виникають при наявності мультиколінеарності факторів, коли більш ніж два фактори зв'язані між собою лінійною залежністю, тобто має місце сукупний вплив факторів один на одного. Наявність мультиколінеарності факторів може означати, що деякі фактори будуть завжди діяти в унісон. У результаті варіація у вихідних даних перестає бути повністю незалежною й не можна оцінити вплив кожного фактору окремо.

Включення в модель мультиколінеарних факторів небажано в чинність наступних наслідків:

1. Утрудняється інтерпретація параметрів множинної регресії як характеристик дії факторів в «чистому» вигляді, тому що фактори корельовані; параметри лінійної регресії втрачають економічний зміст.

2. Оцінки параметрів ненадійні, виявляють більші стандартні помилки й міняються зі зміною обсягу спостережень (не тільки по величині, але й за знаком), що робить модель непридатною для аналізу й прогнозування.

Для оцінки мультиколінеарності факторів може використовуватися визначник матриці парних коефіцієнтів кореляції між факторами. Чим ближче до нуля визначник матриці межфакторної кореляції, тим сильніше мультиколінеарність факторів і ненадійніше результати

множинної регресії. І, навпаки, чим ближче до одиниці визначник матриці межфакторної кореляції, тим менше мультиколінеарність факторів.

Існує ряд підходів подолання сильної межфакторної кореляції. Найпростіший шлях усунення мультиколінеарності складається у виключенні з моделі одного або декількох факторів. Інший підхід пов'язаний з перетворенням факторів, при якому зменшується кореляція між ними.

Одним зі шляхів обліку внутрішньої кореляції факторів є перехід до сполучених рівнянь регресії, тобто до рівнянь, які відображають не тільки вплив факторів, але і їхня взаємодія. Так, якщо $y = f(x_1, x_2, x_3)$ (5.2), те можлива побудова наступного сполученого рівняння:

$$y = a + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + \varepsilon \quad (5.3).$$

Розглянуте рівняння включає взаємодія першого порядку (взаємодія двох факторів). Можливе включення в модель і взаємодій більше високого порядку, якщо буде доведена їхня статистична значимість по F -критерії Фішера, але, як правило, взаємодії третього й більше високих порядків виявляються статистично незначущими.

Відбір факторів, що включаються в регресію, є одним з найважливіших етапів практичного використання методів регресії. Підходи до відбору факторів на основі показників кореляції можуть бути різні. Вони приводять побудову рівняння множинної регресії відповідно до різних методик. Залежно від того, яка методика побудови рівняння регресії прийнята, міняється алгоритм її рішення на ЕОМ.

Найбільш широке застосування одержали наступні методи побудови рівняння множинної регресії:

1. Метод виключення - відсівання факторів з повного його набору.
2. Метод включення - додаткове введення фактору.
3. Кроковий регресійний аналіз - виключення раніше уведеного фактору.

При відборі факторів також рекомендується користуватися наступним правилом: число факторів, що включаються, звичайно в 6-7 разів менше обсягу сукупності, по якій будується регресія. Якщо це співвідношення порушене, то число ступенів волі залишкової дисперсії дуже мало. Це приводить до того, що параметри рівняння регресії виявляються статистично незначущими, а F -критерій менше табличного значення.

5.2. Метод найменших квадратів (МНК) для багатофакторної регресії. Властивості оцінок на основі МНК.

Можливі різні види рівнянь множинної регресії: лінійні й нелінійні. Через чітку інтерпретацію параметрів найбільше широко використовується лінійна функція.

У лінійній множинній регресії

$$\hat{y}_x = a + b_1x_1 + b_2x_2 + \dots + b_mx_m \quad (5.4)$$

параметри при x називаються коефіцієнтами «чистої» регресії. Вони характеризують середню зміну результату зі зміною відповідного фактору на одиницю при незміненому значенні інших факторів, закріплених на середньому рівні.

Розглянемо лінійну модель множинної регресії

$$y = a + b_1x_1 + b_2x_2 + \dots + b_mx_m + \varepsilon. \quad (5.5)$$

Класичний підхід до оцінювання параметрів лінійної моделі множинної регресії заснований на методі найменших квадратів (МНК). МНК дозволяє одержати такі

де σ_y^2 – загальна дисперсія результативної ознаки;

$\sigma_{\text{ост}}^2$ – залишкова дисперсія.

Границі зміни індексу множинної кореляції від 0 до 1. Чим ближче його значення до 1, тим тісніше зв'язок результативної ознаки з усім набором досліджуваних факторів. Величина індексу множинної кореляції повинна бути більше або дорівнює максимальному парному індексу кореляції:

$$R_{yx_1x_2\dots x_m} \geq r_{yx_i(\max)} \quad (i = \overline{1, m}) \quad (5.9).$$

При правильному включенні факторів у регресійну модель величина індексу множинної кореляції буде істотно відрізнятися від індексу кореляції парної залежності. Якщо ж додатково включені в рівняння множинної регресії фактори третьорядні, то індекс множинної кореляції може практично збігатися з індексом парної кореляції (розходження в третій, четвертому знаках). Звідси ясно, що порівнюючи індекси множинної й парної кореляції, можна зробити висновок про доцільність включення в рівняння регресії того або іншого фактору.

Розрахунок індексу множинної кореляції припускає визначення рівняння множинної регресії й на його основі залишкової дисперсії:

$$\sigma_{\text{ост}}^2 = \frac{1}{n} \sum (y - \hat{y}_{x_1x_2\dots x_m})^2. \quad (5.10)$$

Можна користуватися наступною формулою індексу множинної детермінації:

$$R_{yx_1x_2\dots x_m}^2 = 1 - \frac{\sum (y - \hat{y}_{x_1x_2\dots x_m})^2}{\sum (y - \bar{y})^2}. \quad (5.11)$$

Формула індексу множинної кореляції для лінійної регресії одержала назву *лінійного коефіцієнта множинної*

кореляції, або, що того ж саме, *сукупного коефіцієнта кореляції*. Ранжирування факторів, що беруть участь у множинній лінійній регресії, може бути проведене через стандартизовані коефіцієнти регресії (β -коефіцієнти). Ця ж ціль може бути досягнута за допомогою приватних коефіцієнтів кореляції (для лінійних зв'язків). Крім того, частки показника кореляції широко використовуються при рішенні проблеми відбору факторів: доцільність включення того або іншого фактору в модель можна довести величиною показника приватної кореляції.

Приватні коефіцієнти кореляції характеризують тісноту зв'язку між результатом і відповідним фактором при елімініруванні (усуненні впливу) інших факторів, включених у рівняння регресії.

Показники приватної кореляції являють собою відношення скорочення залишкової дисперсії за рахунок додаткового включення в аналіз нового фактору до залишкової дисперсії, що мала місце до введення його в модель.

У загальному виді при наявності m факторів для рівняння

$$y = a + b_1x_1 + b_2x_2 \dots + b_mx_m + \varepsilon \quad (5.12)$$

коефіцієнт приватної кореляції, що вимірює вплив на y фактора x_i , при незмінному рівні інших факторів, можна визначити по формулі:

$$r_{yx_i: x_1x_2 \dots x_{i-1}x_{i+1} \dots x_m} = \sqrt{1 - \frac{1 - R_{yx_1x_2 \dots x_i \dots x_m}^2}{1 - R_{yx_1x_2 \dots x_{i-1}x_{i+1} \dots x_m}^2}}, \quad (5.13)$$

де $R_{yx_1x_2 \dots x_i \dots x_m}^2$ – множинний коефіцієнт детермінації всіх m факторів з результатом;

$R^2_{y \times x_1 x_2 \dots x_{i-1} x_{i+1} \dots x_m}$ – той же показник детермінації, але

без введення в модель фактору x_i .

В економетрії приватні коефіцієнти кореляції звичайно не мають самостійного значення. Їх використовують на стадії формування моделі. Так, будуючи багатофакторну модель, на першому кроці визначається рівняння регресії з повним набором факторів і розраховується матриця приватних коефіцієнтів кореляції. На другому кроці відбирається фактор з найменшої й несуттєвої по t -критерії Стьюдента величиною показника приватної кореляції. Виключивши його з моделі, будується нове рівняння регресії. Процедура триває доти, поки не виявиться, що всі частки коефіцієнти кореляції істотно відрізняються від нуля. Якщо виключено несуттєвий фактор, то множинні коефіцієнти детермінації на двох суміжних кроках побудови регресійної моделі майже не відрізняються друг від друга, $R^2_{m+1} \approx R^2_m$, де m – число факторів.

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

$$F = \frac{S_{\text{факт}}}{S_{\text{ост}}} = \frac{R^2}{1 - R^2} \cdot \frac{n - m - 1}{m}, \quad (5.14)$$

де $S_{\text{факт}}$ – факторна сума квадратів на один ступінь волі;

$S_{\text{ост}}$ – залишкова сума квадратів на один ступінь волі;

R^2 – коефіцієнт (індекс) множинної детермінації;

m – число параметрів при змінних X (у лінійній регресії збігається із числом включених у модель факторів);

n – число спостережень.

Оцінюється значимість не тільки рівняння в цілому, але й фактору, додатково включеного в регресійну модель. Необхідність такої оцінки пов'язана з тим, що не кожний фактор, що увійшов у модель, може істотно збільшувати частку поясненої варіації результативної ознаки. Крім того, при наявності в моделі декількох факторів вони можуть вводитися в модель у різній послідовності. Через кореляцію між факторами значимість того самого фактору може бути різної залежно від послідовності його введення в модель.

5.4. Узагальнений метод найменших квадратів (УМНК).

При порушенні гомоскедастичності й наявності автокореляції помилок рекомендується традиційний метод найменших квадратів (відомий в англійській термінології як метод OLS – Ordinary Least Squares) замінити *узагальненим методом*, тобто *методом GLS* (Generalized Least Squares).

Узагальнений метод найменших квадратів застосовується до перетворених даних і дозволяє одержувати оцінки, які володіють не тільки властивістю несмещенности, але й мають менші вибіркові дисперсії. Зупинимося на використанні УМНК для коректування гетероскедастичности.

Як і раніше, будемо припускати, що середнє значення залишкових величин дорівнює нулю. А от дисперсія їх не залишається незмінної для різних значень фактору, а пропорційна величині K_i , тобто $\sigma_{\varepsilon_i}^2 = \sigma^2 \cdot K_i$ (5.15),

де $\sigma_{\varepsilon_i}^2$ – дисперсія помилки при конкретному i -м значенні фактору; σ^2 – постійна дисперсія помилки при дотриманні передумови про гомоскедастичності залишків; K_i – коефіцієнт пропорційності, що міняється зі зміною величини фактору, що й обумовлює неоднорідність дисперсії.

При цьому передбачається, що σ^2 невідомо, а відносно величин K_i висувуються певні гіпотези, що характеризують структуру гетероскедастичності.

У загальному виді для рівняння

$$y_i = a + bx_i + \varepsilon_i \quad (5.16) \quad \text{при} \quad \sigma_{\varepsilon_i}^2 = \sigma^2 \cdot K_i \quad (5.17)$$

модель прийме вид: $y_i = a + bx_i + \sqrt{K_i} \varepsilon_i$ (5.18)

У ній залишкові величини гетероскедастичні. Припускаючи в них відсутність автокореляції, можна перейти до рівняння з гомоскедастичними залишками, поділивши всі змінні, зафіксовані в ході i -го спостереження, на $\sqrt{K_i}$. Тоді дисперсія залишків буде величиною постійної, тобто $\sigma_{\varepsilon_i}^2 = \sigma^2$ (5.19). Оцінка параметрів нового рівняння з перетвореними змінними приводить до зваженого методу найменших квадратів, для якого необхідно мінімізувати суму квадратів відхилень виду $S(a, b) = \sum_{i=1}^n \frac{1}{K_i} (y_i - a - bx_i)^2$ (5.20).

Якщо перетворені змінні x й y взяти у відхиленнях від середніх рівнів, то коефіцієнт регресії b можна

визначити як
$$b = \frac{\sum \frac{1}{K} \cdot x \cdot y}{\sum \frac{1}{K} \cdot x^2} \quad (5.21)$$

Разом з тим, варто мати на увазі, що нові перетворені змінні одержують нове економічне втримування і їхню регресію має інший зміст, чим регресія за вихідним даними.

Контрольні питання

1. Назвіть основні елементи багатofакторної економетричної моделі? Чим відрізняється багатofакторна економетрична модель від однофакторної?

2. Як визначається в кількісному регресійному аналізі коефіцієнт регресії, детермінації?

3. Назвіть особливості визначення випадкового члена в кількісному регресійному аналізі?

4. Які умови необхідні для отримання більш точного коефіцієнта регресії?

5. В чому особливості УМНК?

Тема 6. Методи дослідження якісних економічних показників

План

6.1. Якісні економічні показники.

6.2. Регресійні моделі з бінарними незалежними змінними.

6.3. Регресійні моделі з бінарними залежними змінними.

Література [2, 5, 9, 15, 17, 18, 25].

6.1. 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*-змінними 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 змогу оцінювати одне рівняння, у якому різні класи спостережень розділяються за допомогою цієї змінної.

Приклад 6.1. Нехай регресійна модель залежності заробітної плати y від деяких кількісних факторів x_1, x_2, \dots, x_m має вигляд

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m + u \quad (6.1)$$

Тоді для вивчення впливу вищої освіти на рівень оплати праці вводять нову змінну A , яка може набувати двох значень: $A = 1$, якщо робітник має вищу освіту, та $A = 0$, якщо не має. Модель, що враховує цей фактор, матиме вигляд

$$y = a_0 + a_1Ax_1 + a_2x_2 + \dots + a_mx_m \quad (6.2)$$

тобто за наявності вищої освіти заробітна плата в середньому становить

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_mx_m + u, \quad \dots \quad (6.3)$$

У такому разі коефіцієнт u 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 u встановлена в процесі перевірки нульової гіпотези $H_0 : u = 0$, 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.

6.2. 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.

Крім того, бінарні змінні використовують також при дослідженні моделей, які описують структурні зміни в економіці. Розглянемо такий приклад.

Приклад 6.2. Нехай досліджується залежність обсягу випущеної підприємством продукції y від обсягу його основного фонду x . Припускається, що після досягнення основним фондом підприємства розміру 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:

$$y = a_0 + a_1x + a_2(x - x)A + u, \quad (6.4)$$

де бінарна змінна $y = 0$ або $y = 1$.

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.

6.3. 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: $y = 1$, 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. 6.1). 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.

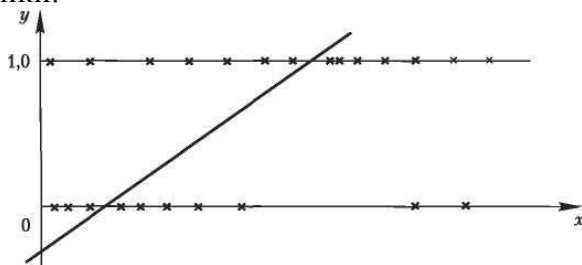


Рис. 6.1

Логічно припустити, що регресійна функція, як і функція розподілу випадкової величини, має S-подібну траєкторію розвитку (рис.6. 2). Практикою перевірено, що функції розподілу доходів можуть бути підпорядковані нормальному чи логістичному закону розподілу.

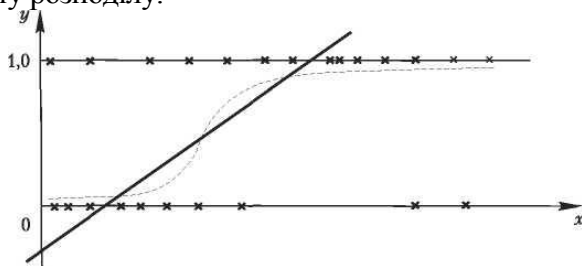


Рис. 6.2.

Означення 1. Регресійна модель з бінарною (дихотомною) залежною змінною, що має нормальний розподіл, називається *пробіт-моделлю*.

Означення 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 питання

1. Запишіть в загальному вигляді структурну форму моделі на основі одночасових рівнянь.
2. Що означає зведена форма моделі? Як її одержати?
3. Дайте визначення рекурсивних систем і запишіть модель на основі рекурсивної системи.
4. Яка система рівнянь називається точно ідентифікованою?
5. Яка система рівнянь називається надідентифікованою?
6. Запишіть умову ідентифікованості системи рівнянь.
7. На основі якого методу можна оцінити параметри моделі, якщо вона складається із системи рекурсивних рівнянь?
8. Який метод оцінки параметрів можна застосувати, коли всі рівняння моделі є точно ідентифікованими?

9. На основі якого методу можна оцінити параметри моделі, якщо вона має надіентифіковані рівняння?

10. Чи можна виконувати оцінку параметрів моделі окремо для групи точно ідентифікованих і надіентифікованих рівнянь?

Тема 7. Дисперсійний аналіз економетричної моделі

План

7.1. Поняття дисперсійного аналізу. 1МНК.

7.2. Приклад дисперсійного аналізу економетричної моделі та прогноз.

Література [15,18, 21, 25, 28].

7.1. Поняття дисперсійного аналізу. 1МНК.

Між оцінками параметрів економетричної моделі та коефіцієнтом кореляції, що характеризує тісноту зв'язку, існує зв'язок. Для простої економетричної моделі його можна записати так:

$$\hat{a} = r \frac{\sigma_Y}{\sigma_X} \quad (7.1) ,$$

де r — коефіцієнт парної кореляції;

σ_Y , σ_X — середньоквадратичне відхилення відповідно залежної і незалежної змінної.

Це співвідношення було покладено в основу алгоритму визначення альтернативної оцінки параметрів моделі за методом 1МНК. Алгоритм має назву *покрокової регресії* і наступні кроки:

Крок 1. Стандартизація (нормалізація) всіх змінних моделі:

$$Y_i^* = \frac{Y_i - \bar{Y}}{\sigma_Y};$$

$$X_{ij}^* = \frac{X_{ij} - \bar{X}_j}{\sigma_{X_j}}.$$

Крок 2. Визначення кореляційної матриці r , елементи якої розраховуються таким чином:

$$r_{YX_j} = \frac{1}{n} Y^* X_j^* ;$$

$$r_{X_k X_j} = \frac{1}{n} X_k^* X_j^* ; \quad k = \overline{1, m}, j = \overline{1, m}.$$

Крок 3. Із усіх елементів матриці вибирається той, якому відповідає . Це означає, що незалежна змінна найтісніше зв'язана з залежною змінною Y . Будується економетрична модель:

$$\hat{Y}^* = \hat{\beta}_j X_j^*$$

Крок 4. Серед інших елементів матриці r_{YX_j} знову вибирається . Якщо даному коефіцієнту кореляції відповідає X_{j+1}^* , то ця змінна вводиться в побудовану раніше економетричну модель; в результаті дістанемо:

$$Y^* = \hat{\beta}_j X_j^* + \hat{\beta}_{j+1} X_{j+1}^* \quad \text{і т.д.}$$

Процес продовжується до тих пір, поки всі незалежні змінні поступово будуть включені в модель. Якщо є обмеження, яке вказує на недоцільність розширення економетричної моделі за рахунок змінних, що залишилися, то процес розрахунку закінчується раніше. Таким обмеженням може бути співвідношення між

коефіцієнтом кореляції чи детермінації, виправленими й не виправленими на число ступеней свободи.

Система нормальних рівнянь у даному алгоритмі:

$$\begin{cases} r_{YX_1} = \hat{\beta}_1 + \hat{\beta}_2 r_{X_1 X_2} + \hat{\beta}_3 r_{X_1 X_3} + \dots + \hat{\beta}_m r_{X_1 X_m} \\ r_{YX_2} = \hat{\beta}_1 r_{X_1 X_2} + \hat{\beta}_2 + \hat{\beta}_3 r_{X_2 X_3} + \dots + \hat{\beta}_m r_{X_2 X_m} \\ \dots \\ r_{YX_m} = \hat{\beta}_1 r_{X_1 X_m} + \hat{\beta}_2 r_{X_2 X_m} + \hat{\beta}_3 r_{X_3 X_m} + \dots + \hat{\beta}_m \end{cases}$$

Позначимо елементи r_{YX_j} через вектор r_{YX} , а інші елементи $r_{X_k X_j}$ — через матрицю r , тоді система рівнянь у матричному вигляді матиме такий вигляд:

$$r \hat{\beta} = r_{YX}$$

$$\text{Звідси } \hat{\beta} = r^{-1} r_{YX}, \quad (7.2)$$

тобто отримаємо альтернативний оператор оцінювання параметрів моделі за методом ІМНК.

Оскільки оцінки параметрів моделі $\hat{\beta}$ відносяться до стандартизованих змінних, то щоб перейти до оцінок параметрів моделі, в якій змінні мають свій початковий вимір, необхідно:

$$\hat{a}_j = \hat{\beta}_j \frac{\sigma_Y}{\sigma_{X_j}}, \quad j = \overline{1, m}; \quad (7.3)$$

$$\hat{a}_0 = \bar{y} - \sum_j \hat{a}_j \bar{x}_j.$$

Множинний коефіцієнт детермінації, який визначає рівень варіації залежної змінної за рахунок незалежних, розраховується таким чином:

$$R^2 = \frac{\sigma_Y^2 - \sigma_U^2}{\sigma_Y^2} \quad (7.4),$$

$$\text{чи } R^2 = \frac{\hat{A}' X' Y}{Y' Y} \cdot \frac{n-1}{m-1}. \quad (7.5)$$

Коефіцієнт детермінації без урахування числа ступеней свободи:

$$\bar{R}^2 = \frac{\hat{A}' X' Y}{Y' Y}$$

Співвідношення між ними дорівнюватиме:

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

Множинний коефіцієнт кореляції $R = \sqrt{R^2}$ характеризує тісноту зв'язку між залежною і незалежними змінними. Множинний коефіцієнт детермінації і кореляції знаходяться на множині

$$R^2 \in]0,1[;$$

$$R \in]0,1[.$$

Якщо оцінка параметрів моделі отримана на основі покрокової регресії, то для визначення коефіцієнта детермінації можна використати такі співвідношення:

$$R^2 = 1 - \frac{|r|}{R_{11}};$$

$$R^2 = \hat{\beta}_1 r_{YX_1} + \hat{\beta}_2 r_{YX_2} + \hat{\beta}_3 r_{YX_3} + \dots + \hat{\beta}_m r_{YX_m},$$

де $|r|$ – визначник матриці r ;

R_{11} — алгебраїчне доповнення першого елемента матриці r .

Гіпотеза про наявність чи відсутність зв'язку між залежною і незалежною змінними може бути перевірена на основі F -критерію:

$$F = \frac{\sigma_Y^2 - \sigma_U^2}{\sigma_U^2} = \frac{\sigma_P^2}{\sigma_U^2} \quad (7.4)$$

Фактичне значення F — критерію порівнюється з табличним при ступенях свободи $n-m$ і $m-1$ і вибраному рівні довіри. Якщо $F_{\text{факт}} > F_{\text{табл}}$, то гіпотеза про суттєвість зв'язку між залежною і незалежними змінними економетричної моделі підтверджується, в протилежному випадку — відкидається.

Альтернативна формула розрахунку F — критерію через коефіцієнт детермінації:

$$F = \frac{R^2 / (m - 1)}{(1 - R^2) / (n - m)} \quad (7.5)$$

Значущість оцінок параметрів моделі можна визначити на основі t — критерію:

$$t_j = \frac{a_j}{\sqrt{\sigma_U^2 c_{jj}}}$$

Значення критерію t_j порівнюється з табличним при вибраному рівні значущості і $n - m$ ступенями свободи. Якщо $t_{\text{факт}} > t_{\text{табл}}$, то відповідний параметр економетричної моделі є достовірним.

На основі t — критерію і стандартної помилки будуються довірчі інтервали для параметрів \hat{a}_j :

$$a_j = \hat{a}_j \pm t_{(\alpha)} \sqrt{\sigma_U^2 c_{jj}}$$

Прогноз залежної змінної на основі економетричної моделі при заданих залежних змінних можна виконати на основі такого співвідношення:

$$\hat{Y}_{\text{пр}} - t_{(\alpha)} S(\hat{y}) \leq Y \leq \hat{Y}_{\text{пр}} + t_{(\alpha)} S(\hat{y})$$

У цьому співвідношенні $S(\hat{y})$ є стандартною помилкою прогнозу:

$$S(\hat{y}) = \sqrt{\sigma_u^2 X_p' (X' X)^{-1} X_p}$$

де X_p — прогнознi значення незалежних змiнних.

7.2. Приклад дисперсійного аналізу економетричної моделі та прогноз.

Визначити коефіцієнти детермінації та кореляції для економетричної моделі, яка побудована в прикладі 2.1. Перевірити гіпотезу про суттєвість зв'язку на основі F - і t -критеріїв. Виконати прогноз витрат на харчування, якщо загальні затрати становитимуть 900 одиниць, а середній склад сім'ї – 8,5.

Розв'язання

Економетрична модель має вигляд:

$$\hat{Y} = 9,596 + 0,1837 X_1 + 6,854 X_2$$

1. Визначимо коефіцієнт детермінації на основі співвідношення:

$$R^2 = 1 - \frac{\sigma_U^2}{\sigma_Y^2}$$

де σ_U^2 , σ_Y^2 — відповідно залишкова й загальна дисперсії.

$$R^2 = 1 - \frac{95,1977}{2434,533} = 0,9609$$

Це значення коефіцієнта детермінації свідчить про те, що варіація витрат на харчування на 96,09% визначається варіацією загальних затрат і складу сім'ї.

2. Коефіцієнт кореляції $R = \sqrt{R^2} = 0,98$ Оскільки коефіцієнт кореляції наближається до одиниці, то це

свідчить, що зв'язок між витратами на харчування, загальними затратами і складом сім'ї є дуже тісним.

3. Визначимо F - критерій (критерій Фішера):

$$F = \frac{\sigma_Y^2 - \sigma_U^2}{\sigma_U^2} = \frac{2434,533 - 95,1977}{95,1977} = 24,5734$$

Порівняємо розраховане значення критерію Фішера з табличним. При ступенях свободи $m-1=2$; $n-m=13$ і рівні довіри $\alpha = 0,05$,

$$F_{\text{табл}} = 19,36.$$

Оскільки $F_{\text{факт}} > F_{\text{табл}}$, то гіпотеза про значущість зв'язку, який описується економетричною моделлю, підтверджується.

4. Розрахуємо t - критерії:

$$t_{\hat{a}_0} = \frac{|\hat{a}_0|}{S_{\hat{a}_0}} = 1,773 ;$$

$$t_{\hat{a}_1} = \frac{|\hat{a}_1|}{S_{\hat{a}_1}} = 9,439 ;$$

$$t_{\hat{a}_2} = \frac{|\hat{a}_2|}{S_{\hat{a}_2}} = 4,908 .$$

Табличне значення t - критерію при ступені свободи $n-m=13$ і рівні довіри $\alpha = 0,05$ дорівнює 2,16.

Враховуючи, що

$$t_{\hat{a}_1} > t_{(0,05)} \quad \text{оцінки параметрів моделі } \hat{a}_1 \text{ і } \hat{a}_2 \text{ є}$$

$$t_{\hat{a}_2} > t_{(0,05)} ,$$

достовірними. Оскільки $t_{\hat{a}_0} < t_{(0,05)}$, то понизимо рівень довіри: $\alpha = 0,10$. У цьому випадку $t_{\text{табл}} = 1,77$.

А це означає, що 10-процентний рівень довіри підтверджує значущість вільного члена моделі.

5. Побудуємо довірчі інтервали для оцінок параметрів моделі:

$$a_j = \hat{a}_j \pm t_{(\alpha)} S \hat{a}_j$$

$$\hat{a}_0 - t_{(0,1)} S \hat{a}_0 \leq a_0 \leq \hat{a}_0 + t_{(0,1)} S \hat{a}_0$$

$$\hat{a}_1 - t_{(0,05)} S \hat{a}_1 \leq a_1 \leq \hat{a}_1 + t_{(0,05)} S \hat{a}_1$$

$$\hat{a}_2 - t_{(0,05)} S \hat{a}_2 \leq a_2 \leq \hat{a}_2 + t_{(0,05)} S \hat{a}_2$$

$$0,01455 \leq a_0 \leq 19,177$$

$$0,14165 \leq a_1 \leq 0,2257$$

$$3,83710 \leq a_2 \leq 9,8706 .$$

6. Розрахуємо прогнозне значення витрат на харчування на основі економетричної моделі.

6.1. Визначимо точковий прогноз витрат на харчування на основі моделі:

$$\hat{Y}_{np} = 9,596 + 0,1837 X_{1np} + 6,854 X_{2np} = 9,596 + 0,1837 \cdot 900 + 6,854 \cdot 8,5 = 233,1678$$

6.2. Знайдемо дисперсію прогнозу:

$$\sigma_{\hat{Y}_{np}}^2 = X'_{np} \text{ var}(\hat{A}) X_{np} = 97,0404$$

6.3. Стандартна помилка прогнозу:

$$S_{\hat{Y}_{np}} = \sqrt{\sigma_{\hat{Y}_{np}}^2} = \sqrt{97,0404} = 9,8509 .$$

6.4. Визначимо довірчі інтервали прогнозного рівня витрат на харчування:

$$\hat{Y}_{np} - t_{(\alpha)} S_{\hat{Y}_{np}} \leq Y_{np} \leq \hat{Y}_{np} + t_{(\alpha)} S_{\hat{Y}_{np}} .$$

При $\alpha = 0,05$, $t_{kp} = 2,16$.

$$233,1678 - 2,16 \cdot 9,8509 \leq Y_{np} \leq 233,1678 + 2,16 \cdot 9,8509 ;$$
$$211,8896 \leq Y_{np} \leq 254,4455 .$$

Таким чином, точковий прогноз витрат на харчування дорівнює 233,1678 одиниць, а інтервальний буде знаходитись у межах від 211,8896 до 254,4455 одиниць.

Контрольні питання

1. Що характеризується поняттям дисперсії в економетрії?
2. В чому полягає основне завдання дисперсійного аналізу.
3. Алгоритм покрокової регресії.

Другий модуль

Тема 8. Мультиколінеарність

План

- 8.1. Визначення мультиколінеарності та її природа.
- 8.2. Практичні наслідки мультиколінеарності.
- 8.3. Тестування наявності та рівня мультиколінеарності. Метод Фаррара - Глобера.
- 8.4. Засоби вилучення мультиколінеарності.

Література [15, 17, 18, 19, 21, 23, 25].

8.1. Визначення мультиколінеарності та її природа.

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 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:*

Наявність мультиколінеарності створює певні проблеми при розробці 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.

8.2. 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.

8.3. Тестування наявності та рівня мультиколінеарності. Метод Фаррара - Глобера.

Найбільш повне дослідження мультиколінеарності можна здійснити на основі алгоритму Фаррара - Глобера. Цей алгоритм включає три види статистичних критеріїв, на основі яких перевіряється мультиколінеарність всього масиву незалежних змінних (χ^2 , χ^2 -квдрат); кожної незалежної змінної зі всіма незалежними змінними (F -критерій) і мультиколінеарність кожної пари незалежних змінних (t -критерій).

Всі ці критерії при порівнянні з їх критичними значеннями дають можливість зробити конкретні висновки відносно наявності чи відсутності мультиколінеарності незалежних змінних.

Опишемо алгоритм Фаррара—Глобера.

Крок 1. Стандартизація (нормалізація) змінних.

Позначимо вектори незалежних змінних економетричної моделі через $X_1, X_2, X_3, \dots, X_m$. Елементи стандартизованих векторів розрахуємо за формулою:

$$X_{ik}^* = \frac{X_{ik} - \bar{X}_k}{\sqrt{n\sigma_{X_k}^2}}, \quad (8.1)$$

де n — число спостережень, ($i = \overline{1, n}$);

m — число незалежних змінних, ($k = \overline{1, m}$);

\bar{X}_k — середня арифметична k — її незалежної змінної;

$\sigma_{X_k}^2$ — дисперсія k -ї незалежної змінної.

Крок 2. Знаходження кореляційної матриці (матриці моментів стандартизованої системи нормальних рівнянь): $R = X^{*'} X^*$,

де X^* — матриця стандартизованих незалежних змінних;

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

Крок 3. Визначення критерію χ^2 (хі-квадрат):

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

де $|R|$ — визначник кореляційної матриці R .

Значення цього критерію порівнюється з табличним при $\frac{1}{2}m(m-1)$ ступенях свободи і рівні значущості α .

Якщо $\chi^2_{\text{факт}} < \chi^2_{\text{табл}}$, в масиві незалежних змінних не існує мультиколінеарності.

Крок 4. Визначення оберненої матриці C :

$$C = R^{-1} = (X^* X^*)^{-1}.$$

Крок 5. Розрахунок F -критеріїв:

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

де c_{kk} — діагональні елементи матриці C . Фактичні значення критеріїв F_k порівнюються з табличними при $n-m$ і $m-1$ ступенях свободи і рівні значущості α . Якщо $F_k_{\text{факт}} > F_{\text{табл}}$, відповідна k -та незалежна змінна мультиколінеарна з іншими.

Коефіцієнт детермінації для кожної змінної розраховується таким чином:

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

Крок 6. Знаходження часткових коефіцієнтів кореляції:

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

де c_{kj} — елемент матриці C , що заходить в k -му рядку і j -му стовпці, $k = \overline{1, m}$, $j = \overline{1, m}$, c_{kk} і c_{jj} — діагональні елементи матриці C .

Крок 7. Розрахунок t критеріїв:
$$t_{kj} = \frac{r_{kj} \sqrt{n-m}}{\sqrt{1-r_{kj}^2}} .$$

Фактичні значення критеріїв t_{kj} порівнюються з табличними при $n-m$ ступенях свободи і рівні значущості α . Якщо $t_{kj \text{ факт}} > t_{\text{табл}}$, між незалежними змінними X_k і X_j існує мультиколінеарність.

8.4. Засоби вилучення мультиколінеарності.

Звичайно, усе залежить від ступеня мультиколінеарності, однак у будь-якому разі можна запропонувати декілька 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**.

Контрольні питання

1. Що означає мультиколінеарність змінних?
2. Ознаки мультиколінеарності.
3. Як впливає наявність мультиколінеарності змінних на оцінку параметрів моделі?

4. Які статистичні критерії використовуються для виявлення мультиколінеарності?
5. Дайте коротку характеристику алгоритму Фаррара — Глобера.
6. На чому ґрунтується метод головних компонентів? Коли він застосовується?
7. Як обчислити головні компоненти і яким умовам вони задовольняють?
8. Як оцінюються параметри моделі за допомогою головних компонентів?

Тема 9. Гетероскедастичність.

План

- 9.1. Визначення гетероскедастичності та її природа.
 - 9.2. Наслідки порушення припущення про гомоскедастичність.
 - 9.3. Знаходження гетероскедастичності та її вилучення.
- Література [15, 17, 18, 19, 21, 23, 25].

9.1. Визначення гетероскедастичності та її природа.

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 (гомоскедастичність залишків).

Означення 1. *Якщо дисперсія залишків стала для кожного спостереження, тобто $M(uu') = \sigma_u^2 E$, то це явище називається гомоскедастичністю:*

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

Означення 2. Якщо дисперсія залишків змінюється для кожного спостереження або групи спостережень, тобто

$$M(uu') = \sigma_u^2 S \quad (9.1),$$

то це явище називається гетероскедастичністю:

Розглянемо питання про доцільність припущення 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 щ навколо її математичного сподівання не залежить від значення факторів x :

Форма гетероскедастичності залежить від знаків і значень коефіцієнтів у залежності

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

9.2. Наслідки порушення припущення про гомоскедастичність:

1) неможливо знайти середньоквадратичне відхилення параметрів регресії, а отже, неможливо оцінити значущість параметрів;

2) неможливо побудувати довірчий інтервал для прогнозних значень $Y_{пр}$;

3) отримані за МНК оцінки параметрів регресії не є ефективними (не мають найменшої дисперсії). Зазначимо, що якщо незважаючи на гетероскедастичність

4) ми використовуємо звичайні процедури перевірки гіпотез, то висновки можуть бути неправильними. Зрозуміло, гетероскедастичність є суттєвою проблемою, а тому потрібно вміти з'ясувати її наявність.

9.3. Знаходження гетероскедастичності та її вилучення.

Як і в разі мультиколінеарності, єдиних правил виявлення гетероскедастичності немає, а є різноманітні тести (критерії): критерій F , параметричний та непараметричний тести Гольдфелда — 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:

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*).

Можливість перевірки припущень про наявність гетероскедастичності залежить від природи вихідних даних. Розглянемо методи перевірки гетероскедастичності для різних вихідних даних.

Перевірка гетероскедастичності на основі критерію F

Цей метод застосовується тоді, коли вихідна сукупність спостережень досить велика. Розглянемо відповідний алгоритм.

Крок 1. Вихідні дані залежної змінної Y розбиваються на k груп ($r = \overline{1, k}$) відповідно до зміни рівня величини Y .

Крок 2. За кожною групою даних обчислюється сума квадратів відхилень:

$$S_r = \sum_{i=1}^k (y_{ir} - \bar{y}_r)^2.$$

Крок 3. Визначається сума квадратів відхилень в цілому по всій сукупності спостережень:

$$\sum_{r=1}^k S_r = \sum_{i=1}^{n_r} \sum_{i=1}^k (y_{ir} - \bar{y}_r)^2.$$

Крок 4. Обчислюється параметр α :

$$\alpha = \prod_{i=1}^k \left(\frac{S_r}{n_r} \right)^{n_r/2} / \left(\frac{\sum_{r=1}^k S_r}{n} \right)^{n/2},$$

де n — загальна сукупність спостережень; n_r — кількість спостережень r -ї групи.

Крок 7. Обчислюється критерій: $\mu = -2 \ln \alpha$,

який наближено відповідає розподілу χ^2 при ступені свободи $k-1$, коли дисперсія всіх спостережень однорідна. Тобто якщо значення μ не менше за табличне значення χ^2 при вибраному рівні довіри і ступені свободи $k-1$, то спостерігається гетероскедастичність.

Параметричний тест Гольдфелда — Квандта

Коли сукупність спостережень невелика, то розглянутий метод не застосовний.

У такому разі Гольдфелд і Квандт запропонували розглянути випадок, коли

$$M(uu') = \sigma_u^2 x_{ij}^2,$$

тобто дисперсія залишків зростає пропорційно до квадрата однієї з незалежних змінних моделі:

$$Y = XA + u.$$

Для виявлення наявності гетероскедастичності згадані вчені склали параметричний тест, в якому потрібно виконати такі кроки.

Крок 1. Упорядкувати спостереження відповідно до величини елементів вектора X_j .

Крок 2. Відкинути c спостережень, які містяться в центрі вектора. Згідно з експериментальними розрахунками автори знайшли оптимальні співвідношення між параметрами c і n , де n — кількість елементів вектора x_j : $\frac{c}{n} = \frac{4}{15}$.

Крок 3. Побудувати дві економетричні моделі на основі 1МНК за двома утвореними сукупностями спостережень $(n-c)/2$ за умови, що $(n-c)/2$ перевищує кількість змінних m .

Крок 4. Знайти суму квадратів залишків за першою (1) і другою (2) моделями S_1 і S_2 :

$S_1 = u_1' u_1$, де u_1 — залишки за моделлю (1);

$S_2 = u_2' u_2$, де u_2 — залишки за моделлю (2).

Крок 7. Обчислити критерій $R^* = \frac{S_2}{S_1}$,

який в разі виконання гіпотези про гомоскедастичність відповідатиме F -розподілу з $(n-c-2m)/2$, $(n-c-2m)/2$ ступенями свободи. Це означає, що обчислене значення R^* порівнюється з табличним значенням F -критерію для ступенів свободи $(n-c-2m)/2$ і $(n-c-2m)/2$ і вибраного рівня довіри. Якщо $R^* \leq F_{\text{табл}}$, то гетероскедастичність відсутня.

$R_{0,01} = 11$. Оскільки $R^* > F_{\text{кр}}$, то вихідні дані мають гетероскедастичність.

Непараметричний тест Гольдфелда - Квандта

Гольдфелд і Квандт для оцінювання наявності гетероскедастичності запропонували також непараметричний тест. Цей тест базується на числі піків у

величини залишків після упорядкування спостережень за x_{ij} .

Закономірність зміни залишків, коли дисперсія є однорідною, — явище гомоскедастичності ілюструє рис. 9.1, а на рис.9.2 спостерігається явище гетероскедастичності.

Цей тест, звичайно, не такий надійний, як параметричний, але він досить простий.

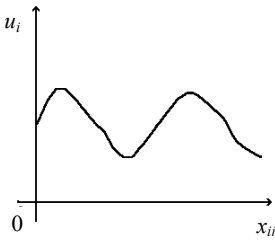


Рис. 9.1.

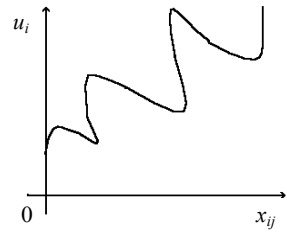


Рис. 9.2.

Зауважимо, що на рис. 9.1 зображено, як змінюються залишки, що мають постійну дисперсію, а на рис. 9.2 — залишки, дисперсія яких змінна для різних груп спостережень.

Тест Глейсера

Ще один тест перевірки гетероскедастичності склав Глейсер. Він запропонував розглядати регресію абсолютних значень залишків $|u_j|$, що відповідають регресії найменших квадратів, як певну функцію від x_j , де x_j — та незалежна змінна, яка відповідає зміні дисперсії σ_u^2 . Для цього використовуються такі види функцій:

- 1) $|u| = a_0 + a_1 x_j$;
- 2) $|u| = a_0 + a_1 x_j^{-1}$;
- 3) $|u| = a_0 + a_1 x_j^{1/2}$ і т.ін.

Рішення про відсутність гетероскедастичності залишків приймається на підставі статистичної значущості

коефіцієнтів a_0 і a_1 . Переваги цього тесту визначаються можливістю розрізнити випадок чистої і змішаної гетероскедастичності.

Чистій гетероскедастичності відповідають значення параметрів $a_0 = 0, a_1 \neq 0$, а змішаній — $a_0 \neq 0, a_1 \neq 0$. Залежно від цього треба користуватись різними матрицями S . Нагадаємо, що $M(uu') = \sigma^2 S$.

Визначення матриці s

Щоб оцінити параметри моделі, коли дисперсії залишків визначаються $M(uu') = \sigma_u^2 S$, потрібно визначити матрицю S .

Спинимось на визначенні матриці S .

Оскільки явище гетероскедастичності пов'язане лише з тим, що змінюються дисперсії залишків, а коваріація між ними відсутня, то матриця S має бути діагональною, а саме:

$$S = \begin{pmatrix} \frac{1}{\lambda_1} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\lambda_2} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\lambda_3} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{1}{\lambda_n} \end{pmatrix} \quad (9.2)$$

Щоб пояснити, чому саме такий вигляд має ця матриця, потрібно ще раз наголосити: за наявності гетероскедастичності для певних вихідних даних одна (або кілька) пояснювальних змінних можуть різко змінюватись від одного спостереження до іншого, тоді як залежна змінна має такі самі коливання, як і для попередніх спостережень.

Але це означає, що дисперсія залишків, яка змінюватиметься від одного спостереження до іншого (чи для групи спостережень), може бути пропорційною до величини пояснювальної змінної X (або до її квадрата), яка

зумовлює гетероскедастичність, або пропорційною до квадрата залишків.

Звідси в матриці S значення λ_i можна обчислити, користуючись гіпотезами:

а) $M(uu') = \sigma_u^2 x_{ij}^2$, тобто дисперсія залишків пропорційна до зміни пояснювальної змінної x_{ij} ;

б) $M(uu') = \sigma_u^2 x_{ij}^2$, тобто зміна дисперсії пропорційна до зміни квадрата пояснювальної змінної (x_{ij}^2);

в) $M(uu') = \sigma_u^2 \{|u|\}^2$, 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: $\lambda_i = \{|u_i|\}^2$, або $\lambda_i = (a_0 - a_1 x_{ij})^2$, або $\lambda_i = (a_0 + a_1 x_o^{-1})^2$.

Оскільки матриця S — симетрична і додатно визначена, то при $S = P'P$, матриця P має вигляд:

$$P = \begin{pmatrix} \frac{1}{\sqrt{\lambda_1}} & 0 & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\lambda_2}} & 0 & \dots & 0 \\ 0 & 0 & \frac{1}{\sqrt{\lambda_3}} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \frac{1}{\sqrt{\lambda_n}} \end{pmatrix}, \quad P^{-1} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & 0 & \dots & 0 \\ 0 & 0 & \sqrt{\lambda_3} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \sqrt{\lambda_n} \end{pmatrix}. \quad (9.3)$$

Контрольні питання

1. Дайте означення гомоскедастичності і гетероскедастичності.

2. Як впливає явище гетероскедастичності на оцінку параметрів моделі?

3. Назвіть методи визначення гетероскедастичності.
4. Як перевіряється гетероскедастичність згідно з критерієм μ ?
5. Як застосовується параметричний тест для визначення гетероскедастичності?
6. У чому сутність непараметричного тесту?
7. Як визначається гетероскедастичність з допомогою регресії залишків?
8. Опишіть методи формування матриці S в умові $M(uu') = \sigma^2 S$.
9. Як використовується матриця S ?
10. Які властивості повинна мати матриця S ?

Тема 10. Автокореляція: поняття, причини виникнення, наслідки.

План

- 10.1. Поняття та визначення автокореляції
Причини виникнення автокореляції.
- 10.2. Методи перевірки наявності автокореляції.
- 10.3. Методи оцінки параметрів моделі з автокорельованими залишками.

Література [15, 17, 18, 19, 21, 23, 25].

**10.1. Поняття та визначення автокореляції
Причини виникнення автокореляції.**

В економетричних дослідженнях часто зустрічаються такі випадки, коли дисперсія залишків є постійною, але спостерігається їх коваріація. Це явище має назву *автокореляції залишків*.

Автокореляція залишків виникає частіше за все тоді, коли економетрична модель будується на основі часових рядів. Якщо існує кореляція між послідовними значеннями деякої незалежної змінної, то буде спостерігатись і

кореляція послідовних значень залишків. Тобто в цьому випадку також порушується гіпотеза, згідно з якою

$$M(uu') = \sigma_U^2 E ,$$

але при гетероскедастичності змінюється дисперсія залишків при відсутності їх коваріації, а при автокореляції — існує коваріація залишків при незмінній дисперсії.

При автокореляції залишків, як і при гетероскедастичності дисперсія залишків запишеться:

$$M(uu') = \sigma_U^2 S ,$$

але матриця S матиме тут зовсім інший вигляд. Запишемо цю матрицю:

$$S = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \rho & \dots & \rho^{n-3} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \dots & 1 \end{pmatrix} \quad (10.1)$$

В даній матриці параметр ρ характеризує коваріацію кожного наступного значення залишків із попереднім. Так, якщо для залишків записати авторегресійну модель першого порядку:

$$u_t = \rho u_{t-1} + \varepsilon_t \quad (10.2)$$

то ρ характеризує силу зв'язку величини залишків у період t від величини залишків у період $t - 1$.

Якщо проігнорувати матрицю S при визначенні дисперсії залишків, і для оцінки параметрів моделі застосувати метод ІМНК, то можливі такі **наслідки**:

1) оцінки параметрів моделі можуть бути незміщеними, але неефективними, тобто вибіркові дисперсії вектора оцінок \hat{A} можуть бути невиправдано великими;

2) статистичні критерії t і F - статистики, які отримані для класичної лінійної моделі, практично 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 вибірккову дисперсію.

10.2. Методи перевірки наявності автокореляції.

Для перевірки наявності автокореляції залишків можна застосувати чотири методи:

1. Критерій Дарбіна—Уотсона.
2. Критерій фон Неймана.
3. Нециклічний коефіцієнт автокореляції.
4. Циклічний коефіцієнт автокореляції.

1. Критерій Дарбіна—Уотсона:

$$DW = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2} . \quad (10.3)$$

Критерій Дарбіна—Уотсона може приймати значення на множині $DW \in [0,4]$. Якщо залишки u_t є випадковими величинами, тобто не автокорельовані, то значення DW знаходиться поблизу 2. При додатній автокореляції $DW < 2$, при від'ємній $DW > 2$.

Значення критерія DW табульовані на інтервалі $DW_1 - DW_2$, де DW_1 — нижня межа, DW_2 — верхня межа. Фактичні значення критерію порівнюються з табличними (критичними) для числа спостережень n і числа незалежних змінних m при вибраному рівні довіри

α . Якщо $DW_{\text{факт}} < DW_1$, залишки мають автокореляцію. Якщо $DW_{\text{факт}} > DW_2$, приймається гіпотеза про відсутність автокореляції. Якщо $DW_1 < DW < DW_2$ конкретних висновків зробити не можна.

2. Критерій фон Нейман

$$Q = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2} \cdot \frac{n}{n-1} \quad (10.4)$$

Звідси $Q = DW \cdot \frac{n}{n-1}$,

при $n \rightarrow \infty$, $Q = DW$. Фактичне значення критерію фон Неймана порівнюється з табличним при вибраному рівні довіри α і заданому числі спостережень. Якщо $Q_{\text{факт}} < Q_{\text{табл}}$, то існує додатня автокореляція.

3. Нециклічний коефіцієнт автокореляції:

$$r^* = \frac{\sum_{t=2}^n (u_t u_{t-1}) - \frac{1}{n-1} \left(\sum_{t=2}^n u_t \right) \left(\sum_{t=2}^n u_{t-1} \right)}{\sqrt{\left[\sum_{t=2}^n u_t^2 - \frac{1}{n-1} \left(\sum_{t=2}^n u_t \right)^2 \right] \left[\sum_{t=2}^n u_{t-1}^2 - \frac{1}{n-1} \left(\sum_{t=2}^n u_{t-1} \right)^2 \right]}}$$

r^* може приймати значення в інтервалі $[-1; +1]$. Від'ємні значення свідчать про від'ємну автокореляцію, додатні — про додатню. Значення, що знаходяться в деякій критичній області біля нуля, свідчать про відсутність автокореляції.

4. Циклічний коефіцієнт автокореляції:

$$r^0 = \frac{\sum_{t=2}^n u_t u_{t-1} + u_n u_1 - \frac{1}{n} \left(\sum_{t=1}^n u_t \right)^2}{\sum_{t=1}^n u_t^2 - \frac{1}{n} \left(\sum_{t=1}^n u_t \right)^2} \quad (10.5)$$

Фактичне значення цього критерію порівнюється з табличним для вибраного рівня довіри і довжини ряду спостережень n . Якщо $r_{\text{факт}}^0 \geq r_{\text{табл}}^0$, то існує автокореляція.

Припускаючи, що $\sum_{t=2}^n u_t \approx \sum_{t=1}^n u_{t-1} \approx 0$,

циклічний коефіцієнт автокореляції можна записати так:

$$r = \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2} \quad (10.6).$$

10.3. Методи оцінки параметрів моделі з автокорельованими залишками.

Оцінку параметрів моделі з автокорельованими залишками можна виконувати на основі чотирьох методів:

- 1) Ейткена;
- 2) перетворення вихідної інформації;
- 3) Кочрена—Оркатта;
- 4) Дарбіна.

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

$$u_t = \rho u_{t-1} + \varepsilon_t \quad (10.7),$$

Ітеративні методи Кочрена—Оркатта і Дарбіна можна застосовувати для оцінки параметрів економетричної моделі і тоді, коли залишки описуються авторегресійною моделлю більш високого ступеня:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t \quad (10.8);$$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \varepsilon_t \quad (10.9)$$

1. Метод Ейткена

Оператор оцінювання цим методом запишеться так:

$$\hat{A} = (X' S^{-1} X)^{-1} X' S^{-1} Y \quad \text{або}$$

$$\hat{A} = (X' V^{-1} X)^{-1} X' V^{-1} Y \quad ,$$

де S^{-1} — матриця, обернена до матриці S .

V^{-1} — матриця, обернена до матриці $V = \sigma_U^2 S$.

Оскільки в матриці S коваріація залишків ρ^S при $S > 2$ наближається до нуля, то матриця, обернена до матриці S , буде мати наступного вигляду:

$$S^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} \quad (10.10)$$

На практиці для розрахунку ρ використовується співвідношення:

$$\rho \approx r = \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2} \quad \text{або}$$

$$\rho \approx r' = \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2} \cdot \frac{n}{n-1}$$

2. Метод перетворення вихідної інформації

Цей метод складається з двох кроків:

1) перетворення вихідної інформації при застосуванні для цього параметра ρ ;

2) застосування методу ІМНК для оцінки параметрів моделі на основі перетворених даних.

Перетворення вихідної інформації виконується за допомогою матриць T_1 або T_2 :

$$T_1 = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

$$T_2 = \begin{pmatrix} -\rho & 1 & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ 0 & 0 & -\rho & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix},$$

тобто замість матриці X використовуємо $T_1 X$ або $T_2 X$, замість вектора Y — $T_1 Y$ або $T_2 Y$.

$$\sqrt{1-\rho^2} Y_1;$$

$$Y_2 - \rho Y_1;$$

$$Y_3 - \rho Y_2;$$

$$Y_4 - \rho Y_3;$$

.....

$$Y_n - \rho Y_{n-1}.$$

3. Метод Кочрена—Оркатта

Цей метод є ітеративним методом наближеного пошуку параметрів a_j і ρ ($j = \overline{0, m}$) і, які мінімізують суму квадратів залишків.

Коли економетрична модель має вигляд:

$$Y_t = a_0 + a_1 X_t + u_t \quad (10.11);$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad t = \overline{1, n}, \quad |\rho| < 1,$$

сума квадратів залишків запишеться так:

$$\sum_{t=2}^n \varepsilon_t = \sum [(Y_t - \rho Y_{t-1}) - a_0(1 - \rho) - a_1(X_t - \rho X_{t-1})]^2.$$

Алгоритм

Крок 1. Довільно вибирається значення $\rho < 1$ ($\rho = r_1$) і підставляється в формулу суми квадратів залишків.

Крок 2. На основі методу ІМНК знаходяться параметри $a_0^{(1)}$ і $a_1^{(1)}$.

Крок 3. Приймавши $a_0 = a_0^{(1)}$ і $a_1 = a_1^{(1)}$, підставимо у формулу суми квадратів залишків і розрахуємо $\rho = r_2$.

Крок 4. Підставивши $\rho = r_2$, розрахуємо параметри $a_0^{(2)}$ і $a_1^{(2)}$ і т.д.

Процедура продовжується доти, доки послідовні значення параметрів не будуть відрізнятися менше ніж на задану величину.

4. Метод Дарбіна

Цей метод базується на простій двокроковій процедурі.

Крок 1. Підставимо значення залишків, яке підкоряється авторегресійній моделі першого порядку

$$u_t = \rho u_{t-1} + \varepsilon_t,$$

$$\text{в економетричну модель } Y_t = a_0 + a_1 X_t + u_t,$$

$$\text{тоді одержимо: } Y_t = a_0 + a_1 X_t + \rho u_{t-1} + \varepsilon_t,$$

$$\text{де } u_{t-1} = Y_{t-1} - a_0 - a_1 X_{t-1}.$$

Звідси

$$Y_t = a_0(1-\rho) + \rho Y_{t-1} + a_1 X_t + a_1(\rho X_{t-1}) + \varepsilon_t.$$

ε_t в даному випадку має скалярну матрицю дисперсій:

$$M(\varepsilon_t \varepsilon_t') = \sigma_{\varepsilon_t}^2 E.$$

На основі методу 1МНК розраховуються оцінки параметрів $\hat{\rho}$, \hat{a}_0 , \hat{a}_1 .

Крок 2. Оцінка параметра $\hat{\rho}$ використовується для перетворення змінних $(Y_t - \hat{\rho} Y_{t-1})$ і $(X_t - \hat{\rho} X_{t-1})$, а метод 1МНК уже застосовується для перетворених даних. Прогноз на основі економетричної моделі з автокорельованими залишками виконується за формулою: $\hat{Y}_{n+1} = X_{n+1} A + \rho \varepsilon_n$.

Контрольні питання

1. Дайте означення автокореляції.
2. Які причини виникнення автокореляції залишків?
3. Як впливає автокореляція залишків на оцінку параметрів економетричної моделі?
4. Чим відрізняється метод оцінювання параметрів за методом Ейткена при автокореляції?
5. Запишіть матриці перетворення вихідної інформації згідно з двокроковою процедурою.
6. В яких випадках при автокореляції залишків доцільніше використовувати методи Кочрена — Оркатта або Дарбіна?
7. Дайте коротку характеристику алгоритму метода Кочрена — Оркатта.
8. Чим відрізняється метод Дарбіна від методу Кочрена — Оркатта?
9. Як записати формулу прогнозу залежної змінної при автокореляції залишків? Чому вона має такий вигляд?

Тема 11 Метод інструментальних змінних.

План

11.1. Властивості оцінок моделі при стохастичних змінних.

11.2. Метод інструментальних змінних.

11.3. Визначення інструментальних змінних.

Література [15, 17,18, 19, 21, 23, 25].

11.1 Властивості оцінок моделі при стохастичних змінних.

У попередніх розділах, розглядаючи модель $Y = XA + u$, ми виходили з припущення, що змінні X є детермінованими і набувають значення з деякої множини фіксованих чисел. Проте, виходячи з економічних досліджень, доцільно замінити це припущення на менш жорстке, згідно з яким змінні X є стохастичними. У такому разі постає запитання, чи справдзуватимуться й тоді основні результати, що стосуються перевірки значущості, довірчих інтервалів і т.ін. За умови, що змінні X мають функцію розподілу, жодним чином не пов'язану з параметрами a і σ_u^2 , і що всі ці змінні розподілені незалежно від залишків u , переважна більшість цих результатів виконуватиметься і для тих економетричних моделей, які мають стохастичну матрицю пояснювальних змінних X .

Нехай $\hat{a}^{(n)}$ — оцінка скалярного параметра a . Верхній його індекс вказує на розмір вибіркової сукупності, на основі якої оцінені ці параметри.

Означення 1. Сукупність оцінок $\hat{a}^{(n)}$ спостережень називається послідовністю оцінок:

$$\{a^{(n)}\} = a^{(1)}, a^{(2)} \dots a^{(n)}.$$

Означення 2. Якщо послідовність математичного сподівання параметрів $\hat{a} = M\{\hat{a}^{(n)}\}$ прямує до деякої константи, то ця константа є асимптотичним сподіванням, тобто $M(\hat{a}) = \lim_{n \rightarrow \infty} M\{a^{(n)}\}$.

Означення 3. Граничне значення послідовності дисперсій для $\hat{a}^{(n)}$ називається асимптотичною дисперсією

$$M[\hat{a}^{(n)} - Ma^{(n)}]^2 = \lim_{n \rightarrow \infty} M[\hat{a}^{(n)} - Ma^{(n)}]^2.$$

Оскільки для $n \rightarrow \infty$ вираз у правій частині може дорівнювати нулю, то дисперсія являє собою єдину точку, а саме $\sigma_{\hat{a}}^2$.

$$S^2(x'x)^{-1} = n^{-1}S^2\left(\frac{1}{n}x'x\right)^{-1}, \quad (11.4),$$

$$\text{де } S^2 = \frac{u'u}{n-m},$$

Дуже часто на практиці змінні X не можуть бути повністю незалежними від u , як це припускалося раніше. Наприклад, однією з пояснювальних змінних може бути лагове значення залежної змінної Y , що може призвести до зміщення оцінки ІМНК для кінцевих вибіркової сукупностей.

11.2. Метод інструментальних змінних

Якщо одна чи більше зі змінних X гранично корелює із залишками, тобто

$$P\lim\left\{\frac{1}{n}(x'u)\right\} \neq 0$$

то це означає, що оцінки ІМНК необґрунтовані. Зауважимо, що навіть коли один елемент вектора $P\lim\left\{\frac{1}{n}(x'u)\right\} \neq 0$, ми можемо дістати всі елементи вектора v необґрунтованими.

Кореляція між пояснювальними змінними і залишками є досить серйозною перепорою для застосування 1МНК. Така кореляція може виникнути з різних причин, але основними є три:

- 1) помилки вимірювання пояснювальних змінних;
- 2) побудова економетричної моделі за системою одночасових рівнянь;
- 3) наявність в економетричній моделі лагових змінних.

До лагових пояснювальних змінних відноситимемо такі змінні, які впливають на залежну змінну через певний проміжок часу. Наприклад, якщо залежна змінна в період t залежить від рівня тієї самої змінної в період $t-1$, то ця змінна входить до переліку пояснювальних змінних моделі, які в такому разі стають стохастичними. Вони включають лагову залежну змінну, яка є стохастичною і має зв'язок із залишками.

При існуванні кореляції між пояснювальними змінними і залишками можна застосувати поширений альтернативний метод оцінювання, який називається методом інструментальних змінних.

Розглянемо модель $y = xa + u$, для якої $P\lim\left\{\frac{1}{n}(x'u)\right\} \neq 0$.

Припустимо, що існує матриця Z порядку $n \times m$, яка має такі властивості: 1) $P\lim\left\{\frac{1}{n}(Z'u)\right\} = 0$;

2) $P\lim\left\{\frac{1}{n}(x'x)\right\} = \sum xx$, (11.10). де матриця $\sum Zx$ —

невироджена і, крім того, існує границя $P\lim\left\{\frac{1}{n}(Z'x)\right\} = \sum Zx$.

Отже, припускається, що змінні Z гранично некорельовані із залишками u , а їх перехресні моменти зі змінними X не всі дорівнюють нулю і створюють невірроджену матрицю. Якщо деякі зі змінних X не

корелюють із залишками u , то їх можна використовувати для формування стовпців матриці Z і знаходити додаткові інструментальні змінні лише для тих стовпців, що залишилися.

Оператор оцінювання вектора a з допомогою інструментальних змінних можна записати так:

$$\hat{a} = (Z'x)^{-1}Z'Y. \quad (11.5)$$

Щоб дістати його, помножимо ліворуч модель на $\frac{1}{n}Z'$: $\frac{1}{n}Z'Y = \frac{1}{n}Z'xa + \frac{1}{n}Z'u$.

Оскільки $P\lim(\frac{1}{n}Z'u) = 0$, то $Z'Y = Z'xa$.

Звідси дістаємо оператор оцінювання який забезпечує визначення обгрунтованої оцінки, у чому можна переконатися, підставивши, маємо: $\hat{a} = a + (Z'x)^{-1}Z'u$.

Асимптотична матриця коваріацій $asy \text{ var}(\hat{a}) = n^{-1}\sigma_u^2 \sum ZX^{-1} \sum ZZ \sum ZX^{-1}$

На практиці обчислюють так: $asy \text{ var}(\hat{a}) = \sigma_u^2 (Z'x)^{-1} (Z'Z) (x'Z)^{-1}$,

$$\text{де } \sigma_u^2 = \frac{u'u}{n-m}.$$

Реальна трудність застосування цього методу полягає в знаходженні змінних, які можна використовувати як інструментальні. Істиний розподіл їх встановити практично неможливо, а тому важко переконатися, що вибрані інструментальні змінні справді не корелюють із залишками. Водночас ці змінні повинні мати досить високу кореляцію зі змінними X , бо в противному разі вибіркові дисперсії для оцінок, здобутих за допомогою інструментальних змінних, будуть дуже великими.

Коротко вимоги до інструментальних змінних Z можна сформулювати так:

- 1) Z тісно пов'язані з X ;
- 2) Z зовсім не пов'язані із залишками u .

11.3. Визначення інструментальних змінних.

Розглядаючи способи визначення інструментальних змінних, скористаємося найпростішими економетричними моделями, які використовувались в різних операторах оцінок.

Оператор оцінювання Вальда

Нехай економетрична модель має вигляд

$$Y_t = a_0 + a_1 x_t + u_t. \quad (11.6)$$

У такому разі, якщо вибіркова сукупність містить парне число спостережень, то матриця інструментальних змінних Z запишеться так:

$$Z' = \begin{pmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ -1 & -1 & -1 & 1 & \dots & 1 \end{pmatrix}.$$

Щоб визначити другий рядок цієї матриці, необхідно виконати такі дії

1. Знайти відхилення кожного елемента вектора X від медіани. Матриця пояснювальних змінних для цієї моделі запишеться у вигляді:

$$x' = \begin{pmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & x_4 & \dots & x_n \end{pmatrix}.$$

Матриця інструментальних змінних на основі даної матриці замість рядка пояснювальної змінної міститиме рядок інструментальної.

2. Величини відхилень, що мають знак «плюс», замінюються на одиниці, величини відхилень, що мають знак «мінус», — на одиниці з цим знаком. Використовуючи оператор оцінювання $\hat{a} = (Z'x)^{-1} Z'Y$,

маємо:

$$\hat{a} = \begin{pmatrix} n & 0 \\ 0 & \frac{n}{2}(\bar{x}_2 - \bar{x}_1) \end{pmatrix} \begin{pmatrix} n\bar{Y} \\ \frac{n}{2}(\bar{Y}_2 - \bar{Y}_1) \end{pmatrix},$$

де \bar{x}_2 і \bar{x}_1 характеризують середні відхилення значень X відповідно вгору і вниз від медіани, а \bar{Y}_2 і \bar{Y}_1 — середні

значення залежної змінної, які відповідають середнім \bar{x}_2 і \bar{x}_1 . Звідси

$$\hat{a} = \left(\frac{\bar{Y} - \bar{Y}_1}{\bar{x}_2 - \bar{x}_1} \right),$$

Це означає, що параметр \hat{a}_1 у моделі подається у вигляді:

$$\hat{a}_1 = \frac{\bar{Y}_2 - \bar{Y}_1}{\bar{x}_2 - \bar{x}_1},$$

Причому $\hat{a}_0 = \bar{Y} - a_1 \bar{x}$. (11.7).

Коли вибірка сукупність містить непарне число спостережень, то перш ніж розпочинати обчислення, необхідно відкинути середнє спостереження.

При загальних припущеннях оцінка, яка здобута методом Вальда, є обґрунтованою, але її вибірка дисперсія може бути досить великою, тобто оцінка є неефективною.

Особливості оцінювання методом Бартлета

Бартлет показав, що ефективність оцінки можна збільшити, якщо розбити впорядковані значення змінної X на три групи однакового розміру. Перша з них містить найменші значення X , друга — середні, а третя — найбільші. Вилучивши середню групу — $n/3$ спостережень, дістанемо оцінку для параметра :

$$a_1 = \frac{\bar{Y}_2 - \bar{Y}_1}{\bar{x}_2 - \bar{x}_1}, \text{ де } \bar{x}_i, \bar{Y}_i \text{ — середні величини для}$$

спостережень, які потрапили в дві крайні групи. Вільний член оцінюється так само. Поділ вибіркової сукупності спостережень на три рівні групи має задовольняти вимоги прикладних досліджень, оскільки немає змоги дістати точну інформацію про закон розподілу значень X .

Оператор оцінювання Дарбіна

Дарбін запропонував упорядковувати значення вектора X в порядку зростання і ввів як інструментальну змінну порядковий номер (ранг), тобто числа 1, 2, 3, 4, ... n . Учений показав, що для великих вибірових сукупностей ефективність застосування такого методу оцінювання досягає майже 96 % від ефективності оцінок 1МНК, а для сукупностей $n = 20$ ефективність застосування такого методу оцінювання становить близько 86 %.

Модель Дарбіна не має вільного члена. Щоб застосувати його метод для оцінювання всіх параметрів моделі, у тому числі й для вільного члена, матриці змінних подамо у вигляді:

$$Z' = \begin{pmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ -1 & -1 & -1 & 1 & \dots & 1 \end{pmatrix} \quad \text{і} \quad x' = \begin{pmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & x_4 & \dots & x_n \end{pmatrix},$$

де x_i характеризують відхилення від середнього значення \bar{x} , які впорядковуються за зростанням.

$$\text{Оператор оцінювання} \quad \hat{a}_1 = \frac{\sum_{i=1}^n iY_i}{\sum_{i=1}^n ix_i}, \quad (11.8)$$

причому

$$\hat{a}_0 = \bar{Y} - \hat{a}_1 \bar{x} \quad (11.9),$$

де i — порядковий номер.

Дисперсія оцінок параметрів

$$\text{asy var}(\hat{a}) = \sigma_u^2 (Z'x)^{-1} (Z'Z)(x'Z)^{-1}.$$

Метод Дарбіна можна застосовувати і тоді, коли модель містить кілька пояснювальних змінних. У такому разі спочатку знаходяться відхилення значень кожної змінної од відповідного середнього значення. Потім ці відхилення упорядковуються за зростанням і кожному з них присвоюється порядковий номер.

Контрольні питання

1. Що таке детерміновані і стохастичні пояснювальні змінні?
2. Чи різнитимуться між собою методи оцінювання параметрів та перевірки їх значущості для стохастичних пояснювальних змінних моделі?
3. Що таке асимптотичні властивості оцінок параметрів моделі? Дайте основні визначення.
4. Покажіть обґрунтованість оцінок параметрів моделі при стохастичних пояснювальних змінних.
5. Коли оцінка параметрів моделі ІМНК стає необґрунтованою?
6. Якщо пояснювальні змінні X корелюють із залишками u , то який метод оцінювання параметрів моделі доцільно застосувати?
7. Запишіть оператор оцінювання за методом інструментальних змінних (МІЗ).
8. Які вимоги ставляться до інструментальних змінних?
9. Опишіть оператор оцінювання Вальда.
10. Назвіть особливості оцінювання методом Бартлета.
11. Дайте характеристику оператору оцінювання Дарбіна.

Тема 12. Моделі розподіленого лагу

План

- 12.1. Поняття лагу і лагових змінних.
- 12.2. Взаємна кореляційна функція.
- 12.3. Лаги залежних і незалежних змінних.
- 12.4. Методи оцінювання та прогнозу.

Література [15, 17, 21, 23, 25, 26, 27, 28].

12.1. Поняття лагу і лагових змінних.

Для багатьох економічних процесів типовим є те, що ефект від впливу деякого фактора на показник, який характеризує процес, виявляється не одразу, а поступово, через деякий період часу. Таке явище називається *лагом* (*запізненням*).

Потреба враховувати лаг при кількісному вимірюванні взаємозв'язку між економічними показниками постає дуже часто. Наприклад, у динамічних моделях необхідно враховувати лаг при визначенні зв'язку між обсягом продукції і капітальними вкладеннями, або частину цього лагу — будівельний.

Кількісне вираження взаємозв'язку між капітальними вкладеннями і введенням основних фондів, між витратами виробничих ресурсів і обсягом виробництва, між доходами і витратами тощо має базуватися на врахуванні впливу запізнення, або лагу. Причому вплив деяких пояснювальних змінних на залежну може проявлятися не лише через певний період часу, а й протягом певного часу, тобто лаг може складатись з кількох часових періодів. У такому разі маємо справу з економетричною моделлю розподіленого лагу.

Нехай економетрична модель розподіленого лагу визначається так:

$$y_t = \sum_{j=0}^{\infty} a_j x_{t-\tau} + u_t, \quad (12.1)$$

де a_j — параметри моделі при лагових змінних; $x_{t-\tau}$ — пояснювальна лагова змінна; τ — період зрушення; u_t — залишки, що розподілені нормально, тобто мають нульове математичне сподівання і стали дисперсію.

Означення 1. *Модель (12.1) називається загальною моделлю нескінченного розподіленого лагу, якщо для неї справджуються такі умови:*

- 1) $a_k a_j \geq 0$, для будь-яких k, j ;
 - 2) $a_j \geq 0, j = 1, 2, 3, \dots; k = 1, 2, 3, \dots$;
 - 3) $\sum_{j=0}^{\infty} a_j = w$, де w — скінченне число;
 - 4) $w_j = \frac{a_j}{w}$
 - 5) $\sum_{j=0}^{\infty} w_j = 1, 0 \leq w_j \leq 1$.
-

Означення 2. *Коефіцієнти $a_j, j = 0, 1, 2, \dots$, називаються коефіцієнтами лагу, а послідовність $a = \{a_j: j = 0, 1, 2, \dots\}$ — структурою лагу.*

Означення 3. *Якщо виконуються умови 1)-5), то величини $w_j, j = 0, 1, 2, \dots$, називаються нормованими коефіцієнтами лагу, а послідовність $w = \left\{ w_j: j = 0, 1, 2, \dots, \sum_{j=0}^{\infty} w_j = 1 \right\}$ — нормованою структурою лагу для моделі (12.1).*

Моделі розподілених лагів можуть задовільно описувати процеси лише в тому разі, коли забезпечена відносна стабільність умов, в яких ці процеси реалізуються. Може йтися про стабільність відповідних індексів цін, процентних ставок за кредити, норми амортизації, термінів будівництва, обсягу та структури ресурсів. Така стабільність далеко не завжди спостерігається для порівняно довгих проміжків часу, протягом яких формується сукупність спостережень. Усе це підводить до побудови *узгаальної моделі розподіленого лагу*, яка містить не лише лагові змінні, а й

інші фактори — пояснювальні змінні z_1, z_2, \dots, z_m , значення яких характеризують поточні умови функціонування економічних систем у період t .

Узагальнена модель розподіленого лагу задаватиметься рівнянням

$$y_t = \sum_{\tau \geq 0} a_\tau x_{t-\tau} + \sum_{s=1}^m b_s z_{t,s} + u_t.$$

Труднощі оцінювання параметрів такої моделі пов'язані з необхідністю враховувати обмеження на параметри a_τ .

12.2. Взаємна кореляційна функція.

Теоретично побудову моделі з розподіленими лагами можна узагальнити на будь-яку кількість незалежних змінних $x_{t-\tau}$. Але практична реалізація такої моделі часто стикається з непереборними труднощами, що зумовлені великою кількістю факторів, істотною обмеженістю часових рядів і складністю їх внутрішньої структури.

Як правило, до моделі входять такі змінні $x_{t-\tau}$, для яких лаги обґрунтовані теоретично і перевірені емпірично. Для обґрунтування лагу чи лагів доцільно використовувати *взаємну кореляційну функцію*. Ця функція характеризує тісноту зв'язку кожного елемента вектора залежної змінної y_t з елементом вектора незалежної x_t , зсунутим один відносно одного на часовий лаг τ .

$$r_{(\tau)} = \frac{(n-\tau) \sum_{t=1}^{n-\tau} y_t x_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} x_{t+\tau}}{\sqrt{\left[(n-\tau) \sum_{t=1}^{n-\tau} y_t^2 - \left(\sum_{t=1}^{n-\tau} y_t \right)^2 \right] \left[(n-\tau) \sum_{t=1}^{n-\tau} x_{t+\tau}^2 - \left(\sum_{t=1}^{n-\tau} x_{t+\tau} \right)^2 \right]}}. \quad (12.2)$$

Для різних значень τ на основі взаємної кореляційної функції можна дістати $n + 1$ значення $r_{(\tau)}$. Якщо $\tau = 0$, то

маємо парний коефіцієнт кореляції. Значення $r_{(\tau)}$ містяться на множині $r_{(\tau)} \in]-1, 1[$. Найбільше значення $r_{(\tau)}$ за модулем (найближче до одиниці) визначає зрушення, або часовий лаг. Якщо серед множини значень $r_{(\tau)}$ є кілька, величини яких наближаються до одиниці, то це означає, що запізнення впливу змінної x_t відбувається протягом певного проміжку часу і в результаті маємо кілька часових лагів для двох взаємопов'язаних часових рядів. Знайшовши часові лаги для визначення взаємозв'язку між економічними показниками, можна побудувати економетричну модель розподіленого лагу.

12.3. Лаги залежних і незалежних змінних.

Лаги незалежних змінних.

Наявність мультиколінеарності між лаговими змінними утруднює побудову економетричної моделі з лаговими змінними. Один зі способів звільнитись від мультиколінеарності — це ввести такі коефіцієнти при лагових змінних, які мали б однаковий знак і для них можна було знайти суму. З врахуванням умов 1)-5), модель з розподіленим лагом набере такого вигляду

$$y_t = a \sum_{j=0}^l w_j x_{t-\tau} + u_t \quad (12.3)$$

Л. Койк запропонував вибрати для зображення лагових коефіцієнтів форму спадної геометричної прогресії $w_j = (1-\lambda)\lambda^j$, де $0 \leq \lambda \leq 1$.

$$\text{Звідси } y_t = w(1-\lambda)x_t + \lambda y_{t-1} + (u_t - \lambda u_{t-1}). \quad (12.4)$$

Якщо через D позначити оператор зрушення, такий, що $Dx_t = x_{t-1}$, $D^2x_t = x_{t-2}$ і т.д., то вираз можна записати так: $w_0 + w_1D + w_2D^2 + \dots = (1-\lambda)(1 + \lambda D + \lambda^2 D^2 + \dots) = \frac{1-\lambda}{1-\lambda D}$.

З урахуванням цього модель матиме вигляд: $y_t = \frac{a(1-\lambda)}{1-\lambda D} x_t + u_t$.

Це припущення, що його зробив Койк, приводить до значних спрощень співвідношення Адже замість оцінки цілого ряду параметрів моделі a_j достатньо дати оцінки лише двох параметрів w і λ у рівнянні, де y_t розглядається як функція x_t і y_{t-1} . Діставши оцінку параметра λ і скориставшись співвідношенням можна обчислити всі вагові коефіцієнти. Середнє значення розподілу λ дорівнює $\frac{\lambda}{1-\lambda}$, тому для геометричного розподілу середній лаг $\bar{\tau} = \frac{\lambda}{1-\lambda}$. Входження до формули лагового значення змінної Y має забезпечити досить добру апроксимацію даної змінної. При цьому слід зауважити, що не завжди лаги розподілятимуться обов'язково за законом Койка, який забезпечує найближчому значенню X найбільшу вагу, а всім наступним — постійно спадні ваги. Якщо можна припустити, що це не так, то тоді лишається кілька перших вагових коефіцієнтів вільними, а для всіх інших використовується закон розподілу Койка.

Наприклад, можна записати

$$\begin{aligned} y_t &= a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \lambda a_2 x_{t-3} + \lambda^2 a_2 x_{t-4} + \dots + u_t = \\ &= a_0 x_t + a_1 x_{t-1} + \frac{a_2}{1-\lambda D} x_{t-2} + u_t, \end{aligned}$$

де перші два коефіцієнти лишаються вільними, а починаючи з a_2 вони спадають геометрично. Використаємо оператор зрушення D для скороченого запису моделі

Рівняння можна подати у вигляді

$$y_t = a_0x_t + (a_1 - \lambda a_0)x_{t-1} + (a_2 - \lambda a_1)x_{t-2} + a_2x_{t-2} + \lambda a_2x_{t-3} + \lambda^2 a_2x_{t-4} + \dots$$

$$\dots + u_t = a_0x_t + a_1x_{t-1} + \frac{a_2}{1 - \lambda D}x_{t-2} + \lambda y_{t-1} + (u_t - \lambda u_{t-1}).$$

Якщо модель має дві пояснювальні змінні, скажімо, X і Z , то розподілені лаги Койка можуть бути використані для кожної з них. Найпростіше припустити, що для обох змінних вибирається однакове значення λ .

Тоді модель розподіленого лагу

$$y_t = a(1 - \lambda)x_t + b(1 - \lambda)z_t + \lambda y_{t-1} + (u_t - \lambda u_{t-1}).$$

Якщо взяти параметри λ різними для різних пояснювальних змінних, то до моделі треба ввести змінні x_t , Z_t , y_t з оператором зрушення $Dx_t = x_{t-1}$, $D^2x_t = x_{t-2}$, $DZ_t = Z_{t-1}$, $D^2Z_t = Z_{t-2}$, $Dy_t = y_{t-1}$, $D^2y_t = y_{t-2}$:

$$y_t = a(1 - \lambda_1)x_t - a\lambda_2(1 - \lambda_1)x_{t-1} + b(1 - \lambda_2)z_t - b\lambda_1(1 - \lambda_2)z_{t-1} + (\lambda_1 + \lambda_2)y_{t-1} - \lambda_1 \cdot \lambda_2 y_{t-2} + [u_t - (\lambda_1 + \lambda_2)u_{t-1} + \lambda_1 \cdot \lambda_2 u_{t-2}].$$

Отже, припущення, зроблене Койком, спричинюється до появи в правій частині рівняння величин y_{t-1} і y_{t-2} . Причому для змінної y_{t-1} слід узяти суму параметрів λ_1 і λ_2 , а для змінної y_{t-2} — їх добуток. Аналогічно діють із залишками u_{t-1} і u_{t-2} .

Лаги залежної змінної

Зі щойно сказаного впливає: коли використовувати схему Койка для економетричної моделі, яка має лагові пояснювальні змінні, то в правій частині моделі серед таких змінних з'являється лагова залежна змінна $y_{t-\tau}$. З її появою стають стохастичними пояснювальні змінні моделі.

До появи в правій частині моделі лагових значень залежної змінної приводять і деякі інші моделі. Добре відомими моделями такого типу є модель **часткового**

коригування і модель **адаптивних сподівань**. Коли відсутнє повне уявлення про об'єкт, його інерційність, то застосовується метод часткового коригування. Розглянемо його.

Нехай $y_t^* = \alpha + \beta x_t$. У цьому рівнянні y_t^* розглядається як оптимальне значення y_t , яке відповідає x_t . Так, наприклад, якщо x_t — дохід, то y_t може визначати величину оптимальних витрат при доході x_t . Нехай величина доходу x_t різко змінюється (збільшується чи зменшується). При цьому споживчі витрати y_t можуть не змінитись адекватно доходу з різних причин: певна інерційність, недостатня інформація, договірні умови і т.ін. Тому в даному разі використаємо коригуючу функцію: $y_t - y_{t-1} = \gamma(y_t^* - y_{t-1}) + u_t$, $0 \leq \gamma \leq 1$,

яка вказує, що протягом поточного періоду часу буде пройдено лише частину відстані між вихідним станом y_{t-1} та оптимальним y_t^* . Об'єднавши дістанемо модель часткового коригування:

$$y_t = \alpha\gamma + \beta\gamma x_t + (1-\gamma)y_{t-1} + u_t, \quad 0 \leq \gamma \leq 1.$$

Ця залежність дуже схожа на кінцеве рівняння Койка. Вона відрізняється від нього наявністю вільного члена і простішою формою залишків.

Недоліком моделі часткового коригування є те, що оптимальне значення y_t^* не завжди визначається лише поточним значенням x_t , а й попередніми значеннями цієї змінної.

Якщо значення x_t змінюється від періоду до періоду, то оптимальне значення також змінюватиметься. Це явище знайшло своє відображення в моделі **адаптивних сподівань**, яка характеризує зв'язок змінної Y з очікуваним рівнем X . Позначимо його через x^* . Маємо $y_t = \alpha + \beta x_t^* + u_t$, де x_t^* — очікуване значення x_t , яке сформоване в поточний

момент часу, u_t — залишки, які можуть бути пов'язані з неточним вимірюванням значення змінної x_t^* .

Оскільки x_t^* — очікуване значення, то слід доповнити модель деякими припущеннями відносно формування очікуваного значення x_t^* . Загальноприйнятими в такому разі є припущення про адаптивні сподівання, які можна записати так: $x_t^* - x_{t-1}^* = \delta(x_t^* - x_{t-1}^*)$, $0 \leq \delta \leq 1$.

Це означає, що змінні, які спостерігатимуться протягом поточного періоду порівняно з очікуваними раніше, ураховуються лише частково, що й відображує у формулі додатне число δ , яке не перевищує одиниці. Щоб перейти до змінних x_t , які фактично спостерігаються, запишемо:

$x_t^* - \lambda x_{t-1}^* = \delta x_t$, де $\lambda = 1 - \delta$ Використовуючи оператор зрушення D , можна записати: $x_t^* = \frac{\delta}{1 - \lambda D} x_t$.

Підставимо це значення x_t^* $y_t = \alpha + \frac{\beta \delta}{1 - \lambda D} x_t + u_t$, помноживши обидві частини на $1 - \lambda D$, дістанемо:

$$y_t - \lambda D y_t = \alpha(1 - \lambda D) + \beta_\sigma x_t + u_t(1 - \lambda D), \text{ або}$$

$$y_t - \lambda y_{t-1} = \alpha(1 - \lambda) + \beta(1 - \lambda)x_t + u_t(1 - \lambda u_{t-1}).$$

Остаточно це рівняння матиме вигляд $y_t = \alpha(1 - \lambda) + \beta(1 - \lambda)x_t + \lambda y_{t-1} + u_t(1 - \lambda u_{t-1})$.

Останнє рівняння є простою моделлю **адаптивних сподівань**. Порівнявши його, побачимо, що воно має такі самі змінні, як і модель **часткового коригування**, відрізняється лише формуванням залишків. Модель **адаптивних сподівань** відрізняється від **схеми Койка** лише наявністю вільного члена.

Остаточні рівняння всіх трьох моделей практично збігаються, бо як у моделі адаптивних сподівань, так і в моделі часткового коригування використовуються вагові коефіцієнти, що спадають за геометричною прогресією.

12.4. Методи оцінювання та прогнозу.

Алгоритм Уоліса.

Розглянемо один з методів оцінювання, запропонований Уолісом трикроковий алгоритм.

Крок 1. Оцінюються параметри моделі

$$y_t = a_0 + a_1 y_{t-1} + a_2 x_t + v_t \quad (12.5),$$

де x_{t-1} використовується як інструментальна змінна для y_{t-1} . Таким чином, визначають: $\hat{A} = (Z'X)^{-1}Z'Y$, де

$$Z = \begin{pmatrix} 1 & x_0 & x_1 \\ 1 & x_1 & x_2 \\ 1 & x_2 & x_3 \\ \dots & \dots & \dots \\ 1 & x_{n-1} & x_n \end{pmatrix} \quad \text{і} \quad X = \begin{pmatrix} 1 & y_0 & x_1 \\ 1 & y_1 & x_2 \\ 1 & y_2 & x_3 \\ \dots & \dots & \dots \\ 1 & y_{n-1} & x_n \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_{n-1} \end{pmatrix}. \quad (12.6)$$

Крок 2. Для залишків цієї моделі $\hat{v} = Y - X\hat{A}$ розраховують коефіцієнт автокореляції першого порядку з

$$\text{урахуванням поправки на зміщення: } r = \frac{\sum_{t=2}^n \hat{v}_t \hat{v}_{t-1}}{\frac{n-1}{\sum_{t=1}^n \hat{v}_t^2} + \frac{m+1}{n}},$$

де $r \approx \rho$.

Крок 3. За допомогою оцінки, здобутої для ρ , формують матрицю:

$$S = \begin{pmatrix} 1 & r & r^2 & \dots & r^{n-1} \\ r & 1 & r & \dots & r^{n-2} \\ r^2 & r & 1 & \dots & r^{n-3} \\ \dots & \dots & \dots & \dots & \dots \\ r^{n-1} & r^{n-2} & r^{n-3} & \dots & 1 \end{pmatrix} \quad (12.7)$$

і обчислюють оцінку вектора \hat{A} узагальненим методом найменших квадратів $\hat{A} = (X'S^{-1}X)^{-1}X'S^{-1}Y$. Проведені Уолісом експерименти показали, що його метод оцінювання

приводить до значно менших величин зміщення і до меншої суми квадратів залишків, ніж застосування методу Ейткена.

Контрольні питання

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?

Тема 13. Системи одночасних структурних рівнянь.

План.

- 13.1. Структурна й наведена форми моделі.
 - 13.2. Проблема ідентифікації.
 - 13.3. Методи оцінки параметрів структурної форми моделі.
- Література [5, 8, 15, 23, 25].

13.1. Структурна й наведена форми моделі.

При використанні окремих рівнянь регресії, наприклад для економічних розрахунків, у більшості випадків передбачається, що аргументи (фактори) можна змінювати незалежно друг від друга. Однак це припущення є дуже грубим: практично зміну один змінної, як правило, не може відбуватися при абсолютній незмінності інших. Її зміна спричинить зміни у всій

період часу (*лаговые змінні*). Структурна форма моделі дозволяє побачити вплив змін будь-якої екзогенної змінної на значення ендогенної змінної. Доцільно в якості екзогенних змінних вибирати такі змінні, які можуть бути об'єктом регулювання. Міняючи їх і управляючи ними, можна заздалегідь мати цільові значення ендогенних змінних. Структурна форма моделі в правій частині містить при ендогенних змінні коефіцієнти b_{ik} й екзогенні змінних – коефіцієнти a_{ij} , які називаються *структурними коефіцієнтами* моделі. Всі змінні в моделі виражені у відхиленнях від середнього рівня, тобто під x мається на увазі $x - \bar{x}$, а під y – відповідно $y - \bar{y}$. Тому вільний член у кожному рівнянні системи (13.2) відсутній.

Використання МНК для оцінювання структурних коефіцієнтів моделі дає, як прийнято вважати в теорії, зміщені й неспроможні оцінки. Тому звичайно для визначення структурних коефіцієнтів моделі структурна форма моделі перетвориться в *наведену форму* моделі. Наведена форма моделі являє собою систему лінійних функцій ендогенних змінних від екзогенних. По своєму виді наведена форма моделі нічим не відрізняється від системи незалежних рівнянь, параметри якої оцінюються традиційним МНК. Застосовуючи МНК, можна оцінити значення ендогенних змінних через екзогенні. Коефіцієнти наведеної форми моделі являють собою нелінійні функції коефіцієнтів структурної форми моделі. Розглянемо це положення на прикладі найпростішої структурної моделі, виразивши коефіцієнти наведеної форми моделі через коефіцієнти структурної моделі. Для структурної моделі виду

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + \varepsilon_2 \end{cases} \quad (13.3)$$

наведена форма моделі має вигляд

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + u_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + u_2. \end{cases} \quad (13.4)$$

З першого рівняння (13.4) можна виразити y_2 в такий спосіб (заради спрощення опускаємо випадкову величину): $y_2 = \frac{y_1 - a_{11}x_1}{b_{12}}$. Підставляючи в друге рівняння

в (13.4), маємо $\frac{y_1 - a_{11}x_1}{b_{12}} = b_{21}y_1 + a_{22}x_2$, звідки

$$y_1 = \frac{a_{11}}{1 - b_{12}b_{21}}x_1 + \frac{a_{22}b_{12}}{1 - b_{12}b_{21}}x_2.$$

Надходячи аналогічно із другим рівнянням системи (13.5), одержимо $y_2 = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}}x_1 + \frac{a_{22}}{1 - b_{12}b_{21}}x_2$,

тобто система (13.4) приймає вид

$$\begin{cases} y_1 = \frac{a_{11}}{1 - b_{12}b_{21}}x_1 + \frac{a_{22}b_{12}}{1 - b_{12}b_{21}}x_2, \\ y_2 = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}}x_1 + \frac{a_{22}}{1 - b_{12}b_{21}}x_2. \end{cases} \quad (13.6)$$

Таким чином, можна зробити висновок про те, що коефіцієнти наведеної форми моделі будуть виражатися через коефіцієнти структурної форми в такий спосіб:

$$\delta_{11} = \frac{a_{11}}{1 - b_{12}b_{21}}, \quad \delta_{12} = \frac{a_{22}b_{12}}{1 - b_{12}b_{21}}, \quad (13.7)$$

$$\delta_{21} = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}}, \quad \delta_{22} = \frac{a_{22}}{1 - b_{12}b_{21}}.$$

Варто помітити, що наведена форма моделі хоча й дозволяє одержати значення ендогенної змінної через

значення екзогенних змінних, але аналітично вона уступає структурній формі моделі, тому що в ній відсутні оцінки взаємозв'язку між ендogenousними змінними.

13.2. Проблема ідентифікації

При переході від наведеної форми моделі до структурного економетрист зіштовхується із проблемою ідентифікації. Ідентифікація - це одиничність відповідності між наведеною й структурною формами моделі.

Структурна модель (2) у повному виді містить $m \cdot (m + n - 1)$ параметрів, а наведена форма моделі в повному виді містить $m \cdot n$ параметрів. тобто у повному виді структурна модель містить більше число параметрів, чим наведена форма моделі. Відповідно $m \cdot (m + n - 1)$ параметрів структурної моделі не можуть бути однозначно визначені з $m \cdot n$ параметрів наведеної форми моделі.

Щоб одержати єдино можливе рішення для структурної моделі, необхідно припустити, що деякі зі структурних коефіцієнтів моделі через слабкий взаємозв'язок ознак з ендogenousної змінної з лівої частини системи дорівнюють нулю. Тим самим зменшиться число структурних коефіцієнтів моделі. Зменшення числа структурних коефіцієнтів моделі можливо й другим шляхом: наприклад, шляхом прирівнювання деяких коефіцієнтів друг до друга, тобто шляхом припущень, що їхній вплив на формовану ендogenousну змінну однаково. На структурні коефіцієнти можуть накладатися, наприклад, обмеження виду

$$b_{ik} + a_{ij} = 0 \quad (13.8)$$

З позиції ідентифікованості структурні моделі можна підрозділити на три види:

- 1) ідентифіковані
- 2) неідентифіковані;

3) поверхідентифіковані

Модель *ідентифікована*, якщо всі структурні її коефіцієнти визначаються однозначно, єдиним образом за коефіцієнтами наведеної форми моделі, тобто якщо число параметрів структурної моделі дорівнює числу параметрів наведеної форми моделі. У цьому випадку структурні коефіцієнти моделі оцінюються через параметри наведеної форми моделі й модель ідентифікована.

Модель *неідентифікована*, якщо число наведених коефіцієнтів менше числа структурних коефіцієнтів, і в результаті структурні коефіцієнти не можуть бути оцінені через коефіцієнти наведеної форми моделі.

Модель *сверхідентифікована*, якщо число наведених коефіцієнтів більше числа структурних коефіцієнтів. У цьому випадку на основі коефіцієнтів наведеної форми можна одержати два або більше значення одного структурного коефіцієнта. У цій моделі число структурних коефіцієнтів менше числа коефіцієнтів наведеної форми. Сверхідентифікована модель на відміну від неідентифікованої практично розв'язувана, але вимагає для цього спеціальних методів обчислення параметрів.

Структурна модель завжди являє собою систему спільних рівнянь, кожне з яких потрібно перевіряти на ідентифікацію. Модель вважається ідентифікованою, якщо кожне рівняння системи ідентифіковано. Якщо хоча б одне з рівнянь системи неідентифіковано, то й всій моделі вважається не ідентифікованою. Поверхідентифікована модель містить хоча б одне поверхідентифіковане рівняння.

Виконання умови ідентифікованості моделі перевіряється для кожного рівняння системи. Щоб рівняння було ідентифікованим, необхідно, щоб число визначених змінних, відсутніх у даному рівнянні, але

присутніх у системі, було дорівнює числу ендогенних змінних у даному рівнянні без одного.

Якщо позначити число ендогенних змінних в i -м рівнянні системи через H , а число екзогенних (визначених) змінних, які втримуються в системі, але не входять у дане рівняння, — через D , то умова ідентифікованості моделі може бути записане у вигляді наступного рахункового правила:

Таблиця 1

$D + 1 = H$	рівняння ідентифіковано
$D + 1 < H$	рівняння неідентифіковане
$D + 1 > H$	рівняння поверхідентифіковане

Для оцінки параметрів структурної моделі система повинна бути ідентифікована або сверхідентифікована.

Розглянуте рахункове правило відображає необхідне, але недостатня умова ідентифікації. Більш точно умови ідентифікації визначаються, якщо накладати обмеження на коефіцієнти матриць параметрів структурної моделі. Рівняння ідентифіковане, якщо по відсутнім у ньому змінним (ендогенним і екзогенним) можна з коефіцієнтів при них в інших рівняннях системи одержати матрицю, визначник якої не дорівнює нулю, а ранг матриці не менше, ніж число ендогенних змінних у системі без одного.

Доцільність перевірки умови ідентифікації моделі через визначник матриці коефіцієнтів, відсутніх у даному рівнянні, але присутніх в інші, пояснюється тим, що можливо ситуацію, коли для кожного рівняння системи виконане рахункове правило, а визначник матриці названих коефіцієнтів дорівнює нулю. У цьому випадку

дотримується лише необхідне, але недостатня умова ідентифікації.

В економетричних моделях часто поряд з рівняннями, параметри яких повинні бути статистично оцінені, використовуються балансові тотожності змінних, коефіцієнти при яких рівні ± 1 . У цьому випадку, хоча сама тотожність і не вимагає перевірки на ідентифікацію, тому що коефіцієнти при змінних у тотожності відомі, у перевірці на ідентифікацію властиво структурних рівнянь системи тотожності беруть участь.

13.3. Методи оцінки параметрів структурної форми моделі

Коефіцієнти структурної моделі можуть бути оцінені різними способами залежно від виду системи одночасних рівнянь. Найбільше поширення в літературі одержали наступні методи оцінювання коефіцієнтів структурної моделі:

- 1) непрямий метод найменших квадратів;
- 2) двухшаговый метод найменших квадратів;
- 3) трехшаговый метод найменших квадратів;
- 4) метод максимальної правдоподібності з повною інформацією;
- 5) метод максимальної правдоподібності при обмеженій інформації.

Розглянемо коротенько сутність кожного із цих методів.

Непрямий метод найменших квадратів (НМНК) застосовується у випадку точно ідентифікованої структурної моделі. Процедура застосування НМНК припускає виконання наступних етапів роботи.

1. Структурна модель перетворюється в наведену форму моделі.

2. Для кожного рівняння наведеної форми моделі звичайним МНК оцінюються наведені коефіцієнти δ_{ij} .

3. Коефіцієнти наведеної форми моделі трансформуються в параметри структурної моделі.

Якщо система свержідентифікована, то КМНК не використовується, тому що він не дає однозначних оцінок для параметрів структурної моделі. У цьому випадку можуть використовуватися різні методи оцінювання, серед яких найпоширеніший і простим є двокроковий метод найменших квадратів (2 МНК). Основна ідея 2 МНК – на основі наведеної форми моделі одержати для поверхідентифікованого рівняння теоретичні значення ендогенних змінних, що втримуються в правій частині рівняння. Далі, підставивши їх замість фактичних значень, можна застосувати звичайний МНК до структурної форми поверхідентифікованої рівняння. Метод одержав назву двухшагового МНК, тому що двічі використовується МНК: на першому кроці при визначенні наведеної форми моделі й знаходженні на її основі оцінок теоретичних значень ендогенної змінної

$$\hat{y}_i = \delta_{i1}x_1 + \delta_{i2}x_2 + \dots + \delta_{in}x_n \quad (13.9)$$

й на другому кроці стосовно до структурного поверхідентифікованого рівнянню при визначенні структурних коефіцієнтів моделі по даним теоретичних (розрахункових) значень ендогенних змінних.

Поверхідентифікована структурна модель може бути двох типів:

- 1) всі рівняння системи поверхідентифіковані;
- 2) система містить поряд зі поверхідентифікованими точно ідентифіковані рівняння.

Якщо всі рівняння системи поверхідентифіковані, то для оцінки структурних коефіцієнтів кожного рівняння

використовується 2 МНК. Якщо в системі є точно ідентифіковані рівняння, то структурні коефіцієнти по них перебувають із системи наведених рівнянь.

Непрямий і двухроковий методи найменших квадратів докладно описані в літературі й розглядаються як традиційні методи оцінки коефіцієнтів структурної моделі. Ці методи досить легко реалізовані.

Метод максимальної правдоподібності розглядається як найбільш загальний метод оцінювання, результати якого при нормальному розподілі ознак збігаються із МНК. Однак при великій кількості рівнянь системи цей метод приводить до досить складних обчислювальних процедур. Тому як модифікація використовується метод максимальної правдоподібності при обмеженій інформації (метод найменшого дисперсійного відношення), розроблений в 1949 р. Т.Андерсоном і Н.Рубиним.

Подальшим розвитком 2 МНК є трьохроковий МНК (3 МНК), запропонований в 1962 р. А. Зельнером і Г.Тейлом. Цей метод оцінювання придатний для всіх видів рівнянь структурної моделі. Однак при деяких обмеженнях на параметри більше ефективним виявляється 2 МНК.

Контрольні питання

1. Які основні причини використання систем одночасних рівнянь?
2. У чому полягає основна розбіжність між структурними рівняннями системи і рівняннями у зведеній 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)?

Глосарій

- **Автокореляція** - це кореляційна залежність рівнів ряду від попередніх значень.
- **Визначені змінні** - це екзогенні змінні системи й лагові ендогенні змінні системи.
- **Відхилення**- це чисто випадковий процес у вигляді дрібномасштабних коливань, що не містить уже детермінованого компонента, який є в регресії.
- **Внутрішня нелінійна регресія** - це істинно нелінійна регресія, що не може бути приведена до лінійної регресії перетворенням змінних і введенням нових змінних.
- **Гетероскедастичність** - порушення сталості дисперсії для всіх спостережень.
- **Гомоскедастичність** - сталість дисперсії для всіх спостережень, або однаковість дисперсії кожного відхилення (залишку) для всіх значень факторних змінних.
- **Дисперсія** - показник варіації, який характеризує відхилення значення показника від його середнього значення.

- **Екзогенні змінні** - це змінні, які визначаються поза системою і є незалежними.
- **Економетрична модель** - це рівняння або система рівнянь, особливим образом залежності, що представляють залежність між результатом і факторами. В основі економетричної моделі лежить розбивка складної й малозрозумілої залежності між результатом і факторами на суму двох наступних компонентів: регресію (регресійний компонент) і випадковий (флуктуаційний) - залишок. Інший клас економетричних моделей утворить тимчасові ряди.
- **Ендогенні змінні** - це залежні змінні від екзогенних, визначаються усередині самої системи.
- **Ефективність оцінки** - це властивість оцінки мати найменшу дисперсію із всіх можливих.
- **Значимість рівняння регресії** - дійсна наявність досліджуваної залежності, а не просто випадковий збіг факторів, що імітує залежність, яка фактично не існує.
- **Залишкова дисперсія** - не пояснена дисперсія, що показує варіацію результату під впливом всіх інших факторів, не врахованих регресією.
- **Інтеркореляція** і пов'язана з нею мультиколінеарність - це наближення до повної лінійної залежності, тісний зв'язок між факторами.
- **Кореляція** - стохастична залежність, що є узагальненням строго детермінованої функціональної залежності за допомогою включення імовірнісної (випадкової) компоненти.
- **Коефіцієнт детермінації** - показник тісноти стохастичного зв'язку в загальному випадку нелінійної регресії.
- **Коефіцієнт довіри** - це коефіцієнт, що зв'яже лінійною залежністю граничну і середню помилки,

з'ясовує зміст граничної помилки, що характеризує точність оцінки, і є аргументом розподілу (найчастіше, інтеграла ймовірностей). Саме ця ймовірність і є ступінь надійності оцінки.

- **Коефіцієнт довіри (нормоване відхилення)** - результат ділення відхилення від середнього на стандартне відхилення, змістовно характеризує ступінь надійності отриманої оцінки.
- **Коефіцієнт лінійної кореляції** - показник тісноти стохастичного зв'язку між фактором і результатом у випадку лінійної регресії.
- **Коефіцієнт регресії** - коефіцієнт при факторній змінній у моделі лінійної регресії.
- **Критерій Ст'юдента** - перевірка значимості окремих коефіцієнтів регресії й значимості коефіцієнта кореляції.
- **Критерій Фішера** - спосіб статистичної перевірки значимості рівняння регресії, при якому розрахункове (фактичне) значення F-критерію рівняється з його критичним (теоретичним) значенням.
- **Лагові змінні** - це змінні, стосовні до попередніх моментів часу.
- **Лінійна регресія** - це зв'язок (регресія), що представлена рівнянням прямої лінії й виражає найпростішу лінійну залежність.
- **Метод інструментальних змінних** - це різновид МНК. Використовується для оцінки параметрів моделей, описуваних декількома рівняннями. Головна властивість - часткова заміна непридатної пояснюючої змінної на таку змінну, котра некоррелирована з випадковим членом. Ця змінна називається інструментальною і приводить до одержання заможних оцінок параметрів.

- **Метод найменших квадратів (МНК)** - спосіб наближеного знаходження (оцінювання) невідомих коефіцієнтів (параметрів) регресії. Цей метод заснований на вимозі мінімізації суми квадратів відхилень значень результату, розрахованих по рівнянню регресії, і спостережених значень результату.
- **Множинна лінійна регресія** - це множинна регресія, що представляє лінійний зв'язок по кожному факторі.
- **Множинна регресія** - регресія із двома й більше факторними змінними.
- **Модель ідентифікована** - модель, у якій всі структурні коефіцієнти однозначно визначаються за коефіцієнтами наведеної форми моделі.
- **Модель рекурсивних рівнянь** - модель, що містить залежні змінні (результативні) одних рівнянь у ролі фактору, виявляючись у правій частині інших рівнянь.
- **Наведена форма системи** - форма, що, на відміну від структурної, уже містить одні тільки лінійно залежні від екзогенних змінних ендогенні змінні. Зовні нічим не відрізняється від системи незалежних рівнянь.
- **Незміщена оцінка** - оцінка, середнє якої дорівнює самій оцінюваній величині.
- **Нульова гіпотеза** - припущення про те, що результат не залежить від фактору (коефіцієнт регресії дорівнює нулю).
- **Пояснена дисперсія** - показник варіації результату, обумовленою регресією.
- **Результативна змінна** - змінна, котра статистично залежить від факторної змінної, або пояснюючої (регресора).

- **Розрахункове значення F-критерію**- значення, що одержують діленням поясненої дисперсії на 1 ступінь волі на залишкову дисперсію на 1 ступінь волі.
- **Регресія (залежність)** - це усереднена (згладжена), тобто вільна від випадкових дрібномасштабних коливань (флуктуацій), квазидетерминированная зв'язок між пояснюва_ змінною (змінними) і пояснюючої змінної (змінними). Цей зв'язок виражається формулами, які характеризують функціональну залежність і не містять явно стохастических (випадкових) змінних, які свій вплив тепер роблять як результуючий вплив, що приймає вид чисто функціональної залежності.
- **Регресор** (пояснююча змінна, факторна змінна) - це незалежна змінна, статистично пов'язана з результуючої змінної. Характер цього зв'язку й вплив зміни (варіації) регресора на результат досліджуються в економетриці.
- **Рівень значимості** - величина, що показує, яка ймовірність помилкового висновку при перевірці статистичної гіпотези за статистичним критерієм.
- **Система взаємозалежних рівнянь** - це система одночасних або взаємозалежних рівнянь. У ній ті самі змінні виступають одночасно як залежні в одних рівняннях і в той же час незалежні в інших. Це структурна форма системи рівнянь. До неї не застосуємо МНК.
- **Система зовні не зв'язаних між собою рівнянь** - система, що характеризується наявністю одних тільки кореляцій між залишками (помилками) у різних рівняннях системи.
- **Специфікація моделі** - визначення істотних факторів і виявлення мультиколінеарності.

- **Стандартна помилка** - середньоквадратичне (стандартне) відхилення. Воно зв'язане зі середньою помилкою і коефіцієнтом довіри.
- **Ступінь волі** - це величини, що характеризують число незалежних параметрів і необхідні для знаходження по таблицях розподілів їхніх критичних значень.
- **T-відношення (t-критерій)** - відношення оцінки коефіцієнта, отриманої за допомогою МНК, до величини стандартної помилки оцінюваної величини.
- **Тренд** - основна тенденція розвитку, плавна стійка закономірність зміни рівнів ряду.
- **Узагальнений метод найменших квадратів (УМНК)** - метод, що не вимагає сталості дисперсії (гомоскедастичності) залишків, але припускає пропорційність залишків загальному множнику (дисперсії). Таким чином, це зважений МНК.
- **Фіктивні змінні** - це змінні, які відбивають сезонні компоненти ряду для будь-якого одного періоду.
- **Часовий ряд** - це послідовність значень ознаки (результативного змінного), прийнятих протягом послідовних моментів часу або періодів.

СПИСОК ВИКОРИСТАНОЇ ТА РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

Основна література

1. Айвазян С.А. Прикладная статистика и основы эконометрики: Учебник для вузов/ С. А. Айвазян, В. С. Мхитарян - М.: ЮНИТИ, 1998. - 1022 с.
2. Айвазян С.А. Основы эконометрики: Учебник . Т.2 / С. А. Айвазян. - 2-е изд.; испр. - М.: ЮНИТИ, 2001. - (Прикладная статистика. Основы эконометрики. В 2-х т.).
3. Бородич С. А. Эконометрика: Учеб. Пособие / С. А. Бородич — Минск: Новое знание, 2001.- 408 с.
4. Грубер Й. Економетрія: Вступ до множинної регресії та економетрії: / Й. Грубер- К: Нічлава, 1999.
5. Домбровский В. В. Эконометрика: Учебник / В. В. Домбровский. - М.: Новый учебник, 2004. - 342с
6. Джонстон Дж. Эконометрические методы./ Дж. Джонстон — М.: Статистика, 1990 — 444 с.
7. Доугерти К. Введение в эконометрику: Пер. с англ./ К. Доугерти — М.: ИНФРА-М, 1997. - 402с.
8. Емельянов А. С. Эконометрия и прогнозирование/ А. С. Емельянов — М.: Экономика, 1999. -267 с.
9. Слейко В. Основы економетрії / Слейко В. — Львів: "Марка Лтд", 1995. — 191с.
10. Замков О. О., Толстопятенко А. В., Черемних Ю. Н.

Математические методы в экономике: Учебник/ О. О. Замков, А. В. Толстопятенко, Ю. Н. Черемних / Под общ. ред. А. В. Сидоровича— М.: Дело и Сервис, 2001. — 368 с.

11. Корольов О. А. Економетрія: Навч. посіб. / О. А. Корольов — К: Європейський ун-т, 2002.- 660 с.

12. Кремер Н.Ш. Эконометрика: Учебник / Н. Ш. Кремер, Б. А. Путко; [ВЗФЭИ]; Под ред. Н.Ш. Кремера. - М.: ЮНИТИ, 2002. - 311с

13. Лате О. Введение в эконометрию. — М.: Прогрессе, 1964. — 360 с.

14. 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.

15. Лещинський О.Л. Економетрія: Навч. посібник для студ. вищ. навч. закл. / О. Л. Лещинський, В. В. Рязанцева, О. О. Юнькова. — К.: МАУП, 2003. — 208 с.

16. Лук'яненко І. Г., Краснікова Л. І. Економетрика: Підручник/ І. Г. Лук'яненко, Л. І. Краснікова — К.: Т-во "Знання", КОО, 1998. - 494 с.

17. Наконечний С. І., Терещенко Т. О., Романюк Т. П. Економетрія: Навч. посіб. - К: КНЕУ, 2004. - 352 с.

18. Тихомиров Н.П. Эконометрика: Учебник / Н. П. Тихомиров, Е. Ю. Дорохина. - М.: Экзамен, 2003. - 512с.

19. Толбатов Ю. А. Эконометрика: Підруч. для студ. екон. спец. вищ. навч. закл./ Ю. А. Толбатов — К.: Четверта хвиля, 2002. — 320 с.

Додаткова література

1. Айвазян С.А. Теория вероятностей и прикладная статистика: Учебник. Т.1 / С. А. Айвазян, В. С. Мхитарян. - 2-е изд.; испр. - М.: ЮНИТИ, 2001. - 656с. - (Прикладная

- статистика. Основы эконометрики. В 2-х т., Т.1).
2. Балдин К.В. Эконометрика: Учебное пособие / К. В. Балдин, О. Ф. Быстров, М. М. Соколов. - 2-е изд.; перераб. и доп. - М.: ЮНИТИ, 2004. - 254с. .
 3. Дрейпер Н., Смит Г. Прикладной регрессионный анализ / Н. Дрейпер, Г. Смит — М.: Финансы и статистика, 1986. — Т. 1 — 365 с; Т. 2 — 379 с.
 4. Клейнер Г.Б. Эконометрические зависимости: Принципы и методы построения / Г. Б. Клейнер, С. А. Смоляк. - М.: Наука, 2000. - 104с.
 5. Корольов О. А. Економетрія в задачах, ситуаціях та проблемах для студентів спеціальності «Фінанси і кредит»: Конспект лекцій і практикум: У 3 ч. —К: КДТЕУ, 1997.
 6. Носко В.П. Эконометрика. Элементарные методы и введение в региональный анализ временных рядов = Econometrics. / В. П. Носко. - М.: ИЭПП, 2004. – 501 с.
 7. Орлова И. В. Экономико - математическое моделирование. Практическое пособие по решению задач / И. В. Орлова; ВЗФЭИ. - М.: Вузовский учебник, 2004. – 144 с.
 8. Практикум по эконометрике: Учебное пособие / Под ред. Елисейевой И.И. - М.: Финансы и статистика, 2004. – 192 с.
 9. Фомин Б. С. Эконометрические теории и модели международных экономических отношений / Фомин Б. С. — М.: Мысль, 1997 — 268 с.

Навчально-методичне видання

ФЕДЬКО Яна Вікторівна

ЕКОНОМЕТРІЯ

*Курс лекцій для студентів
денної та заочної форм навчання
спеціальностей «Прикладна статистика», «Менеджмент»*

За редакцією автора
Комп'ютерний макет – Федько Я. В.
Коректор – Безгодова Н. С.

Здано до склад. 20.12.2010 р. Підп. до друку 19.01.2011 р.
Формат 60×84 1/16. Папір офсет. Гарнітура Times New Roman.
Друк ризографічний. Ум. друк. арк. 8,72. Наклад 100 прим. Зам № 5.

Видавець і виготовлювач
Видавництво Державного закладу
«Луганський національний університет імені Тараса Шевченка»
вул. Оборонна, 2, м. Луганськ, 91011. Тел./факс: (0642) 58-03-20
e-mail: alma-mater@list.ru
Свідоцтво суб'єкта видавничої справи ДК № 3459 від 09.04.2009 р.