

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ
УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ
УНІВЕРСИТЕТ**

Руська Р. В.

ЕКОНОМЕТРИКА

Навчальний посібник

Тернопіль 2012

Руська Р. В. Економетрика : навчальний посібник / Р. В. Руська. – Тернопіль : Тайп, 2012. – 224с.

Затверджено на засіданні кафедри економіко-математичних методів, протокол №6 від 03 лютого 2012 р.

Рецензенти:

О. Б. Альохін – доктор економічних наук, професор кафедри «Інформаційні системи у менеджменті» Одеського національного політехнічного університету;

В. І. Єлейко – доктор економічних наук, професор, завідувач кафедри «Економетрії та статистики» Львівської комерційної академії

О. М. Ляшенко – доктор економічних наук, професор, завідувач кафедри «Економічної кібернетики» Тернопільського національного економічного університету

У навчальному посібнику розглянуто основні завдання економетричних досліджень і методи їх розв'язання, визначається місце та значення дисципліни «Економетрика», її зв'язок з іншими дисциплінами. Розглянуто основні принципи побудови економетричних моделей, оцінювання невідомих параметрів, перевірки моделей та аналізу отриманих результатів. Моделювання динаміки економічних процесів розглянуто на прикладі дистрибутивно-лагових і авто регресійних моделей. Економетричний аналіз проводиться за допомогою MS Excel

Посібник розрахований на студентів економічних спеціальностей, що вивчають економетрику. Вона може бути корисною також при розв'язанні широкого кола прикладних проблем, які виникають у практичній роботі.

ЗМІСТ

ВСТУП.....	8
ПРЕДМЕТ ТА МЕТОД ЕКОНОМЕТРИКИ.....	12
1.1 ПРЕДМЕТ ТА МЕТОД ЕКОНОМЕТРИКИ.....	12
1.2 ІСТОРИЧНІ ВІДОМОСТІ.....	13
1.3 ПРИКЛАДИ МОДЕЛЕЙ ТА МЕТОДІВ, ЯКІ НОСЯТЬ ХАРАКТЕР ЕКОНОМЕТРИЧНИХ ДОСЛІДЖЕНЬ.....	15
1.4 ЗНАЧЕННЯ КУРСУ ТА ВЗАЄМОЗВ'ЯЗОК З ІНШИМИ ЕКОНОМІЧНИМИ ДИСЦИПЛІНАМИ.....	19
1.5 МАТЕМАТИЧНА МОДЕЛЬ ТА ОСНОВНІ ЕТАПИ ЇЇ ПОБУДОВИ.....	20
1.6 ТЕОРЕТИЧНІ ОСНОВИ МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ ТА КЛАСИФІКАЦІЯ МОДЕЛЕЙ.....	22
1.7 ОСНОВНІ ЯКІСНІ ТА КІЛЬКІСНІ ХАРАКТЕРИСТИКИ СИСТЕМИ ОБЛІКУ ЯК ОБ'ЄКТУ МОДЕЛЮВАННЯ.....	24
ОДНОФАКТОРНА ЛІНІЙНА ЕКОНОМЕТРИЧНА МОДЕЛЬ.....	27
2.1 РЕГРЕСІЙНА ЕКОНОМЕТРИЧНА МОДЕЛЬ.....	27
2.2 ЗНАХОДЖЕННЯ СТАТИСТИЧНИХ ОЦІНОК ПАРАМЕТРІВ МЕТОДОМ НАЙМЕНШИХ КВАДРАТІВ.....	32
МЕТОДИКА ВИЗНАЧЕННЯ ПАРАМЕТРІВ ОДНОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ.....	39
СТАТИСТИЧНА ПЕРЕВІРКА ОЦІНОК ОДНОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ.....	44
3.1 СТАНДАРТНА ПОХИБКА ОЦІНКИ ЗА РІВНЯННЯМ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ.....	44
3.2 КОЕФІЦІЄНТ ДЕТЕРМІНАЦІЇ ТА КОЕФІЦІЄНТ КОРЕЛЯЦІЇ.....	48
3.3 ЗАГАЛЬНІ ВІДОМОСТІ ПРО СТАТИСТИЧНІ ОЦІНКИ. НЕЗМІЩЕНІСТЬ І ЕФЕКТИВНІСТЬ ОЦІНОК МНК.....	50
3.4 ПЕРЕВІРКА НУЛЬОВИХ ГІПОТЕЗ. ПОБУДОВА ІНТЕРВАЛІВ ДОВІР'Я РІВНЯННЯ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ.....	56
3.4 ПЕРЕВІРКА НУЛЬОВИХ ГІПОТЕЗ І ДОВІРЧИ ІНТЕРВАЛИ ПАРАМЕТРІВ. ПЕРЕВІРКА МОДЕЛІ НА АДЕКВАТНІСТЬ.....	62
МЕТОДИКА ПЕРЕВІРКИ ОЦІНОК ОДНОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ.....	65
ОДНОФАКТОРНІ НЕЛІНІЙНІ ЕКОНОМЕТРИЧНІ МОДЕЛІ.....	70
4.1 КРИВІ ЗРОСТАННЯ. ЗВЕДЕННЯ НЕЛІНІЙНИХ МОДЕЛЕЙ ДО ЛІНІЙНИХ.....	70
4.2 ПРИКЛАДИ ЗАСТОСУВАННЯ ЕКСПОТЕНЦІЙНИХ ФУНКЦІЙ.....	74
МЕТОДИКА ЗНАХОДЖЕННЯ ОЦІНОК НЕЛІНІЙНИХ ОДНОФАКТОРНИХ МОДЕЛЕЙ.....	77

КЛАСИЧНА ЛІНІЙНА БАГАТОФАКТОРНА МОДЕЛЬ	80
5.1 ЛІНІЙНА БАГАТОФАКТОРНА ЕКОНОМЕТРИЧНА МОДЕЛЬ.	80
5.2 МНК ДЛЯ БАГАТОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ.	81
5.3 КОЕФІЦІЄНТ МНОЖИННОЇ КОРЕЛЯЦІЇ ТА ДЕТЕРМІНАЦІЇ. ОЦІНЕНИЙ КОЕФІЦІЄНТ ДЕТЕРМІНАЦІЇ.....	83
МЕТОДИКА ПОБУДОВИ БАГАТОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ.....	84
МАТРИЧНИЙ ПІДХІД ДО ЛІНІЙНОЇ БАГАТОФАКТОРНОЇ МОДЕЛІ	89
6.1 ПОСТАНОВКА ЗАДАЧІ В МАТРИЧНІЙ ФОРМІ ТА ОСНОВНІ ПРИПУЩЕННЯ МНК ДЛЯ ЗАГАЛЬНОГО ВИПАДКУ. МНК В МАТРИЧНІЙ ФОРМІ.	89
6.2 ДИСПЕРСІЙНО-КОВАРІАЦІЙНА МАТРИЦЯ VAR(a) МАТРИЦЯ КОРЕЛЯЦІЇ.	92
6.3 ПЕРЕВІРКА МОДЕЛІ НА АДЕКВАТНІСТЬ. ПЕРЕВІРКА НУЛЬОВИХ ГІПОТЕЗ І ДОВІРЧІ ІНТЕРВАЛИ ПАРАМЕТРІВ.	96
6.4 ПЕРЕВІРКА НУЛЬОВОЇ ГІПОТЕЗИ СТОСОВНО КОЕФІЦІЄНТА МНОЖИННОЇ КОРЕЛЯЦІЇ. ПРОГНОЗУВАННЯ ЗА ЕКОНОМЕТРИЧНОЮ МОДЕЛЛЮ.	98
МЕТОДИКА ЗАСТОСУВАННЯ МАТРИЧНОГО МЕТОДУ ДО БАГАТОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ.....	99
ЧАСОВІ РЯДИ І ПРОГНОЗУВАННЯ.....	108
7.1 КЛАСИФІКАЦІЯ ЧАСОВИХ РЯДІВ.	108
7.2 СТАЦІОНАРНІ ЧАСОВІ РЯДИ ТА ЇХ ХАРАКТЕРИСТИКИ, АВТОКОРЕЛЯЦІЙНА ФУНКЦІЯ.....	111
7.3 АНАЛІТИЧНЕ ВИРІВНЮВАННЯ ЧАСОВОГО РЯДУ. ПРОГНОЗУВАННЯ НА ОСНОВІ МОДЕЛІ ЧАСОВИХ РЯДІВ.	114
МЕТОДИКА ДОСЛІДЖЕННЯ ЧАСОВОГО РЯДУ	118
МУЛЬТИКОЛІНЕАРНІСТЬ.....	125
8.1 МУЛЬТИКОЛІНЕАРНІСТЬ І ЇЇ НАСЛІДКИ.	125
8.2 ДОСЛІДЖЕННЯ МУЛЬТИКОЛІНЕАРНОСТІ.....	127
8.3 СПОСОБИ УСУНЕННЯ МУЛЬТИКОЛІНЕАРНОСТІ.	129
МЕТОДИКА ПЕРЕВІРКИ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ НА МУЛЬТИКОЛІНЕАРНІСТЬ.....	132
ГЕТЕРОСКЕДАСТИЧНІСТЬ.....	139
9.1 ПОНЯТТЯ ГЕТЕРОСКЕДАСТИЧНОСТІ. МЕТОДИ ВИЗНАЧЕННЯ ГЕТЕРОСКЕДАСТИЧНОСТІ.	139
9.2 УЗАГАЛЬНЕНИЙ МЕТОД НАЙМЕНШИХ КВАДРАТІВ.	143
МЕТОДИКА ПЕРЕВІРКИ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ НА НАЯВНІСТЬ ГЕТЕРОСКЕДАСТИЧНОСТІ/	146
АВТОКОРЕЛЯЦІЯ	153

10.1 ПРИЧИНИ ВИНИКНЕННЯ АВТОКОРЕЛЯЦІЇ В ЕКОНОМЕТРИЧНИХ МОДЕЛЯХ. ПЕРЕВІРКА НАЯВНОСТІ АВТОКОРЕЛЯЦІЇ.....	153
10.2 ОЦІНКА ПАРАМЕТРІВ МОДЕЛІ З АВТОКОРЕЛЬОВАНИМИ ЗАЛИШКАМИ.	156
МЕТОДИКА ПЕРЕВІРКИ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ НА НАЯВНІСТЬ АВТОКОРЕЛЯЦІЇ.....	164
АВТОРЕГРЕСІЙНІ І ДИСТРИБУТИВНО-ЛАГОВІ МОДЕЛІ.	168
11.1 ПРИРОДА АВТОРЕГРЕСІВНИХ МОДЕЛЕЙ.....	168
11.2 МЕТОДИ ОЦІНЮВАННЯ ПАРАМЕТРІВ ЛАГОВОЇ МОДЕЛІ.	174
МЕТОДИКА РОЗВ'ЯЗАННЯ ЗАДАЧ	188
DUMMY-ЗМІННІ.....	190
12.1 ПРИРОДА DUMMY-ЗМІННИХ.	190
12.2 РЕГРЕСІЯ ОДНІЄЇ КІЛЬКІСНОЇ ТА ОДНІЄЇ ЯКІСНОЇ ЗМІННОЇ ДВОХ КЛАСІВ, АБО КАТЕГОРІЙ.	191
12.3 РЕГРЕСІЯ КІЛЬКІСНОЇ ЗМІННОЇ ТА ОДНІЄЇ ЯКІСНОЇ ЗМІННОЇ З БІЛЬШ НІЖ ДВОМА КЛАСАМИ.	195
МЕТОДИКА РОЗВ'ЯЗАННЯ ЗАДАЧ	198
ЛІТЕРАТУРА	208
ДОДАТКИ.....	210
ДОДАТОК 1	210
ДОДАТОК 2	217

ВСТУП

Економічна теорія вивчає причинно-наслідкові відносини між явищами і процесами на якісному рівні. Однак існуючі явища і процеси пов'язані між собою, і знаходження залежності і взаємозв'язків між ними шляхом створення математичних моделей і наступного їхнього кількісного опису дозволяє глибше зрозуміти існуючі закономірності.

Математичне моделювання одержало широке поширення в різних областях знань: механіці, фізиці, медицині, біології, хімії, у тому числі й в економіці. Розрізняють детерміністичне і стохастичне моделювання. Якщо у відношенні величин, аналізованих детерміністичними моделями, передбачається їхня стабільність, а випадковими відхиленнями зневажають, відносячи їх тільки на рахунок помилок спостережень і вимірів, то в основі стохастичних моделей лежить випадковий характер величин, оцінюваний імовірними методами. Тут задача полягає у відшуканні тенденцій, що виявляються у випадкових відхиленнях.

Детерміністичні моделі описують закономірності, що виявляються в одиночному, у кожному окремо узятому елементі сукупності. Такій закономірності властива тверда механічна причинність, що конкретно визначає поведінку кожної одиниці сукупності. Вона одержала назву динамічної або закономірності з твердою детермінацією. Типовими прикладами динамічних закономірностей можуть служити закони класичної механіки. У динамічних закономірностях зв'язок між причиною і наслідком може бути виражений цілком точно у вигляді конкретних математичних формул. Кожному набору значень пояснюючих змінних завжди відповідає визначене значення пояснюваної змінної. Такий зв'язок називається функціональним. Детерміністична модель служить вираженням функціонального зв'язку.

У централізованій плановій економіці цілком вдовольнялись детерміністичними моделями. Заздалегідь був відомий результат, і теорія призначалася для його обґрунтування. В основному використовувалися

балансові або оптимізаційні моделі: міжгалузевого балансу і лінійного програмування.

Однак в умовах ринкової економіки результат заздалегідь невідомий, і не враховувати фактор випадковості просто неможливо. Економічні явища і процеси являють собою результат багатьох одночасно і сукупно діючих причин. При розгляді зв'язків між ними головні причини, що обов'язково приводять до даного наслідку, необхідно відрізнити від другорядних. Останні ускладнюють і спотворюють дію істотних у даному аспекті причин. Крім того, причини можуть мати і непередбачений характер.

Щоб відокремити істотні фактори, що діють на досліджуваний об'єкт, від другорядних і випадкових, спостереження повинні бути багаторазовими, масовими. Закономірності, що виявляються при масових спостереженнях, називають статистичними. Статистичні закономірності також причино обумовлені, як і динамічні, тільки причин може бути безліч, вони взаємно переплетені і діють у різних напрямках. Імовірність одержання конкретного результату тут дорівнює нулю. У подібних ситуаціях можна лише знайти інтервал, у який попадає значення досліджуваного показника з наперед заданою імовірністю. Виявлення статистичних закономірностей, визначення інтервальних оцінок невідомих параметрів і перевірка різних гіпотез здійснюються методами математичної статистики.

Стохастичні моделі описують закономірності, які обумовлені одночасною дією на об'єкт багатьох факторів і які проявляються чітко тільки при масових спостереженнях. До найбільш розповсюджених методів побудови стохастичних моделей відносяться методи, об'єднані під загальною назвою — багатовимірний статистичний аналіз, зокрема — кореляційний і регресійний аналізи. Практика показує, що стохастичні моделі, одержані за допомогою кореляційного і регресійного аналізів, мають перевагу при кількісному описі причинно-наслідкових відносин в економіці і соціальній сфері в порівнянні з детерміністичними моделями. Виявлення кількісних співвідношень у вигляді регресії дає можливість краще зрозуміти природу досліджуваного явища. А це,

у свою чергу, дозволяє впливати на виявлені фактори, втручатися у відповідний економічний процес з метою одержання потрібних результатів.

Класичний регресійний аналіз описує економічні процеси за допомогою одного рівняння регресії. Це рівняння не функціональне, а стохастичне. У ньому кожному наборові пояснюючих змінних може відповідати відразу кілька значень пояснюваної змінної. У рівнянні повинні бути присутні тільки істотні пояснюючі змінні. Неконтрольовані або невраховані фактори, а також помилки виміру включаються у випадковий член (випадкове відхилення). Передбачається, що пояснюючі змінні не випадкові і не корелюють між собою, а випадкова складова має діагональну дисперсійно-коваріаційну матрицю з рівними діагональними елементами (дисперсіями).

Як показали подальші дослідження, опису економічних процесів за допомогою одного рівняння регресії явно недостатньо через безліч переплетень причин і наслідків. Для більш адекватного відображення реальних взаємовідносин в економічних процесах необхідно застосовувати систему регресійних рівнянь. Застосування обґрунтованих тестів для перевірки гіпотези про вигляд дисперсійно-коваріаційної матриці випадкового відхилення показало, що розраховані значення випадкового члена рівняння регресії в багатьох випадках (особливо при аналізі часових рядів) відкидають основні припущення класичного регресійного аналізу. Ідея про взаємозв'язок між економічними змінними, а також припущення про загальний вигляд дисперсійно-коваріаційної матриці випадкового члена привело до створення нового типу стохастичних моделей, що стали називатися економетричними.

Уперше термін "економетрія!" (Öconometrie) з'явився в німецькій книзі по бухгалтерському обліку, автор якої розумів під ним теорію бухгалтерії. В економічну науку цей термін ввів в оборот норвезький статистик Рагнар Фріш (згодом лауреат Нобелівської премії в області економіки) у доповіді, опублікованій ним у 1928 р., для позначення самостійної галузі наукових досліджень. За ініціативою Р. Фріша 29 грудня 1930 р. була створена економетрична спілка, яка стала видавати щомісячний журнал «Економетрика»

("Econometrica"). Основним завданням спілка оголосила «розвиток економічної теорії в її зв'язку зі статистикою і математикою». Коло задач, розв'язуваних економетрикою, безупинно розширюється. Економісти, переконавшись у вірогідності одержуваних економетристами результатів, досить швидко визнали економетрику як самостійну наукову дисципліну. Наведемо думки визнаних авторитетів в економіці і економетриці.

«Економетрика займається визначенням конкретних кількісних закономірностей, що спостерігаються в економічному житті, застосовуючи для цієї мети статистичні методи» (О. Ланге).

«Економетрика — могутній інструмент для виявлення найбільш стійких характеристик у поведженні реальних економічних одиниць» (Е. Маленво).

«Економетрика дозволяє проводити кількісний аналіз реальних економічних явищ, ґрунтуючись на сучасному розвитку теорії і спостереженнях, зв'язаних з методами одержання висновків» (Самуельсен).

«Основна задача економетрики — наповнити емпіричним змістом апріорні економічні міркування» (Л. Клейн).

«Економісти використовують кількісні дані для спостереження за ходом розвитку економіки, її аналізу і прогнозів. Набір статистичних методів, що використовуються для цих цілей, називається в сукупності економетрикою» (Ц. Гріліхес).

«Сучасна університетська економічна освіта тримається на трьох китах: макроекономіці, мікроекономіці і економетриці » (В. Л. Макаров).

Економетрія, економетрика — один з напрямків економіко-математичних методів аналізу, що полягає в статистичному вимірюванні (оцінюванні) параметрів, які характеризують деяку економічну концепцію про взаємозв'язок і розвиток об'єкта або явища, і в застосуванні таким чином економетричних моделей для конкретних економічних висновків.

ПРЕДМЕТ ТА МЕТОД ЕКОНОМЕТРИКИ.

1.1 ПРЕДМЕТ ТА МЕТОД ЕКОНОМЕТРИКИ.

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.

«Предметом економетрики є економіка в кількісному аспекті. Об'єкт вивчення — зв'язки між народногосподарчим ансамблем і його найпростішими компонентами і відповідно поводження останніх» (Т. Шателес).

«Мета економетрики — емпіричний висновок з економічних законів. Економетрика доповнює теорію, використовуючи реальні дані для перевірки й уточнення відношень, що постулюються» (Е. Маленво).

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 обґрунтованих управлінських рішень.

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

Основні етапи економетричного аналізу

Процес економетричного моделювання складається з таких кроків:

- 1) вибір конкретної форми аналітичної залежності між економічними показниками (специфікація моделі) па підставі відповідної економічної теорії;
- 2) збирання та підготовка статистичної інформації;
- 3) оцінювання параметрів моделей;

4) перевірка адекватності моделі та достовірності її параметрів;

5) застосування моделі для прогнозування розвитку економічних процесів з метою подальшого керування ними.

1.2 ІСТОРИЧНІ ВІДОМОСТІ.

На початку XX 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. Паретто, у якій досліджувалися доходи населення в різних країнах. У ній вперше була застосована так звана крива Паретто, 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 увійшла 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. Клейп (1980) за розробку складних економетричних моделей та їх застосування для аналізу кон'юктурних 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. Макфедеп відзначені Нобелівською премією за розробку мікроеконометрії та методів статистичного аналізу.

1.3 ПРИКЛАДИ МОДЕЛЕЙ ТА МЕТОДІВ, ЯКІ НОСЯТЬ ХАРАКТЕР ЕКОНОМЕТРИЧНИХ ДОСЛІДЖЕНЬ.

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

У загальному випадку рівняння в економетричній моделі має вигляд

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

де Y – результат або залежна змінна, 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 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 містить ще щонайменше деякі коефіцієнти, що поєднують змінні у певних співвідношеннях і визначають структуру рівняння. Ці коефіцієнти називаються параметрами моделі.

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

Існує відмінність між змінними та параметрами моделі. Змінні – це економічні величини, що можуть набувати певних значень з деякої множини допустимих величин. Параметри – це сталі коефіцієнти. Хоча вони не завжди відомі, та все ж у будь-якій ситуації вони мають фіксоване значення. Параметри

можна назвати "незмінними" (інколи відомими, інколи невідомими), що пов'язують змінні в рівняннях. Ці рівняння, а отже, і параметри визначають структуру моделі вони вказують на характер припустимих співвідношень між змінними.

Параметри чимось подібні до незалежних (заданих ззовні) змінних, однак між ними є важливі відмінності. Припускається, що параметри залишаються незмінними протягом усього періоду спостереження, а екзогенні змінні, безумовно, мають змінюватися з часом. Саме змінювання екзогенних змінних приводить модель у рух, зумовлює перехід системи до нового стану.

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

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

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

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

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

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

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

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

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

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

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

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

1.4 ЗНАЧЕННЯ КУРСУ ТА ВЗАЄМОЗВ'ЯЗОК З ІНШИМИ ЕКОНОМІЧНИМИ ДИСЦИПЛІНАМИ.

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.

1.5 МАТЕМАТИЧНА МОДЕЛЬ ТА ОСНОВНІ ЕТАПИ ЇЇ ПОБУДОВИ.

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.

Процес побудови моделі складається з таких етапів:

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

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

За визначенням, будь-яка економічна модель є абстрактною, а отже, неповною. Це пов'язано з тим, що для виокремлення закономірностей функціонування економічного об'єкта потрібно абстрагуватися від інших факторів, які хоч і мають незначний вплив, однак у сукупності можуть визначати не лише відхилення в поведінці об'єкта, а й його поведінку. Звичайно вважають, що всі фактори, невраховані явно в моделі, мають незначний результуючий вплив на процес чи явище, що досліджується. Склад урахованих факторів і їх структура коригуються в процесі вдосконалення моделі.

1.6 ТЕОРЕТИЧНІ ОСНОВИ МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ ТА КЛАСИФІКАЦІЯ МОДЕЛЕЙ.

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

Макроекономічні моделі описують економіку загалом, пов'язуючи між собою узагальнені матеріальні та фінансові показники: ВВП, споживання, інвестиції, зайнятість, відсоткову 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 етап економічного об'єкта в певний момент чи період часу, а динамічні моделі вивчають взаємозв'язки економічних змінних у часі. Змінні, що вивчаються в динаміці, у статичних моделях мають фіксоване значення. Однак динамічна модель не зводиться до простої суми статичних моделей, а описує взаємодію сил, що рухають економіку.

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

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

Предметом економетричного дослідження є прикладні стохастичні економічні моделі, тобто загальні економічні моделі, у яких модельні

коефіцієнти набувають конкретних числових значень залежно від використаної статистичної інформації.

1.7 ОСНОВНІ ЯКІСНІ ТА КІЛЬКІСНІ ХАРАКТЕРИСТИКИ СИСТЕМИ ОБЛІКУ ЯК ОБ'ЄКТУ МОДЕЛЮВАННЯ.

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

Економічні дані звичайно поділяють на два види: перехресні дані та часові ряди. Перехресними є дані за якимось економічним показником, що отримані для різних однотипних об'єктів (фірм, регіонів). Причому дані отримано в один і той самий момент часу або часова приналежність несуттєва. Часові ряди характеризують один і той самий об'єкт, але в різні моменти часу. Наприклад, дані бюджетних досліджень населення в певний момент часу є перехресними, а динаміка рівня інфляції за певний період відображується часовими рядами. Послідовні значення часових рядів можуть бути пов'язані між собою певними залежностями, спостерігаються деякі закономірності у відхиленнях від загальної тенденції розвитку чи виявляються часові зсуви показників (часові лаги) Тому методи обробки таких даних дещо відрізняються від методів, що застосовуються для обробки перехресних даних

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

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

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

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

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

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

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

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

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

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

ОДНОФАКТОРНА ЛІНІЙНА ЕКОНОМЕТРИЧНА МОДЕЛЬ.

2.1 РЕГРЕСІЙНА ЕКОНОМЕТРИЧНА МОДЕЛЬ.

Економетрія вивчає кількісні закономірності та взаємозв'язки між економічними величинами. Для вимірювання взаємозв'язків між економічними показниками будують економетричну модель.

Економетрична модель - це рівняння чи система рівнянь, що описує кореляційно-регресійний зв'язок між економічними показниками, один чи кілька з яких є залежними змінними, інші — незалежними.

У загальному вигляді економетрична модель запишеться так:

$$y = f(x_1, x_2, x_3, \dots, x_m, u)$$

де y - залежна змінна; x_j , ($j=1, 2, \dots, m$) – незалежні змінні; u — стохастична складова, або

$$y_s = f(x_{s1}, x_{s2}, x_{s3}, \dots, x_{sm}, u_s), s=1, 2, \dots, k,$$

де u – стохастична складова s -го рівняння, тобто ця економетрична модель складається з k рівнянь.

Незалежні змінні моделі називаються *факторами*, наперед заданими змінними. Залежні змінні називаються *показниками* або *результуючими ознаками*. Інколи вживаються і інші назви "регресор", "регресанд".

Економетрична модель, що будується на основі системи рівнянь, крім регресійних функцій, може включати алгебраїчні рівняння.

Побудова будь-якої економетричної моделі, незалежно від того, на якому рівні і для яких показників вона будується, здійснюється як послідовність певних кроків.

Крок 1. Встановлення причинно-наслідкового зв'язку між досліджуваними економічними показниками.

Крок 2. Вибір найбільш істотних ознак і встановлення вигляду і форми функції зв'язку.

Крок 3. Знаходження параметрів функції зв'язку.

Крок 4. Оцінка достовірності отриманих результатів.

Крок 5. Прогнозування значень результуючої ознаки та їх аналіз.

Економетрична модель базується на єдності двох аспектів – теоретичного, якісного аналізу взаємозв'язків та емпіричної інформації. Теоретична інформація знаходить своє відображення в специфікації моделі.

Специфікація моделі - це аналітична форма економетричної моделі. На основі досліджуваних чинників вона складається з певного виду функції чи функцій, що використовуються для побудови моделей, має ймовірнісні характеристики, які притаманні стохастичним залишкам моделі.

Лінійні функції найпоширеніші в економетричному моделюванні.

Маючи на увазі, що вибір аналітичної форми економетричної моделі не може розглядатись без конкретного переліку незалежних змінних, *специфікація моделі передбачає відбір чинників для економетричного дослідження.*

При цьому в процесі такого дослідження можна кілька разів повертатись до етапу специфікації моделі, уточнюючи перелік незалежних змінних та вид функції, що застосовується. Адже коли вид функції та її складові не відповідають реальним процесам, то йдеться про помилки специфікації.

Помилки специфікації моделі можуть бути такими:

- 1) ігнорування при побудові економетричної моделі істотного фактора;
- 2) введення в модель незалежної змінної, яка не є істотною для вимірюваного зв'язку;
- 3) використання не відповідних математичних форм залежності.

Перша з цих помилок призводить до зміщення оцінок, причому зміщення буде тим більшим, чим більша кореляція між введеними та відсутніми в моделі змінними, а напрям зміщення залежить від знака оцінок параметрів при введених змінних і від характеру кореляції між введеними та відсутніми змінними. Оцінки параметрів також будуть зміщеними (у такому разі вони вищі), тому застосування способів перевірки їх значущості може спричинитися до хибних висновків щодо значень параметрів генеральної сукупності.

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

Третя помилка специфікації. Припускається, що залежна змінна є лінійною функцією від деякої пояснювальної змінної, тоді як насправді тут краще підійшла б квадратична, кубічна чи якась поліноміальна залежність вищого порядку. У цьому разі наслідки такі самі, як і при першій помилці специфікації, тобто оцінки параметрів моделі матимуть зміщення.

Питання про вибір найкращої форми залежності має базуватися на перевірці ступеня узгодженості виду функції з вхідними даними спостережень.

Адекватність побудованої моделі можна встановити, аналізуючи залишки моделі. Вони обчислюються як різниці між фактичними значеннями залежної змінної і обчисленими за моделлю. Щоб перевірити, чи має розподіл залишків не випадковий характер, можна скористатися критерієм Дарбіна-Уотсона.

Побудова економетричних моделей базується на технології моделювання, яка характерна для здійснення регресійного аналізу. Регресія починається там, де є ряд спостережень над досліджуваними величинами

$$\begin{array}{cccccc} y & y_1 & y_2 & \dots & y_n \\ x & x_1 & x_2 & \dots & x_n \end{array}$$
 – який є рядом даних, а x і y - випадкові величини, при

цьому x_i, y_i – їх практичні реалізації або можливі значення, n - кількість спостережень.

При вимірюванні кількісних ознак можуть бути отримані два типи рядів даних – динамічні та варіаційні.

Динамічний ряд – це послідовність спостережень за процесом або явищем у рівновіддалені проміжки часу. Якщо x_i - значення деякої ознаки економічного процесу в i -й проміжок часу, то динамічний ряд $x_1, x_2, x_3, \dots, x_n$ можна отримати, вимірюючи значення цієї ознаки в рівновіддалені проміжки часу.

Варіаційні ряди – це ряди даних, які показують кількісну міру певної ознаки у всіх об'єктів однієї сукупності, наприклад, оцінки за екзамен студентів однієї групи.

Першим кроком до побудови регресійної моделі є *ідентифікація змінних* (спостережуваних величин. Треба визначити яка із спостережуваних величин є ознакою (залежною величиною, пояснюваною змінною, функцією), а які величини є незалежними (аргументами, пояснюючими змінними, факторами). Якщо ми записуємо, що

$$y = f(x) \quad (2.1)$$

то вважаємо, що змінні є ідентифіковані і y , в цьому випадку, пояснювана змінна, x - пояснююча.

Другим етапом до побудови регресійної моделі є *специфікація моделі*, а саме: на основі даних ряду досліду потрібно визначити аналітичну форму зв'язку (2.1).

Найчастіше специфікацію проводять з допомогою хмарки точок (діаграми розсіювання). На осі абсцис відзначають значення незалежної змінної (x), на осі ординат – значення залежної змінної (y).

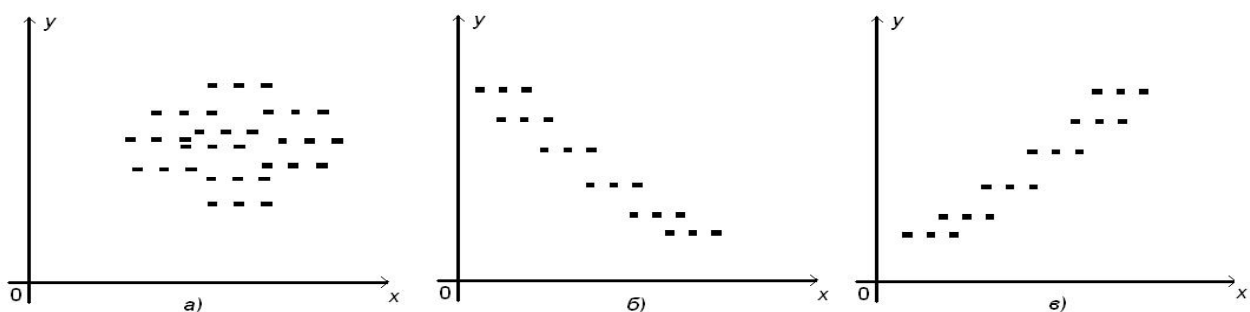


Рис. 2.1. Діаграма розсіювання: а) зв'язок між y та x відсутній; б) зв'язок - лінійний, спадний; в) зв'язок - лінійний, зростаючий

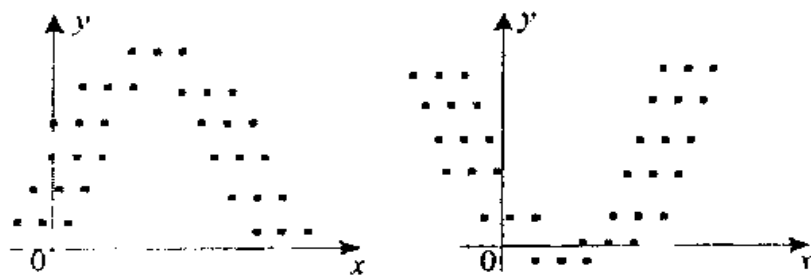


Рис. 2.2. Діаграми розсіювання - зв'язок між y та x параболічний

За виглядом діаграми розсіювання можна висунути гіпотезу про лінійність чи не лінійність зв'язку між змінними.

Третій етап – оцінка параметрів моделі.

Четвертий – аналіз моделі по залишках.

Припустимо, що в результаті специфікації ми переконалися, що залежність між y та x – лінійна, тоді (2.1) набере вигляду:

$$y = a_0 + a_1x + u \quad (2.2)$$

де: a_0, a_1 – невідомі детерміновані параметри; y - вектор спостережень за залежною змінною, $y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{bmatrix}$ — вектор спостережень за незалежною змінною, $x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ x_n \end{bmatrix}$; u - випадкова складова, збурення, $u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ \dots \\ u_n \end{bmatrix}$.

Рівняння регресії (2.2) - називається *економетричною моделлю*, якщо y та x є кількісними показниками деяких економічних явищ чи процесів. Таким чином, рівняння регресії перетворюється в економетричне, якщо воно налагоджує зв'язок між кількісними показниками об'єктів економіки.

Існують пояснення введення в модель (2.2) випадкового доданку u . В рівняннях регресії він вводиться лише по одній причині, через помилки виміру. В економетрії причини складніші:

1) в економетричних моделях на ознаку діє така кількість факторів, яка значно перевищує кількість дослідів, замірів, об'єм. По тій причині з ситуації виходять так: з усіх змінних виділяють 1, 2 чи 3 значущих, а решту об'єднують в одну, яку називають випадковою величиною. Зрозуміло, що решта змінних діє в різних напрямках, тому ми можемо вважати її середню дію рівною 0.

$$M \left(\frac{1}{x} \right) = 0$$

Крім того за центральною граничною теоремою Ляпунова u нормально розподілена величина. *Центральна гранична теорема Ляпунова:* якщо випадкова величина x представляє собою суму дуже великого числа взаємно незалежних випадкових величин, вплив яких на всю суму як завгодно малий, то x має розподіл близький до нормального).

- 2) полягає в тому, що предметом економетрії є людське суспільство з його багатогранністю смаків і уподобань, які неможливо нічим, крім випадкової величини відобразити.
- 3) полягає у похибці вимірювання.

2.2 ЗНАХОДЖЕННЯ СТАТИСТИЧНИХ ОЦІНОК ПАРАМЕТРІВ МЕТОДОМ НАЙМЕНШИХ КВАДРАТІВ.

Задача економетрії полягає в точковому або інтервальному оцінюванні параметрів економетричної моделі. Практично моделі (2.2) не існує, її потрібно оцінити деяким рівнянням

$$\hat{y} = \hat{a}_0 + \hat{a}_1 x \quad (2.3)$$

де \hat{a}_0 та \hat{a}_1 називаються статистичними оцінками параметрів a_0 та a_1 і знаходяться на основі вибірових даних. Якщо розглянути (2.3) як пряму на площині, то \hat{a}_0 – перетин з віссю ординат, \hat{a}_1 – нахил прямої до осі абсцис. \hat{a}_1 також визначає зміну результативного показника при зміні x на одиницю.

Значення \hat{y} , показують середнє значення залежної змінної y при заданому x_i у припущенні, що єдиною причиною зміни y є змінна x , а випадкова збурена змінна u прийняла значення, рівне нулеві. Розкид спостережених значень змінної y довкола \hat{y} зумовлений впливом множини неврахованих факторів.

Різниця між y_i і розрахунковим \hat{y} , називається *залишком* (відхиленням), який

дає числову оцінку значення збурення u . Отже, він визначається $u_i = y_i - \hat{y}_i$ $i=1, \dots, n$. Чим менше значення u_i , тим краще підібрана пряма. На цьому ґрунтується метод найменших квадратів, до якого ми переходимо.

Взагалі, існує необмежена кількість прямих $\hat{y} = \hat{a}_0 + \hat{a}_1 x$, які можна провести через множину спостережуваних точок. Щоб знати, яку пряму вибрати, потрібно користуватись певним критерієм. Згідно методу найменших квадратів (МНК) потрібно провести пряму таким чином, щоб сума квадратів залишків була мінімальною. Тобто ідея МНК полягає в побудові цільової функції, яка мінімізує суму квадратів відхилень $u_i = y_i - \hat{y}_i$ (див. рис. 2.3).

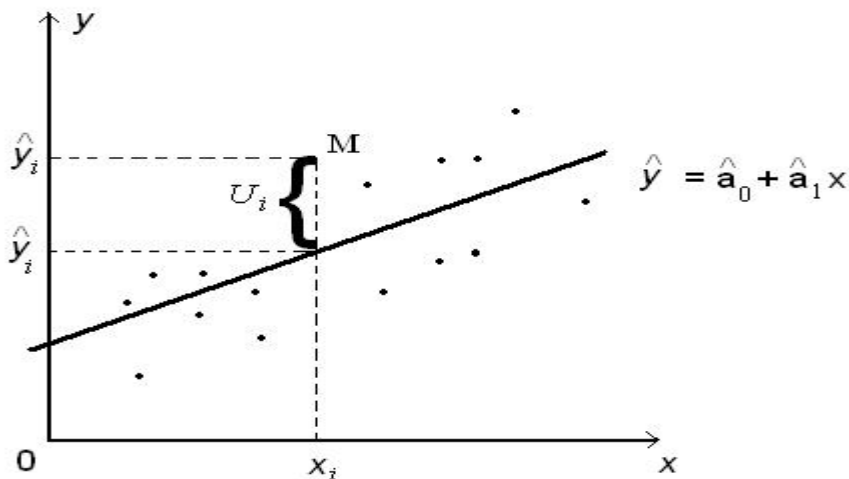


Рис. 2.3. Відхилення теоретичних значень від фактичних

Згідно критерію маємо для всієї сукупності точок:

$$\sum_{i=1}^n u_i^2 = \sum_{i=1}^n \left(y_i - \hat{a}_0 - \hat{a}_1 x \right)^2 = f\left(\hat{a}_0, \hat{a}_1 \right) \rightarrow \min$$

Щоб дослідити функцію двох змінних на \min , знаходимо її часткові похідні і прирівнюємо їх до нуля.

$$\frac{d\left(\sum_{i=1}^n u_i^2\right)}{d \hat{a}_0} = -2 \sum_{i=1}^n \left(y_i - \hat{a}_0 - \hat{a}_1 x_i\right) = 0$$

$$\frac{d\left(\sum_{i=1}^n u_i^2\right)}{d \hat{a}_1} = -2 \sum_{i=1}^n x_i \left(y_i - \hat{a}_0 - \hat{a}_1 x_i\right) = 0$$

Звідки отримаємо систему лінійних рівнянь:

$$\begin{cases} \sum_{i=1}^n y_i = n \hat{a}_0 + \hat{a}_1 \sum_{i=1}^n x_i \\ \sum_{i=1}^n y_i x_i = \hat{a}_0 \sum_{i=1}^n x_i + \hat{a}_1 \sum_{i=1}^n x_i^2 \end{cases} \quad (2.4)$$

яка називається *нормальною системою МНК*. Розв'язавши систему одержимо статистичні оцінки. Зокрема, якщо систему розв'язати методом Крамера, то отримаємо:

$$\hat{a}_0 = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n y_i x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2}, \quad \hat{a}_1 = \frac{n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \quad (2.5)$$

Другий спосіб, без розв'язання системи. Поділимо перше рівняння системи (2.4) на n

$$\frac{\sum_{i=1}^n y_i}{n} = \hat{a}_0 + \hat{a}_1 \frac{\sum_{i=1}^n x_i}{n}$$

і введемо позначення для середніх арифметичних $\bar{y} = \frac{\sum_{i=1}^n y_i}{n}$, $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$. Тоді

рівняння набуде вигляду

$$\bar{y} = \hat{a}_0 + \hat{a}_1 \bar{x} \quad (2.6)$$

Таким чином, оціночна пряма проходить через точку, координати якої є середні арифметичні вибірових даних. Віднімемо (2.6) від (2.3):

$\hat{y}_i - \bar{y} = \hat{a}_1 (x_i - \bar{x})$ тобто отримаємо

$$\hat{y}_i = \bar{y} + \hat{a}_1 (x_i - \bar{x}) \quad (2.6')$$

Нехай позначимо прирости (відхилення від середнього арифметичного) через

$$y_i - \bar{y} = \Delta y_i, \quad x_i - \bar{x} = \Delta x_i, \quad \text{тоді} \quad \hat{y}_i = \bar{y} + \hat{a}_1 \Delta x_i.$$

Згідно критерію маємо

$$\sum_{i=1}^n u_i^2 = \sum_{i=1}^n \left(y_i - \bar{y} - \hat{a}_1 \Delta x_i \right)^2 = \sum_{i=1}^n \left(\Delta y_i - \hat{a}_1 \Delta x_i \right)^2 = f(\hat{a}_1) \rightarrow \min$$

Щоб дослідити функцію на \min , знаходимо її похідну по \hat{a}_1 і прирівнюємо

її до нуля:

$$\frac{d\left(\sum_{i=1}^n u_i^2\right)}{d\hat{a}_1} = -2 \sum_{i=1}^n \Delta x_i \left(\Delta y_i - \hat{a}_1 \Delta x_i \right) = 0$$

Звідси отримаємо

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

Отже, статистичну оцінку \hat{a}_1 можна обчислити за формулою

$$\hat{a}_1 = \frac{\sum_{i=1}^n \Delta x_i \Delta y_i}{\sum_{i=1}^n \Delta x_i^2} \quad (2.7)$$

а оцінку \hat{a}_0 згідно (2.6)

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x} \quad (2.7')$$

За формулами (2.7) і (2.7') обчислюються статистичні оцінки по МНК через прирости (відхилення від середніх арифметичних). Формулу (2.7) можна записати у вигляді

$$\hat{a}_1 = \frac{\text{cov}(x, y)}{\text{var}(x)},$$

оскільки, за означенням, коефіцієнт коваріації між x та y дорівнює

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

а дисперсія змінної x визначається за формулою

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

Якщо економетрична модель задана в матричній формі:

$$Y = Xa + u, \quad (2.8)$$

де Y – вектор значень залежної змінної; X – матриця незалежних змінних розміром $n \times m$ (n – число спостережень, m – кількість незалежних змінних); a – вектор параметрів моделі; u – вектор залишків.

Застосовуємо метод найменших квадратів для оцінки параметрів моделі, якщо виконуються такі умови:

1) математичне сподівання залишків дорівнює нулю, тобто

$$M(u) = 0;$$

2) значення u_i , вектора залишків u незалежні між собою і мають постійну дисперсію, тобто

$$M(uu^t) = \sigma^2 E,$$

де E – одинична матриця;

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

$$M(x^t u) = 0;$$

4) незалежні змінні моделі утворюють лінійно незалежну систему векторів, або, іншими словами, незалежні змінні не повинні бути мультиколінеарними, тобто

$$|X^t X| \neq 0$$

Це означає, що матриця X має повний ранг.

Перша умова очевидна. Адже коли математичне сподівання залишків не дорівнює нулю, то це означає, що існує систематичний вплив на залежну змінну, а до модельної специфікації не введено всіх основних незалежних

змінних. Якщо ця передумова не виконується, то йдеться про помилку специфікації.

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

Третя умова передбачає незалежність між залишками і пояснювальними змінними, яка порушується насамперед тоді, коли економетрична модель будується на базі одночасових структурних рівнянь або має лагові змінні. Тоді для оцінювання параметрів моделі використовуються, як правило, дво- або трикроковий методи найменших квадратів.

Четверта умова означає, що всі пояснювальні змінні, які входять до економетричної моделі, мають бути незалежними між собою. Проте очевидно, що в економіці дуже важко вирізнити такий масив незалежних (пояснювальних) змінних, які були б зовсім не пов'язані між собою. Тоді щоразу необхідно з'ясовувати, чи не впливатиме залежність пояснювальних змінних на оцінку параметрів моделі. Це явище називають *мультиколінеарністю* змінних, що призводить до ненадійності оцінки параметрів моделі, робить їх чутливими до вибраної специфікації моделі та до конкретного набору даних.

Нехай у результаті спостережень за об'єктом отримані такі дані:

x	x_1	x_2	...	x_n
y	y_1	y_2	...	y_n

Оцінимо методом найменших квадратів параметри моделі $Y=XA+u$, де

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{pmatrix}$$

Рівняння моделі подамо у вигляді: $u = Y - Xa$. Тоді суму квадратів залишків u можна записати так:

$$\sum_{i=1}^n u_i^2 = u^t u = (Y - Xa)^t (Y - Xa) = Y^t Y - 2a^t X^t Y + a^t X^t X a$$

В точці екстремуму для функціонала $u^t u$ виконується умова:

$$\frac{d(u^t u)}{da} = -2X^t Y + 2X^t X a = 0$$

або

$$X^t X a = X^t Y. \quad (2.9)$$

Тут X^t – матриця, транспонована до матриці X . Звідси

$$a = (X^t X)^{-1} X^t Y. \quad (2.10)$$

Рівняння (2.9) дає матричну форму запису системи рівнянь, а формула (2.10) показує, що вектор a — розв'язок системи таких рівнянь.

Неважко показати, що отримані оцінки параметрів a , що обчислені за (2.10), мінімізують суму квадратів залишків u .

Якщо незалежні змінні в матриці X взяті як відхилення кожного значення від свого середнього, то матрицю $X^t X$ називають *матрицею моментів*. У цьому випадку числа, що розміщені на її головній діагоналі, характеризують величину дисперсій незалежних змінних, інші елементи відповідають взаємним коваріаціям.

МЕТОДИКА ВИЗНАЧЕННЯ ПАРАМЕТРІВ ОДНОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ

Задача 1. На основі даних про прибуток підприємства та сумою заробітної плати, виплаченої на підприємстві

Х-сума зарплати, тис. грн	3,2	2,5	2,1	1,62	1,35	1,33	1,17	1,25	1,1	0,9
у-прибуток підприємства тис. грн./на одного працюючого	105,4	114,9	120,3	150,1	152,4	163,3	171,4	185,1	195,1	

необхідно:

1) провести специфікацію моделі;

2) розрахувати оцінки \hat{a}_0 та \hat{a}_1 методом найменших квадратів:

- за системою нормальних рівнянь;
- через відхилення від середніх;
- матричним методом.

Розв'язання:

1) Специфікацію моделі здійснюємо за допомогою діаграми розсіювання. У прямокутній декартовій системі координат xOy будуємо точки $(x_i; y_i)$, $i=1,2,\dots,9$, координати яких визначаються за табл..1. (див рис.1).

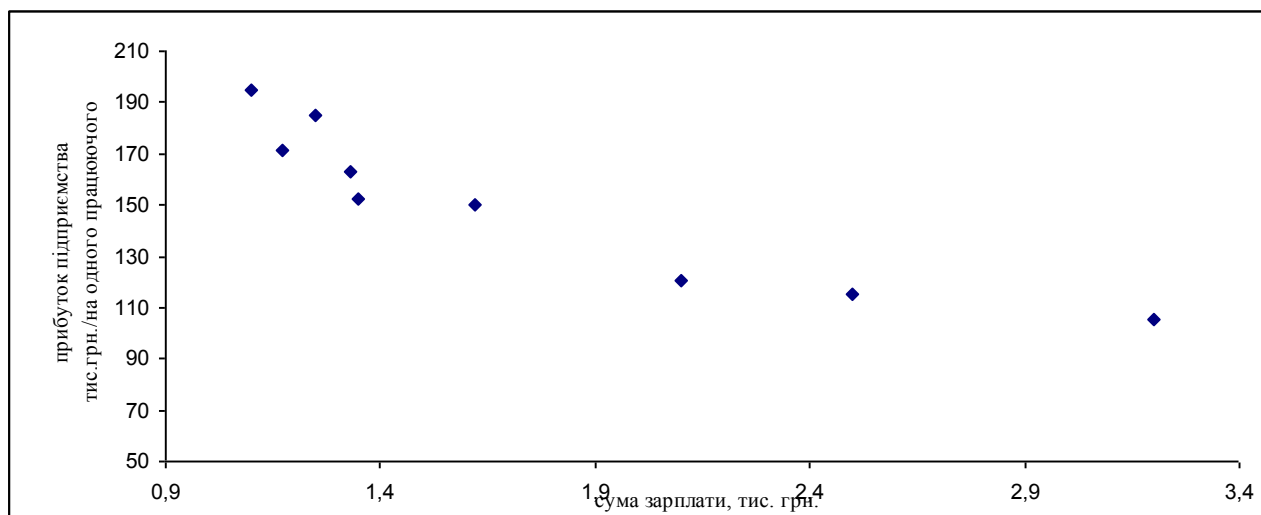


Рис. 2.4 Специфікація моделі

Переконавшись, з графіку, що залежність між факторами x і показником y лінійна, знаходимо оціночне рівняння $\hat{y} = \hat{a}_0 + \hat{a}_1 x$.

2) Оцінки параметрів \hat{a}_0 та \hat{a}_1 методу найменших квадратів:

за системою нормальних рівнянь:

$$\begin{cases} \sum_{i=1}^n y_i = n \hat{a}_0 + \hat{a}_1 \sum_{i=1}^n x_i \\ \sum_{i=1}^n y_i x_i = \hat{a}_0 \sum_{i=1}^n x_i + \hat{a}_1 \sum_{i=1}^n x_i^2 \end{cases}$$

Будуємо розрахункову таблицю:

№	x_i	y_i	$x_i y_i$	x_i^2
1	3,2	105,4	337,28	10,24
2	2,5	114,9	287,25	6,25
3	2,1	120,3	252,63	4,41
4	1,62	150,1	243,162	2,6244
5	1,35	152,4	205,74	1,8225
6	1,33	163,3	217,189	1,7689
7	1,17	171,4	200,538	1,3689
8	1,25	185,1	231,375	1,5625
9	1,1	195,1	214,61	1,21
Σ	15,62	1358	2189,774	31,2572

Система рівнянь у числах набуде вигляду:

$$\begin{cases} 1358 = 9 \hat{a}_0 + \hat{a}_1 \cdot 15,62 \\ 2189,774 = \hat{a}_0 \cdot 15,62 + \hat{a}_1 \cdot 31,2572 \end{cases}$$

Її можна розв'язати довільним методом. Розв'яжемо за допомогою методу Крамера. Оцінки будуть розраховуватися за формулами:

$$\hat{a}_0 = \frac{\sum_{i=1}^n y_i \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \sum_{i=1}^n y_i x_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}, \quad \hat{a}_1 = \frac{n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \text{ тоді}$$

$$\hat{a}_0 = \frac{1358 \cdot 31,2572 - 15,62 \cdot 2189,774}{9 \cdot 31,2572 - 15,62^2} = 220,8122$$

$$\hat{a}_1 = \frac{9 \cdot 2189,774 - 15,62 \cdot 1358}{9 \cdot 31,2572 - 15,62^2} = -40,2887.$$

Через відхилення від середніх

Ці ж оцінки знайдемо не розв'язуючи системи нормальних рівнянь. Згідно МНК через відхилення від середніх значень

$$\hat{a}_1 = \frac{\sum_{i=1}^n \Delta x_i \Delta y_i}{\sum_{i=1}^n \Delta x_i^2}, \quad \hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}$$

Будуємо розрахункову таблицю

№	x_i	y_i	\bar{x}_i	\bar{y}_i	Δx_i	Δy_i	Δx_i^2	$\Delta x_i \Delta y_i$
1	3,2	105,4			1,464444	-45,4889	2,1445975	-66,616
2	2,5	114,9			0,764444	-35,9889	0,5843753	-27,5115
3	2,1	120,3			0,364444	-30,5889	0,1328198	-11,148
4	1,62	150,1	1,735556	150,8889	-0,11556	-0,78889	0,0133531	0,09116
5	1,35	152,4			-0,38556	1,511111	0,1486531	-0,58262
6	1,33	163,3			-0,40556	12,41111	0,1644753	-5,0334
7	1,17	171,4			-0,56556	20,51111	0,3198531	-11,6002
8	1,25	185,1			-0,48556	34,21111	0,2357642	-16,6114
9	1,1	195,1			-0,63556	44,21111	0,4039309	-28,0986
Σ	15,62	1358					4,1478222	-167,11

$$\hat{a}_1 = \frac{-167,11}{4,1478222} = -40,2887,$$

$$\hat{a}_0 = 150,8889 + 40,2887 \cdot 1,735556 = 220,8122.$$

Матричний метод

$$x = \begin{pmatrix} 1 & 3,2 \\ 1 & 2,5 \\ 1 & 2,1 \\ 1 & 1,62 \\ 1 & 1,35 \\ 1 & 1,33 \\ 1 & 1,17 \\ 1 & 1,25 \\ 1 & 1,1 \end{pmatrix} \text{-матриця пояснюючої змінної,} \quad y = \begin{pmatrix} 105,4 \\ 114,9 \\ 120,3 \\ 150,1 \\ 152,4 \\ 163,3 \\ 171,4 \\ 185,1 \\ 195,1 \end{pmatrix} \text{-вектор}$$

результативної змінної. Проведемо наступні обчислення

$$\begin{bmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3,2 & 2,5 & 2,1 & 1,62 & 1,35 & 1,33 & 1,17 & 1,25 & 1,1 \end{pmatrix} = \begin{pmatrix} 9 & 15,62 \\ 15,62 & 31,26 \end{pmatrix}$$

$$\begin{bmatrix} 1 & \cdot \\ \cdot & \cdot \end{bmatrix} \begin{bmatrix} \cdot & \cdot \\ \cdot & \cdot \end{bmatrix} = \begin{pmatrix} -0,837 & -0,418 \\ -0,418 & 0,241 \end{pmatrix}.$$

$$\begin{bmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3,2 & 2,5 & 2,1 & 1,62 & 1,35 & 1,33 & 1,17 & 1,25 & 1,1 \end{pmatrix} = \begin{pmatrix} 1358 \\ 2189,8 \end{pmatrix}.$$

Тоді вектор оцінок

$$\hat{a} = \begin{pmatrix} -0,837 & -0,418 \\ -0,418 & 0,241 \end{pmatrix} \cdot \begin{pmatrix} 1358 \\ 2189,8 \end{pmatrix} = \begin{pmatrix} 220,8122 \\ -40,2887 \end{pmatrix}.$$

Отже економетрична модель має вигляд: $\hat{y} = 220,81 - 40,29x$.

СТАТИСТИЧНА ПЕРЕВІРКА ОЦІНОК ОДНОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ.

3.1 СТАНДАРТНА ПОХИБКА ОЦІНКИ ЗА РІВНЯННЯМ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ.

Застосовуючи метод найменших квадратів ми знайшли оцінки \hat{a}_0 і \hat{a}_1 економетричної моделі у вигляді одного рівняння:

$$y = a_0 + a_1x + u.$$

Ми сподіваємося, що оцінки \hat{a}_0 та \hat{a}_1 якнайкраще відображають істинні значення a_0 і a_1 економетричної моделі. Однак, так як ми інтуїтивно висували гіпотезу про лінійний характер зв'язку і, крім того, користувалися для знаходження оцінок \hat{a}_0 і \hat{a}_1 лише вибіркою із генеральної сукупності, необхідно визначити похибки знайдених оцінок. Зауважимо, що перевірка знайдених оцінок являється трудомістким завданням, але саме вона є основним завданням економетрії.

Нормальні рівняння МНК дають можливість розрахувати оцінки \hat{a}_0 та \hat{a}_1 навіть у тому випадку, коли гіпотеза $y = a_0 + a_1x + u$ вибрана не зовсім вдало. Тому необхідно навчитися відрізняти "добрі" оцінки від "поганих".

Існують загально прийняті етапи перевірки оцінок для визначення такої різниці.

Основна ідея аналізу оцінок базується на тому, що значення змінної y визначаються двома компонентами:

1. Систематичною складовою $a_0 + a_1x$;
2. Випадковою складовою u .

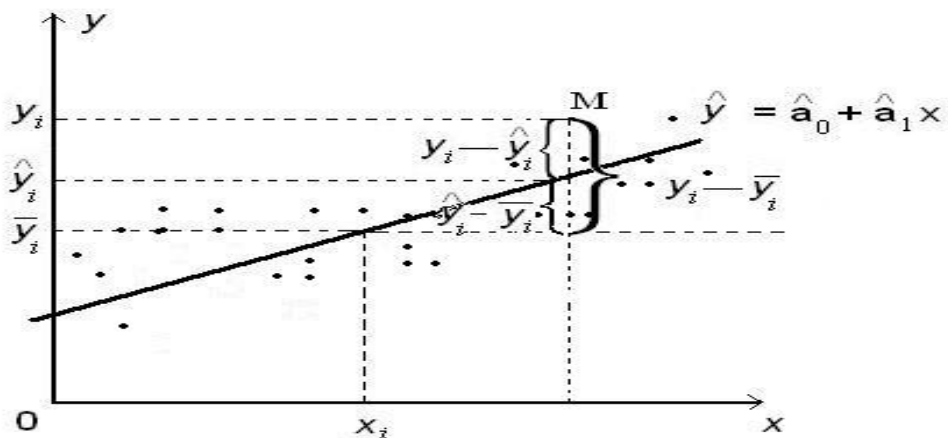


Рис.3.1. Розклад відхилення ендогенної змінної на систематичну і випадкову складову

З рисунку 3.1 видно, що

$$y_i - \bar{y} = (y_i - \hat{y}_i) + (\hat{y}_i - \bar{y}) \quad (3.1)$$

Різницю $y_i - \bar{y}$ називають загальним відхиленням.

Різницю $y_i - \bar{y}$ – поясненим відхиленням, адже його можна пояснити (обрахувати) маючи оціночну пряму (\bar{y} незмінна величина, а \hat{y}_i обчислюється підстановкою x_i в оціночну пряму).

Різницю $y_i - \hat{y}_i$ – непоясненим відхиленням, адже його не можна пояснити маючи оціночну пряму (якщо x_i змінюється то величина \hat{y}_i обчислюється, а y_i не обчислюється з оціночної прямої).

Піднесемо обидві частини (3.1) до квадрату та просумуємо за всіма індексами.

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

Оскільки

$$\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{y}) = \sum_{i=1}^n \left(\hat{a}_0 + \hat{a}_1 x_i - (\hat{a}_0 + \hat{a}_1 \bar{x}) \right) (y_i - \hat{y}_i) =$$

то

$$\hat{a}_1 \sum_{i=1}^n x_i (y_i - \hat{y}_i) + \hat{a}_1 \sum_{i=1}^n \bar{x} (y_i - \hat{y}_i) = 0$$

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (y_i - \bar{y})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.2)$$

$$СКЗ = СКП + СКН$$

де: $СКЗ = \sum_{i=1}^n (y_i - \bar{y})^2$ – загальна сума квадратів; $СКП = \sum_{i=1}^n (y_i - \bar{y})^2$ – пояснена

сума квадратів; $СКН = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ – непояснена сума квадратів. Поділимо (3.2)

на n , отримаємо вираз:

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

де:

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

$$\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n} = \sigma_{пояс}^2 \text{ – пояснена дисперсія;}$$

$$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} = \sigma_{непоясн}^2 \text{ – непояснена дисперсія.}$$

$$\text{Отже, } \sigma_{заг}^2 = \sigma_{пояс}^2 + \sigma_{непоясн}^2 \quad (3.3)$$

Якщо вважати незмінною $\sigma_{заг}^2$, то чим менша $\sigma_{непоясн}^2$, тим більша $\sigma_{пояс}^2$ і тим меншими будуть відхилення даних вибірки від оціночної прямої.

Кожній сумі квадратів з (3.2) ставиться у відповідність число, яке називають ступенем вільності. Воно показує, скільки незалежних елементів інформації, що утворилися з елементів $y_1, y_2, y_3, \dots, y_n$ необхідно для розрахунку суми квадратів.

Для отримання $СКЗ$ використовують числа $y_1 - \bar{y}, y_2 - \bar{y}, y_3 - \bar{y}, \dots, y_n - \bar{y}$. Ці числа мають властивість

$\sum_{i=1}^n (y_i - \bar{y})^2 = 0$. Тому серед них незалежними будуть $n-1$ чисел. Звідси ступінь

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

Наступну суму $СКП = \sum_{i=1}^n (y_i - \bar{y})^2$, використовуючи співвідношення попередньої теми $\hat{y}_i - \bar{y} = \hat{a}_1 (x_i - \bar{x})$ можна записати у вигляді:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \hat{a}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$$

Отже, $\sum_{i=1}^n (y_i - \bar{y})^2$ утворюється, використанням однієї одиниці незалежної інформації \hat{a}_1 , тому ступінь вільності її дорівнює 1.

Сума квадратів $СКН = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ матиме $n-2$ ступені вільності. Вона обраховується як різниця між кількістю спостережень n і оцінюваних параметрів (їх у випадку лінійної економетричної моделі $2 - \hat{a}_0, \hat{a}_1$).

Число, що утворюється діленням суми квадратів на відповідний ступінь вільності, називається середнім квадратом. Середні квадрати обчислюються тільки для $СКП$ і $СКН$:

$$\overline{СКП} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{1}$$

$$\overline{СКН} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}$$

Додатній корінь з $\overline{СКН}$ називається стандартною похибкою оцінки за рівнянням економетричної моделі:

$$\sigma_{yx} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2}} \quad (3.4)$$

Стандартна похибка оцінки за рівнянням економетричної моделі є мірою не поясненої варіації в $\sigma_{заг}$. Якщо стандартна похибка дорівнює нулю, то це означає, що $\sigma_{непоясн}=0$ і всі дані y_i , лежать на оціночній прямій, тобто зв'язок між y та x функціональний.

Найбільше значення стандартна похибка має, коли оцінка \hat{a}_1 оціночного рівняння дорівнює 0 і саме оціночне рівняння має вигляд $y = \hat{a}_0$, де $\hat{a}_0 = \bar{y}$, тобто оціночна пряма є прямою паралельною осі OX , віддаленою від початку координат на величину середнього значення результативної змінної. В даному випадку $\sigma_{заг}^2$ складається тільки з $\sigma_{непоясн}^2$ і $\sigma_{заг}^2 = \sigma_{непоясн}^2$.

Отже, інтервал зміни σ_{yx} : $0 \leq \sigma_{yx} \leq \sigma_{заг}$, σ_{yx} на практиці застосовується при побудові довірчих інтервалів (інтервальних оцінок), які визначають область ймовірних значень y , при відповідних значеннях x , що розглянемо нижче.

3.2 КОЕФІЦІЄНТ ДЕТЕРМІНАЦІЇ ТА КОЕФІЦІЄНТ КОРЕЛЯЦІЇ.

Розглянемо рівність

$$\sigma_{заг}^2 = \sigma_{посн}^2 + \sigma_{непоясн}^2$$

Якщо розділити обидві частини рівності на $\sigma_{заг}^2$, то отримаємо:

$$1 = \frac{\sigma_{посн}^2}{\sigma_{заг}^2} + \frac{\sigma_{непоясн}^2}{\sigma_{заг}^2} \quad (3.5)$$

Перша частина $\frac{\sigma_{посн}^2}{\sigma_{заг}^2}$ є пропорцією поясненої дисперсії у загальній дисперсії, тобто доля дисперсії, яку можна пояснити через регресійну лінію.

Друга $\frac{\sigma_{непоясн}^2}{\sigma_{заг}^2}$ являє собою частину дисперсії, яку не можна пояснити через регресійну лінію. Коефіцієнтом детермінації називається частина дисперсії, яку можна пояснити через регресійний зв'язок. Коефіцієнт детермінації позначається d або r^2 .

$$d = r^2 = \frac{\sigma_{\text{посн}}^2}{\sigma_{\text{заг}}^2} \quad (3.6)$$

Ця величина не залежить від одиниць вимірювання y і з формули (3.5) випливає, що коефіцієнт детермінації завжди додатній і знаходиться в межах $0 \leq d \leq 1$. Коефіцієнт детермінації використовується при визначенні адекватності моделі і також вимірює щільність зв'язку між показниками.

Коефіцієнт кореляції $r = r_{yx}$ дає можливість кількісно оцінити щільність зв'язку між y та x . Він розраховується за формулою:

$$r = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \text{var}(y)}} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.7)$$

де $\text{cov}(x, y)$ – коефіцієнт коваріації між x та y , $\text{var}(x)$ – дисперсія змінної x , $\text{var}(y)$ – дисперсія змінної y .

Коефіцієнт кореляції є відносною мірою зв'язку між двома факторами. Значення його лежать в межах від -1 до 1 $-1 \leq r \leq 1$. Додатне значення коефіцієнта кореляції свідчить про існування прямого зв'язку між змінними, від'ємне – про зворотній зв'язок. Якщо коефіцієнт кореляції прямує за абсолютною величиною до 1 , то між змінними x та y існує сильний зв'язок ($r \rightarrow \pm 1$ – щільність зв'язку велика); а якщо коефіцієнт кореляції прямує до 0 ($r \rightarrow 0$) – то між змінними x та y лінійного зв'язку немає.

Перевіримо чи існує зв'язок між коефіцієнтом кореляції і коефіцієнтом детермінації. Для цього проведемо наступні перетворення коефіцієнта детермінації:

$$d = \frac{\sigma_{\text{посн}}^2}{\sigma_{\text{заг}}^2} = \frac{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}}{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}} = \frac{\frac{1}{n} \sum_{i=1}^n \left(\hat{a}_0 + \hat{a}_1 x_i - \left(\hat{a}_0 + \hat{a}_1 \bar{x} \right) \right)^2}{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2} = \frac{\hat{a}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \hat{a}_1^2 \frac{\sigma_x^2}{\sigma_y^2}$$

Проведемо наступні перетворення коефіцієнта кореляції (врахувавши, що оцінка \hat{a}_1 обчислюється по формулі $\hat{a}_1 = \frac{\text{cov}(\hat{y}, y)}{\text{var}(\hat{y})}$):

$$r = \frac{\text{cov}(\hat{y}, y)}{\sqrt{\text{var}(\hat{y}) \text{var}(y)}} = \frac{\text{cov}(\hat{y}, y)}{\sigma_x \sigma_y} = \frac{\text{cov}(\hat{y}, y)}{\sigma_x^2} \cdot \frac{\sigma_x}{\sigma_y} = \hat{a}_1 \frac{\sigma_x}{\sigma_y} \quad (3.8)$$

З проведених перетворень видно, що коефіцієнт детермінації дорівнює квадрату коефіцієнта кореляції.

$$d = r^2 \quad (3.9)$$

З формули (3.8) випливає, що знак коефіцієнта кореляції r завжди збігається з знаком оцінки \hat{a}_1 , оскільки значення σ_x і σ_y завжди додатні. Крім того значення коефіцієнта кореляції r залежить від нахилу оціночної прямої \hat{a}_1 та середніх квадратичних відхилень σ_x і σ_y .

З формул (3.9) і (3.5) випливає, що коефіцієнт кореляції можна обчислити за формулою:

$$r = \pm \sqrt{d} = \pm \sqrt{\frac{\sigma_{\text{посн}}^2}{\sigma_{\text{заг}}^2}} = \sqrt{1 - \frac{\sigma_{\text{непосн}}^2}{\sigma_{\text{заг}}^2}} \quad (3.10)$$

В тих випадках, коли всі точки лежать на оціночній прямій то $\sigma_{\text{непосн}}^2 = 0$ і $r = 1$. Якщо ж $\sigma_{\text{заг}}^2 = \sigma_{\text{непосн}}^2$, то $r = 0$ (оціночна пряма є горизонтальною лінією). Чим ближче вибіркові значення наближаються до оціночної прямої, тим ближче r наближається до 1 і, навпаки, чим далі вибіркові значення віддаляються від оціночної прямої, тим ближче r наближається до 0. Таким чином, коефіцієнт кореляції характеризує «якість» підбору оціночного рівняння.

3.3 ЗАГАЛЬНІ ВІДОМОСТІ ПРО СТАТИСТИЧНІ ОЦІНКИ. НЕЗМІЩЕНІСТЬ І ЕФЕКТИВНІСТЬ ОЦІНОК МНК.

У разі узагальненої економетричної моделі

$$y = a_0 + a_1 x + u$$

випадкова величина u є не спостережуваною величиною, і тому можна зробити лише деякі припущення щодо її поведінки та закону розподілу.

Припущення 1. Математичне сподівання випадкової величини u для всіх спостережень дорівнює нулю:

$$M(u_i) = 0, \quad i = \overline{1, n}.$$

Припущення 2. Дисперсія випадкової величини є однакою для всіх спостережень:

$$M(u_i^2) = \text{var}(u_i) = \sigma^2.$$

Припущення 3. Випадкові величини незалежні між собою (значення випадкового члена u в i -му спостереженні не залежить від того, які значення він приймав в j -му спостереженні):

$$M(u_i u_j) = 0, \quad i \neq j$$

або

$$\text{cov}(u_i, u_j) = 0, \quad i \neq j$$

Припущення 4. Випадкова змінна розподілена незалежно від змінної x :

$$M(u_i x_i) = 0 \text{ або } \text{cov}(u_i, x_i) = 0.$$

Припущення 5. Випадкова змінна «розподілена по нормальному закону з нульовим математичним сподіванням і постійною дисперсією.

Якщо випадковий доданок нормально розподілений, то так будуть розподілені і коефіцієнти регресії a_0, a_1 .

Перш ніж перейти до інтервального аналізу оцінок економетричної моделі у вигляді одного рівняння нагадаємо деякі поняття з математичної статистики та теорії ймовірностей та доведемо властивості оцінок.

Розрізняють наступні оцінки оцінюваних параметрів:

- точкова;
- інтервальна;
- незміщена;
- ефективна;
- спроможна.

Точковою оцінкою оцінюваного параметра називається знайдене значення оцінки (в нашому випадку це \hat{a}_0, \hat{a}_1).

Інтервальною оцінкою оцінюваного параметра називається інтервал, в межах якого з наперед заданою імовірністю лежить істинне значення параметра.

Незміщеною оцінкою оцінюваного параметра a називається така оцінка \hat{a} , математичне сподівання якої дорівнює оцінюваному параметру, тобто

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

Незміщеність - це мінімальна вимога, яка ставиться до оцінок параметрів a . Якщо оцінка незміщена, то при багаторазовому повторенні випадкової вибірки попри те, що для окремих вибірок, можливо, були помилки оцінки, середнє значення цих помилок дорівнює нулю.

Якщо ж математичне сподівання не дорівнює оцінюваному параметру a , то оцінка \hat{a} називається зміщеною. Різниця між математичним сподіванням оцінки і значенням оціненого параметра $\theta = M(\hat{a}) - a$ називається зміщенням оцінки.

Не можна плутати *помилку* оцінки з її зміщенням. *Помилка* дорівнює $\hat{a} - a$ і є випадковою величиною, а зміщення - величина стала.

Дуже важливою властивістю оцінки вважається її обґрунтованість.

Вибіркова оцінка \hat{a} параметрів a називається обґрунтованою (спроможною), якщо при довільному $\varepsilon > 0$ справджується співвідношення

$$\lim_{n \rightarrow \infty} P\{|\hat{a} - a| < \varepsilon\} = 1$$

Іншими словами, оцінка обґрунтована, коли вона задовольняє закон великих чисел. Обґрунтованість оцінки означає, що чим більші будуються вибірки, тим більша ймовірність того, що помилка оцінки не перевищуватиме достатньо малої величини ε .

Ефективною оцінкою оцінюваного параметра a називається така оцінка \hat{a} , яка при заданому об'ємі вибірки має найменшу дисперсію (σ^2). Ефективна оцінка забезпечує стійкість економетричної моделі на предмет відсутності систематичних додатних чи від'ємних відхилень розрахункових значень від фактичних при малих виходах за границі вибірки.

Оцінка a параметрів \hat{a} називається інваріантною, якщо для довільно заданої функції g оцінка параметрів функції $g(a)$ подається у вигляді $g(\hat{a})$.

Перш ніж доводити, що оцінки \hat{a}_0 , і \hat{a}_1 є незміщені, покажемо, що вони лінійно залежать від вибірових значень y_i .

Проведемо перетворення оцінки \hat{a}_1 . Згідно

$$\hat{a}_1 = \frac{\sum_{i=1}^n \Delta x_i \Delta y_i}{\sum_{i=1}^n \Delta x_i^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} - \frac{\bar{y} \sum_{i=1}^n (x_i - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Так як $\sum_{i=1}^n (x_i - \bar{x}) = 0$ то

$$\hat{a}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{i=1}^n (x_i - \bar{x})^2} = \sum_{i=1}^n w_i y_i \quad (3.11)$$

де $w_i = \frac{x_i - \bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}$. Для величини w_i справджуються тотожності:

$$\sum_{i=1}^n w_i = 0, \quad \sum_{i=1}^n w_i^2 = \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad \sum_{i=1}^n w_i (x_i - \bar{x}) = \sum_{i=1}^n w_i x_i = 1.$$

Здійснюємо перетворення над \hat{a}_0

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x} = \sum_{i=1}^n \left(\frac{1}{n} - \bar{x} w_i \right) y_i \quad (3.12)$$

Лінійна залежність від вибірових значень y_i показана. Для доведення незміщеності знайдемо математичне сподівання оцінки \hat{a}_1 :

$$\begin{aligned} M \hat{\epsilon}_1 &\stackrel{\text{def}}{=} M \left(\sum_{i=1}^n w_i y_i \right) = M \left[\sum_{i=1}^n w_i \left(\epsilon_0 + a_1 x_i + u_i \right) \right] = \\ &= M \left[a_0 \sum_{i=1}^n w_i + a_1 \sum_{i=1}^n w_i x_i + \sum_{i=1}^n w_i u_i \right] = M \left[a_0 \cdot 0 + a_1 \cdot 1 + \sum_{i=1}^n w_i u_i \right] = \\ &= M \left(\epsilon_0 \right) + \sum_{i=1}^n w_i M \left(\epsilon_i \right) \stackrel{\text{def}}{=} a_1 \end{aligned}$$

Так як математичне сподівання оцінки \hat{a}_1 дорівнює оцінюваному параметру a_0 , то \hat{a}_1 є незміщеною оцінкою.

Знайдемо математичне сподівання оцінки \hat{a}_0 проробивши перед тим наступні перетворення:

$$\begin{aligned} \hat{a}_0 &= \sum_{i=1}^n \left(\frac{1}{n} - \bar{x} w_i \right) y_i = \sum_{i=1}^n \left(\frac{1}{n} - \bar{x} w_i \right) \left(\epsilon_0 + a_1 x_i + u_i \right) \stackrel{\text{def}}{=} a_0 - a_0 \bar{x} \sum_{i=1}^n w_i + \\ &+ a_1 \bar{x} - a_1 \bar{x} \sum_{i=1}^n w_i x_i + \sum_{i=1}^n \left(\frac{1}{n} - \bar{x} w_i \right) u_i = a_0 + \sum_{i=1}^n \left(\frac{1}{n} - \bar{x} w_i \right) u_i \\ M \hat{\epsilon}_0 &\stackrel{\text{def}}{=} M \left[a_0 + \sum_{i=1}^n \left(\frac{1}{n} - \bar{x} w_i \right) u_i \right] = M \left(\epsilon_0 \right) + M \left[\sum_{i=1}^n \left(\frac{1}{n} - \bar{x} w_i \right) u_i \right] = \\ &= a_0 + \sum_{i=1}^n \left(\frac{1}{n} - \bar{x} w_i \right) \cdot M \left(\epsilon_i \right) \stackrel{\text{def}}{=} a_0 \end{aligned}$$

Отже, \hat{a}_0 і \hat{a}_1 є незміщеними оцінками.

Якщо оцінка незміщена, то при багаторазовому повторенні випадкової вибірки, навіть якщо для окремих вибірок, можливо, були помилки оцінки (помилка дорівнює $\hat{a}_1 - a_1$ є випадковою величиною), середнє значення цих помилок дорівнює нулю. Різниця між математичним сподіванням оцінки та кількісним значенням оціненого параметра $M \hat{\epsilon}_1 \stackrel{\text{def}}{=} a_1$ називається зміщенням оцінки. Зміщення є сталою величиною.

Знайдемо дисперсії оцінок.

$$\sigma_{\hat{a}_1}^2 = M \left[\left(\hat{a}_1 - a_1 \right)^2 \right] = M \left[\sum_{i=1}^n \left(w_i u_i \right)^2 \right] = \sigma_u^2 \sum_{i=1}^n \left(w_i \right)^2 = \frac{\sigma_u^2}{\sum_{i=1}^n \left(x_i - \bar{x} \right)^2}$$

Отже,

$$\sigma_{\hat{a}_1}^2 = \frac{\sigma_u^2}{\sum_{i=1}^n \left(x_i - \bar{x} \right)^2} \quad (3.13)$$

де $\sigma_u^2 = \frac{\sum_{i=1}^n \left(y_i - \hat{y}_i \right)^2}{n-2} = \frac{\sum u^2}{n-2}$ – оцінка дисперсії випадкової величини u .

Дисперсія \hat{a}_0 :

$$\begin{aligned} \sigma_{\hat{a}_0}^2 &= M \left[\left(\hat{a}_0 - a_0 \right)^2 \right] = \sigma_u^2 \sum_{i=1}^n \left(\frac{1}{n} - \bar{x} w_i \right)^2 = \sigma_u^2 \left(\frac{1}{n} - \bar{x}^2 \sum_{i=1}^n \left(w_i \right)^2 - \frac{2\bar{x}}{n} \sum_{i=1}^n w_i \right) \\ \sigma_{\hat{a}_0}^2 &= \frac{\sum_{i=1}^n x_i^2 \sigma_u^2}{n \sum_{i=1}^n \left(x_i - \bar{x} \right)^2} \end{aligned} \quad (3.14)$$

Можна довести, що дисперсії $\sigma_{\hat{a}_0}^2$ і $\sigma_{\hat{a}_1}^2$ є найменшими величинами в класі незміщених оцінок, тобто оцінки \hat{a}_0 і \hat{a}_1 є ефективними.

Примітка. При розв'язуванні контрольних і лабораторних робіт для спрощення розрахунків можна використовувати для обчислень дисперсій $\sigma_{\hat{a}_0}^2$ і

$\sigma_{\hat{a}_1}^2$ наступні формули:

$$\sigma_{\hat{a}_0}^2 = \frac{\sum_{i=1}^n x_i^2 \sigma_{\text{непоясн}}^2}{n \sum_{j=1}^n \left(x_j - \bar{x} \right)^2}, \quad \sigma_{\hat{a}_1}^2 = \frac{\sigma_{\text{непоясн}}^2}{\sum_{j=1}^n \left(x_j - \bar{x} \right)^2}$$

3.4 ПЕРЕВІРКА НУЛЬОВИХ ГІПОТЕЗ. ПОБУДОВА ІНТЕРВАЛІВ ДОВІР'Я РІВНЯННЯ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ.

При формулюванні задачі досвідчений дослідник інтуїтивно передбачає результат. Передбачуваний результат називають *гіпотезою*. Далі дослідник складає план експерименту, за яким виконує його. Аналіз даних експерименту визначає істинність чи хибність гіпотези з певним ступенем надійності висновку. Таким є попереднє окреслення *задач перевірки гіпотез*. Тут важливо, щоб формулювання гіпотез здійснювалося до одержання даних експерименту.

Зазначимо також, що розмежування задач оцінювання і задач перевірки гіпотез є умовним. Можливим є їх перетин. Адже обидва класи задач належать до загальної теорії статистичного висновку.

Статистична гіпотеза — це деяке твердження (припущення) щодо властивостей даної генеральної сукупності об'єктів. Властивості сукупності об'єктів описуються ознаками, які підпорядковуються відповідним законам розподілу. Тому статистична гіпотеза є передбаченням про значення параметрів заданого розподілу ознаки або про тип закону розподілу цієї ознаки. Статистична гіпотеза доповнюється альтернативною (конкуруючою) гіпотезою H_1 . Форма запису статистичних гіпотез така:

H_0 : "зміст нульової гіпотези";

H_1 "зміст альтернативної гіпотези".

Приклад . Відомо, що висока концентрація нітратів в овочах є шкідливою. Сформулювати відповідні гіпотези перед перевіркою.

Перед дослідженням овочів на вміст нітратів і утворенням вибірки вважаємо, що вже є сформульовані гіпотези:

H_0 : вміст нітратів в овочах не шкідливий для здоров'я споживача;

H_1 : концентрація нітратів висока і це може мати шкідливі наслідки.

Перевіряють завжди гіпотезу H_0 , яку називають *нульовою гіпотезою*. Гносеологія терміна належить філософії. Адже один з методів філософського аналізу — це наведення доказів для анулювання гіпотез.

Як і довільна гіпотеза, статистична гіпотеза може бути істинною або хибною. Як тільки з'являється факт, який заперечує істинність гіпотези, вона відхиляється як хибна. Таким чином, перевірка статистичної гіпотези — це процедура виявлення її хибності. Нульова гіпотеза відхиляється при експериментальному доказі хибності її формулювання. До того ж нульову гіпотезу не вважають істинною, коли дані експерименту їй не заперечують. Тоді висновок експерименту звучить таким чином: немає підстав відхилити нульову гіпотезу.

Перевірку статистичної гіпотези здійснюють на основі вибіркової сукупності. Нехай вибірка має об'єм n . Тоді X_1, X_2, \dots, X_n — випадкові величини ознаки, які будуть виміряні в запланованому експерименті. Перевірку гіпотези здійснюють за допомогою *критерію перевірки* K . Критерій перевірки — випадкова величина, яка є статистикою вибірки. Закон розподілу критерію вважають відомим. Значення критерію, які не суперечать змісту нульової гіпотези, належать до *області допустимих значень* O . Решта значень критерію належать до *критичної області* W . Якщо значення критерію K належать до критичної області, то нульова гіпотеза H_0 відхиляється на користь альтернативної. Точки критерію, які відділяють область допустимих значень від критичної області, називають *критичними точками* k^* . Звичайно мають місце випадки, зображені на рис. 3.2, де:

- а) правостороння критична область $W \in [k^*; \infty)$;
- б) лівостороння критична область $W \in (-\infty; k^*]$;
- в) двостороння критична область $W \in (-\infty; k_1^*] \cup [k_2^*; \infty)$.

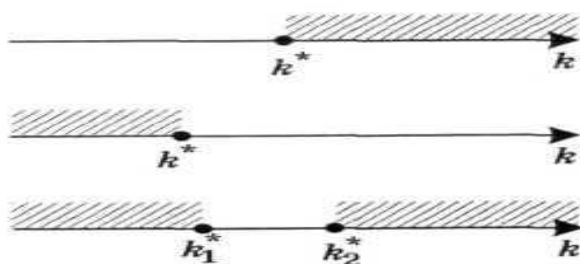


Рис. 3.1

Рис. 3.2 Области значень

Перевірка статистичної гіпотези здійснюється на основі методології статистичного висновку за даними вибірки. У цьому розумінні загальноживаним є термін *статистична перевірка гіпотез*. Статистичний висновок про властивості генеральної сукупності за даними вибірки не може бути категоричним, а тільки має певний ступінь надійності. При перевірці статистичної гіпотези за даними вибірки дослідник завжди ризикує прийняти неправильне рішення. Тому метод перевірки гіпотези доповнюється показниками ризику прийняття хибного рішення.

Розрізняють два типи помилок, яких припускаються при статистичній перевірці гіпотез: відхилення гіпотези H_0 , коли вона є дійсною, — *помилка першого роду*; прийняття гіпотези H_0 , коли вона є хибною, — *помилка другого роду*.

Ймовірність помилки першого роду називають *рівнем значущості* критерію. Рівень значущості α означає ймовірність відхилення правильної гіпотези:

$$\alpha = \Pr \{ \text{відхиляється } H_0 | H_0 \text{ дійсна} \}.$$

Величину $1-\alpha$ називають *надійністю* критерію. Надійність критерію означає ймовірність не допустити помилки першого роду. Понижуючи (зменшуючи) рівень значущості, підвищуємо надійність критерію в тому, щоб не відхилити істинну нульову гіпотезу. Ймовірність помилки другого роду називають *оперативною характеристикою* критерію відносно альтернативної гіпотези. Оперативна характеристика β — ймовірність прийняти хибну нульову гіпотезу:

$$\beta = \Pr \{ \text{приймається } H_0 | H_1 \text{ дійсна} \}.$$

Величину $1-\beta$ називають *потужністю* критерію. Потужність критерію означає ймовірність не допустити помилки другого роду:

$$1-\beta = \Pr \{ \text{відхиляється } H_0 | H_1 \text{ дійсна} \}.$$

Різні варіанти при перевірці гіпотез зведені у табл. 3.1:

Таблиця 3.1

Істина	Статистичний висновок	Тип помилки	Ймовірність
Гіпотеза H_0 істинна	Гіпотеза H_0 не заперечується	Немає помилки	$\Pr \mathcal{K} \in O H_0 \stackrel{\text{I}}{=} 1 - \alpha$
	Гіпотеза H_0 відхиляється	Помилка першого роду	$\Pr \mathcal{K} \in W H_0 \stackrel{\text{I}}{=} \alpha$
Гіпотеза H_0 хибна	Гіпотеза H_0 не заперечується	Помилка другого роду	$\Pr \mathcal{K} \in O H_1 \stackrel{\text{II}}{=} \beta$
	Гіпотеза H_0 відхиляється	Немає помилки	$\Pr \mathcal{K} \in W H_1 \stackrel{\text{II}}{=} 1 - \beta$

Оскільки виключити помилки першого і другого роду неможливо в принципі, тому в кожній конкретній задачі намагаються їх мінімізувати. Помилки першого і другого роду є конкуруючими. Можна ухилитись від помилки першого роду, понижуючи рівень значущості α . Але тоді зростає ймовірність помилки другого роду. Одночасне зменшення ймовірностей помилок першого та другого роду можливе тільки при збільшенні об'єму вибірки.

Отже, нульова гіпотеза не відхиляється, поки немає вагомих доказів її хибності. Доти, доки відхилення від змісту нульової гіпотези зумовлені тільки випадковістю, вона не відхиляється. Для неї підходить статус "презумпції невинності". У цьому розумінні позиція нульової гіпотези є пасивною. Альтернативну гіпотезу нерідко ще називають дослідницькою. Тільки тоді, коли дослідник дає вагомі докази того, що нульова гіпотеза хибна, вона відхиляється на користь альтернативної. Позиція альтернативної гіпотези активна тому, що вона вимагає доказів. Зауважимо ще, що ніколи не говорять про відхилення альтернативної гіпотези.

Оскільки в економетрії статистичні дані створюються випадковими факторами, то більшість досліджень слід перевіряти на деякі припущення або гіпотези про джерела цих даних.

Основне припущення називається нульовою гіпотезою H_0 . В загальному випадку вона полягає в тому, що між параметром вибірки і параметром генеральної сукупності немає суттєвих різниць. Інше, яке його заперечує –

альтернативною гіпотезою H_1 . Альтернативна гіпотеза стверджує, що між параметром вибірки і параметром генеральної сукупності є суттєві різниці.

Називаємо, що ми допустилися помилки першого роду, якщо відкинули нульову гіпотезу, якщо насправді вона вірна. Допустилися помилки другого роду, якщо прийняли нульову гіпотезу, якщо насправді вона не вірна.

При перевірці гіпотез встановлюють рівень значущості γ , іншими словами ймовірність того, що H_0 відкидається, коли вона в дійсності вірна.

Нульова гіпотезу стосовно коефіцієнта кореляції формулюється таким чином: коефіцієнт кореляції генеральної сукупності рівний нулю ($H_0 : r_{\text{ген}}=0$), суттєвої різниці між вибірковим коефіцієнтом кореляції і коефіцієнтом кореляції генеральної сукупності немає.

Альтернативна гіпотеза $H_1: r_{\text{ген}} \neq 0$, між змінними y та x є суттєвий зв'язок.

За результатами вибірки обчислюється статистика

$$t_{emn} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

яка має розподіл Стюдента з $k=n-2$ ступенями вільності. Для заданого рівня значущості γ і k ступенів вільності знаходимо табличне значення $t_{кр}$ за таблицями розподілу Стюдента. Якщо $|t_{emn}| > t_{кр}$ то із заданим рівнем значущості γ гіпотезу H_0 слід відкинути, прийняти альтернативну гіпотезу H_1 про існування залежності між цими випадковими величинами, і вибірковий коефіцієнт кореляції r вважати статистично значущим.

Надійність оцінки визначається імовірністю з якою стверджується, що побудований за результатами вибірки довірчий інтервал містить невідомий параметр генеральної сукупності. Імовірність інтервальної оцінки параметра називають довірчою і позначають p . Тоді можна сподіватися, що при множині спостережень параметр генеральної сукупності буде правильно оцінений (довірчий інтервал покриє дійсне значення цього параметра) з надійністю p . Ризик помилки визначається рівнем значущості γ , який називається довірчим рівнем даного інтервалу.

Якщо параметр генеральної сукупності θ , а його оцінка $\hat{\theta}$, то довірчий інтервал визначається формулою:

$$P \left\{ -k\sigma_{\hat{\theta}} \leq \theta \leq \hat{\theta} + k\sigma_{\hat{\theta}} \right\} = 1 - \gamma$$

де $\sigma_{\hat{\theta}}$ – середньоквадратичне відхилення $\hat{\theta}$, k – довірчий множник, який вказує частку стандартного відхилення, яка повинна бути врахована, щоб із заданою імовірністю p довірчий інтервал $\left[-k\sigma_{\hat{\theta}}; \theta + k\sigma_{\hat{\theta}} \right]$ покрити параметр генеральної сукупності.

Знайдемо довірчий інтервал рівняння економетричної моделі, який визначають область ймовірних значень y , при відповідних значеннях x . Для цього вводять поняття граничної похибки оцінки за рівнянням економетричної моделі

$$\Delta_{yx} = \sigma_{yx} \cdot t_p$$

де t_p – імовірнісний коефіцієнт, значення якого при заданому рівні ймовірності знаходиться за таблицею нормального розподілу (t_p є коренем рівняння $2\Phi(t_p) = p$ де $\Phi(t_p)$ – функція Лапласа).

Нижню межу інтервалу довір'я оціночного рівняння одержують шляхом віднімання граничної похибки від значень \hat{y} по оціночній прямій, а верхню – через додавання цієї ж величини до значень \hat{y} .

Таким чином фактичні значення реальних значень y_i повинні задовольняти нерівності:

$$\hat{y}_i - \Delta_{yx} \leq y_i \leq \hat{y}_i + \Delta_{yx}$$

Замінімо \hat{y} через $\hat{a}_0 + \hat{a}_1 x_i$, тоді одержимо довірчий інтервал у вигляді:

$$\hat{a}_0 + \hat{a}_1 x_i - \Delta_{yx} \leq y_i \leq \hat{a}_0 + \hat{a}_1 x_i + \Delta_{yx}$$

$$\left[\hat{a}_0 - \Delta_{yx}; \hat{a}_1 x_i \right] \leq y_i \leq \left[\hat{a}_0 + \Delta_{yx}; \hat{a}_1 x_i \right]$$

3.4 ПЕРЕВІРКА НУЛЬОВИХ ГІПОТЕЗ І ДОВІРЧІ ІНТЕРВАЛИ ПАРАМЕТРІВ. ПЕРЕВІРКА МОДЕЛІ НА АДЕКВАТНІСТЬ.

Часто здійснюється перевірка нульових гіпотез стосовно параметрів. Розглянемо нульові гіпотези стосовно параметрів a_0 і a_1 .

Стосовно a_0 :

$H_0: a_0=0$. Альтернативна $H_1: a_0 \neq 0$.

Стосовно a_1 :

$H_0: a_1=0$. Альтернативна $H_1: a_1 \neq 0$.

Емпіричне значення відношення t_{emn} для перевірки нульової гіпотези стосовно параметру a_1 знаходиться за формулою:

$$t_{emn} = \frac{|\hat{a}_1|}{\sigma_{\hat{a}_1}}$$

стосовно a_0 :

$$t_{emn} = \frac{|\hat{a}_0|}{\sigma_{\hat{a}_0}}$$

Емпіричне значення оцінки t_{emn} порівнюють з критичним, знайденим за таблицями Стюдента, для заданого рівня значущості γ і $k=n-2$ ступенів вільності. Якщо $|t_{emn}| > t_{кр}$ то із заданим рівнем значущості γ гіпотезу H_0 слід відкинути і відповідну оцінку вважають статистично значущою. Якщо $|t_{emn}| \leq t_{кр}$ оцінка є статистично не значущою.

Побудова довірчих інтервалів для параметрів a_0 і a_1 проводиться шляхом визначення граничної похибки оцінок \hat{a}_0 та \hat{a}_1 . Вони знаходяться аналогічно довірчим інтервалам рівняння економетричної моделі. Спочатку знаходимо граничні похибки оцінок за формулою

$$\Delta_{\hat{a}_0} = \sigma_{\hat{a}_0} t_p, \quad \Delta_{\hat{a}_1} = \sigma_{\hat{a}_1} t_p$$

де t_p – імовірнісний коефіцієнт, який знаходиться за таблицями нормального розподілу; $\sigma_{\hat{a}_0}^2$, $\sigma_{\hat{a}_1}^2$ – дисперсії оцінок \hat{a}_0 та \hat{a}_1 , які визначаються за формулами

$$\sigma_{\hat{a}_0}^2 = \frac{\sum_{i=1}^n x_i^2 \cdot \sigma_{\text{непоясн}}^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}, \quad \sigma_{\hat{a}_1}^2 = \frac{\sigma_{\text{непоясн}}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Довірчий інтервал для a_0 :

$$\hat{a}_0 - \Delta_{\hat{a}_0} \leq a_0 \leq \hat{a}_0 + \Delta_{\hat{a}_0}$$

Довірчий інтервал для a_1 :

$$\hat{a}_1 - \Delta_{\hat{a}_1} \leq a_1 \leq \hat{a}_1 + \Delta_{\hat{a}_1}$$

Уже було сказано, що з допомогою коефіцієнта детермінації перевіряється адекватність (відповідність) побудованої економетричної моделі реальній дійсності. Іншими словами, дається відповідь на запитання, чи справді зміна значення y лінійно залежить саме від зміни значення x , а не відбувається під впливом випадкових факторів. Якщо значення коефіцієнта детермінації близьке до 1, то можна вважати, що побудована модель є адекватна. У випадку якщо значення коефіцієнта детермінації близьке до 0, то можна вважати, що побудована модель є неадекватна. В такому випадку відсутній лінійний зв'язок між змінними. Для визначення адекватності прийнятої економетричної моделі реальній дійсності у випадку, коли значення d має нечітко виражене граничне значення, наприклад, 0.6, 0.55, 0.46 і т.д., користуються найчастіше критерієм Фішера (F-критерій). Він дозволяє оцінити, чи значно нахил \hat{a}_1 відрізняється від нуля, тобто перевірити гіпотезу $H_0 : a_1=0$. Тому що, якщо \hat{a}_1 незначно відрізняється від нуля, тоді з:

$$\hat{y}_i = \hat{a}_0 + \hat{a}_1 x_i = \bar{y} + \hat{a}_1 (x_i - \bar{x})$$

Тобто перевірка H_0 дасть можливість визначити як краще апроксимувати дані: середнім значенням $\hat{y}_i = \bar{y}$ чи моделлю $\hat{y}_i = \hat{a}_0 + \hat{a}_1 x_i$.

Альтернативна гіпотеза полягає в тому, що $a_1 \neq 0$ і позначається $H_1: a_1 \neq 0$.

Перевірка моделі на адекватність за F- критерієм Фішера проходить наступним чином:

1) На першому кроці розраховуємо величину F- відношення:

$$F = \frac{\overline{СКП}}{\overline{СКН}} = \frac{\sum_{i=1}^n (Q_i - \bar{y})^2}{1} = \frac{(n-2) \sum_{i=1}^n (Q_i - \bar{y})^2}{\sum_{i=1}^n (Q_i - \hat{y}_i)^2} = \frac{r^2}{1-r^2} \cdot \frac{n-2}{1}$$

2) На другому - задаємо рівень значущості γ (або $\gamma \cdot 100\%$). Якщо, наприклад, ми задаємо рівень значущості 0,05 (або 5%), то це означає, що ми можемо помилитися в 5% випадків, а в 95% – наші висновки будуть правильними.

3) Третій крок передбачає знаходження критичного значення $F_{кр}$ за статистичними таблицями F- розподілу Фішера з $(1, n-2)$ ступенями вільності і вибраному рівні значущості γ .

4) На четвертому кроці порівнюємо величину розрахованого F з $F_{кр}$. Якщо $F > F_{кр}$, то ми відкидаємо H_0 з ризиком помилитися не більше ніж в $\gamma\%$ випадків, і приймаємо, що побудоване рівняння економетричної моделі адекватне реальній дійсності. В протилежному випадку $F \leq F_{кр}$ – H_0 приймаємо і вважаємо, що побудована модель неадекватна. Тоді необхідно, можливо, будувати нелінійну модель або ввести додаткові фактори.

Зауваження. Для лінійної економетричної моделі нульова гіпотеза $H_0: a_1=0$ аналогічна перевірці адекватності моделі за F- критерієм Фішера.

Якщо рівняння економетричної моделі адекватне реальній дійсності, то по ньому можна здійснювати прогноз. *Прогнозування* – це наукове передбачення імовірнісних шляхів розвитку явищ і процесів для більш-менш віддаленого майбутнього. Прогноз буває точковий та інтервальний. Точковий прогноз в $n+1$ періоді одержуємо безпосередньою підстановкою значення незалежної змінної x_{n+1} в рівняння економетричної моделі:

$$\hat{y}_{n+1} = \hat{a}_0 + \hat{a}_1 x_{n+1}$$

Прогнозне значення \hat{y}_{n+1} є оцінкою дійсного значення змінної y_{n+1} . Інтервал, у який з певною заданою імовірністю $p=1-\gamma$ потрапляє дійсне значення залежної змінної визначається за формулою:

$$\hat{y}_{n+1} \pm t_{\gamma} \cdot \sigma_{yx} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

(Нижню межу інтервалу довіри часто називають песимістичним прогнозом, верхню – оптимістичним).

На практиці більше застосовується побудова інтервалів довіри для математичного сподівання y_{n+1} , тобто для $M y_{n+1} = \hat{a}_0 + \hat{a}_1 x_{n+1}$. Для $\gamma \cdot 100\%$ – ного рівня значущості він має вигляд:

$$\hat{y}_{n+1} \pm t_{\gamma} \cdot \sigma_{yx} \cdot \sqrt{\frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

МЕТОДИКА ПЕРЕВІРКИ ОЦІНОК ОДНОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ

Задача. На основі даних попередньої задачі потрібно:

- обчислити загальну, пояснену і непояснену дисперсії;
- інтервал довір'я ($p=0,9$) рівняння економетричної моделі;
- дати геометричну інтерпретацію оціночних рівнянь;
- коефіцієнти детермінації та кореляції;
- інтервал довір'я ($p=0,9$) параметрів \hat{a}_0, \hat{a}_1 ;
- перевірити нульову гіпотезу щодо коефіцієнта кореляції r ($p=0,95$);
- перевірити нульову гіпотезу щодо кутового коефіцієнта \hat{a}_1 ($p=0,95$);
- перевірити адекватність прийнятої економетричної моделі експериментальним даним ($p=0,95$).

Розв'язання:

Знаходимо загальну, пояснену і непояснені дисперсії.

Необхідні обчислення приведені в таблиці:

№	x_i	y_i	Δy_i	$\left(\Delta y_i\right)^2$	\hat{y}_i	$\hat{y}_i - \bar{y}$	$\left(\hat{y}_i - \bar{y}\right)^2$	$y_i - \hat{y}_i$	$\left(y_i - \hat{y}_i\right)^2$
1	3,2	105,4	-45,48	2069,29	91,8883	-59,0006	3481,0697	13,5117	182,5661
2	2,5	114,9	-35,98	1295,2	120,0904	-30,7985	948,54683	-5,1904	26,94027
3	2,1	120,3	-30,58	935,681	136,2059	-14,683	215,59048	-15,9059	252,9973
4	1,62	150,1	-0,788	0,62234	155,5445	4,655585	21,674475	-5,44447	29,6423
5	1,35	152,4	1,5111	2,28345	166,4224	15,53354	241,29085	-14,0224	196,6285
6	1,33	163,3	12,411	154,035	167,2282	16,33931	266,97318	-3,9282	15,43078
7	1,17	171,4	20,511	420,705	173,6744	22,78551	519,17942	-2,2744	5,172885
8	1,25	185,1	34,211	1170,4	170,4513	19,56241	382,68794	14,6487	214,5844
9	1,1	195,1	44,211	1954,62	176,4946	25,60572	655,65286	18,60539	346,1606
Σ	15,62	1358		8002,78	1358		6732,6658		1270,123

$$\frac{\sum_{i=1}^n \left(\Delta y_i\right)^2}{n} = \sigma_{заг}^2; \quad \frac{\sum_{i=1}^n \left(\hat{y}_i - \bar{y}\right)^2}{n} = \sigma_{пояс}^2; \quad \frac{\sum_{i=1}^n \left(y_i - \hat{y}_i\right)^2}{n} = \sigma_{непоясн}^2$$

$$\sigma_{заг}^2 = \frac{8002,78}{9} = 889,1988;$$

$$\sigma_{пояс}^2 = \frac{6732,6658}{9} = 748,074;$$

$$\sigma_{непоясн}^2 = \frac{1270,123}{9} = 141,1248 \text{ а відповідні їм середньоквадратичні відхилення}$$

$$\sigma_{заг} = \sqrt{889,1988} = 29,82; \quad \sigma_{пояс} = \sqrt{748,074} = 27,35; \quad \sigma_{непоясн} = \sqrt{141,1248} = 11,88$$

Інтервал довір'я ($p=0,9$) рівняння економетричної моделі:

Граничну похибку оцінки за рівнянням економетричної моделі знаходимо по формулі $\Delta_{yx} = t_p \sigma_{непоясн}$, де t_p – імовірнісний коефіцієнт, який при заданих рівнях імовірності p знаходиться за таблицями нормального розподілу. При цьому розв'язуємо рівняння $2\Phi(t_p)=p$, де $\Phi(t_p)$ – функція Лапласа. Для $p=0,9$ ($t_p=1,65$). Одержуємо: $\Delta_{yx} = 1,65 \cdot 11,88 = 19,6$. Довірчий інтервал знаходимо з нерівності:

$$\hat{y}_i - 19,6 \leq y_i \leq \hat{y}_i + 19,6$$

$$220,81 - 40,29x_i - 19,6 \leq y_i \leq 220,81 - 40,29x_i + 19,6$$

$$201,21 - 40,29x_i \leq y_i \leq 240,41 - 40,29x_i$$

Для наочного уявлення одержаних розрахунків будуюмо графіки: фактичних даних y , оціночного рівняння \hat{y}_i , довірчого інтервалу.

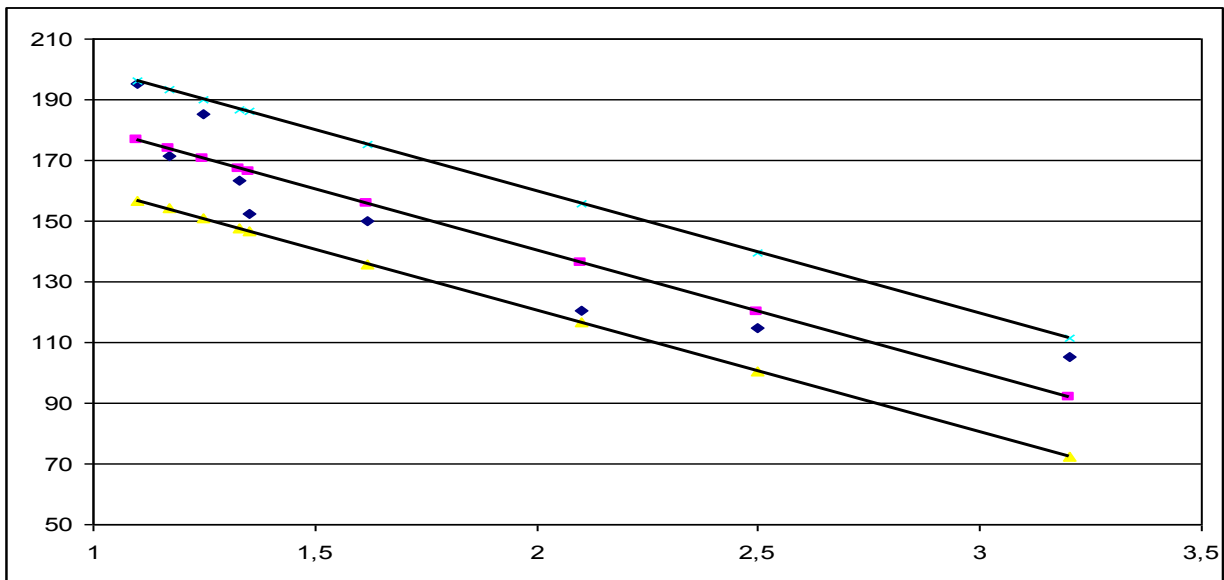


Рис.3.4 Відхилення теоретичних значень від фактичних

Коефіцієнти детермінації та кореляції

Знайдемо значення коефіцієнта детермінації за формулою:

$$d = r^2 = 1 - \frac{\sigma_{\text{непоясн}}^2}{\sigma_{\text{заг}}^2} = 1 - \frac{141,1248}{889,1988} = 0,16$$

і коефіцієнта кореляції за формулою

$$r = \pm \sqrt{1 - \frac{\sigma_{\text{непоясн}}^2}{\sigma_{\text{заг}}^2}} = -\sqrt{0,16} = -0,4.$$

Оскільки $d=0,16$, то це означає, що оціночна пряма пояснює 16% загальної дисперсії y . Дисперсія зумовлена випадковою складовою u (невраховані фактори, помилки виміру, суб'єктивний чинник) складає 84%.

Коефіцієнт кореляції $r = -0,4$ показує, що між прибутком підприємства і заробітною платою існує слабкий обернений лінійний зв'язок. Знак коефіцієнта кореляції повинен співпадати з знаком \hat{a}_1 .

Інтервали довір'я ($p=0,9$) параметрів a_0, a_1 .

Переходимо до знаходження довірчих інтервалів для параметрів a_0 та a_1 . Знаходимо аналогічно довірчим інтервалам оцінки за рівнянням економетричної моделі. Граничні похибки оцінок обраховуємо за формулою

$$\Delta_{\hat{a}_0} = t_p \sigma_{\hat{a}_0}, \quad \Delta_{\hat{a}_1} = t_p \sigma_{\hat{a}_1}$$

де $\sigma_{\hat{a}_0}$, $\sigma_{\hat{a}_1}$ – середньоквадратичні відхилення оцінок \hat{a}_0 , \hat{a}_1 , які визначаємо за формулами

$$\sigma_{\hat{a}_0} = \sqrt{\frac{\sum_{i=1}^n x_i^2 \sigma_{\text{непохи}}^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}} = \sqrt{\frac{31,2572 \cdot 141,1248}{9 \cdot 4,1478222}} = \sqrt{118,17} = 10,87$$

$$\sigma_{\hat{a}_1} = \sqrt{\frac{\sigma_{\text{непохи}}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} = \sqrt{\frac{141,1248}{4,1478222}} = \sqrt{34,02} = 5,83.$$

$$\Delta_{\hat{a}_0} = 1,65 \cdot 10,87 = 17,94, \quad \Delta_{\hat{a}_1} = 1,65 \cdot 5,83 = 9,62,$$

отже довірчий інтервал параметра a_0

$$\hat{a}_0 - \Delta_{\hat{a}_0} \leq a_0 \leq \hat{a}_0 + \Delta_{\hat{a}_0}$$

$$220,8122 - 17,94 \leq a_0 \leq 220,8122 + 17,94$$

$$202,81 \leq a_0 \leq 238,75$$

довірчий інтервал параметра a_1

$$\hat{a}_1 - \Delta_{\hat{a}_1} \leq a_1 \leq \hat{a}_1 + \Delta_{\hat{a}_1}$$

$$-40,2887 - 9,62 \leq a_1 \leq -40,2887 + 9,62$$

$$-49,91 \leq a_1 \leq -30,66$$

Перевірка нульової гіпотези щодо коефіцієнта кореляції r ($p=0,95$).

Коефіцієнт кореляції r знайдено на основі вибірових даних і є випадковою величиною. Здійснимо перевірку нульової гіпотези: коефіцієнт кореляції в генеральній сукупності дорівнює нулю (відсутній кореляційний зв'язок між x і y в генеральній сукупності) і необхідно дослідити сумісність коефіцієнта кореляції із нашої вибірки r з даною гіпотезою. Нульова гіпотеза $H_0: r_{ген}=0$, альтернативна – $H_1: r_{ген} \neq 0$. Для вибірки обчислюється статистика

$$t_{емп} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{-0,4 \cdot \sqrt{9-2}}{\sqrt{1-0,16}} = -1,1547$$

Яка має розподіл Стюдента з $k=n-2=9-2=7$ ступенями вільності. Для заданої імовірності $p=0,95$ і $k=7$ ступенів вільності знаходимо табличне значення t -статистики, яке становить $t_{кр}=2,365$.

Якщо $|t_{емп}| \geq t_{кр}$, то із заданою надійністю p гіпотезу H_0 слід відкинути і прийняти альтернативну H_1 про існування залежності між цими випадковими величинами.

В даному випадку емпіричне значення менше критичного, тому приймається нульова гіпотеза. В 95% вибірок з генеральної сукупності коефіцієнт кореляції дорівнює нулю.

Перевірка нульової гіпотези щодо кутового коефіцієнта \hat{a}_1 ($p=0,95$).

Знаходимо емпіричне значення відношення $t_{емп}$ для перевірки нульової гіпотези за формулою:

$$t_{емп} = \frac{|\hat{a}_1|}{\sigma_{\hat{a}_1}} = \frac{40,3}{5,83} = 6,91, \quad t_{кр} = 2,365.$$

Емпіричне значення t більше критичного, тому нульова гіпотеза відхиляється. Кутовий коефіцієнт \hat{a}_1 розрахований по вибірці вважається статистично значущим з імовірністю 0,95.

Перевірка адекватності прийнятої економетричної моделі експериментальним даним ($p=0,95$).

Для цього скористаємося F-критерієм Фішера. Розрахункове значення критерію:

$$F = \frac{r^2}{1-r^2} \cdot \frac{n-m-1}{m} = \frac{0,16}{1-0,16} \cdot \frac{9-1-1}{1} = 1,33$$

Табличне значення критерію для імовірності $p=0,95$ і числа ступенів вільності $k_1=m=1$, $k_2=n-m-1=9-2=7$ дорівнює 5,59.

Оскільки $F_{табл} > F_{роз}$ то з надійністю $p=0,95$ економетричну модель можна вважати неадекватною, і проводити економічний аналіз не варто.

ОДНОФАКТОРНІ НЕЛІНІЙНІ ЕКОНОМЕТРИЧНІ МОДЕЛІ

4.1 КРИВІ ЗРОСТАННЯ. ЗВЕДЕННЯ НЕЛІНІЙНИХ МОДЕЛЕЙ ДО ЛІНІЙНИХ.

Криві зростання описують різні тенденції економічних процесів, наприклад, життєвий цикл товару, процес нагромадження капіталу, маркетингові зусилля фірм тощо. Економічна практика вже накопичила певний досвід і певні типи кривих, які найчастіше використовуються в макро- та мікроекономічних дослідженнях. До таких кривих відносяться:

експотенційна:

$$y = a_0 a_1^x$$

степенева (мультиплікативна):

$$y = a_0 x^{a_1}$$

зворотна:

$$y = a_0 + a_1 \frac{1}{x}$$

квадратична:

$$y = a_0 + a_1 x + a_2 x^2$$

модифікована експонента:

$$y = a_0 a_1^x + a_2$$

крива Гомерця:

$$y = e^{a_0 a_1^x + a_2}$$

логістична крива:

$$y = \frac{1}{a_0 a_1^x + a_2}$$

де $f(x)$ — одна з функцій зростання; ε — випадкова величина.

Як і у випадку з простою лінійною регресією, основне завдання полягає в розрахунку невідомих параметрів кривих зростання і подальшому аналізі обраної моделі. Оцінку невідомих параметрів проводять по-різному: експоненційні функції шляхом логарифмічних перетворень зводять до лінійної регресії, квадратичні функції зводять до багатofакторної регресії, для інших використовують ітеративні методи, метод трьох точок, метод Тейла тощо. В цьому розділі ми особливу увагу звернемо на функції зростання, які шляхом перетворень зводяться до лінійної регресії. Для таких функцій зберігається вся

методологія досліджень, яку ми детально розглядали у випадку простої лінійної регресії. Крім того, ми розглянемо найпоширеніші в сучасному економічному аналізі криві, які не зводяться шляхом елементарних перетворень до лінійної регресії, але параметри яких можна легко розрахувати спеціальними спрощеними методами.

Спираючись на метод найменших квадратів, ми можемо одержати чисельні значення оцінок параметрів для всіх залежностей і обчислити в кожному випадку значення функції помилок, тобто суму квадратів різниць значень результативної змінної, що спостерігалися, і що розраховувалися за відповідною формулою. Найбільш адекватною досліджуваному процесу природно вважати ту модель, для якої ця сума буде мінімальною.

Особливістю регресійних моделей є те, що параметри a_0, a_1, \dots, a_m і відхилення u входять у них лінійно. Не всі залежності задовольняють цій вимозі. Інші моделі нелінійні щодо параметрів і їх необхідно шляхом перетворень зводити до лінійних моделей. Покажемо як здійснювати оцінку параметрів. Розглянемо функцію $y = a_0 + a_1x + a_2x^2$. В цій функції робимо заміну $x=t_1, x^2=t_2$ тоді квадратична функція зветься до лінійної відносно факторів t_1, t_2 . На практиці часто застосовуються крім методів заміни методи логарифмування і знаходження оберненого значення.

У випадку логарифмічної функції регресії $\hat{y} = a_0 + a_1 \ln x$ одержуємо:

$$\begin{cases} n\hat{a}_0 + \hat{a}_1 \sum_{i=1}^n \ln x_i = \sum_{i=1}^n y_i \\ a_0 \sum_{i=1}^n \ln x_i + \hat{a}_1 \sum_{i=1}^n \ln^2 x_i = \sum_{i=1}^n y_i \ln x_i \end{cases} \quad (4.1)$$

Якщо обрано гіперболічну залежність $\hat{y} = a_0 + a_1x^{-1}$ то, поклавши $t=x^{-1}$, для визначення оцінок \hat{a}_0 і \hat{a}_1 приходимо до системи рівнянь

$$\begin{cases} n\hat{a}_0 + \hat{a}_1 \sum_{i=1}^n \frac{1}{t_i} = \sum_{i=1}^n y_i \\ \hat{a}_0 \sum_{i=1}^n \frac{1}{t_i} + \hat{a}_1 \sum_{i=1}^n \left(\frac{1}{t_i}\right)^2 = \sum_{i=1}^n y_i \frac{1}{t_i} \end{cases} \quad (4.2)$$

Якщо вихідна модель характеризується показниковою функцією $\hat{y} = a_0 \cdot a_1^x$, то покладемо $y = a_0 \cdot a_1^x \cdot e^u$ і після логарифмування одержуємо лінійну регресійну модель: $y^* = a_0^* + a_1^* x + u$, $y^* = \ln y$, $a_0^* = \ln a_0$, $a_1^* = \ln a_1$.

Далі маємо

$$\begin{cases} n\hat{a}_0^* + \hat{a}_1^* \sum_{i=1}^n x_i = \sum_{i=1}^n \ln y_i \\ \hat{a}_0^* \sum_{i=1}^n x_i + \hat{a}_1^* \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i \ln y_i \end{cases} \quad (4.3)$$

Нехай вихідна модель описується степеневою функцією $\hat{y} = a_0 \cdot x_i^{a_1}$. З метою отримання лінійної регресійної моделі покладемо $y = a_0 \cdot x_i^{a_1} \cdot e^u$. Отже, $\ln y = \ln a_0 + a_1 \ln x + u$ і рівняння $y^* = a_0^* + a_1 \ln x + u$ $y^* = \ln y$, $a_0^* = \ln a_0$ стає лінійним щодо параметрів \hat{a}_0^* і a_1 відхилення u . Система рівнянь для оцінок \hat{a}_0^* і \hat{a}_1 набуває вигляду

$$\begin{cases} n\hat{a}_0^* + \hat{a}_1 \sum_{i=1}^n \ln x_i = \sum_{i=1}^n \ln y_i \\ \hat{a}_0^* \sum_{i=1}^n x_i + \hat{a}_1 \sum_{i=1}^n \ln^2 x_i = \sum_{i=1}^n \ln x_i \ln y_i \end{cases} \quad (4.4)$$

Якщо вихідна модель описується функцією $\hat{y} = e^{a_0 + \frac{a_1}{x}}$, то, з метою отримання лінійної регресійної моделі, покладемо $y = e^{a_0 + \frac{a_1}{x} + u}$

Після логарифмування одержуємо $y^* = a_0 + \frac{a_1}{x} + u$ ($y^* = \ln y$).

Далі маємо

$$\begin{cases} n\hat{a}_0 + \hat{a}_1 \sum_{i=1}^n \frac{1}{x_i} = \sum_{i=1}^n \ln y_i \\ \hat{a}_0 \sum_{i=1}^n \frac{1}{x_i} + \hat{a}_1 \sum_{i=1}^n \frac{1}{x_i^2} = \sum_{i=1}^n \frac{1}{x_i} \cdot \ln y_i \end{cases} \quad (4.5)$$

Припустимо, що моделююча функція $\hat{y} = \frac{1}{a_0 + a_1 x}$ тоді зв'язок між змінними y і x будемо шукати у вигляді $y = \frac{1}{a_0 + a_1 x + u}$, що приводить до наступної перетвореної (лінійної) моделі $\frac{1}{y} = a_0 + a_1 x + u$. Тепер

$$\begin{cases} n\hat{a}_0 + \hat{a}_1 \sum_{i=1}^n x_i = \sum_{i=1}^n \frac{1}{y_i} \\ \hat{a}_0 \sum_{i=1}^n x_i + \hat{a}_1 \sum_{i=1}^n x_i^2 = \sum_{i=1}^n \frac{x_i}{y_i} \end{cases} \quad (4.6)$$

Якщо моделююча функція $\hat{y} = \frac{1}{a_0 + a_1 \ln x}$, то покладаємо

$y = \frac{1}{a_0 + a_1 \ln x + u}$ відповідно лінійна модель $\frac{1}{y} = a_0 + a_1 \ln x + u$. Далі маємо:

$$\begin{cases} n\hat{a}_0 + \hat{a}_1 \sum_{i=1}^n \ln x_i = \sum_{i=1}^n \frac{1}{y_i} \\ a_0 \sum_{i=1}^n \ln x_i + \hat{a}_1 \sum_{i=1}^n \ln^2 x_i = \sum_{i=1}^n \frac{\ln x_i}{y_i} \end{cases} \quad (4.7)$$

У випадку залежності $\hat{y} = \frac{x}{a_0 + a_1 x}$ або $\hat{y} = \frac{1}{a_1 + \frac{a_0}{x}}$ маємо $y = \frac{1}{a_1 + \frac{a_0}{x} + u}$,

$\frac{1}{y} = a_1 + \frac{a_0}{x} + u$. Система нормальних рівнянь, що відповідає лінійній моделі,

така:

$$\begin{cases} n\hat{a}_1 + \hat{a}_0 \sum_{i=1}^n \frac{1}{x_i} = \sum_{i=1}^n \frac{1}{y_i} \\ \hat{a}_1 \sum_{i=1}^n \frac{1}{x_i} + \hat{a}_0 \sum_{i=1}^n \frac{1}{x_i^2} = \sum_{i=1}^n \frac{1}{x_i y_i} \end{cases}$$

4.2 ПРИКЛАДИ ЗАСТОСУВАННЯ ЕКСПОТЕНЦІЙНИХ ФУНКЦІЙ.

Експоненційна функція може набирати різних чисельних еквівалентних форм:

$$y = a_0 a_1^x, \text{ основна форма } a_1 > 0 \quad (4.8)$$

$$y = a_0 \cdot e^{b_1 x}, a_1 \text{ замінюємо на } e^{b_1 x}, \text{ де } b_1 = \ln(a_1) \quad (4.9)$$

$$y = a_0 \cdot (1-r)^x, a_1 \text{ замінюємо на } (1-r) \text{ де } r = a_1 - 1 \quad (4.10)$$

$$y = e^{b_0 + b_1 x} a_0 \text{ замінюємо на } e^{b_0} \text{ та } a_1 \text{ на } e^{b_1} \text{ де } b_0 = \ln(a_0) \text{ і } b_1 = \ln(a_1) \quad (4.11)$$

$$y = 10^{a + bx} a_0 \text{ замінюємо на } 10^a \text{ та } a_1 \text{ на } 10^b \text{ де } a = \log(a_0) \text{ і } b = \log(a_1) \quad (4.12)$$

Усі ці форми використовуються на практиці для опису різних економічних процесів, наприклад, форму (4.10) найчастіше використовують у фінансах. У цьому разі r інтерпретується як норма річного відсотка. Розглянемо декілька прикладів застосування експоненційної функції у бізнесі та фінансах.

Приклад 1. Припустимо, що капітал C_0 знаходиться у банку протягом t років із річним відсотком r . Він змінюється відповідно до функції (4.10), тобто через t років капітал дорівнюватиме:

$$C_t = C_0 (1+r)^t$$

З експоненційної форми (4.9) легко отримати так зване **правило 70**, поширене в фінансових розрахунках. Правило 70 дає значення часу t , через яке змінна подвоїть своє значення відносно часу 0. Припустимо, що у початковій період часу ми маємо капітал C_0 , а в період часу t — капітал C_t . Тоді подвоєння капіталу можна записати, переходячи від форми (4.10) до форми (4.9):

$$C_t = 2C_0 = C_0 e^{b_1 t}, \text{ звідки } \Rightarrow t = \frac{\ln 2}{b_1} \approx \frac{0,6931}{b_1} \approx \frac{0,70}{b_1}$$

Отже, капітал подвоїться через $\frac{0,70}{b_1}$ років. Правило має назву 70, тому

що $\ln 2$ приблизно дорівнює 0,70.

Приклад 2. *Приведення витрат до поточного часу*

Порівнюючи різні інвестиційні проекти, ми повинні порівняти в першу чергу кошти, які на них витрачаються, та майбутні прибутки, які вони приносять. Найпростіше це зробити, оцінюючи майбутні витрати та прибутки у вартості поточного року. Для ілюстрації розглянемо приклад. Заощаджуючи в банку A гривень у поточному році при ставці процента r , через t років ми отримаємо:

$$B = Ae^{rt} \quad (4.13)$$

Відповідно B гривень, отриманих у t році при ставці відсотка r , у поточному році коштуватимуть:

$$A = Be^{-rt} \quad (4.14)$$

Величина називається приведеними витратами B гривень через t років до поточного часу (при нормі відсотка r).

Дискретною версією виразу (4.13) є відома формула

$$B = A \left(1 + r \right)^t$$

яку ми розглядали у першому прикладі, а зворотною до неї відповідно:

$$A = B / \left(1 + r \right)^t = B \left(1 + r \right)^{-t}$$

З дискретною версією легше працювати, коли t – є цілим числом. Якщо t набуває неперервних значень, перевагу слід віддати виразу (4.13) або (4.14).

Приведені витрати можуть також визначатись для потоку платежів. При нормі відсотка r приведені витрати платежу B_1 який ми мали б через t_1 років, платежу B_2 через t_2 років, ..., платежу B_n відповідно через t_n років мають вигляд:

$$A = B_1 e^{-rt_1} + B_2 e^{-rt_2} + \dots + B_n e^{-rt_n} \quad (4.15)$$

Приклад 3. Довічна рента

Довічна рента – це послідовність однакових платежів у рівновіддалені інтервали для певного проміжку часу. Приведені витрати ренти A , яка виплачується наприкінці кожного з N років при постійній нормі відсотка r , що нараховується неперервно, дорівнює:

$$PV = Ae^{-r1} + Ae^{-r2} + \dots + Ae^{-rN} = A \left(e^{-r1} + e^{-r2} + \dots + e^{-rN} \right) \quad (4.16)$$

де PV — приведені витрати.

Враховуючи, що вираз $(a+\dots+a^n)(1-a)=a-a^{n+1}$, його легко переписати у вигляді:

$$a + \dots + a^n = \frac{a(1-a^{n+1})}{1-a} \quad (4.17)$$

Якщо у виразі (4.16) припустити, що $a=e^{-r}$, а $n=N$, тоді приведені витрати довічної ренти можна записати у вигляді:

$$PV = A \frac{e^{-r}(1-e^{-rN})}{1-e^{-r}} = \frac{A(1-e^{-rN})}{e^r-1} \quad (4.18)$$

За припущення, що ми будемо отримувати ренту довічно, тобто нескінченну кількість років $N \rightarrow \infty$, вираз (4.18) набуде вигляду:

$$PV = \frac{A}{e^r-1}$$

за припущення, що $e^{-rN} \rightarrow 0, N \rightarrow \infty$.

Якщо ж відсоток нараховується щорічно, приведену довічну ренту до поточного часу краще подавати у відомому вигляді:

$$PV = \frac{A}{1+r} + \dots + \frac{A}{1+r^N}$$

Якщо ми скористаємось формулою (4.17) та покладемо $a = \frac{1}{1+r}, n=N$, то

отримаємо:

$$PV = A \frac{1}{\frac{r}{1+r}} \left(1 - \left(\frac{1}{1+r} \right)^N \right) = \frac{A}{r} \left(1 - \left(\frac{1}{1+r} \right)^N \right) \quad (4.19)$$

Якщо $N \rightarrow \infty$ то вираз (4.19) остаточно набере вигляду: $PV = \frac{A}{r}$.

Взагалі експоненційні криві використовуються для опису швидко зростаючих або спадаючих економічних процесів. При цьому, якщо $(\beta > 0, b_1 > 0)$ – функція зростає до нескінченності, якщо $(\beta < 0, b_1 < 0)$ – функція спадає до 0.

Параметр b можна інтерпретувати як коефіцієнт зростання у часі.

МЕТОДИКА ЗНАХОДЖЕННЯ ОЦІНОК НЕЛІНІЙНИХ ОДНОФАКТОРНИХ МОДЕЛЕЙ

Задача. На основі даних про прибуток підприємства та сумою заробітної плати, виплаченої на підприємстві

X-сума зарплати, тис. грн	3,2	2,5	2,1	1,62	1,35	1,33	1,17	1,25	1,1	0,9
y-прибуток підприємства тис. грн./на одного працюючого	105,4	114,9	120,3	150,1	152,4	163,3	171,4	185,1	195,1	

Підібрати криву, яка найкраще описує тенденцію з допомогою Excel (лінія тренду). Для кращого підбору лінії тренду побудувати декілька варіантів кривих: лінійна, квадратична, поліноміальна (ступінь 3), логарифмічна, степенева, експотенційна.

Розв'язання.

Будуємо графік даних. Для цього можна використати алгоритм додатку (вибираємо тип діаграми **Точечная**, вид точкової діаграми 1), забираємо рамку і фон. Отримуємо діаграму розсіювання:

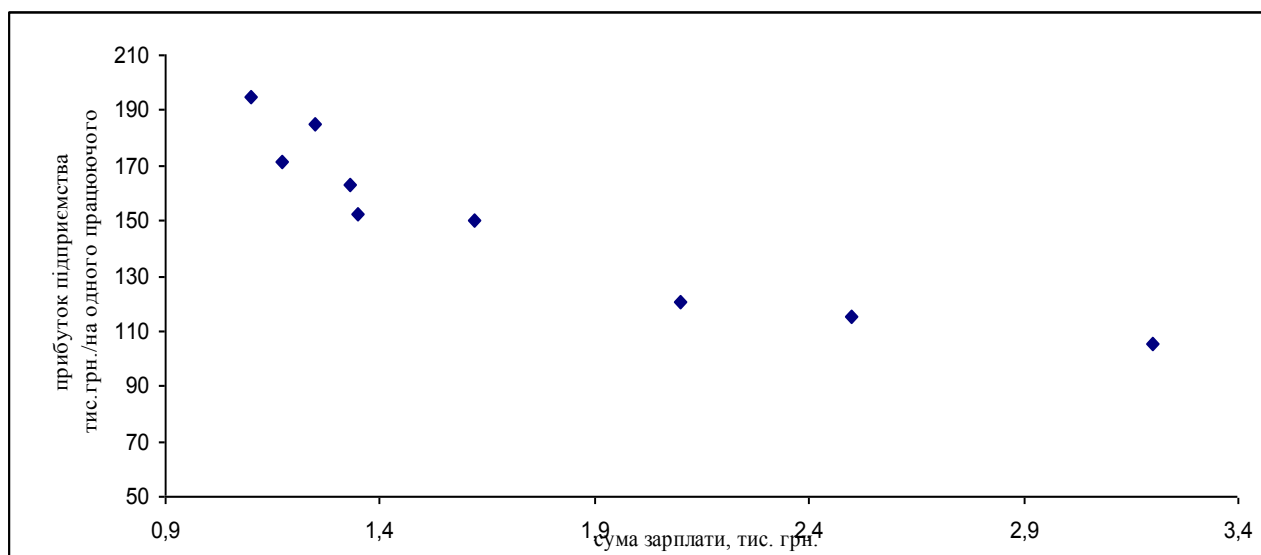


Рис.4.1 Діаграма розсіювання

Правою кнопкою миші натискаємо по одній із точок діаграми і вибираємо її списку, що появився **Добавить** **линию** **тренда**, ЛК.

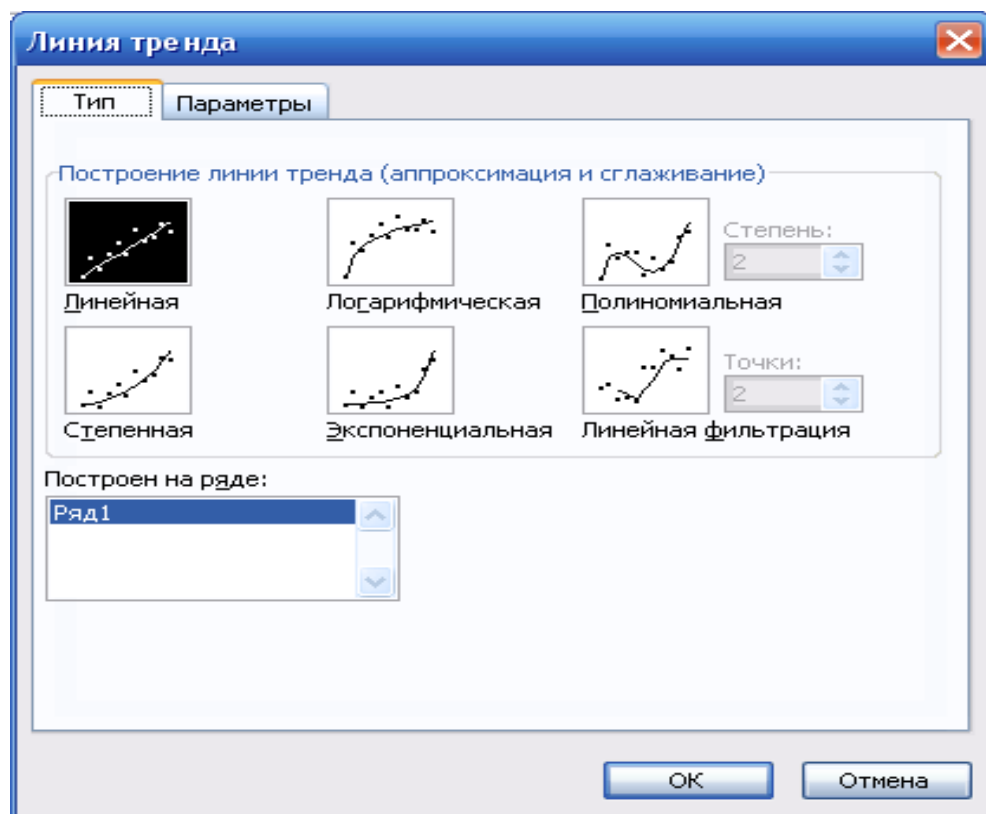


Рис. 4.2 Діалогове вікно

У вікні, що появилось вибираємо тип лінії тренду, а в **Параметрах** відмічаємо **Показывать уравнение на диаграмме** і **Поместить на диаграмму величину R²**. ОК.

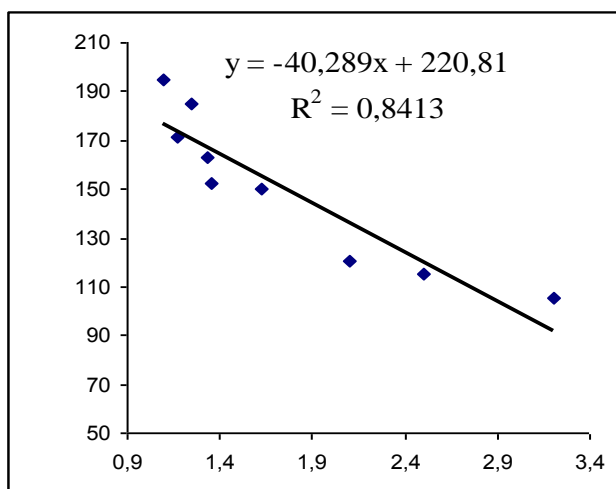


Рис. 4.3 Лінійна залежність

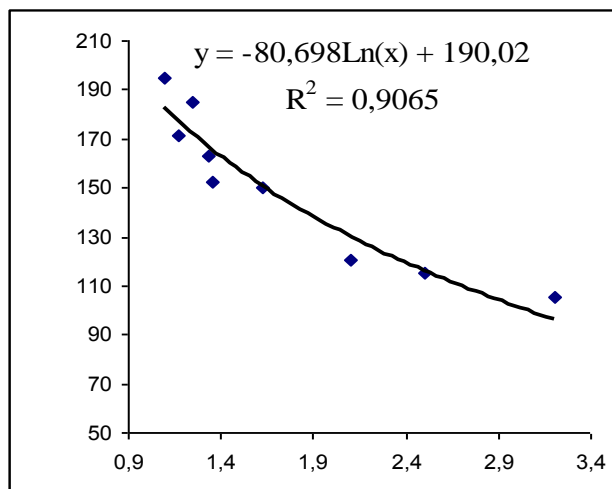


Рис. 4.4 Логарифмічна залежність

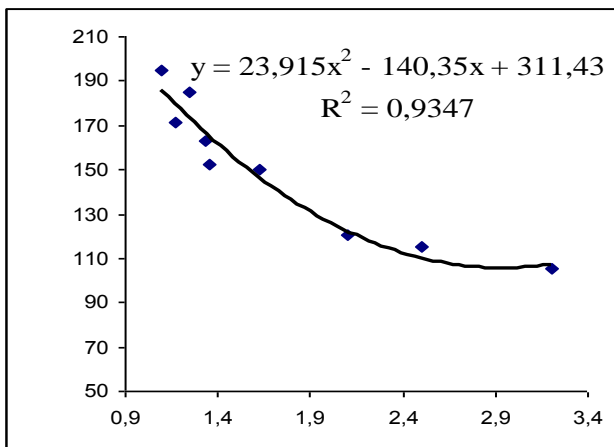


Рис. 4.5 Параболічна залежність

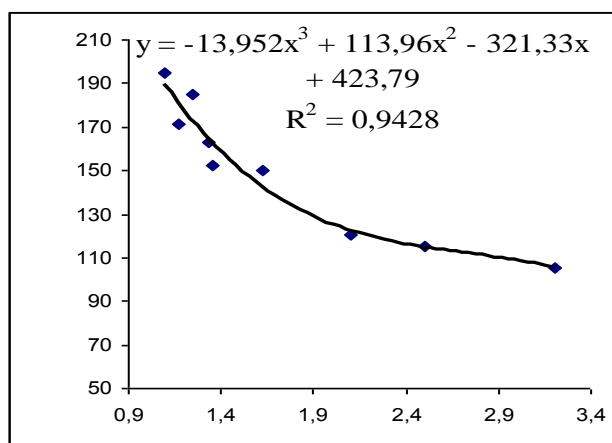


Рис. 4.6 Поліномінальна залежність (ступінь 3)

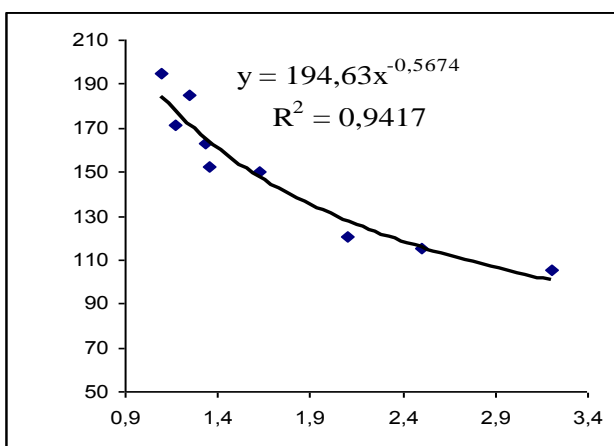


Рис. 4.7 Степенева залежність

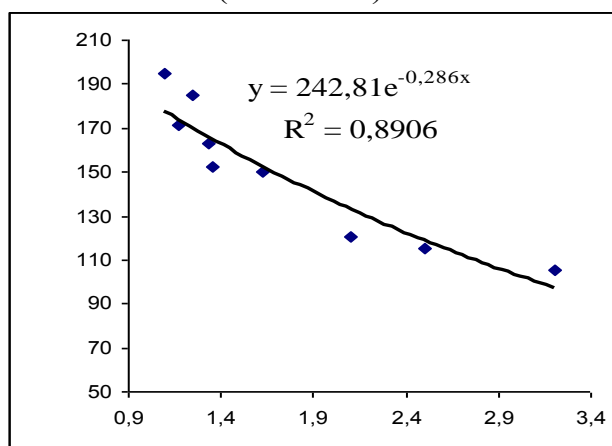


Рис. 4.8 Експотенційна залежність

Для вибору моделі дивимось на значення коефіцієнтів. Найкраще апроксимує статистичні дані поліном третього степеня $y = -13,952x^3 + 113,96x^2 - 321,33x + 423,79$ з коефіцієнтом детермінації $R^2 = 0,9428$, оскільки із всіх нелінійних моделей він має найбільше значення.

КЛАСИЧНА ЛІНІЙНА БАГАТОФАКТОРНА МОДЕЛЬ

5.1 ЛІНІЙНА БАГАТОФАКТОРНА ЕКОНОМЕТРИЧНА МОДЕЛЬ.

На практиці економічний процес змінюється під впливом багатьох різноманітних факторів, які треба вміти виявити та оцінити. Наприклад: обсяг продажу на фірмі залежить не тільки від витрат на рекламу, а й на обсяг продажу впливає частина ринку, яку отримує фірма, якість продукції, імідж марки продукту, середня заробітна плата населення в регіоні продажу, кількість населення, кількість безробітних та інші фактори.

Якщо змінна y залежить від незалежних змінних x_1, x_2, \dots, x_m і зв'язок між ними лінійний, то економетрична модель матиме вигляд:

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

де y – залежна змінна; x_1, x_2, \dots, x_m – незалежні змінні; a_0, a_1, \dots, a_m – невідомі детерміновані параметри; u – випадкова складова збурення, $u = \{u_1, u_2, \dots, u_m\}$.

Оціночне рівняння для даної моделі таке:

$$\hat{y} = \hat{a}_0 + \hat{a}_1x_1 + \hat{a}_2x_2 + \dots + \hat{a}_mx_m \quad (5.2)$$

де $\hat{a}_0, \hat{a}_1, \hat{a}_2, \dots, \hat{a}_m$ – оцінки невідомих параметрів a_0, a_1, \dots, a_m .

При використанні МНК для знаходження оцінок параметрів лінійної багатофакторної економетричної моделі як і для моделі з двома змінними використовуються деякі припущення. Насамперед вони стосуються випадкової змінної. Оскільки u є неспостережуваною величиною то ми можемо зробити лише деякі припущення щодо її поведінки та закону розподілу. Вони зберігаються такі ж самі як і для моделі з двома змінними, але дещо модифікуються.

Припущення 1.

Математичне сподівання випадкової величини u для всіх спостережень дорівнює нулю:

$$M(u_i) = 0, \quad i = \overline{1, n}.$$

Припущення 2. Дисперсія випадкової величини є однакова для всіх спостережень (властивість гомоскедастичності):

$$M \left(\epsilon_i^2 \right) = \text{var} \left(\epsilon_i \right) = \sigma^2.$$

Припущення 3. Випадкові величини незалежні між собою (значення випадкового члена u в i -му спостереженні не залежить від того, які значення він приймав в j -му спостереженні):

$$M \left(\epsilon_i u_j \right) = 0, \quad i \neq j$$

або

$$\text{cov} \left(\epsilon_i, u_j \right) = 0, \quad i \neq j$$

Припущення 4. Випадкова змінна розподілена незалежно від змінної x :

$$M \left(\epsilon_i u_i \right) = 0 \text{ або } \text{cov} \left(\epsilon_i, u_i \right) = 0.$$

Припущення 5. Випадкова змінна u розподілена по нормальному закону з нульовим математичним сподіванням і постійною дисперсією.

Припущення 6. Між факторами x_i, x_j – відсутня мультиколінеарність, тобто фактори незалежні між собою.

5.2 МНК ДЛЯ БАГАТОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ.

Нехай маємо ряд спостережень за залежною змінною $y = \{y_1, y_2, \dots, y_m\}$ та за незалежними змінними $x_1 = \{x_{11}, x_{12}, \dots, x_{1m}\}$, $x_2 = \{x_{21}, x_{22}, \dots, x_{2m}\}$, $x_m = \{x_{m1}, x_{m2}, \dots, x_{mm}\}$. Як у випадку лінійної моделі з двома змінними, знайдемо оцінки невідомих параметрів таким чином, щоб сума квадратів відхилень фактичних даних від розрахункових була мінімальною:

$$\begin{aligned} \sum_{i=1}^n u_i^2 &= \sum_{i=1}^n \left(\epsilon_i - \hat{y}_i \right)^2 = \sum_{i=1}^n \left(y_i - \hat{a}_0 - \hat{a}_1 x_{1i} - \dots - \hat{a}_m x_{mi} \right)^2 \\ &= f \left(\hat{a}_0, \hat{a}_1, \dots, \hat{a}_m \right) \rightarrow \min \end{aligned} \quad (5.3)$$

Щоб дослідити функцію $f \left(\hat{a}_0, \hat{a}_1, \dots, \hat{a}_m \right)$ на \min , знаходимо її часткові похідні і прирівнюємо їх до нуля. Одержуємо систему нормальних рівнянь:

$$\left\{ \begin{array}{l} \frac{d\left(\sum_{i=1}^n u_i^2\right)}{d \hat{a}_0} = -2 \sum_{i=1}^n \left(y_i - \hat{a}_0 - \hat{a}_1 x_{1i} - \dots - \hat{a}_m x_{mi} \right) = 0 \\ \frac{d\left(\sum_{i=1}^n u_i^2\right)}{d \hat{a}_1} = -2 \sum_{i=1}^n x_{1i} \left(y_i - \hat{a}_0 - \hat{a}_1 x_{1i} - \dots - \hat{a}_m x_{mi} \right) = 0 \\ \dots \\ \frac{d\left(\sum_{i=1}^n u_i^2\right)}{d \hat{a}_m} = -2 \sum_{i=1}^n x_{mi} \left(y_i - \hat{a}_0 - \hat{a}_1 x_{1i} - \dots - \hat{a}_m x_{mi} \right) = 0 \end{array} \right. \quad (5.4)$$

Або після перетворень

$$\left\{ \begin{array}{l} n \hat{a}_0 + \hat{a}_1 \sum_{i=1}^n x_{1i} + \hat{a}_2 \sum_{i=1}^n x_{2i} + \dots + \hat{a}_m \sum_{i=1}^n x_{mi} = \sum_{i=1}^n y_i \\ \hat{a}_0 \sum_{i=1}^n x_{1i} + \hat{a}_1 \sum_{i=1}^n x_{1i}^2 + \hat{a}_2 \sum_{i=1}^n x_{2i} x_{1i} + \dots + \hat{a}_m \sum_{i=1}^n x_{mi} x_{1i} = \sum_{i=1}^n y_i x_{1i} \\ \dots \\ \hat{a}_0 \sum_{i=1}^n x_{mi} + \hat{a}_1 \sum_{i=1}^n x_{1i} x_{mi} + \hat{a}_2 \sum_{i=1}^n x_{2i} x_{mi} + \dots + \hat{a}_m \sum_{i=1}^n x_{mi}^2 = \sum_{i=1}^n y_i x_{mi} \end{array} \right. \quad (5.5)$$

Система (5.5) є системою нормальних рівнянь МНК для знаходження оцінок $\hat{a}_0, \hat{a}_1, \hat{a}_2, \dots, \hat{a}_m$.

У випадку, коли $m=2$, тобто у залежить від x_1 і x_2 , система (5.5) набуде вигляду:

$$\left\{ \begin{array}{l} n \hat{a}_0 + \hat{a}_1 \sum_{i=1}^n x_{1i} + \hat{a}_2 \sum_{i=1}^n x_{2i} = \sum_{i=1}^n y_i \\ \hat{a}_0 \sum_{i=1}^n x_{1i} + \hat{a}_1 \sum_{i=1}^n x_{1i}^2 + \hat{a}_2 \sum_{i=1}^n x_{2i} x_{1i} = \sum_{i=1}^n y_i x_{1i} \\ \hat{a}_0 \sum_{i=1}^n x_{2i} + \hat{a}_1 \sum_{i=1}^n x_{1i} x_{2i} + \hat{a}_2 \sum_{i=1}^n x_{2i}^2 = \sum_{i=1}^n y_i x_{2i} \end{array} \right. \quad (5.6)$$

Розв'язавши систему (5.6) отримаємо оцінки $\hat{a}_0, \hat{a}_1, \hat{a}_2$, а значить рівняння оціночної площини $\hat{y} = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2$.

Оціночна площина, знайдена з допомогою МНК, проходить через точку координати якої є середні значення (\bar{x}_1, \bar{y}_1) , (\bar{x}_2, \bar{y}_2) . У цьому можна переконатися розділивши перше рівняння системи (5.6) на n .

5.3 КОЕФІЦІЄНТ МНОЖИННОЇ КОРЕЛЯЦІЇ ТА ДЕТЕРМІНАЦІЇ. ОЦІНЕНИЙ КОЕФІЦІЄНТ ДЕТЕРМІНАЦІЇ.

В економетричній моделі з 2-ма змінними (однофакторній) ми дали поняття коефіцієнта кореляції r і коефіцієнта детермінації d . У випадку економетричної моделі з кількістю змінних більшою ніж 2, вводиться поняття коефіцієнта множинної кореляції R . Він визначається як коефіцієнт кореляції між y та \hat{y} по формулі:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (5.7)$$

Квадрат коефіцієнта множинної кореляції, як і у випадку моделі з двома змінними, називають коефіцієнтом детермінації R^2 . Його можна визначити також по формулі:

$$R^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2 + \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5.8)$$

або

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5.8')$$

Коефіцієнт детермінації R^2 має важливу властивість, а саме він – неспадна функція від кількості факторів, які входять до моделі. Тобто якщо в модель додати новий фактор, то це збільшить значення коефіцієнта R^2 . Це видно з

формули (5.8'). Знаменник не залежить від кількості факторів, а чисельник $\sum_{i=1}^n (y_i - \hat{y}_i)^2$ – спадає чи хоча б не зростає. Якщо порівнюються дві моделі з однаковою залежною змінною, але різною кількістю факторів x , то, звичайно, вибирається та, яка має більше значення R^2 . Тому при порівнянні коефіцієнтів детермінації краще обрахувати оцінений коефіцієнт детермінації, який враховує кількість факторів x і має вигляд:

$$\bar{R}^2 = 1 - \frac{\sum_{i=1}^n u_i^2}{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}} \quad (5.9)$$

Якщо вже відомий коефіцієнт детермінації R^2 , то оцінений можна розрахувати за формулою:

$$\bar{R}^2 = 1 - R^2 \frac{n-1}{n-m-1} \quad (5.10)$$

З (5.10) випливає, що при зростанні кількості факторів оцінений коефіцієнт детермінації зростає повільніше, ніж звичайний коефіцієнт детермінації R^2 . Тому при виборі кращої моделі потрібно вибирати ту в якій більший оцінений коефіцієнт детермінації. Зауважимо, що R^2 приймає значення в інтервалі $[0;1]$, а оцінений коефіцієнт детермінації може бути і негативним (тоді R^2 прямує до нуля) і якщо $R^2 = 1$, то \bar{R}^2 теж дорівнює одиниці.

МЕТОДИКА ПОБУДОВИ БАГАТОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ

Задача. Маючи статистичні дані про результати господарської діяльності підприємств побудувати багатофакторну лінійну модель залежності показника y (чистий прибуток підприємств, млн. грн.) від двох факторів: x_1 (основні фонди, млн. грн.), x_2 (оборотні фонди, млн. грн.).

y	13	8	15	7	18	21	25	26	28	30	33	38
x ₁	21	19	18	15	24	23	21	24	25	26	25	29
x ₂	16	7	11	9	15	17	19	21	20	22	21	20

Визначити оцінки \hat{a}_0 , \hat{a}_1 , \hat{a}_2 з допомогою МНК, вважаючи, що економетрична модель лінійна. Знайти ці ж оцінки, застосовуючи формули відхилення від середніх.

Розв'язання. Оцінки знаходимо розв'язавши систему нормальних рівнянь, виведених із допомогою МНК

$$\begin{cases} n\hat{a}_0 + \hat{a}_1 \sum_{i=1}^n x_{1i} + \hat{a}_2 \sum_{i=1}^n x_{2i} = \sum_{i=1}^n y_i \\ \hat{a}_0 \sum_{i=1}^n x_{1i} + \hat{a}_1 \sum_{i=1}^n x_{1i}^2 + \hat{a}_2 \sum_{i=1}^n x_{2i}x_{1i} = \sum_{i=1}^n y_i x_{1i} \\ \hat{a}_0 \sum_{i=1}^n x_{2i} + \hat{a}_1 \sum_{i=1}^n x_{1i}x_{2i} + \hat{a}_2 \sum_{i=1}^n x_{2i}^2 = \sum_{i=1}^n y_i x_{2i} \end{cases}$$

Як і в попередніх роботах розрахунки сум, які входять у систему нормальних рівнянь, зводимо в таблицю.

№	y	x ₁	x ₂	x ₁ ²	x ₂ ²	yx ₁	yx ₂	x ₁ x ₂
1	13	21	16	441	256	273	208	336
2	8	19	7	361	49	152	56	133
3	15	18	11	324	121	270	165	198
4	7	15	9	225	81	105	63	135
5	18	24	15	576	225	432	270	360
6	21	23	17	529	289	483	357	391
7	25	21	19	441	361	525	475	399
8	26	24	21	576	441	624	546	504
9	28	25	20	625	400	700	560	500
10	30	26	22	676	484	780	660	572
11	33	25	21	625	441	825	693	525
12	38	29	20	841	400	1102	760	580
сума	262	270	198	6240	3548	6271	4813	4633

Випишемо систему нормальних рівнянь, підставляючи обраховані суми

$$\begin{cases} 12\hat{a}_0 + \hat{a}_1 270 + \hat{a}_2 198 = 262 \\ \hat{a}_0 270 + \hat{a}_1 6240 + \hat{a}_2 4633 = 6271 \\ \hat{a}_0 198 + \hat{a}_1 4633 + \hat{a}_2 3548 = 4813 \end{cases}$$

Розв'язуємо систему рівнянь з допомогою матричного методу.

Матриця коефіцієнтів системи нормальних рівнянь

$$\begin{pmatrix} 12 & 270 & 198 \\ 270 & 6240 & 4633 \\ 198 & 4633 & 3548 \end{pmatrix}$$

Обернена до неї матриця

$$\begin{pmatrix} 3,830524 & -0,2306 & 0,087358 \\ -0,2306 & 0,01914 & -0,01212 \\ 0,087358 & -0,01212 & 0,011239 \end{pmatrix}$$

Розв'язки системи нормальних рівнянь, отримані множенням оберненої матриці на вектор вільних членів

$$\begin{pmatrix} 3,830524 & -0,2306 & 0,087358 \\ -0,2306 & 0,01914 & -0,01212 \\ 0,087358 & -0,01212 & 0,011239 \end{pmatrix} \cdot \begin{pmatrix} 262 \\ 6271 \\ 4813 \end{pmatrix} = \begin{pmatrix} -22,06851 \\ 1,255773 \\ 0,948301 \end{pmatrix}$$

Отже, розв'язок системи нормальних рівнянь буде таким:

$$\hat{a}_0 = -22,06851, \quad \hat{a}_1 = 1,255773, \quad \hat{a}_2 = 0,948301$$

Багатофакторна лінійна модель матиме вигляд:

$$\hat{y} = -22,06851 + 1,255773x_1 + 0,948301x_2.$$

Розрахуємо \hat{a}_0 , \hat{a}_1 , \hat{a}_2 оцінки за формулами відхилення від середніх.

Формули відхилення від середніх отримуємо виконанням нескладних алгебраїчних перетворень системи нормальних рівнянь.

$$\hat{a}_1 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_{1i} - \bar{x}_1) \sum_{i=1}^n (x_{2i} - \bar{x}_2) - \sum_{i=1}^n (y_i - \bar{y})(x_{2i} - \bar{x}_2) \sum_{i=1}^n (x_{1i} - \bar{x}_1)}{\sum_{i=1}^n (x_{1i} - \bar{x}_1) \sum_{i=1}^n (x_{2i} - \bar{x}_2) - \left(\sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) \right)^2}$$

$$\hat{a}_2 = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_{2i} - \bar{x}_2) \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 - \sum_{i=1}^n (y_i - \bar{y})(x_{1i} - \bar{x}_1) \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2 - \left(\sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2) \right)^2}$$

$$\hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}_1 - \hat{a}_2 \bar{x}_2$$

Розрахунки зводимо в таблицю.

Отримані необхідні результати підставляємо у формули і обраховуємо оцінки.

$$\hat{a}_1 = \frac{376 \cdot 281 - 490 \cdot 178}{165 \cdot 281 - 178^2} = 1,255773$$

$$\hat{a}_2 = \frac{490 \cdot 165 - 376 \cdot 178}{165 \cdot 281 - 178^2} = 0,948301$$

$$\hat{a}_0 = 21,8 - 1,255773 \cdot 22,5 - 0,948301 \cdot 16,5 = -22,06851$$

Результати двох методів співпали. Отже оцінки знайдено вірно.

N°	y_i	x_{1i}	x_{2i}	\bar{y}	\bar{x}_{1i}	\bar{x}_{2i}	$y_i - \bar{y}$	$x_{1i} - \bar{x}_{1i}$	$x_{2i} - \bar{x}_{2i}$	$(x_{1i} - \bar{x}_{1i})^2$	$(x_{2i} - \bar{x}_{2i})^2$	$(x_{1i} - \bar{x}_{1i})(x_{2i} - \bar{x}_{2i})$	$(y_i - \bar{y})(x_{1i} - \bar{x}_{1i})$	$(y_i - \bar{y})(x_{2i} - \bar{x}_{2i})$
1	13	21	16				-8,83	-1,5	-0,5	2,25	0,25	0,75	13,25	4,42
2	8	19	7				-13,8	-3,5	-9,5	12,25	90,25	33,25	48,41	131,42
3	15	18	11				-6,83	-4,5	-5,5	20,25	30,25	24,75	30,75	37,58
4	7	15	9				-14,8	-7,5	-7,5	56,25	56,25	56,25	111,25	111,25
5	18	24	15	21,8	22,5	16,5	-3,8	1,5	-1,5	2,25	2,25	-2,25	-5,75	5,75
6	21	23	17				-0,8	0,5	0,5	0,25	0,25	0,25	-0,42	-0,42
7	25	21	19				3,167	-1,5	2,5	2,25	6,25	-3,75	-4,75	7,92
8	26	24	21				4,167	1,5	4,5	2,25	20,25	6,75	6,25	18,75
9	28	25	20				6,167	2,5	3,5	6,25	12,25	8,75	15,42	21,58
10	30	26	22				8,167	3,5	5,5	12,25	30,25	19,25	28,58	44,92
11	33	25	21				11,17	2,5	4,5	6,25	20,25	11,25	27,92	50,25
12	38	29	20				16,17	6,5	3,5	42,25	12,25	22,75	105,08	56,58
Σ	262	270	198							165	281	178	376	490

$$a = \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_m \end{bmatrix} - \text{вектор невідомих параметрів}; \quad U = \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix} - \text{вектор випадкової}$$

величини.

Виходячи з позначень, маємо:

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} & \dots & x_{m1} \\ 1 & x_{12} & x_{22} & \dots & x_{m2} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & \dots & x_{mn} \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_m \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix} \quad (6.4)$$

У матричному вигляді основні припущення лінійної багатofакторної економетричної моделі матимуть вигляд:

Припущення 1. Математичне сподівання i -го значення ($i=1,2,\dots,n$) випадкової величини u_i , дорівнює нулю:

$$M \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix} = \begin{bmatrix} M \langle u_1 \rangle \\ M \langle u_2 \rangle \\ \dots \\ M \langle u_n \rangle \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

Припущення 2. Випадкові величини незалежні між собою:

$$M \langle U' U' \rangle = M \begin{bmatrix} 1 & & & \\ & u_2 & & \\ & & \dots & \\ & & & u_n \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix}$$

де U' – транспонований вектор-стовпець випадкових величин (вектор-рядок).

Припущення 3. Вектор випадкових величин u_i , має нормальний закон розподілу з нульовим математичним сподіванням і постійною дисперсією.

Припущення 4. Матриця X утворюється з фіксованих елементів, тобто не випадкова.

Припущення 5. Між факторами x_i, x_j відсутня мультиколінеарність, тобто фактори незалежні між собою. Це означає, що стовпці матриці X лінійно незалежні, тобто не знайдеться таких чисел $\gamma_0, \gamma_1, \dots, \gamma_m$, серед яких не всі дорівнюють нулю, щоб виразити один стовпець як лінійну комбінацію інших:

$$\gamma_0 \cdot 1 + \gamma_1 x_{1i} + \dots + \gamma_m x_{mi} = 0 \text{ або } \gamma' \cdot X = 0.$$

Для того щоб знайти оцінки параметрів a запишемо економетричну модель у матричному вигляді, використавши попередні позначення і наступні:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{bmatrix} \text{ – вектор розрахункових значень; } \hat{a} = \begin{bmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \dots \\ \hat{a}_m \end{bmatrix} \text{ – вектор оцінок}$$

параметрів; $U = \begin{bmatrix} u_1 \\ u_2 \\ \dots \\ u_n \end{bmatrix}$ – вектор відхилень фактичних даних від розрахункових.

Лінійна багатofакторна економетрична модель має вигляд:

$$\hat{Y} = X\hat{a} + U \tag{6.5}$$

Вектор оцінок параметрів знайдемо методом найменших квадратів, мінімізуючи суму квадратів залишків:

$$\sum_{i=1}^n u_i^2 = u'u = (Y - X\hat{a})'(Y - X\hat{a}) = Y'Y - 2\hat{a}'X'Y + \hat{a}'X'X\hat{a} \rightarrow \min \tag{6.6}$$

де символ штрих (') означає операцію транспонування.

Зауваження. При перетворенні (6.6) враховані властивості транспонованих матриць: $(X\hat{a})' = \hat{a}'X'$; $\hat{a}'X'Y = Y'X\hat{a}$.

Знайдемо частинну похідну виразу (6.6) за компонентами вектора \hat{a} і прирівняємо її до нуля:

$$\frac{du(\hat{a})}{d\hat{a}} = -2X'Y + 2X'X\hat{a} = 0$$

Звідси отримуємо систему рівнянь в матричній формі:

$$X'X\hat{a} = X'Y \quad (6.7)$$

(6.7) є системою нормальних рівнянь МНК для знаходження оцінок $\hat{a}_0, \hat{a}_1, \hat{a}_2, \dots, \hat{a}_m$ в матричній формі. Якщо визначник матриці $X'X$ не дорівнює нулю ($\det(X'X) \neq 0$), то існує обернена матриця $(X'X)^{-1}$ і розв'язок системи (6.7) буде вектор-стовпець

$$\hat{a} = (X'X)^{-1} X'Y \quad (6.8)$$

Рівняння (6.8) є фундаментальним результатом для визначення невідомих параметрів у матричному вигляді.

6.2 ДИСПЕРСІЙНО-КОВАРІАЦІЙНА МАТРИЦЯ $VAR(a)$ МАТРИЦЯ КОРЕЛЯЦІЇ.

В економетричній моделі $Y = Xa + U$ вектор U і залежний від нього вектор Y є випадкові змінні. Оскільки \hat{a} ми визначаємо з виразу $\hat{a} = (X'X)^{-1} X'Y$ в який входить Y , то \hat{a} також можна вважати випадковою функцією оцінювання параметрів моделі. Тому для характеристики \hat{a} необхідно знати не тільки математичне сподівання, а й дисперсію, коваріацію.

У матричній формі легко знайти дисперсії параметрів $\hat{a}_0, \hat{a}_1, \hat{a}_2, \dots, \hat{a}_m$ та встановити коваріації між двома попарними їхніми значеннями, тобто між \hat{a}_i , та \hat{a}_j при $i \neq j$. Ці значення утворюють дисперсійно-коваріаційну матрицю

$$\begin{aligned}
& M \begin{bmatrix} \hat{a}_0 - a_0 \\ \hat{a}_1 - a_1 \\ \vdots \\ \hat{a}_m - a_m \end{bmatrix} \\
&= \begin{bmatrix} M(\hat{a}_0 - a_0) & M(\hat{a}_0 - a_0)(\hat{a}_1 - a_1) & \dots & M(\hat{a}_0 - a_0)(\hat{a}_m - a_m) \\ M(\hat{a}_0 - a_0)(\hat{a}_1 - a_1) & M(\hat{a}_1 - a_1) & \dots & M(\hat{a}_1 - a_1)(\hat{a}_m - a_m) \\ \dots & \dots & \dots & \dots \\ M(\hat{a}_m - a_m)(\hat{a}_0 - a_0) & M(\hat{a}_m - a_m)(\hat{a}_1 - a_1) & \dots & M(\hat{a}_m - a_m) \end{bmatrix} \\
&= \begin{bmatrix} \sigma_{a_0}^2 & \sigma_{\hat{a}_0 \hat{a}_1} & \dots & \sigma_{\hat{a}_0 \hat{a}_m} \\ \sigma_{\hat{a}_1 \hat{a}_0} & \sigma_{a_1}^2 & \dots & \sigma_{\hat{a}_1 \hat{a}_m} \\ \dots & \dots & \dots & \dots \\ \sigma_{\hat{a}_m \hat{a}_0} & \sigma_{\hat{a}_m \hat{a}_1} & \dots & \sigma_{a_m}^2 \end{bmatrix}
\end{aligned}$$

На головній діагоналі цієї матриці знаходяться дисперсії оцінок, а поза діагоналю – коваріації оцінок. Позначається дисперсійно-коваріаційна матриця $\text{var}(\hat{a})$ і визначається за формулою:

$$\text{var}(\hat{a}) = \hat{\sigma}_u^2 (X'X)^{-1} \quad (6.9)$$

де $\hat{\sigma}_u^2$ – оцінена дисперсія випадкової величини, яка визначається за формулою:

$$\hat{\sigma}_u^2 = \frac{\sum_{i=1}^n u_i^2}{n - m - 1} \quad (6.10)$$

або

$$\hat{\sigma}_u^2 = \frac{Y'Y - aX'Y}{n - m - 1} = \frac{u'u}{n - m - 1} \quad (6.11')$$

Для вимірювання сили лінійних зв'язків різних пар змінних використовується коефіцієнт парної кореляції. Причому для будь-яких двох факторів x_i і x_j коефіцієнт кореляції лежить в межах:

$$-1 \leq r_{x_i x_j} \leq 1$$

Якщо $r_{x_i x_j} > 0$, то випадкові величини x_i і x_j зв'язані додатною кореляцією (при зростанні однієї друга теж зростає); якщо $r_{x_i x_j} < 0$, то випадкові величини x_i і x_j зв'язані від'ємною кореляцією (при зростанні однієї друга спадає).

Коефіцієнт парної кореляції між змінними x_k і x_j можна визначити за формулою

$$r_{x_k x_j} = \frac{\sum_{i=1}^n (x_{ki} - \bar{x}_k) (x_{ji} - \bar{x}_j)}{\sqrt{\sum_{i=1}^n (x_{ki} - \bar{x}_k)^2 \cdot \sum_{i=1}^n (x_{ji} - \bar{x}_j)^2}} \quad (6.12)$$

Матриця, складена з парних коефіцієнтів кореляції, називається кореляційною матрицею

$$\mathbf{R} = \begin{bmatrix} r_{yy} & r_{yx_1} & \dots & r_{yx_m} \\ r_{x_1 y} & r_{x_1 x_1} & \dots & r_{x_1 x_m} \\ \dots & \dots & \dots & \dots \\ r_{x_m y} & r_{x_m x_1} & \dots & r_{x_m x_m} \end{bmatrix}$$

Слід відмітити що кореляційна матриця є симетричною відносно головної Діагоналі ($r_{x_i x_k} = r_{x_k x_i}$) і по головній діагоналі стоять одиниці ($r_{x_k x_k} = 1$).

Якщо обмежитись розглядом економетричної моделі $y = a_0 + a_1 x_1 + a_2 x_2 + u$ що будується по вибірці n значень змінних y , x_1 і x_2 , то розглядаються три коефіцієнти парної кореляції

- r_{yx_1} – це коефіцієнт парної кореляції між y і x_1 , коефіцієнт кореляції

моделі $y = a_0 + a_1 x_1 + u$;

- r_{yx_2} – це коефіцієнт парної кореляції між y і x_2 , коефіцієнт кореляції

моделі $y = a_0' + a_1' x_2 + u$;

- $r_{x_1 x_2}$ - це коефіцієнт парної кореляції між x_1 , і x_2 , коефіцієнт кореляції

моделі $x_1 = a_0'' + a_1'' x_2 + u$.

Ці коефіцієнти розраховуються за такими формулами:

$$r_{yx_1} = \frac{\sum_{i=1}^n (y_i - \bar{y}) (x_{1i} - \bar{x}_1)}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \cdot \sum_{i=1}^n (x_{1i} - \bar{x}_1)^2}} \quad r_{yx_2} = \frac{\sum_{i=1}^n (y_i - \bar{y}) (x_{2i} - \bar{x}_2)}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \cdot \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2}}$$

$$r_{x_1 x_2} = \frac{\sum_{i=1}^n (x_{1i} - \bar{x}_1) (x_{2i} - \bar{x}_2)}{\sqrt{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2 \cdot \sum_{i=1}^n (x_{2i} - \bar{x}_2)^2}} \quad (6.13)$$

Матриця кореляції наступна

$$\mathbf{R} = \begin{bmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{x_1y} & 1 & r_{x_1x_2} \\ r_{x_2y} & r_{x_2x_1} & 1 \end{bmatrix}$$

Крім коефіцієнтів парної кореляції вводяться через коефіцієнти частинної кореляції. Частинною кореляцією між факторами x_i і x_j називається кореляційна залежність між цими факторами при фіксованих значеннях інших факторів.

Розглянемо знаходження коефіцієнтів частинної кореляції при числі незалежних змінних $m=2$. Коефіцієнт частинної кореляції $r_{yx_2x_1}$ показує, яку долю дисперсії змінної y , що не пояснила змінна x_1 , пояснить введення в модель нової змінної x_2 . $r_{yx_1x_2}$ показує, яку долю дисперсії y , що не пояснила змінна x_2 пояснить введення в модель нової змінної x_1 .

Коефіцієнти частинної кореляції визначається за формулами через коефіцієнти парної кореляції:

$$r_{yx_2x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{1 - r_{yx_1}^2} \cdot \sqrt{1 - r_{x_1x_2}^2}} \quad r_{yx_1x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{1 - r_{yx_2}^2} \cdot \sqrt{1 - r_{x_1x_2}^2}} \quad (6.14)$$

Для $m=3$ маємо наступні коефіцієнти частинної кореляції:

$$r_{yx_1x_2x_3} = \frac{r_{yx_1x_3} - r_{yx_2x_3} \cdot r_{x_1x_2x_3}}{\sqrt{1 - r_{yx_2x_3}^2} \cdot \sqrt{1 - r_{x_1x_2x_3}^2}} \quad r_{yx_2x_1x_3} = \frac{r_{yx_2x_3} - r_{yx_1x_3} \cdot r_{x_2x_1x_3}}{\sqrt{1 - r_{yx_1x_3}^2} \cdot \sqrt{1 - r_{x_2x_1x_3}^2}}$$

$$r_{yx_3x_1x_2} = \frac{r_{yx_3x_2} - r_{yx_1x_2} \cdot r_{x_3x_1x_2}}{\sqrt{1 - r_{yx_1x_2}^2} \cdot \sqrt{1 - r_{x_3x_1x_2}^2}}$$

Для будь-якого числа пояснюючих змінних маємо:

$$r_{yx_1x_2 \dots x_m} = \frac{r_{yx_1x_3 \dots x_m} - r_{yx_2x_3 \dots x_m} \cdot r_{x_1x_2x_3 \dots x_m}}{\sqrt{1 - r_{yx_2x_3 \dots x_m}^2} \cdot \sqrt{1 - r_{x_1x_2x_3 \dots x_m}^2}}$$

Отже, щоб обчислити коефіцієнт частинної кореляції m -го порядку потрібно обчислити спочатку коефіцієнти $(m-1)$ -го порядку.

Маючи коефіцієнт парної кореляції можна обчислити коефіцієнт детермінації (а значить і коефіцієнт множинної кореляції) по формулі:

$$R^2 = \frac{r_{yx}^2 + r_{yx_2}^2 - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} \quad (6.15)$$

6.3 ПЕРЕВІРКА МОДЕЛІ НА АДЕКВАТНІСТЬ. ПЕРЕВІРКА НУЛЬОВИХ ГІПОТЕЗ І ДОВІРЧІ ІНТЕРВАЛИ ПАРАМЕТРІВ.

Перевірка адекватності прийнятої економетричної моделі реальній дійсності проводиться аналогічно лінійній моделі з однією пояснюючою змінною. Для цього користуються найчастіше критерієм Фішера (F-критерієм).

Нульова гіпотеза має вигляд:

$$H_0: a_0 = a_1 = \dots = a_m = 0.$$

Альтернативна гіпотеза наступна:

H_1 : не всі a_j ($j=0, 1 \dots m$) дорівнюють нулю.

Якщо нульова гіпотеза відхиляється і приймається альтернативна гіпотеза, то це означає, що включені до моделі фактори пояснюють змінну y .

Величина F - відношення для багатфакторної моделі з m та $n-m-1$ ступенями вільності наступна:

$$F = \frac{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{m}}{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-m-1}} = \frac{(n-m-1) \sum_{i=1}^n (y_i - \bar{y})^2}{m \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6.16)$$

де m – кількість факторів, що ввійшли в модель, n – кількість спостережень у вибірці.

Або якщо відоме значення коефіцієнта детермінації то F-відношення таке:

$$F = \frac{(n-m-1) R^2}{m(-R^2)} \quad (6.16')$$

Після обчислення F-відношення Фішера знаходимо $F_{кр}(m; n-m-1; \gamma)$, яке є критичним значенням F при заданому рівні значущості γ (або $\gamma \cdot 100\%$) та відповідно m і $n-m-1$ ступенях вільності.

Якщо $F > F_{кр}$, то ми відкидаємо H_0 з ризиком помилитися не більше ніж в $\gamma\%$ випадків, і приймаємо, що побудоване рівняння економетричної моделі адекватне реальній дійсності. В протилежному випадку $F \leq F_{кр}$ – H_0 приймаємо і вважаємо, що побудована модель неадекватна. Тоді необхідно, можливо, будувати нелінійну модель або ввести додаткові фактори.

Для розгляду значущості оцінок параметрів розглядаються нульові гіпотези:

$H_0: a_j = 0$, тобто пояснююча змінна x_j не впливає суттєво на y .

Альтернативна – $H_1: a_j \neq 0$, тобто пояснююча змінна x_j впливає суттєво на y .

Емпіричне значення відношення t_j для перевірки нульової гіпотези стосовно параметрів a_j знаходиться за формулою:

$$t_j = \frac{|\hat{a}_j|}{\sigma_{\hat{a}_j}} \quad (6.17)$$

де $\sigma_{\hat{a}_j}$ – стандартна похибка оцінки \hat{a}_j ; $\sigma_{\hat{a}_j} = \sqrt{\sigma_{\hat{a}_j}^2}$. ($\sigma_{\hat{a}_j}^2$ беремо з матриці $\text{var}(\hat{a})$, яку знаходимо по формулі (6.9).

Емпіричне значення оцінки t_j порівнюють з критичним, знайденим за таблицями Стюдента, для заданого рівня значущості γ і $k=n-m-1$ ступенів вільності. Якщо $|t_j| \leq t_{кр}$ то гіпотеза про рівність нулю параметра a_j – в генеральній сукупності не відхиляється, оцінка \hat{a}_j є статистично незначущою; $|t_{емп}| > t_{кр}$ то із заданим рівнем значущості γ гіпотезу H_0 слід відхилити і відповідну оцінку вважають статистично значущою.

Побудова довірчих інтервалів для параметрів a_j проходить аналогічно довірчим інтервалам параметрів рівняння економетричної моделі з двома змінними. Спочатку знаходимо граничні похибки оцінок за формулою

$$\Delta_{\hat{a}_j} = t_{k,\gamma} \cdot \sigma_{\hat{a}_j}$$

де $t_{k,\gamma}$ – імовірнісний коефіцієнт, який знаходиться за таблицями Стюдента при рівні значущості γ і ступенях вільності $k=n-m-1$.

Довірчий інтервал, в межах якого при заданому рівні значущості γ або надійності $1-\gamma$, міститься невідомий параметр a_j генеральної сукупності:

$$\hat{a}_j - \Delta_{\hat{a}_j} \leq a_j \leq \hat{a}_j + \Delta_{\hat{a}_j} \text{ або}$$

$$\left[\hat{a}_j - \Delta_{\hat{a}_j}; \hat{a}_j + \Delta_{\hat{a}_j} \right]$$

6.4 ПЕРЕВІРКА НУЛЬОВОЇ ГІПОТЕЗИ СТОСОВНО КОЕФІЦІЄНТА МНОЖИННОЇ КОРЕЛЯЦІЇ. ПРОГНОЗУВАННЯ ЗА ЕКОНОМЕТРИЧНОЮ МОДЕЛЛЮ.

Так як коефіцієнт множинної кореляції R є вибірковою характеристикою, то при перевірці якості побудованої моделі проводять оцінку його значущості. За результатами вибірки обчислюється статистика

$$t_{розр} = \frac{R\sqrt{N-M-1}}{\sqrt{1-R^2}}$$

яка має розподіл Стюдента. Для заданого рівня значущості γ і ступенях вільності $k=n-m-1$ знаходимо табличне значення $t_{кр}$ за таблицями розподілу Стюдента. Обчислене значення $t_{розр}$ порівнюють з $t_{кр}$. Прийняття чи відхилення гіпотези про значущість коефіцієнта множинної кореляції проводиться аналогічно як для економетричної моделі з двома невідомими. Якщо $|t_{розр}| > t_{кр}$ то із заданим рівнем значущості у гіпотезу H_0 слід відхилити, прийняти альтернативну гіпотезу H_1 про існування залежності між залежною і незалежними змінними.

Значущість коефіцієнта множинної кореляції можна оцінювати на основі проведення процедури перевірки значущості коефіцієнта детермінації, оскільки вони зв'язані між собою: $R = \sqrt{R^2}$.

Як було зазначено в попередньому розділі, якщо рівняння економетричної моделі адекватне реальній дійсності, то по ньому можна

здійснювати прогноз. Прогноз буває точковий та інтервальний. Точковий прогноз одержуємо безпосередньою підстановкою значень незалежних змінних в $(n+k)$ -період у рівняння економетричної моделі:

$$\hat{y}_{n+k} = \hat{a}_0 + \hat{a}_1 x_{1,n+k} + \dots + \hat{a}_m x_{m,n+k}$$

або в матричній формі:

$$\hat{Y}_{n+k} = X_{n+k} \hat{a}$$

де X_{n+k} – матриця очікуваних пояснювальних змінних.

Прогнозне значення \hat{y}_{n+k} є оцінкою дійсного значення змінної y_{n+k} .

Інтервал, у який з певною заданою імовірністю $p=1-\gamma$ потрапляє дійсне значення залежної змінної визначається за формулою:

$$\hat{Y}_{n+k} - t_\gamma \hat{\sigma}_u \sqrt{1 + X'_{n+k} (X'_{n+k} X_{n+k})^{-1} X_{n+k}} \leq Y_{n+k} \leq \hat{Y}_{n+k} + t_\gamma \hat{\sigma}_u \sqrt{1 + X'_{n+k} (X'_{n+k} X_{n+k})^{-1} X_{n+k}}$$

де X'_{n+k} вектор значень з m факторів у період $n+k$.

Прогнозний інтервал для математичного сподівання для $\gamma \cdot 100\%$ -го рівня значущості має вигляд:

$$\hat{Y}_{n+k} - t_\gamma \hat{\sigma}_u \sqrt{X'_{n+k} (X'_{n+k} X_{n+k})^{-1} X_{n+k}} \leq M_{n+k} \leq \hat{Y}_{n+k} + t_\gamma \hat{\sigma}_u \sqrt{X'_{n+k} (X'_{n+k} X_{n+k})^{-1} X_{n+k}}$$

МЕТОДИКА ЗАСТОСУВАННЯ МАТРИЧНОГО МЕТОДУ ДО БАГАТОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ

Задача. Маючи статистичні дані про результати господарської діяльності підприємств побудувати багатофакторну лінійну модель залежності показника y (чистий прибуток підприємств, млн. грн.) від двох факторів: x_1 (основні фонди, млн. грн.), x_2 (оборотні фонди, млн. грн.), за 12 місяців.

y	13	8	15	7	18	21	25	26	28	30	33	38
x_1	21	19	18	15	24	23	21	24	25	26	25	29
x_2	16	7	11	9	15	17	19	21	20	22	21	20

Визначити оцінки \hat{a}_0 , \hat{a}_1 , \hat{a}_2 з допомогою матричного методу, вважаючи, що економетрична модель лінійна. Знайти матрицю дисперсій

оцінок. Обчислити коефіцієнти парної, частинної та множинних кореляцій. Перевірити адекватність прийнятої економетричної моделі за критерієм Фішера. Перевірити значущість оцінок параметрів за допомогою t-критерію. Побудувати довірчі інтервали для параметрів \hat{a}_0 , \hat{a}_1 , \hat{a}_2 . Перевірити значущість коефіцієнта множинної кореляції. Здійснити прогноз за побудованою моделлю.

Розв'язання.

Знайдемо оцінки параметрів, використовуючи матричні операції за формулою:

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

Порядок знаходження оцінок параметрів економетричної моделі.

Матриця X знаходиться в блоці A2:C13. Знаходимо транспоновану матрицю X' в блоці A16:L18, використовуючи в категорії **Ссылки и массивы** вбудовану функцію **ТРАНСП**(A2:C13).

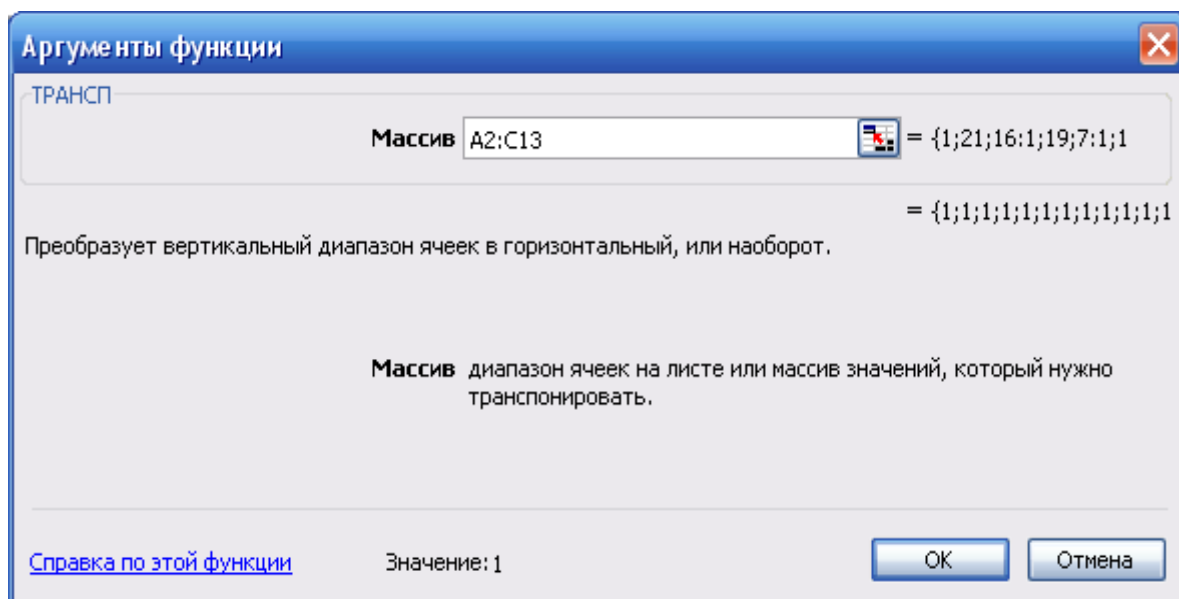


Рис. 6.1 Діалогове вікно **ТРАНСП**

Знаходимо добуток матриць $X'X$ в блоці B21:D23.

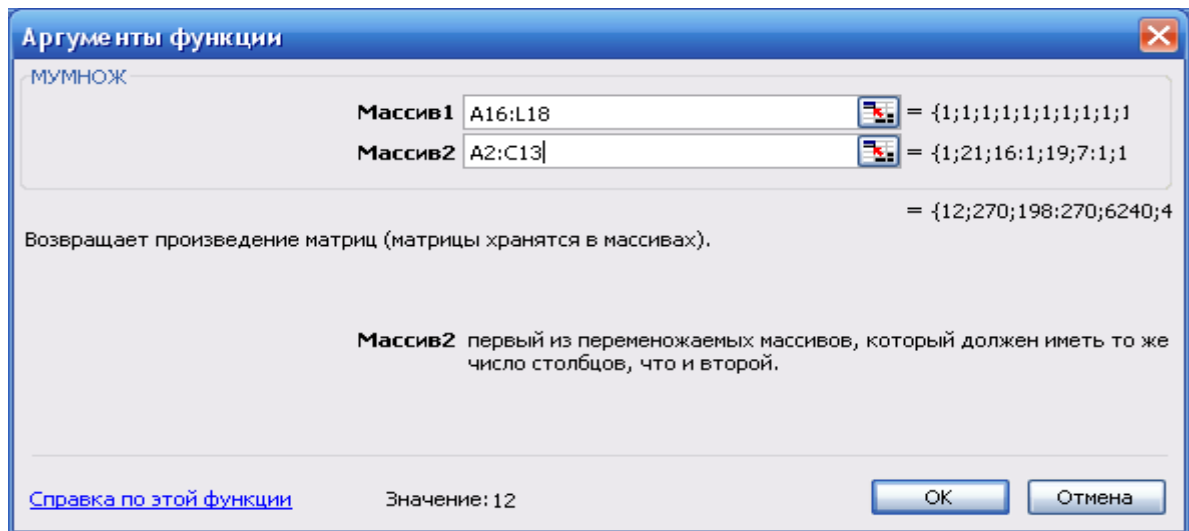


Рис. 6.2 Діалогове вікно **МУМНОЖ**

Обернену матрицю $(X'X)^{-1}$ знаходимо в блоці B26:D28, використовуючи вбудовану математичну функцію **МОБР**.

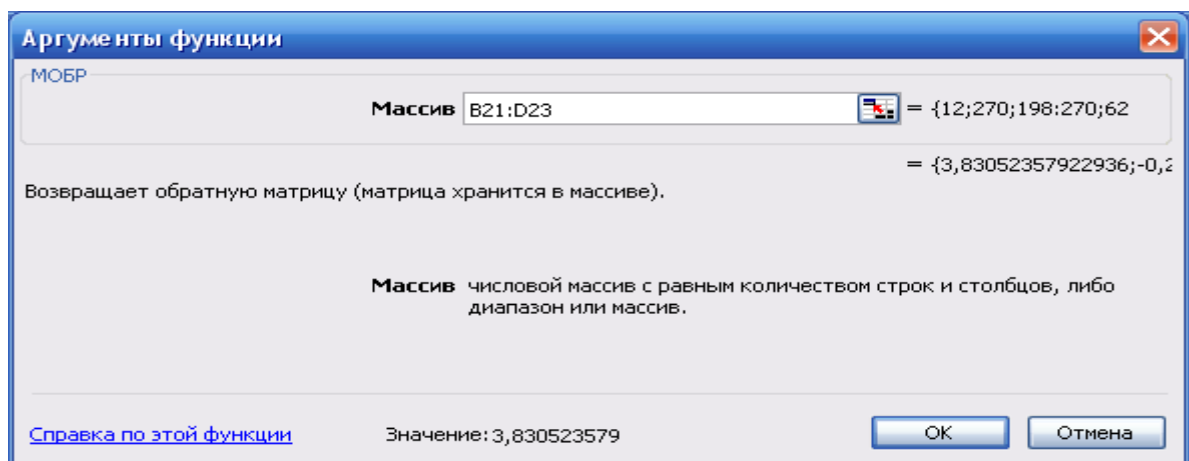


Рис. 6.3 Діалогове вікно **ТРАНСП**

Знаходимо добуток матриць $X'Y$ в блоці G26:G28, використовуючи вбудовану математичну функцію **МУМНОЖ**.

Оцінки вектора знаходимо в блоці C31:C33, використовуючи вбудовану математичну функцію **МУМНОЖ** (блок даних матриці $(X'X)^{-1}$; блок даних матриці $X'Y$).

Знаходимо матрицю дисперсій оцінок :

$$\text{var}(\hat{a}) = \hat{\sigma}_u^2 (X'X)^{-1} = \begin{bmatrix} \sigma_{a_0}^2 & \sigma_{\hat{a}_0\hat{a}_1} & \dots & \sigma_{\hat{a}_0\hat{a}_m} \\ \sigma_{\hat{a}_1\hat{a}_0} & \sigma_{a_1}^2 & \dots & \sigma_{\hat{a}_1\hat{a}_m} \\ \dots & \dots & \dots & \dots \\ \sigma_{\hat{a}_m\hat{a}_0} & \sigma_{\hat{a}_m\hat{a}_1} & \dots & \sigma_{a_m}^2 \end{bmatrix}, \hat{\sigma}_u^2 = \frac{u'u}{n-m-1}$$

Microsoft Excel - Книга1.xls

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка Введите вопрос

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		x1	x2	y	урозр	u							
2		1	21	16	13	19,47554							
3		1	19	7	8	8,429284							
4		1	18	11	15	10,96672							
5		1	15	9	7	5,302794							
6		1	24	15	18	22,29456							
7		1	23	17	21	22,93539							
8		1	21	19	25	22,32044							
9		1	24	21	26	27,98436							
10		1	25	20	28	28,29184							
11		1	26	22	30	31,44421							
12		1	25	21	33	29,24014							
13		1	29	20	38	33,31493							
14													
15				Транспонована матриця									
16		1	1	1	1	1	1	1	1	1	1	1	1
17		21	19	18	15	24	23	21	24	25	26	25	29
18		16	7	11	9	15	17	19	21	20	22	21	20
19													
20		Добуток матриць											
21			12	270	198								
22		(X'X) =	270	6240	4633								
23			198	4633	3548								
24													
25		Обернена матриця											
26			3,830524	-0,2306	0,087358			262					
27		(X'X) ⁻¹ =	-0,2306	0,01914	-0,01212		X'Y =	6271					
28			0,087358	-0,01212	0,011239			4813					
29													
30		Вектор оцінок											
31				-22,0685									
32		â =		1,255773									
33				0,948301									
34													
35		Оціночне рівняння											
36													
37	u =	-6,47554	-0,42928	4,033285	1,697206	-4,29456	-1,93539	2,679558	-1,98436	-0,29184	-1,44421	3,759864	4,685073
38													
39	u'u =	132,8288		σ _u ² =	14,75876								
40													
41			56,53378	-3,40343	1,289293								
42	var(â) =		-3,40343	0,282488	-0,17894								
43			1,289293	-0,17894	0,165674								
44													
45	r _{yx1} =	0,894997		r _{yx2x1} =	0,61313								
46	r _{yx2} =	0,893755		r _{yx1x2} =	0,618723		R ² =	0,875822					
47	r _{x1x2} =	0,826656											
48													

Готово NUM

Рис. 6.4 Оцінювання і аналіз економетричної моделі

Діагональні елементи матриці визначають дисперсії оцінок \hat{a}_0 , \hat{a}_1 , \hat{a}_2 . Інші елементи взаємну коваріації цих оцінок. З допомогою діагональних елементів матриці $\text{var}(\hat{a})$ знаходяться граничні похибки оцінок для заданого рівня імовірності.

Перш ніж досліджувати економетричну модель з трьома змінними $\hat{y} = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2$, необхідно розрахувати три парних моделі наступного виду:

$$y = a'_0 + a'_1 x_1$$

$$y = a''_0 + a''_1 x_2$$

$$x_1 = a'''_0 + a'''_1 x_2$$

Для кожної моделі розраховуємо коефіцієнти кореляції, які називаються коефіцієнтами кореляції нульового порядку чи коефіцієнтами парної кореляції.

Використаємо для обчислення коефіцієнтів парної кореляції вбудовану функцію **КОРРЕЛ**.

Порядок обчислення:

Майстер функцій f_x :

Категорія **Статистические**;

Функція **КОРРЕЛ**;

Далее;

В перший масив заносимо статистичні дані першої змінної, в другий – другої;

Готово.

Зауваження. В Excel є можливість отримати кореляційну матрицю з допомогою пакету **Анализ данных (Корреляция)**, вбудованого в Сервис. Значення коефіцієнтів парної кореляції знаходяться в комірках B45:B47.

Матриця має вигляд:

$$\underline{r} = \begin{pmatrix} 1 & 0,895 & 0,894 \\ 0,895 & 1 & 0,83 \\ 0,894 & 0,83 & 1 \end{pmatrix}$$

З допомогою даної матриці *проведемо аналіз дисперсій змінних моделі*. Значення коефіцієнта кореляції $r_{yx1} = 0,895$ показує, що 89,5% загальної

дисперсії змінної y пояснюється економетричною моделлю $y = a'_0 + a'_1 x_1$. Непояснена дисперсія, викликана випадковою складовою u складає 10,5%.

Коефіцієнт кореляції $r_{yx2} = 0,894$ показує, що 89,4% загальної дисперсії змінної y пояснюється економетричною моделлю $y = a''_0 + a''_1 x_2$. Непояснена дисперсія, викликана випадковою складовою u складає 10,6%.

Очевидно, якщо обмежитися розрахунками тільки перших моделей, то слід вибрати першу модель, яка пояснює більшу частку загальної дисперсії змінної y .

Значення $r_{x1x2} = 0,83$ показує, що 83% загальної дисперсії змінної x_1 пояснюється економетричною моделлю $x_1 = a'''_0 + a'''_1 x_2$. Іншими словами, пояснююча змінна x_1 корелює з пояснюючою змінною x_2 . Велике значення коефіцієнта кореляції між пояснюючими змінними економетричної моделі свідчить про явище мультиколінеарності, яке полягає в тому, що вплив пояснюючої змінної x_1 на результативну y можна вважати опосередкованим через зміну x_2 і навпаки.

Здійснивши аналіз дисперсій на основі коефіцієнтів нульового порядку, можна поставити питання в іншому аспекті: якщо $r_{yx1} = 0,895$, то непоясненою залишається 10,5% дисперсії. Якщо в модель ввести ще одну пояснюючу змінну x_2 , то необхідно дізнатися яку частку непоясненої змінної x_1 пояснить введення змінної x_2 . Визначити цю частку можна з допомогою коефіцієнта частинної кореляції чи коефіцієнта кореляції першого порядку.

Обчислимо коефіцієнти частинної кореляції по формулах (6.13):

Значення коефіцієнтів частинної кореляції знаходяться в комірках E45:E46. Отже, якщо $r_{yx2x1} = 0,61$, то з 10,5% дисперсії змінної y , непоясненої змінної x_1 пояснює введення ще однієї змінної x_2 . Із 10,6% дисперсії змінної y , непоясненої змінної x_2 пояснює введення ще однієї змінної x_1 .

Отже, аналіз дисперсій, який базується на матриці коефіцієнтів нульового порядку та коефіцієнтів частинної кореляції дає важливу інформацію при розрахунку економетричної моделі $\hat{y} = \hat{a}_0 + \hat{a}_1 x_1 + \hat{a}_2 x_2$.

Тепер визначаємо коефіцієнт *множинної детермінації*, який показує частку загальної дисперсії змінної y , що пояснює оціночна площа.

Коефіцієнт *множинної кореляції* обчислимо через коефіцієнти парної кореляції по формулі (6.15). Обчислене значення коефіцієнта *множинної кореляції* знаходяться в комірці Н46, і $R^2 = 0,88$.

Перевіримо адекватність прийнятої економетричної моделі за критерієм Фішера.

$$F = \frac{(k - m - 1) R^2}{m(1 - R^2)} = \frac{(2 - 2 - 1) \cdot 0,88}{2 \cdot (1 - 0,88)} = 33, F_{кр}(2;9;0,05) = 4,26.$$

Оскільки $F > F_{кр}$, то з надійністю 95% економетричну модель $\hat{y} = -22,06851 + 1,255773x_1 + 0,948301x_2$ можна вважати адекватною експериментальним даним і на підставі прийнятої моделі можна проводити економічний аналіз.

Перевіримо значущість оцінок параметрів за допомогою t -критерію.

$$t_j = \frac{|\hat{a}_j|}{\sigma_{\hat{a}_j}}, \text{ де } \sigma_{\hat{a}_j} - \text{стандартна похибка оцінки } \hat{a}_j; \sigma_{\hat{a}_j} = \sqrt{\sigma_{\hat{a}_j}^2}. (\sigma_{\hat{a}_j}^2 \text{ беремо}$$

з матриці $\text{var}(\hat{\mathbf{a}})$.

$$\sigma_{\hat{a}_0} = \sqrt{56,53} = 7,52, \quad \sigma_{\hat{a}_1} = \sqrt{0,282} = 0,53, \quad \sigma_{\hat{a}_2} = \sqrt{0,166} = 0,41.$$

$t_{\hat{a}_0} = 2,93$ $t_{\hat{a}_1} = 2,37$ $t_{\hat{a}_2} = 2,31$, $t_{кр} = 3,250$, для $\gamma = 0,99$ і $k = 9$ ступенів вільності. Оскільки всі $|t_j| \leq t_{кр}$ то гіпотеза про рівність нулю параметрів a_j – в генеральній сукупності не відхиляється, оцінки \hat{a}_0 , \hat{a}_1 , \hat{a}_2 є статистично незначущими.

Побудуємо довірчі інтервали для параметрів \hat{a}_0 , \hat{a}_1 , \hat{a}_2 .

Знаходимо граничні похибки оцінок за формулою: $\Delta_{\hat{a}_j} = t_{кр} \sigma_{\hat{a}_j}$, отримаємо:

$$\Delta_{\hat{a}_0} = 3,25 \cdot 7,52 = 24,44, \quad \Delta_{\hat{a}_1} = 3,25 \cdot 0,53 = 1,72, \quad \Delta_{\hat{a}_2} = 3,25 \cdot 0,41 = 1,33.$$

Довірчий інтервал в межах якого при заданому рівні значущості $\gamma = 0,99$ і $k = 9$ ступенів вільності містяться невідомі параметри a_0 , a_1 , a_2 генеральної сукупності:

$$\hat{a}_0 - \Delta_{\hat{a}_0} \leq a_0 \leq \hat{a}_0 + \Delta_{\hat{a}_0}$$

$$\hat{a}_1 - \Delta_{\hat{a}_1} \leq a_1 \leq \hat{a}_1 + \Delta_{\hat{a}_1}$$

$$\hat{a}_2 - \Delta_{\hat{a}_2} \leq a_2 \leq \hat{a}_2 + \Delta_{\hat{a}_2}$$

Підставляючи дані отримуємо:

$$-46,51 \leq a_0 \leq 2,37$$

$$-0,47 \leq a_1 \leq 2,98$$

$$0,38 \leq a_2 \leq 2,28$$

Перевіримо значущість коефіцієнта множинної кореляції.

За результатами вибірки обчислюємо t -статистику, яка має розподіл Стьюдента.

Для заданого рівня значущості $p=0,99$ і $k=9$ ступенів вільності $t_{кр}=3,250$.

$$t_{розр} = \frac{R\sqrt{N-M-1}}{\sqrt{1-R^2}} = \frac{0,93 \cdot 3}{\sqrt{1-0,88}} = 8,05. \text{ Оскільки } |t_{розр}| > t_{кр} \text{ то із заданим рівнем}$$

значущості у гіпотезу H_0 слід відхилити, прийняти альтернативну гіпотезу H_1 про існування залежності між залежною і незалежними змінними.

Прогнозування за економетричною моделлю.

Визначимо точкове прогнозоване значення показника, коли $x_1=30$, а $x_2=28$: $\hat{y}_0 = -22,06851 + 1,255773 \cdot 30 + 0,948301 \cdot 28 = 42,16$.

Визначимо прогнозований інтервал математичного сподівання для заданого рівня значущості $p=99\%$ і $k=9$ ступенів вільності $t_{кр}=3,250$

$$\hat{Y}_0 - t_\gamma \hat{\sigma}_u \sqrt{X'_0 (X'X)^{-1} X_0} \leq M \leq \hat{Y}_0 + t_\gamma \hat{\sigma}_u \sqrt{X'_0 (X'X)^{-1} X_0}$$

$$X'_0 (X'X)^{-1} X_0 = \begin{pmatrix} 30 & 28 \end{pmatrix} \begin{pmatrix} 3,831 & -0,23 & 0,09 \\ -0,23 & 0,019 & -0,012 \\ 0,09 & -0,012 & 0,011 \end{pmatrix} \begin{pmatrix} 1 \\ 30 \\ 28 \end{pmatrix} = 0,5549$$

$$42,16 - 3,25 \cdot 3,84 \sqrt{0,5549} \leq M \leq 42,16 + 3,25 \cdot 3,84 \sqrt{0,5549}$$

$$32,86 \leq M \leq 51,46$$

Інтервальний прогноз індивідуального значення y_{13} за формулою:

$$\hat{Y}_{n+k} - t_\gamma \hat{\sigma}_u \sqrt{1 + X'_{n+k} (X'X)^{-1} X_{n+k}} \leq Y_{n+k} \leq \hat{Y}_{n+k} + t_\gamma \hat{\sigma}_u \sqrt{1 + X'_{n+k} (X'X)^{-1} X_{n+k}}$$

де X'_{n+k} вектор значень з m факторів у період $n+k$. Здійснимо прогноз на наступний місяць, отримуємо:

$$42,16 - 3,25 \cdot 3,84 \sqrt{1 + 0,5549} \leq Y_0 \leq 42,16 + 3,25 \cdot 3,84 \sqrt{1 - 0,5549}$$

$$26,60 \leq Y_0 \leq 57,72$$

Отже, з імовірністю $p=0,99$ прогноз математичного сподівання потрапляє в інтервал $[32,86;51,46]$, а прогноз індивідуального значення в інтервал $[26,60;57,72]$.

Економічна інтерпретація: якщо в прогнозований період основні фонди становили 30 млн. грн., а оборотні фонди – 28 млн. грн., то середній прибуток потрапляє в інтервал $[32,86;51,46]$. Водночас окреме значення прибутку знаходиться в інтервалі $[26,60;57,72]$.

ЧАСОВІ РЯДИ І ПРОГНОЗУВАННЯ

7.1 КЛАСИФІКАЦІЯ ЧАСОВИХ РЯДІВ.

Часовий ряд або *ряд динаміки* становлять розміщені в хронологічному порядку числові показники досліджуваного явища. Числові значення показників називаються *рівнянням ряду*.

Часові ряди представляють у вигляді таблиць і графіків (табл. 7.1).

Таблиця 7.1

t	t_1	t_2	...	t_i	...	t_n
y	y_1	y_2	...	y_i	...	y_n

За характером показників часові ряди поділяють на моментні ряди – вимірювання в деякий момент часу. Наприклад, кардіограма, концентрація лікарського препарату в крові після його вживання, рівень шуму в аудиторії під час пари; та інтервальні ряди – наприклад, щомісячні показники товарообігу в магазині.

Рівні часового ряду мають:

- відноситись до даного процесу;
- визначатись однією методикою;
- їх потрібно подавати в одних одиницях вимірювання.

Часовий ряд називається детермінованим, якщо його рівні однозначно визначається за відомою функціональною залежністю:

$$y_i = \varphi(t_i), \quad i = \overline{1, n} \quad (7.1)$$

Для дослідження часових рядів основною метою є виявлення і статистична оцінка основної тенденції розвитку досліджуваного процесу і відхилення від нього.

Основні етапи аналізу часових рядів:

- графічне представлення і опис поведінки часового ряду;
- виділення і видалення закономірних (невипадкових) складових часового ряду (тренду) сезонних і циклічних складових;

- згладжування і фільтрація (знищення низько- або високочастотних складових часового ряду);
- дослідження випадкової складової часового ряду, побудова і перевірка адекватної математичної моделі для її опису;
- прогнозування розвитку досліджуваного процесу на основі даного динамічного (часового) ряду;
- дослідження взаємозв'язку між різними часовими рядами.

У Excel існують три основних підходи до задач прогнозування і перспективних оцінок: із застосуванням ковзного середнього, регресії і експоненціального згладжування. Метод ковзного середнього полягає в тому, що розрахунок показника на прогнозований момент часу будується шляхом усереднення значень цього показника за кілька попередніх моментів часу.

Середній рівень моментного часового ряду з рівними проміжками часу між рівнями визначається за формулою:

$$\bar{y} = \frac{1}{n} \left(\frac{y_1 + y_n}{2} + \sum_{i=1}^n y_i \right) \quad (7.3)$$

Середній рівень інтервального ряду з рівними проміжками часу між рівнями визначається середнім арифметичним рівнів ряду динаміки:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (7.4)$$

Мінливість рівнів ряду відносно середнього рівня характеризується середньоквадратичним відхиленням:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}} \quad (7.5)$$

Використовується також безрозмірний показник мінливості, який називається коефіцієнтом варіації:

$$v = \frac{\sigma}{\bar{y}} \quad (7.6)$$

Коефіцієнт варіації можна використовувати для порівняння мінливості рядів динаміки з різними середніми рівнями. Часто його виражають у відсотках.

Нехай ряд динаміки представлено у табл. 7.2. суть методу усереднення полягає у заміні рівнів ряду середніми значеннями трьох, п'яти, семи рівнів. Метод усереднення ще називають методом згладжування ряду динаміки, оскільки усереднені рівні мають меншу мінливість. За згладженим рядом динаміки краще проглядається форма тренда.

Таблиця 7.2

<i>t</i>	1	2	...	<i>i</i>	...	<i>n</i>
<i>y</i>	y_1	y_2	...	y_i	...	y_n

Формула вирівнювання часового ряду методом усереднення за трьома рівнями ($m=3$) така:

$$\bar{y}_i = \frac{1}{3} (y_{i-1} + y_i + y_{i+1}) \quad i = 2, 3, \dots, n-1. \quad (7.7)$$

Усереднені за (7.7) рівні доповнюються початковим і кінцевим вирівняними рівнями, які визначаються за формулами:

$$\bar{y}_1 = \frac{1}{6} (y_1 + y_2 - y_3)$$

$$\bar{y}_n = \frac{1}{6} (y_{n-2} + 2y_{n-1} + 5y_n) \quad (7.8)$$

Формула вирівнювання часового ряду способом усереднення за п'ятьма рівнями ($m=5$) має такий вигляд:

$$\bar{y}_i = \frac{1}{5} (y_{i-2} + y_{i-1} + y_i + y_{i+1} - y_{i-2}) \quad i = 3, 4, \dots, n-2 \quad (7.9)$$

Усереднені за (7.8) рівні доповнюються першим, другим, передостаннім і останнім вирівняними рівнями, які визначаються за формулами:

$$\bar{y}_1 = \frac{1}{5} (y_1 + 2y_2 + y_3 - y_4)$$

$$\bar{y}_2 = \frac{1}{10} (y_1 + 3y_2 + 2y_3 + y_4)$$

$$\bar{y}_{n-1} = \frac{1}{10} (y_{n-3} + 2y_{n-2} + 3y_{n-1} + 4y_n)$$

$$\bar{y}_n = \frac{1}{5} (y_{n-3} + y_{n-2} + 2y_{n-1} + 3y_n) \quad (7.10)$$

Лінія, що з'єднує усереднені точки, є емпіричною лінією тренда. За формою емпіричні лінії обирають відповідну модель тренда.

Вирівнювання ряду динаміки методом усереднення можна проводити за сімома і більше рівнями. Тоді емпірична лінія тренда буде ще більш плавною, але тоді зникатимуть характерні для цього ряду динаміки гармонічні коливання рівнів.

Метод регресії ґрунтується на побудові лінії, що «у середньому» найменш ухиляється від масиву значень, який задає поведінку базової лінії.

Метод експоненціального згладжування враховує відхилення попереднього прогнозу від реального показника.

7.2 СТАЦІОНАРНІ ЧАСОВІ РЯДИ ТА ЇХ ХАРАКТЕРИСТИКИ, АВТОКОРЕЛЯЦІЙНА ФУНКЦІЯ.

Часовий ряд називається строгим стаціонарним, якщо сумісний розподіл імовірності n у спостережень y, y_1, y_2, \dots, y_n таке ж як і в $n - \tau$ спостережень $y_{1+\tau}, y_{2+\tau}, \dots, y_{n+\tau}$.

Іншими словами властивості строго стаціонарних рядів $y(t)$ не залежать від моменту t , тобто закон розподілу і його числові характеристики не залежать від t . Відповідно математичне сподівання:

$$M_y \stackrel{\text{def}}{=} a, \quad \sigma_y \stackrel{\text{def}}{=} \sigma, \quad \bar{y}_t = \frac{\sum_{i=1}^n y_i}{n}$$

$$\sigma_t^2 = \frac{\sum_{i=1}^n (y_i - \bar{y}_t)^2}{n} \quad (7.11)$$

Прикладом стаціонарного часового ряду, в якому $M(t)=0$, а помилки $e(t)$ – не корельовані, є «білий шум».

Тіснота зв'язку між послідовностями досліджень числового ряду y, y_1, y_2, \dots, y_n і ряду $y_{1+\tau}, y_{2+\tau}, \dots, y_{n+\tau}$ вони зсунуті один відносно одного на τ одиниць (або, як кажуть, з лагом τ) визначається коефіцієнт кореляції:

$$R_\tau = \frac{M \left[(y_t - a)(y_{t+\tau} - a) \right]}{\sigma_x(t) \cdot \sigma_x(t + \tau)} = \frac{M \left[(y_t - a)(y_{t+\tau} - a) \right]}{\sigma^2}$$

$$M(y_t) = M(y_{t+\tau}) = a$$

$$\sigma_y(t) = \sigma_y(t + \tau) = \sigma$$

Коефіцієнт кореляції R_τ вимірює кореляцію між членами одного і того ж ряду і його називають коефіцієнтом автокореляції, а залежність $R(t)$ – автокореляційною функцією.

Автокореляційна функція $R(t)$ залежить тільки від лага τ , причому $R(\tau)=R(-\tau)$.

Взаємозв'язок рівнів ряду динаміки з відповідними рівнями того ж ряду називають автокореляцією.

Нехай аналізується ряд динаміки, який має n рівнів:

$$y_1; y_2; \dots; y_n \quad (7.12)$$

З ряду (7.12) утворимо два ряди, які мають по $(n-\tau)$ рівнів, $\tau=1,2,\dots$. один ряд складається з рівнів ряду (7.12) від першого до $(n-\tau)$:

$$y_t = y_1; y_2; \dots; y_{n-\tau} \quad (7.13)$$

Другий ряд складається з рівнів ряду (7.12) від $(n+\tau)$ до n –го:

$$y_{t+\tau} = y_{1+\tau}; y_{2+\tau}; \dots; y_n \quad (7.14)$$

Ряди динаміки (7.13) та (7.14) зведемо в одну таблицю (табл. 7.3) таким чином:

Таблиця 7.3

t	1	2	...	i	...	$n-\tau$
y_t	y_1	y_2	...	y_i	...	$y_{n-\tau}$
$y_{t+\tau}$	$y_{1+\tau}$	$y_{2+\tau}$...	$y_{i+\tau}$...	y_n

τ – набуває значень: $1, 2, \dots, k$, $k < \frac{n}{4}$.

Мірою автокореляції служить коефіцієнт автокореляції. Коефіцієнт автокореляції – це коефіцієнт кореляції рядів y_t та $y_{t+\tau}$, які зведені у таблиці 7.3.

Обчислюють коефіцієнт автокореляції за формулою:

$$r_{\tau} = \frac{\sum_{i=1}^{n-\tau} (y_i - \bar{y}_t)(y_{i+\tau} - \bar{y}_{t+\tau})}{\sqrt{\sum_{i=1}^{n-\tau} (y_i - \bar{y}_t)^2} \sqrt{\sum_{i=1}^{n-\tau} (y_{i+\tau} - \bar{y}_{t+\tau})^2}} \quad (7.15)$$

де

$$\bar{y}_t = \frac{1}{n-\tau} \sum_{i=1}^{n-\tau} y_i; \quad \bar{y}_{t+\tau} = \frac{1}{n-\tau} \sum_{i=1}^{n-\tau} y_{i+\tau}.$$

На практиці будемо використовувати еквівалентну до (7.15), але більш зручну для обчислень таку формулу:

$$r_{\tau} = \frac{\overline{y_t \cdot y_{t+\tau}} - \bar{y}_t \cdot \bar{y}_{t+\tau}}{\sqrt{\left(\overline{y_t^2} - \bar{y}_t^2 \right) \left(\overline{y_{t+\tau}^2} - \bar{y}_{t+\tau}^2 \right)}} \quad (7.16)$$

де

$$\overline{y_t y_{t+\tau}} = \frac{1}{n-\tau} \sum_{i=1}^{n-\tau} y_i y_{i+\tau}; \quad \overline{y_t^2} = \frac{1}{n-\tau} \sum_{i=1}^{n-\tau} y_i^2; \quad \overline{y_{t+\tau}^2} = \frac{1}{n-\tau} \sum_{i=1}^{n-\tau} y_{i+\tau}^2$$

Алгоритм розрахунку величин у формулі (7.16) задається в табл. 7.4:

Таблиця 7.4

t	y_t	$y_{t+\tau}$	$y_t y_{t+\tau}$	y_t^2	$y_{t+\tau}^2$
1	-	$y_{1+\tau}$	$y_1 y_{1+\tau}$	y_1^2	$y_{1+\tau}^2$
2	y_2	$y_{2+\tau}$	$y_2 y_{2+\tau}$	y_2^2	$y_{2+\tau}^2$
...
$n-\tau$	$y_{n-\tau}$	y_n	$y_n y_{n-\tau}$	$y_{n-\tau}^2$	y_n^2
$\frac{1}{n-\tau} \sum$	\bar{y}_t	$\overline{y_{t+\tau}}$	$\overline{y_t \cdot y_{t+\tau}}$	\bar{y}_t^2	$\overline{y_{t+\tau}^2}$

Сукупність коефіцієнтів автокореляції r_τ з лагами $\tau=1,2,\dots$ дає уявлення про структуру рівнів ознаки в досліджуваному процесі. Графічне зображення сукупності коефіцієнтів автокореляції точками $(r_\tau; \tau)$ на площині, з'єднаних відрізками прямих у вигляді многокутника, називають корелограмою. Корелограма дає важливу інформацію про закономірності досліджуваного процесу протягом періоду спостереження, а також є корисною для прогнозування.

Для стаціонарного часового ряду із збільшенням лагу τ взаємозв'язок членів часового ряду y_t і $y_{t+\tau}$ зменшується (слабкий) і автокореляційна функція R_τ спадає по абсолютній величині, хоча для її вибіркового коефіцієнта автокореляції $r(\tau)$, особливо при невеликій кількості пар спостережень $(n-\tau)$, властивість монотонного спадання по абсолютній величині при зростанні τ може порушитися.

7.3 АНАЛІТИЧНЕ ВИРІВНЮВАННЯ ЧАСОВОГО РЯДУ. ПРОГНОЗУВАННЯ НА ОСНОВІ МОДЕЛІ ЧАСОВИХ РЯДІВ.

Для вирівнювання часового ряду необхідно вибрати відповідну функцію: лінійну, експоненційну, логістичну, Гомперця і так далі.

При виборі відповідної функції використовують змістовний аналіз, який допоможе встановити характер часового процесу, графічне зображення часового ряду.

При виборі поліноміальної функції $f(t)=\beta_0+\beta_1x_1+\beta_2x_2^2+\beta_3x_3^3+\dots+\beta_nx_n^n$ можна застосовувати метод послідовних різниць, який полягає в обрахуванні різниць 1-го порядку і порядок різниць, при яких вони будуть майже однакові, приймається

$$\Delta_t=y_t-y_{t-1}; \Delta_t^{2(\text{пор})}=\Delta_t-\Delta_{t-1} \text{ за степiнь номiналу.}$$

З двох функцій перевага надається тій, при якій менша сума квадратів відхилень фактичних даних від розрахункових на основі цих функцій.

Зауваження: так як для будь-якого ряду із m точок можна підібрати поліноміальний $(n-1)$ степiнь, який буде проходити через всі точки і відповідно з мінімальною – нульовою – сумою квадратів відхилень.

В цьому випадку не слід говорити про основну тенденцію, враховуючи випадковий характер даних точок і перевагу слід віддати більш простим функціям. Вибір математичної моделі тренда визначається насамперед виглядом емпіричної лінії тренда. Для визначення основної тенденції найчастіше використовують метод найменших квадратів (МНК). Значення часового ряду y_t розглядається як залежна змінна, а час – як пояснювальна змінна.

Моделюється ряд динаміки, (табл. 7.2). для лінійного тренда рівняння (7.2) має такий вираз:

$$y_t = \beta_0 + \beta_1 t_i + e_t \quad (i=1,2,\dots,n) \quad (7.17)$$

$e(t)$ – випадкова величина розподілена за нормальним законом.

Пряма найменших квадратів матиме вигляд:

$$\hat{y}_t = b_0 + b_1 t \quad (7.18)$$

Система нормальних рівнянь для визначення МНК-оцінок b_0 та b_1 у (7.18) має такий вид:

$$\begin{cases} b_0 n + b_1 \sum_{t=1}^n t = \sum_{t=1}^n y_t \\ b_0 \sum_{t=1}^n t + b_1 \sum_{t=1}^n t^2 = \sum_{t=1}^n y_t \cdot t \end{cases} \quad (7.19)$$

Розв'язком системи (7.19) є такі вирази:

$$b_1 = \frac{\overline{yt} - \bar{y}\bar{t}}{\overline{t^2} - \bar{t}^2}; \quad b_0 = \bar{y} - b_1 \bar{t} \quad (7.20)$$

Оскільки рівні ряду визначені для часової послідовності $\{y_t\}_{t=1,2,\dots,n}$, то використаємо:

$$\begin{aligned} \sum_{i=1}^n t_i &= \frac{n(n+1)}{2}; \quad \bar{t} = \frac{n+1}{2} \\ \sum_{i=1}^n t_i^2 &= \frac{n(n+1)(n+1)}{6}; \quad \overline{t^2} = \frac{(n+1)(n+1)}{6} \end{aligned} \quad (7.21)$$

Основна задача аналізу часового ряду полягає в прогнозуванні, на основі його розвитку вивчаємо процес зміни ряду. Виходячи з того, що тенденція розвитку встановлена в минулому може бути застосована на майбутній період. При цьому ми вважаємо, що $e(t)$ задовольняє умовам регресійного аналізу.

Критична точка визначається з таблиці Фішера для F -розподілу.

$$F^* = F(p=1-\alpha; k_1=1; k_2=n-2). \quad (7.22)$$

Якщо емпіричне значення критерію більше від критичного, то є підстави стверджувати про наявність лінійної тенденції у ряді динаміки.

Дисперсія ряду динаміки відносно прямої найменших квадратів визначається:

$$\sigma^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-2} \quad (7.23)$$

Дисперсія параметрів лінійного тренда розраховується:

$$\sigma_{b_1}^2 = \frac{12\sigma^2}{n(n^2-1)} \quad (7.24)$$

$$\sigma_{b_0}^2 = \frac{2(n+1)\bar{t}^2}{n(n-1)} = \bar{t}^2 \sigma_{b_1}^2 \quad (7.25)$$

Оскільки ми розглядали точковий та інтервальний прогноз значень залежної змінної y для лінійної економетричної моделі (ЛЕМ) і нелінійної економетричної моделі (НЕМ), то аналогічно ми можемо побудувати прогноз і для рядів, а саме точковий та інтервальний.

$$b_0 + b_1 t - t^* \sigma \cdot \sqrt{\frac{1}{n} + \frac{\bar{t}^2}{n(\bar{t}^2 - \bar{t}^2)}} \leq \beta_0 + \beta_1 t \leq b_0 + b_1 t + t^* \sigma \cdot \sqrt{\frac{1}{n} + \frac{\bar{t}^2}{n(\bar{t}^2 - \bar{t}^2)}} \quad (7.26)$$

де

$$t^* = t \left(p = 1 - \frac{\alpha}{2}; \quad k = n - 2 \right) - \text{коефіцієнт Стюдента.} \quad (7.27)$$

Для окремого прогнозованого рівня ряду динаміки при $t=t_0$ з імовірністю $(1-\alpha)$ вірогідний проміжок є таким:

$$\left[b_0 + b_1 t_0 - t^* \sigma \cdot \sqrt{1 + \frac{1}{n} + \frac{\bar{t}^2}{n(\bar{t}^2 - \bar{t}^2)}}; b_0 + b_1 t_0 + t^* \sigma \cdot \sqrt{1 + \frac{1}{n} + \frac{\bar{t}^2}{n(\bar{t}^2 - \bar{t}^2)}} \right] \quad (7.28)$$

Порівняльний аналіз цих методів наведено у табл.7.5.

Таблиця 7.5

Метод	Як запустити	Переваги та недоліки
Ковзного середнього	Сервис – Анализ данных – Скользящее среднее	Швидкий, але не досить точний метод виявлення загальних тенденцій часового ряду. Не дає прогнозу, що виходить за межі відомих даних
Регресії – лінійні прогнози	Вставка – Функция – Статистические Тенденция	Більш точний метод, на основі якого можна робити прогнозування поведінки часового ряду (для простих часових рядів)
Регресії – нелінійні прогнози	Вставка – Функция – Статистические - Рост	
Згладжування	Сервис – Анализ данных – Экспоненциальное сглаживание	Даний метод самокорегується, тобто враховує відхилення попереднього прогнозу від реального показника

МЕТОДИКА ДОСЛІДЖЕННЯ ЧАСОВОГО РЯДУ

Задача. Маємо дані за 8 років

t , рік	1	2	3	4	5	6	7	8
y_t , попит	213	171	291	309	317	362	351	361

Потрібно знайти середнє значення середнього квадратичного відхилення, коефіцієнт автокореляції для лагів $\tau=1;2$ і частинний коефіцієнт кореляції 1-го порядку. Знайти рівняння невинпадкової величини (тренду) для часового ряду y_t припускаючи, що тренд лінійний. Провести згладжування часового ряду y_t методом ковзання середніх і використовуючи середнє арифметичне з інтервалом згладжування $m=3$ роки. Дати точкову та інтервальну оцінку середнього та індивідуального значення попиту на деякий товар на момент $t=9$ з імовірністю $p=0,05$, вважаючи, що тренд лінійний $e(t)$ відповідає класичній моделі.

Розв'язання:

$$\bar{y}_t = \frac{213+171+291+309+317+362+351+361}{8} = 296,88$$

$$\sigma_t^2 = \overline{y_t^2} - (\bar{y}_t)^2$$

$$\overline{y_t^2} = \frac{213^2 + 171^2 + 291^2 + 309^2 + 317^2 + 362^2 + 351^2 + 361^2}{8} = 92478,38$$

$$\sigma_t^2 = 92478,38 - (296,88)^2 = 4343,61$$

$$\sigma_t = \sqrt{4343,61} = 65,31$$

Знайдем коефіцієнт автокореляції $r(\tau)$ часового ряду $\tau=1$.

y_t	213	171	291	309	317	362	351
$y_{t+\tau}$	171	291	309	317	362	351	361

Для цього знайдемо:

$$\sum y_t = 213+171+291+309+317+362+351 = 2014$$

$$\sum y_t^2 = 213^2 + 171^2 + 291^2 + 309^2 + 317^2 + 362^2 + 351^2 = 609506$$

$$\sum y_{t+\tau} = 171 + 291 + 309 + 317 + 362 + 351 + 361 = 2162$$

$$\sum y_{t+\tau}^2 = 171^2 + 291^2 + 309^2 + 317^2 + 362^2 + 351^2 + 361^2 = 694458$$

$$\sum y_t \cdot y_{t+\tau} = 213 \cdot 171 + 171 \cdot 291 + 291 \cdot 309 + 309 \cdot 317 + 317 \cdot 362 + 362 \cdot 351 + 351 \cdot 361 = 642383$$

$$\begin{aligned} r(\tau) &= \frac{(n-\tau) \sum_{t=1}^{n-\tau} y_t y_{t+\tau} - \sum_{t=1}^n y_t \sum_{t=1}^n y_{t+\tau}}{\sqrt{(n-\tau) \sum_{t=1}^{n-\tau} y_t^2 - (\sum_{t=1}^{n-\tau} y_t)^2} \cdot \sqrt{(n-\tau) \sum_{t=1+\tau}^n y_{t+\tau}^2 - (\sum_{t=1+\tau}^{n-\tau} y_{t+\tau})^2}} = \\ &= \frac{9 \cdot 642383 - 2014 \cdot 2162}{\sqrt{9 \cdot 609506 - (2014)^2} \cdot \sqrt{9 \cdot 694458 - (2162)^2}} = \\ &= \frac{5781447 - 4354268}{\sqrt{5485554 - 4056196} \cdot \sqrt{6250122 - 4674244}} = \\ &= \frac{1427179}{\sqrt{1429358} \cdot \sqrt{1575878}} = \frac{1427179}{1195,56 \cdot 1255,34} = \frac{1427179}{1500834,29} = 0,725 \end{aligned}$$

Якщо $\tau=2$, то беремо 6 спостережень, оскільки ряд спадає з $n-\tau=8-2=6$; якщо $\tau=2$, то $r(2)=0,842$.

Для визначення частинного коефіцієнта кореляції 1-го $r_{\text{част}}=r_{02\dots1}$ між членами ряду y_t і y_{t+2} при виключенні впливу y_{t+1} спочатку знайдемо (аналогічно до попереднього) коефіцієнт автокореляції між членами ряду y_{t+1} і y_{t+2} .

$$r(1,2)=0,825$$

$$r_{02\dots1} = \frac{0,842 - 0,725 \cdot 0,825}{\sqrt{1 - 0,725^2} \cdot \sqrt{1 - 0,825^2}} = 0,627$$

Знайдемо рівняння невинуваткової величини (тренду) для часового ряду y_t припускаючи, що тренд лінійний.

$$\sum_{t=1}^8 t = \frac{8 \cdot 9}{2} = 36; \quad \sum_{t=1}^8 t^2 = \frac{8 \cdot 9 \cdot 17}{6} = 204$$

$$\sum_{t=1}^8 y_t = 213 + 171 + 291 + 309 + 317 + 362 + 351 + 361 = 2375$$

$$\sum_{t=1}^8 y_t^2 = 213^2 + 171^2 + 291^2 + 309^2 + 317^2 + 362^2 + 351^2 + 361^2 = 739827$$

$$\sum_{t=1}^8 y_t \cdot t = 213 \cdot 1 + 171 \cdot 2 + 291 \cdot 3 + 309 \cdot 4 + 317 \cdot 5 + 362 \cdot 6 + 351 \cdot 7 + 361 \cdot 8 = 11766$$

$$\begin{cases} 8b_0 + 36b_1 = 237 \\ 36b_0 + 204b_1 = 11766 \end{cases} \quad y_t = 181,32 + 25,679t$$

Тобто попит щорічно зростає в середньому на 25,679 (залежить від b_1).

Перевіримо значимість отриманого рівняння тренду за F-критерієм (Фішера) на 5-ти %-му рівні значимості (імовірність $p=0,05$).

$$F_{\text{стат}} = \frac{Q_R(n-2)}{Q_e} = \frac{27695,3 \cdot 6}{7053,6} = 23,56$$

$$F_{\text{табл}}(0,05; 1,6) = 4,46;$$

Оскільки емпіричне значення критерію більше від критичного, то є підстави стверджувати про наявність лінійної тенденції у ряді динаміки.

Проведемо згладжування часового ряду y_t методом ковзання середніх і використовуємо середнє арифметичне з інтервалом згладжування $m=3$ роки.

$$\tilde{y}_2 = \frac{1}{3}(y_1 + y_2 + y_3) = \frac{1}{3}(213 + 171 + 291) = 225$$

$$\tilde{y}_3 = \frac{1}{3}(y_2 + y_3 + y_4) = \frac{1}{3}(171 + 291 + 309) = 241$$

t	1	2	3	4	5	6	7	8
\tilde{y}_t	–	225	241	305	329,3	336,3	358	–

Дамо точкову та інтервальну оцінку середнього та індивідуального значення попиту на деякий товар на момент $t=9$ з імовірністю $p=0,05$. При цьому вважаємо, що тренд лінійний $e(t)$ відповідає класичній моделі.

$$\hat{y}_t = 181,32 + 25,679t$$

Оцінимо математичне сподівання

$M_{t=9} = \bar{y}(9)$ – *групова середня*

$$\hat{y}_{t=9} = 181,32 + 25,639 \cdot 9 = 412,4$$

$$\sigma^2 = \frac{7059,2}{8-2} = 1176,5$$

$$\sigma_{y_{t=9}}^2 = 1176,5 \cdot \left(\frac{1}{8} + \frac{(9-4,5)^2}{42} \right) = 714,3$$

$$\sigma_{y_{t=9}} = \sqrt{714,3} = 26,73$$

$$\bar{t} = \frac{\sum t}{n} = \frac{36}{8} = 4,5$$

$$\sum_{t=1}^n (t - \bar{t})^2 = \sum_{t=1}^n t^2 - \frac{(\sum t)^2}{n} = 204 - \frac{36^2}{8} = 42$$

$$t_{0,95,6} = 2,45$$

$$412,4 - 2,45 \cdot 26,73 \leq \bar{y}(9) \leq 412,4 + 2,45 \cdot 26,73$$

$$346,9 \leq \bar{y}(9) \leq 477,9$$

Дисперсія індивідуальної оцінки:

$$\sigma_{y_{t_0=9}}^2 = 1176,5 \left(1 + \frac{1}{8} + \frac{(9-4,5)^2}{42} \right) = 1890,8$$

$$\sigma_{y_{t_0=9}} = \sqrt{1890,8} = 43,48$$

$$412,4 - 2,45 \cdot 43,48 \leq y^* \leq 412,4 + 2,45 \cdot 43,48$$

$$305,9 \leq y^* \leq 518,9$$

З надійністю 0,95 середнє значення попиту на товар на 9-й рік буде лежати в межах від 346,9 до 477,9 одиниць, а його індивідуальне значення знаходиться в межах від 305,9 до 518,9 одиниць.

Розглянемо використання прогностичних засобів Excel на прикладі.

Припустимо, що курс грн./євро в попередні 18 днів змінювався так, як показано в таблиці. Існує потреба прогнозу цього курсу на 19-й день від початку спостережень.

дні	1	2	3	4	5	6	7	8	9	
Курс УН/€	6,210	6,174	6,153	6,184	6,025	5,998	6,053	6,072	6,045	
дні	10	11	12	13	14	15	16	17	18	19
Курс УН/€	6,112	6,098	6,145	6,156	5,990	6,056	6,087	6,123	6,098	?

Застосуємо до зазначених даних різні методи прогнозування. Почнемо з ковзного середнього. Метод полягає в тому, що розрахунок показника на прогнозований момент часу будується шляхом усереднення значень показника за кілька попередніх моментів часу. У меню **Сервіс** вибираємо команду **Анализ данных**, появляється діалогове вікно, показане нижче на рис. 7.1.

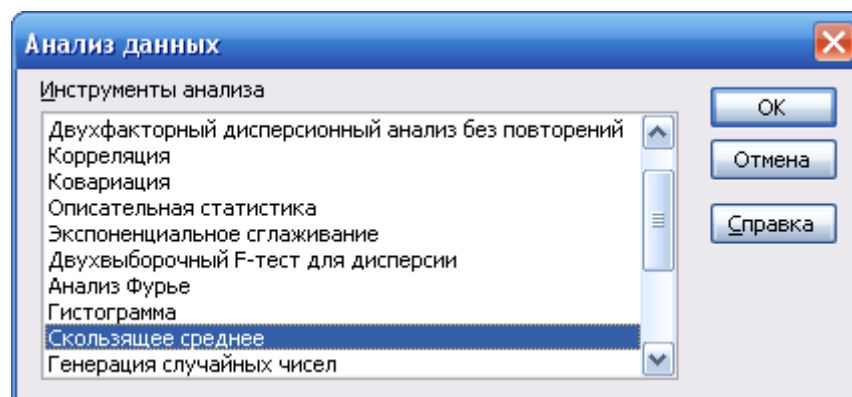


Рис. 7.1 Діалогове вікно **Анализ данных**

Вибравши команду **Скользящее среднее** і натиснувши **ОК**, переходимо до наступного вікна (рис.7.2), і заповнюємо його так, як показано на рисунку.

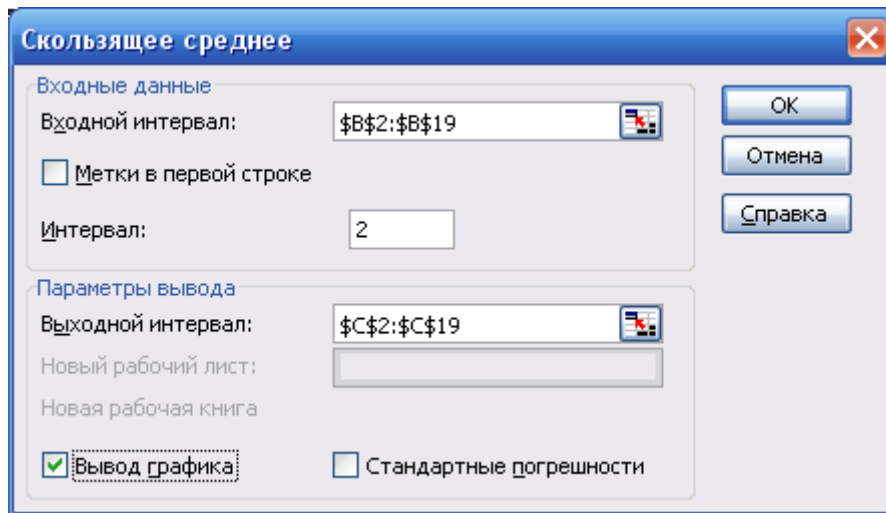


Рис. 7.2 Діалогове вікно **Скользящее среднее**

Вхідний інтервал **\$B\$2-\$B\$19** показує масив комірок, у яких міститься інформація для визначення базової лінії (у нашому випадку – щоденні значення курсу грн./євро). Виберемо інтервал 2, на якому буде обчислюватися середнє значення (зазвичай це 2-3 точки) і вкажемо за вихідний масив комірок **\$C\$2-\$C\$19**, де будуть розташовуватися значення, обчислені даним методом (прогноз 1).

Зверніть увагу, що в першій комірці **C2** з'явиться символ #НД. Це зумовлено тим, що ковзне середнє обчислюється на часовому інтервалі, рівному 2, але для комірки **C2** немає попередніх значень базової лінії, тому воно не може бути обчислене доти, поки не буде відомо двох значень базової лінії, на підставі яких будується прогноз. Природно, що якби ми вибрали інтервал рівний трьом, то для двох перших комірок прогнозні значення не можуть бути обчислені і них знаходиться символ #НД.

У графі **E** розмістимо результати прогнозування методом експоненціального згладжування (прогноз 2). Вибравши в команді **Анализ данных** пункт **Экспоненциальное сглаживание**, вкажемо ті самі параметри, що й у попередньому випадку. Фактор загасання можна не ставити, він буде виставлений автоматично.

	A	B	C	D	E	F	G	H	I
1	дні	курс	прогноз1	(відхилення1) ²	прогноз2	(відхилення2) ²	прогноз3	(відхилення3) ²	
2	1	6,21	#Н/Д		#Н/Д		6,13	0,00642	
3	2	6,174	6,192	0,00032	6,21	0,0013	6,126	0,00228	
4	3	6,153	6,164	0,00011	6,185	0,00101	6,123	0,00093	
5	4	6,184	6,169	0,00024	6,163	0,00046	6,119	0,00424	
6	5	6,025	6,105	0,00632	6,178	0,02328	6,115	0,00815	
7	6	5,998	6,012	0,00018	6,071	0,0053	6,112	0,01291	
8	7	6,053	6,026	0,00076	6,02	0,0011	6,108	0,00302	
9	8	6,072	6,063	0,00009	6,043	0,00084	6,104	0,00104	
10	9	6,045	6,059	0,00018	6,063	0,00034	6,101	0,0031	
11	10	6,112	6,079	0,00112	6,05	0,00378	6,097	0,00022	
12	11	6,098	6,105	0,00005	6,094	0,00002	6,093	0,00002	
13	12	6,145	6,122	0,00055	6,097	0,00234	6,09	0,00306	
14	13	6,156	6,151	0,00003	6,13	0,00065	6,086	0,00489	
15	14	5,99	6,073	0,00689	6,148	0,02507	6,082	0,00854	
16	15	6,056	6,023	0,00109	6,038	0,00034	6,079	0,00052	
17	16	6,087	6,072	0,00024	6,05	0,00134	6,075	0,00014	
18	17	6,123	6,105	0,00032	6,076	0,00221	6,071	0,00266	
19	18	6,098	6,111	0,00016	6,109	0,00012	6,068	0,00091	
20	19(прогноз)		6,098		6,101		6,064		
21									
22									
23	дисперсія			0,0011		0,00409		0,0035	
24									
25									
26									
27									
28									
29									
30									
31									
32									
33									

Рис. 7.3 Результати прогнозування курсу грн./євро

Проаналізувавши дисперсії, найменше значення виявилось для методу ковзного середнього, а найбільше – для методу експотенціального згладжування. Прогноз курсу на 19-й день можна представити у вигляді: «середнє значення із усіх прогнозів, отриманих різними методами» плюс-мінус «найбільше за абсолютною величиною відхилення отриманих прогнозних величин від середнього значення», тобто в нашому випадку $6,088 \pm 0,024$ грн. за євро.

МУЛЬТИКОЛІНЕАРНІСТЬ

8.1 МУЛЬТИКОЛІНЕАРНІСТЬ І ЇЇ НАСЛІДКИ.

Термін "мультиколінеарність" означає, що в багатofакторній регресійній моделі дві або більше, на перший погляд, незалежні змінні (факторів) насправді пов'язані між собою лінійною залежністю, тобто мають високий ступінь кореляції $r_{x_i, x_j} \rightarrow 1, i \neq j$

Якщо в регресії присутнє явище мультиколінеарності, то не виконується одна з чотирьох умов, а саме не виконується умова $|X^T X| \neq 0$, яка необхідна для оцінювання параметрів загальної лінійної моделі методом найменших квадратів. За наявності мультиколінеарності має місце $|X^T X| \rightarrow 0$, отже, як випливає з рівняння з (2.7), при визначенні параметрів моделі має місце ділення на величини, близькі до нуля і оцінки параметрів моделі можуть дуже сильно відхилятися від реальних.

Економічні показники, які входять до економетричної моделі як пояснювальні змінні, на практиці дуже часто пов'язані між собою, це може стати перешкодою для оцінювання параметрів моделі методом найменших квадратів та істотно вплинути на якість економетричного моделювання.

Основні наслідки мультиколінеарності.

1. Падає точність оцінювання, яка виявляється так:

- а) помилки деяких конкретних оцінок стають занадто великими;
- б) помилки досить корельовані одна з одною;
- в) дисперсії оцінок параметрів різко збільшуються.

2. Оцінки параметрів деяких змінних моделі можуть бути незначущими через наявність їх взаємозв'язку з іншими змінними, а не тому, що вони не впливають на залежну змінну. У такому разі множина вибіркового даних не дає змоги цей вплив виявити.

3. Оцінки параметрів стають досить чутливими до обсягів сукупності спостережень. Збільшення сукупності спостережень іноді може спричинитися до істотних змін в оцінках параметрів. Тому в економетричних дослідженнях

важливо з'ясувати, чи існують між факторними змінними взаємозв'язки і наскільки вони суттєві.

Існує багато методів дослідження мультиколінеарності. Основна частина цих методів полягає в дослідженні матриці коефіцієнтів парної кореляції:

$$r = \begin{pmatrix} r_{x_1x_1} & r_{x_1x_2} & r_{x_1x_3} & r_{x_1x_4} & \dots & r_{x_1x_m} \\ r_{x_2x_1} & r_{x_2x_2} & r_{x_2x_3} & r_{x_2x_4} & & r_{x_2x_m} \\ & & & & & \\ & & & & & \\ r_{x_mx_1} & r_{x_mx_2} & r_{x_mx_3} & r_{x_mx_4} & & r_{x_mx_m} \end{pmatrix} \quad (8.1)$$

Проте, коли до моделі входять більш як дві пояснювальні змінні, вивчення питання про мультиколінеарність не може обмежуватись інформацією, що її містять елементи кореляційної матриці. Явище мультиколінеарності в жодному разі не зводиться лише до існування парної кореляції між незалежними змінними.

Більш загальна перевірка передбачає знаходження визначника (детермінанта) кореляційної матриці r , який ще називається детермінантом кореляції і позначається $|r|$. Числові значення детермінанта кореляційної матриці задовольняють умову: $|r| \in [0, 1]$.

Якщо $|r|=0$, то існує повна мультиколінеарність, а коли $|r|=1$, мультиколінеарність відсутня. Чим ближче $|r|$ до нуля, тим певніше можна стверджувати, що між пояснювальними змінними існує мультиколінеарність. Незважаючи на те, що на числове значення $|r|$ впливає дисперсія факторних змінних, цей показник можна вважати точковою мірою рівня мультиколінеарності.

Якщо в економетричній моделі знайдено мале значення параметра A_k при високому рівні частинного коефіцієнта детермінації R_i^2 і при цьому F-критерій істотно відрізняється від нуля, то це також свідчить про наявність мультиколінеарності.

Коли коефіцієнт частинної детермінації R_i^2 , який обчислено для регресійних залежностей між однією пояснювальною змінною та іншими, має

значення, яке близьке до одиниці, то можна говорити про наявність мультиколінеарності.

Нехай при побудові економетричної моделі на основі покрокової регресії введення нової пояснювальної змінної істотно змінює оцінку параметрів моделі при незначному підвищенні (або зниженні) коефіцієнтів кореляції чи детермінації. Тоді ця змінна перебуває, очевидно, у лінійній залежності від інших, які було введено до моделі раніше.

Усі ці ознаки мультиколінеарності мають один спільний недолік: жодна з них чітко не розмежовує випадки, коли мультиколінеарність істотна і коли нею можна знехтувати.

8.2 ДОСЛІДЖЕННЯ МУЛЬТИКОЛІНЕАРНОСТІ.

В економетричних задачах для дослідження наявності мультиколінеарності широко застосовується метод Фаррара-Глобера. Цей метод тріступеневий, тобто має три види статистичних критеріїв, за якими перевіряється мультиколінеарність:

- 1) усього масиву незалежних змінних (χ^2 – «хі»-квадрат);
- 2) кожної незалежної змінної з рештою змінних (F-критерій);
- 3) кожної пари незалежних змінних (t -критерій).

Усі ці критерії при порівнянні з їх критичними значеннями дають змогу робити конкретні висновки щодо наявності чи відсутності мультиколінеарності незалежних змінних. Метод Фаррара-Глобера зручно подати у вигляді алгоритму:

Крок 1. Стандартизація (нормалізація) змінних.

Позначимо вектори незалежних змінних економетричної моделі через x_1, x_2, \dots, x_m . Елементи стандартизованих векторів обчислимо за формулами:

$$x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sigma_{x_k}} \qquad x_{ik}^* = \frac{x_{ik} - \bar{x}_k}{\sqrt{n\sigma_{x_k}^2}} \qquad (8.2)$$

де n - число спостережень ($i = 1, 2, \dots, n$); m — число незалежних змінних, ($k=1, 2, \dots, m$); \bar{x}_k — середнє арифметичне k -ї незалежної змінної; σ_{x_k} — дисперсія k -ї незалежної змінної.

Крок 2.

Знаходження кореляційної матриці

$$1) r = \frac{1}{n} X^{*'} X^* \quad 2) R = X^{*'} X^* \quad (8.3)$$

де X^* – матриця стандартизованих незалежних (пояснювальних) змінних; $X^{*'}$ – матриця, транспонована до матриці X^* .

Крок 3.

Визначник критерію χ^2 :

$$\chi^2 = \left[n - 1 - \frac{1}{6} (m + 5) \right] \ln|r| \quad (8.4)$$

$|r|$ – визначник кореляційної матриці r .

Значення цього критерію порівнюється з табличним при $m(m-1)/2$ ступені вільності і рівні значущості α . Якщо $\chi_{факт}^2 > \chi_{табл}^2$ то в масиві незалежних змінних існує мультиколінеарність.

Крок 4. Визначення оберненої матриці:

$$r = \left(\frac{1}{n} X^{*'} X^* \right)^{-1} \quad (8.5)$$

Крок 5. Обчислення F – критеріїв

$$F_k = z_{kk} - 1 \frac{n - m}{m - 1}$$

де z_{kk} – діагональні елементи матриці r^{-1} . Фактичні значення критеріїв порівнюються з табличними при $\kappa_1 = m - 1$ і $\kappa_2 = n - m$ ступенях вільності та рівні значущості α . Якщо $F_{факт} > F_{табл}$ то відповідна k -та незалежна змінна мультиколінеарна з іншими.

Коефіцієнт детермінації для кожної змінної

$$R_{x_k}^2 = 1 - \frac{1}{z_{kk}}$$

Крок 6. Знаходження частинних коефіцієнтів кореляції:

$$r_{kj,12\dots m} = -\frac{z_{kj}}{\sqrt{z_{kk} \cdot z_{jj}}}$$

де z_{kk} – діагональні елементи матриці r^{-1} , z_{kj} – елемент матриці r^{-1} , що міститься в k -му рядку j -у стовбці.

Крок 7. Обчислення t – критеріїв

$$t_{kj} = \frac{r_{kj,12\dots m} \sqrt{n-m-1}}{1 - r_{kj,12\dots m}^2} \quad (8.6)$$

Фактичні значення критеріїв t_{kj} порівнюються з табличними при рівні значущості α і $n-m-1$ ступенях вільності. Якщо $t_{kj} > t_{kj\text{таб}}$, то між незалежними змінними x_k і x_j існує мультиколінеарність.

8.3 СПОСОБИ УСУНЕННЯ МУЛЬТИКОЛІНЕАРНОСТІ.

Найпростіше позбутися мультиколінеарності в економетричній моделі можна, відкинувши одну зі змінних мультиколінеарної пари. Але на практиці вилучення якогось чинника часто суперечить логіці економічних зв'язків. Тоді можна перетворити певним чином незалежні змінні моделі:

- а) узяти відхилення від середньої;
- б) замість абсолютних значень узяти відносні;
- в) замінити фактор X_j^* , який має тісний зв'язок з X , на фактор $X_j^* = X_i - X_j$, після чого перевірити наявність мультиколінеарності між факторами X_i і X_j^* ;
- г) стандартизувати пояснювальні змінні тощо.

За наявності мультиколінеарності змінних потрібно звертати увагу й на специфікацію моделі. Іноді заміна однієї функції іншою, якщо це не суперечить апріорній інформації, дає змогу уникнути явища мультиколінеарності.

Іноді просте збільшення спостережень у моделі пом'якшує проблему мультиколінеарності. Проте отримати додаткові дані не завжди легко, оскільки на практиці це вимагає значних витрат. Крім того, потрібно впевнитися, що

економічна структура, пов'язана з новими спостереженнями, буде подібна до початкової структури.

Коли жодний з розглянутих способів не дає змоги позбутися мультиколінеарності, то для її виправлення використовують статистичні методи, такі як факторний аналіз, метод головних компонентів, гребенева регресія.

Метод головних компонентів

Цей метод призначений для оцінювання моделей великого розміру, а також для оцінки параметрів моделі, якщо до неї входять мультиколінеарні змінні. Ідея методу полягає в тому, щоб перетворити множину змінних X на нову множину попарно некорельованих змінних (головних компонент), серед яких перша відповідає максимально можливій дисперсії, а друга - максимально можливій дисперсії в підпросторі, який є ортогональним до першого і т. д. Тоді в одержаній системі головні компоненти будуть вже не лише некорельованими, але й впорядкованими за величиною їх дисперсій, причому перша головна компонента має найбільшу дисперсію, друга - трохи меншу і т.д., остання компонента - найменшу дисперсію.

Головні компоненти Z_j обчислюються як добутки матриці нормалізованих пояснювальних змінних на власні вектори a_j матриці $X^T X$, тобто $Z_j = X a_j$. Звідси сума елементів вектора дорівнює $Z_j^T Z_j = a_j^T X^T X a_j$.

Щоб знайти головні компоненти Z_j необхідно максимізувати вектор $Z_j^T Z_j = a_j^T X^T X a_j$ за таких умов:

- $a_j^T a_k = 1, k=j$;
- $a_j^T a_k = 0, k \neq j$;

Перша умова нормалізує вектор a_j , щоб він не став дуже великим, а друга умова забезпечує відсутність кореляції між Z_j і Z_k , бо коваріація між ними подається у вигляді $a_j^T X^T X a_k = \lambda_k a_j^T a_k$ і дорівнює нулю лише тоді, коли $a_j^T a_k = 0$. Для розв'язування цієї задачі будується функція Лагранжа:

$$L = a_j^T X a_k - \lambda_k \left(a_j^T a_k - 1 \right) - \mu \left(a_j^T a_k \right) \rightarrow \max \quad (8.7)$$

де λ_k, μ - множники Лагранжа.

З необхідних умов екстремуму $\frac{dL}{d\lambda_k} = 0$, $\frac{dL}{d\mu} = 0$, отримаємо рівняння

$$\begin{aligned} X^T X \vec{a}_j &= \lambda_j a_j, \text{ розв'язавши яке, знайдемо всі власні вектори матриці} \\ X^T X \vec{a}_j. \end{aligned}$$

Головні компоненти $Z_j = X a_j$ попарно не корельовано, а їх дисперсії визначаються за формулами

$$Z_j^T Z_j = \lambda_j, \quad j = 1, 2, \dots, m;$$

$$\sum_{j=1}^m \lambda_j = Z_1^T Z_1 + Z_2^T Z_2 + \dots + Z_m^T Z_m.$$

Співвідношення $\lambda_j / \sum_{j=1}^m \lambda_j$ характеризують пропорційний внесок кожного з векторів до загальної варіації змінних X , а оскільки ці компоненти ортогональні, сума всіх внесків дорівнює одиниці.

Визначивши всі головні компоненти, знаходимо їх зв'язок із залежною змінною Y . Тоді за допомогою оберненого перетворення повертаємося від параметрів моделі з головними компонентами до знаходження оцінок параметрів змінних X .

Подамо метод головних компонентів у вигляді алгоритму:

Крок 1. Нормалізація всіх пояснювальних змінних:

$$X_{ij}^* = \frac{X_{ij} - \bar{X}_j}{\sigma_{x_j}}, \quad i = 1, 2, \dots, n; \quad j = 1, 2, \dots, m \quad (8.8)$$

Крок 2. Обчислення кореляційної матриці

$$r = \frac{1}{n} X^{*T} X^*. \quad (8.9)$$

Крок 3. Знаходження характеристичних чисел матриці r з рівняння

$$|r - \lambda E| = 0, \quad (8.10)$$

де E – одинична матриця розміром $m \times m$.

Крок 4. Власні значення λ_k упорядковуються за абсолютним рівнем вкладу кожної головної компоненти до загальної дисперсії.

Крок 5. Обчислення власних векторів s_k розв'язуванням системи рівнянь

$$\mathbf{A} - \lambda E \mathbf{S} = 0 \quad (8.11)$$

за таких умов:

$$s_j^T s_k = \begin{cases} 0, & k \neq j \\ 1, & k = j \end{cases}$$

Крок 6. Знаходження головних компонентів - векторів

$$Z = XS \quad (8.12)$$

де $S = [s_1, s_2, \dots, s_m]$ - матриця власних векторів.

Головні компоненти мають задовольняти умови:

$$\sum_{i=0}^n Z_{ki} = 0, \quad \frac{1}{n} Z_k^T Z_k = \lambda_k, \quad k = 1, 2, \dots, m, \quad s_j^T s_k = 0, \quad k \neq j.$$

Крок 7. Визначення параметрів моделі з головними компонентами

$Y = ZB$:

$$\hat{B} = Z^{-1}Y. \quad (8.13)$$

Крок 8. Знаходження параметрів вихідної моделі $\hat{Y} = X\hat{A}$:

$$\hat{A} = S\hat{B} \quad (8.14)$$

МЕТОДИКА ПЕРЕВІРКИ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ НА МУЛЬТИКОЛІНЕАРНІСТЬ

Задача. Відома статистика витрат оборонного бюджету деякої країни за період 1972-1991рр. Вважаючи зв'язок між витратами оборонного бюджету (Y) та факторами – обсягом ВВП (X_1), надходженнями від військових продаж (X_2) обсягом продаж аерокосмічної промисловості (X_3) – лінійним, перевірити модель на мультиколінеарність, використовуючи тест Феррара-Глобера.

Рік	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Y	78,3	74,5	77,8	85,6	89,1	97,5	105,1	117,7	135,9	162,1
X_1	1185,9	1326,4	1434,2	1549,2	1718	1918,3	2163,9	2417,8	2633,1	2937,7

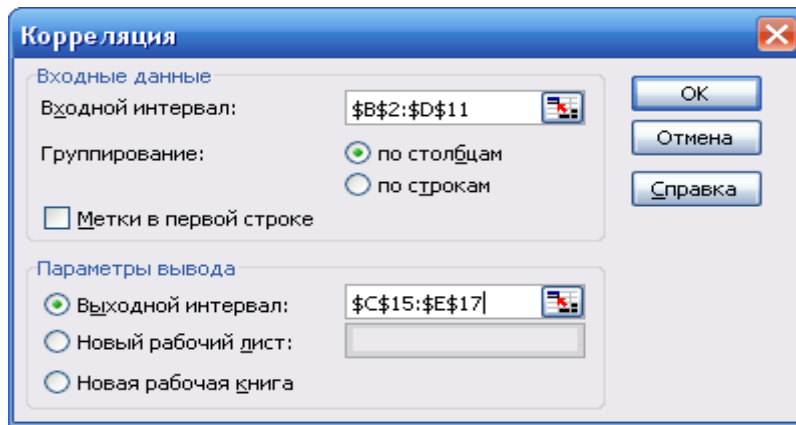


Рис. 8.2 Диалоговое окно **Корреляция**

Для знаходження знаходимо визначник матриці

$$\mathbf{R} = \begin{bmatrix} 1 & 0,702472 & 0,702091 \\ 0,702472 & 1 & 0,999997 \\ 0,702091 & 0,999997 & 1 \end{bmatrix} = 3,37756\text{E}-06$$

Визначник зручно обчислити з допомогою вбудованої математичної функції **МОПРЕД** (блок кореляційної матриці).

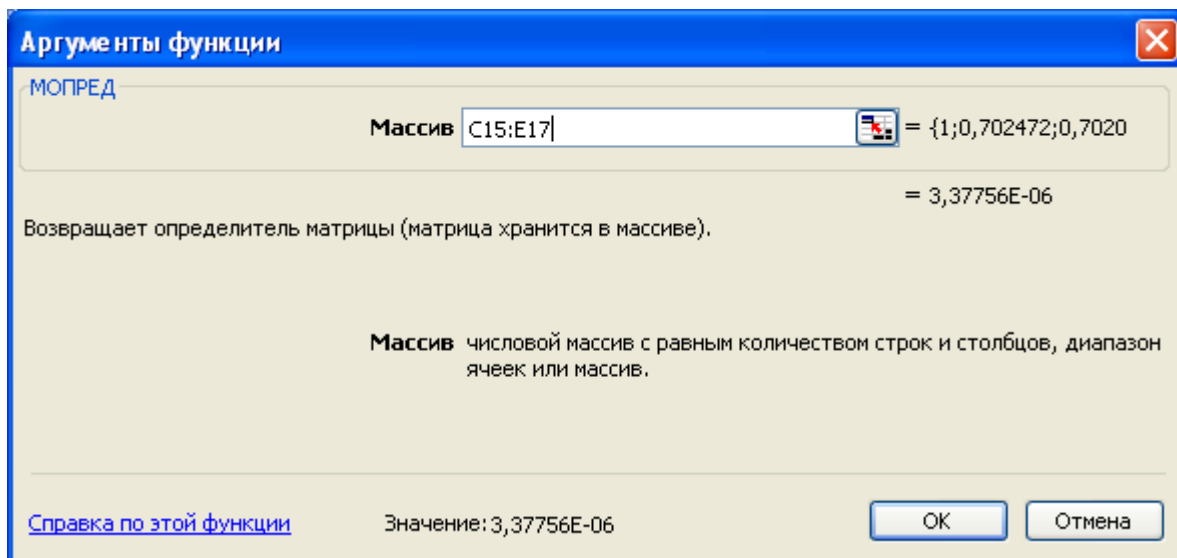


Рис. 8.3 Диалоговое окно пошуку визначника матриці

Розрахункове значення χ^2 знаходимо за формулою:

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

$$\ln \det \mathbf{R} = \ln 3,37756 \cdot 10^{-6} = -12,5984$$

$$\chi^2 = - \left[10 - 1 - \frac{1}{6} \cdot 3 + 5 \right] \cdot (-12,5984) = 90,2882$$

Для довірчої імовірності $p = 0,95$ і числа ступенів вільності $K = \frac{1}{2} m (n - 1) = 3$,

табличне значення $\chi_{\alpha, p}^2(3) = 7,8$. Так як розрахункове значення $\chi^2 > \chi_{\alpha, p}^2$, то

з надійністю 0,95 можна стверджувати, що існує загальна мультиколінеарність.

Знаходимо обернену матрицю до кореляційної з допомогою вбудованої функції МОБР.

$$\mathbf{R}^{-1} = \begin{bmatrix} 2,058123 & 116,595 & 112,1497 \\ -113,663 & 150128,5 & -150048 \\ 112,2178 & -150048 & 149969,9 \end{bmatrix}$$

Використовуючи діагональні елементи матриці \mathbf{R}^{-1} , обчислюємо F-критерії:

$$F_k = z_{ii} - 1 \frac{n - m}{m - 1}$$

де: z_{ii} – елементи матриці \mathbf{R}^{-1} .

$$F_1 = z_{11} - 1 \frac{n - m}{m - 1} = 2,058123 - 1 \frac{10 - 3}{3 - 2} = 3,703;$$

$$F_2 = z_{22} - 1 \frac{n - m}{m - 1} = 150128,5 - 1 \frac{10 - 3}{3 - 2} = 150127,5;$$

$$F_3 = z_{33} - 1 \frac{n - m}{m - 1} = 149969,9 - 1 \frac{10 - 3}{3 - 2} = 149968,9.$$

Для рівня значущості $p = 0,05$ і ступенів вільності $k_1 = m - 1 = 3 - 1 = 2$, $k_2 = n - m = 10 - 3 = 7$ табличне значення F-критерію дорівнює $F_{\text{таб}} = 4,74$. оскільки $F_1 < F_{\text{таб}}$, $F_2 > F_{\text{таб}}$, $F_3 > F_{\text{таб}}$, то один фактор мультиколінеарний з двома іншими.

Визначимо наявність попарної мультиколінеарності. Знайдемо частинні коефіцієнти кореляції, що характеризують тісноту зв'язку між двома змінними за умови, що третя не впливає на цей зв'язок, та t – статистику.

t – статистику пари факторів розраховуємо за формулою:

$$t_{ij} = \frac{r_{ij}^* \sqrt{n-m-1}}{\sqrt{1-r_{ij}^{*2}}}$$

де:

$$r_{ij}^* = \frac{-z_{ij}}{\sqrt{z_{ii}z_{jj}}}, \quad z_{ij}, z_{ii}, z_{jj} - \text{елементи матриці } \mathbf{R}^{-1}.$$

Результати

Кореляційний аналіз				t – статистика			
r_{12}	$\frac{-z_{12}}{\sqrt{z_{11}z_{22}}}$	$\frac{113,663}{\sqrt{2,058123 \cdot 150128,5}}$	0,204	t_{12}	$\frac{r_{12} \sqrt{n-m-1}}{\sqrt{1-r_{12}^2}}$	$\frac{0,204 \cdot \sqrt{10-3-1}}{\sqrt{1-0,204^2}}$	0,564
r_{13}	$\frac{-z_{13}}{\sqrt{z_{11}z_{33}}}$	$\frac{-112,2178}{\sqrt{2,058123 \cdot 149969,9}}$	- 0,202	t_{13}	$\frac{r_{13} \sqrt{n-m-1}}{\sqrt{1-r_{13}^2}}$	$\frac{-0,202 \cdot \sqrt{10-3-1}}{\sqrt{1-0,202^2}}$	-0,557
r_{23}	$\frac{-z_{23}}{\sqrt{z_{22}z_{33}}}$	$\frac{150048}{\sqrt{150128,5 \cdot 149969,9}}$	0,99	t_{23}	$\frac{r_{23} \sqrt{n-m-1}}{\sqrt{1-r_{23}^2}}$	$\frac{0,99 \cdot \sqrt{10-3-1}}{\sqrt{1-0,99^2}}$	206185

Табличне значення t – критерію при рівні значущості $p=0,95$ і ступені вільності $k=n-m-1=10-3-1=7$ дорівнює $t_{\text{табл}}=2,365$. значення $t_{12} < t_{\text{табл}}$, $t_{13} < t_{\text{табл}}$ тобто між факторами x_1 і x_2 та x_1 і x_3 відсутня мультиколінеарність, а між x_2 та x_3 – існує, оскільки $t_{23} > t_{\text{табл}}$. Виключаємо з розгляду один з факторів – x_2 .

Наступним кроком буде знаходження кореляційної матриці між факторами x_1 і x_3 .

$$\mathbf{R}^{-1} = \begin{bmatrix} 1 & 0,702472 \\ 0,702472 & 1 \end{bmatrix} = 0,506533 \quad \mathbf{R}^{-1} = \begin{bmatrix} 1,974205 & -1,38682 \\ -1,38682 & 1,974205 \end{bmatrix}$$

$$\chi^2 = - \left[10 - 1 - \frac{1}{6} \cdot 2 + 5 \right] \cdot \ln 0,51 = 5,101$$

Для довірчої імовірності $p=0,95$ і числа ступенів вільності $K=2$, табличне значення $\chi_{lp}^2(0,95;2) \approx 6$. Так як розрахункове значення $\chi^2 < \chi_{lp}^2$, то з

надійністю 0,95 можна стверджувати, що мультиколінеарність відсутня. На відсутність мультиколінеарності вказує і t – статистика: $r_{12}=0,204$; $t_{12}=0,564$, а значення t – критерію при рівні значущості $p=0,95$ і ступені вільності $k=n-m-1=10-2-1=7$ дорівнює $t_{\text{табл}}=2,365$, значення $t_{12} < t_{\text{табл}}$.

Знаходимо оцінки параметрів економетричної моделі з використанням функції **ЛИНЕЙН**.

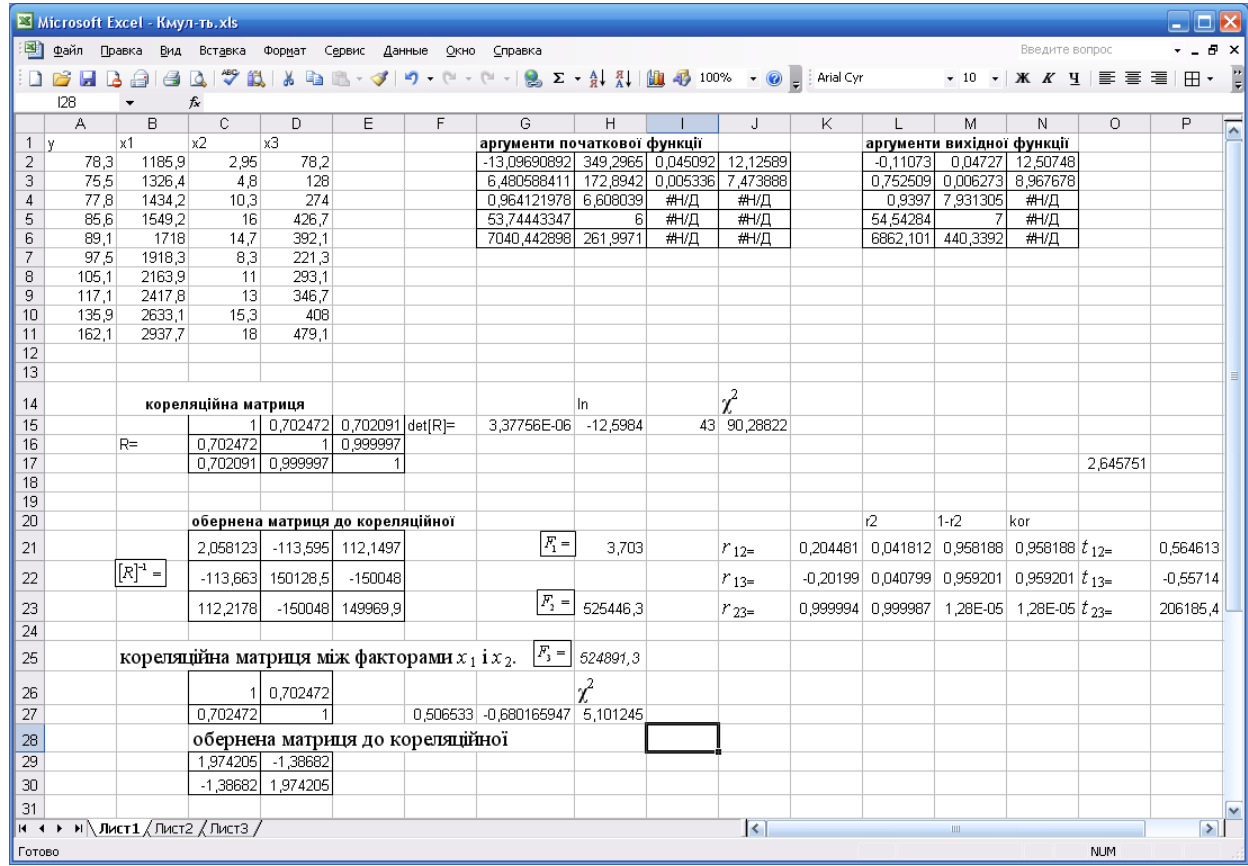


Рис. 8.4 Оцінювання і аналіз економетричної моделі

Модель має вигляд:

$$y = 12,5 + 0,05x_1 - 0,11x_2$$

Перевіримо адекватність прийнятої моделі експериментальним даним. Для цього скористаємось критерієм Фішера. Розрахункове значення критерію $F_{\text{роз}} = 54,54$, а критичне для $k_1 = m - 1 = 2 - 1 = 1$, $k_2 = n - m - 1 = 10 - 3 = 7$ $F_{\text{таб}} = 5,59$. Оскільки

$F_{роз} > F_{таб}$, то з надійністю $p = 0,95$ прийняту економетричну модель можна вважати адекватною експериментальним даним.

ГЕТЕРОСКЕДАСТИЧНІСТЬ

9.1 ПОНЯТТЯ ГЕТЕРОСКЕДАСТИЧНОСТІ. МЕТОДИ ВИЗНАЧЕННЯ ГЕТЕРОСКЕДАСТИЧНОСТІ.

Одним з основних припущень моделі класичної лінійної регресії є припущення про сталість дисперсії випадкової величини u (гомоскедастичність), що можна записати у вигляді $M(u^2) = \sigma_u^2 E$.

У багатьох економетричних дослідженнях виявляється, що припущення про сталу дисперсію випадкової змінної не реалізується. Оскільки випадкова величина u виражає вплив на залежну змінну як помилок в її вимірюванні так і неврахованих факторів, то є підстави для допущення щодо зміни дисперсії u з часом. Помилки вимірювання з плином часу мають тенденцію до накопичення, тому їхня величина збільшується. У такому разі дисперсія u збільшується із зростанням кількості значень X . З іншого боку, техніка вибірки та інші методи збору статистичних даних постійно вдосконалюються і тому похибки вимірювання внаслідок цього можуть зменшуватися. Тоді і величина дисперсії σ_u^2 також зменшуватиметься. Щодо неврахованих змінних, то вони можуть змінюватися в однаковому з X напрямку, викликаючи збільшення відхилення спостережень від лінії регресії.

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

Якщо існує гетероскедастичність залишків, то це приводить до того, що оцінки параметрів моделі методом найменших квадратів будуть незміщеними, обґрунтованими, але неефективними.

Можливість перевірки припущень про наявність гетероскедастичності залежить від природи вихідних даних. Розглянемо методи перевірки гетероскедастичності, в яких робляться різні припущення про залежність між дисперсією випадкової величини і факторами.

Перевірка гетероскедастичності на основі критерію μ . Цей метод застосовується тоді, коли вихідна сукупність спостережень досить велика. Розглянемо відповідний алгоритм.

Крок 1. Вихідні дані залежної змінної Y розбиваються на k груп ($r=1,2,\dots,k$) відповідно до зміни рівня величини Y .

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

$$S_r = \sum_{i=1}^{n_r} (y_{ir} - \bar{Y}_r)^2 \quad (9.1)$$

Крок 3. обчислюється критерій:

$$\mu = -2 \ln \left[\frac{\prod_{r=1}^k \left(\frac{S_r}{n_r} \right)^{\frac{n_r}{2}}}{\left(\frac{1}{n} \sum_{r=1}^k S_r \right)^{\frac{n}{2}}} \right] \quad (9.2)$$

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

Якщо значення μ не менше за табличне значення критерію χ^2 при вибраному наперед рівню довіри і наявному ступеню вільності $k-1$, то спостерігається гетероскедастичність.

Параметричний тест Гольдфельда-Квандта. Даний тест застосовується до великих вибірок. Спостережень має бути хоча б удвічі більше, ніж оцінюваних параметрів. Припускається, що випадкова величина u незалежна і має нормальний розподіл.

Тест Гольдфельда-Квандта складається з декількох кроків.

Крок 1. Упорядкувати спостереження незалежної змінної X в порядку зростання або спадання її значень. У випадку багатьох факторних змінних вибираємо одну з них і впорядковуємо решту відповідно до значень вибраної змінної. Якщо важко відразу визначити змінну X для впорядкування, то почергово проводимо впорядкування за кожною змінною окремо і в кожному випадку застосовуємо тест Гольдфельда-Квандта.

Крок 2. Відкинути c спостережень, які містяться в центрі вектора X . Згідно з експериментальними розрахунками, автори методу знайшли оптимальні співвідношення між параметрами c і n , де n – кількість елементів вектора X_j :

$$c = \frac{4n}{15} \quad (9.3)$$

Залишок $(n-c)$ спостережень ділиться на дві підвибірки однакового розміру $(n-c)/2$, одна з яких включає малі значення X , а інша — великі значення.

Крок 3. Побудувати дві економетричні моделі на основі методу найменших квадратів за двома утвореними сукупностями спостережень. Обчислити суму квадратів залишків за першою і другою моделями:

$\sum u_1^2 = u_1' u_1$ – сума квадратів залишків для першої підвибірки із $(n-c-2m)/2$ ступенями вільності, де u_1 — залишки за першою моделлю, m — загальна кількість оцінюваних параметрів у моделі;

$\sum u_2^2 = u_2' u_2$ – сума квадратів залишків для другої підвибірки із $(n-c-2m)/2$ ступенями вільності, де u_2 – залишки за другою моделлю.

Крок 4. Обчислити критерій

$$F^* = \frac{\sum u_2^2}{\sum u_1^2} \quad (9.4)$$

який в разі виконання гіпотези про гомоскедастичність відповідатиме F - розподілу з $\gamma_1 = \gamma_2 = (n-c-2m)/2$ ступенями вільності, де n – загальна кількість спостережень, c – кількість центральних неврахованих спостережень, m – кількість оцінюваних параметрів у кожній моделі.

Обчислене значення F^* порівнюється з табличним значенням F -критерію для ступенів вільності $k_1 = (n-c-2m)/2$, $k_2 = (n-c-2m)/2$ і вибраного рівня довіри. Якщо $F^* < F_{табл}$, то гетероскедастичність відсутня.

Тест Глейзера. Цей тест для перевірки гетероскедастичності склав Глейзер. Розглянемо його на прикладі простої лінійної регресії.

Крок 1. Знаходимо невідомі параметри простої лінійної регресії методом найменших квадратів та обчислюємо залишки u .

Крок 2. Будуємо регресію абсолютних значень залишків $|u_i|$, що відповідають регресії найменших квадратів, як певну функцію від незалежної змінної x_j . Фактична форма цієї регресії невідома, тому до неї можна підбирати різні форми кривих. Глейзер запропонував такі залежності:

$$\begin{aligned} 1) |u| &= a_0 + a_1 x_j \\ 2) |u| &= a_0 + a_1 x_j^{-1} \\ 3) |u| &= a_0 + a_1 x_j^j \end{aligned} \tag{9.5}$$

тощо.

Рішення про відсутність гетероскедастичності залишків приймається на підставі статистичної значущості коефіцієнтів a_0 і a_1 , використовуючи будь-який стандартний тест перевірки на значущість параметрів. 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 – $a_0 \neq 0, a_1 \neq 0$.

У випадку багатofакторної регресії на першому кроці знаходимо залишки u для регресії, що містить всі фактори. На другому кроці будуємо залежності між абсолютними величинами знайдених залишків та залежною змінною y .

Тест рангової кореляції Спірмена.

Дана процедура є найбільш простою та доступною серед множини аналітичних методів. Її можна використовувати як для малих, так і для великих вибірок. Основу алгоритму даного тесту складає обчислення коефіцієнта рангової кореляції Спірмена:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n} \tag{9.6}$$

де d_i – різниця між рангами, які властиві двом характеристикам i -го об'єкта;
 n – кількість об'єктів, що ранжуються.

Алгоритм: припустимо, що $y_i = b_0 + b_1x_i + u_i$.

Крок 1. Будуємо рівняння регресії $y_i = \hat{b}_0 + \hat{b}_1x$.

Крок 2. Знаходимо відхилення u_i .

Крок 3. Нехтуючи знаком u_i , тобто розглядаємо $|u_i|$, ранжуємо їх та у зростаючому чи спадаючому порядку.

Крок 4. Знаходимо d_i –різницю між рангами, а також d_i^2 .

Крок 5. Знаходимо коефіцієнт r_s .

Крок 6. Перевіряємо значущість отриманого коефіцієнта рангової кореляції за t –критерієм Стьюдента. Розрахункове значення t –критерія знаходимо за формулою:

$$t = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}} \quad (9.7)$$

Де n – кількість спостережень; $k=n-2$ - число ступенів вільності.

Крок 7. Знаходимо табличне значення $t_{табл}$ –критерію.

Крок 8. Якщо $t_{розн} > t_{табл}$, то підтверджується гіпотеза про гетероскедастичність. Якщо дана нерівність не виконується, то має місце гомоскедастичність.

9.2 УЗАГАЛЬНЕНИЙ МЕТОД НАЙМЕНШИХ КВАДРАТІВ.

Коли на базі будь-якого тесту встановлено гетероскедастичність, для її вилучення трансформують початкову модель так, щоб залишки мали постійну дисперсію. Далі невідомі параметри трансформованої моделі розраховують методом найменших квадратів. Трансформація моделі проводиться залежно від специфічної форми гетероскедастичності, тобто від форми залежності між дисперсією σ_u^2 та значеннями незалежних змінних:

$$\sigma_u^2 = f(\mathbf{x}_i)$$

Економетрична модель, якій притаманна гетероскедастичність, є узагальненою моделлю, і для оцінювання її параметрів слід скористатися узагальненим методом найменших квадратів. Розглянемо цей метод.

Нехай задано економетричну модель

$$Y = Xa + u, \quad (9.8)$$

коли $M(u'u) = \sigma_u^2 S$,

де S - симетрична додатно визначена матриця.

Розглянемо такі випадки гетероскедастичності:

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

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

в) $M(u'u) = \sigma_u^2 |x_{ij}|$, тобто дисперсія залишків пропорційна до зміни квадрата залишків за модулем.

Задача полягає в знаходженні оцінок елементів вектора A в моделі. Для цього використовується матриця S , за допомогою якої трансформується вихідна інформація. Ця ідея була покладена в основу узагальненого методу найменших квадратів (методу Ейткена).

Оскільки S – додатно визначена матриця, то вона може бути зображена як добуток PP^T , де матриця P є не виродженою, тобто:

$S = PP^T$, коли $P^{-1}SP^{-1} = E$. Введемо позначення $Y^* = P^{-1}Y$; $X^* = P^{-1}X$; $u^* = P^{-1}u$.

Тоді трансформована модель матиме вигляд:

$$y^* = X^*A + u^*. \quad (9.9)$$

Неважко показати, що отримана модель задовольняє умову незмінності дисперсії залишків. Оцінимо параметри трансформованої моделі на основі методу найменших квадратів.

Тоді отримаємо

$$\hat{A} = (X'X)^{-1}X'Y = (X'S^{-1}X)^{-1}X'S^{-1}Y \quad (9.10)$$

Оцінка параметрів A , яку знайдено за допомогою (9.10), є оцінкою узагальненого методу найменших квадратів (методу Ейткена). Матриця S визначається за формулою

$$S = \begin{pmatrix} \lambda_1^{-1} & 0 & 0 & \dots & 0 \\ 0 & \lambda_2^{-1} & 0 & \dots & 0 \\ 0 & 0 & \lambda_3^{-1} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda_N^{-1} \end{pmatrix} \quad (9.11)$$

Де значення обчислюються залежно від випадку гетероскедастичності:

$$\text{а) } \lambda_i = \frac{1}{X_{ij}}; \text{ б) } \lambda_i = \frac{1}{X_{ij}^2}; \text{ в) } \lambda_i = |u_i|^2$$

Модель узагальненого методу найменших квадратів іноді специфікується у вигляді

$$Y = XA + u,$$

$$M(u) = 0,$$

$$M(u'u) = V$$

де $V = \sigma_u^2 S$ – відома симетрична додатньо визначена матриця. Тоді вираз для оцінки параметрів, згідно з методом Ейткена, запишеться так:

$$\hat{A} = (X'V^{-1}X)^{-1}X'V^{-1}Y \quad (9.12)$$

Дисперсії визначаються за формулами:

$$\sigma_y^2 = \frac{Y'S^{-1}Y}{n-1}; \quad \sigma_A^2 = \frac{\hat{A}X'S^{-1}Y}{m-1}; \quad \sigma_u^2 = \frac{u'S^{-1}u}{n-m}. \quad (9.13)$$

Прогноз:

Коли параметри економетричної моделі оцінюються узагальненим методом найменших квадратів, проблема прогнозування потребує спеціального дослідження.

Розглянемо економетричну модель $Y = XA + u$, коли $M(u) = 0$,

$$M(u'u) = V, \text{ де } V = \sigma_u^2 S$$

Задача зводиться до того, щоб передбачити значення залежної змінної Y_0 для заданого вектора X_0 . Можна записати

$$Y_0 = X_0 A + u_0 \quad (9.14)$$

де u_0 – невідоме значення відхилень у прогностичний період. Нехай для u_0

$$M(u_0) = 0, \quad M(u_0 u_0') = \sigma^2 E$$

$$W = M(u u') = \begin{pmatrix} M(u_1 u_1) \\ M(u_1 u_2) \\ \dots \\ M(u_n u_n) \end{pmatrix}$$

де W – вектор коваріацій поточних і прогностичних значень залишків.

Сформулюємо лінійний прогноз: $Y_{np} = c'Y$,

де c – n -вимірний вектор, який має мінімізувати дисперсію прогнозу:

$$\sigma_{np}^2 = M \left(\left[\begin{matrix} Y_{np} \\ -Y_0 \end{matrix} \right] \left[\begin{matrix} Y_{np} \\ -Y_0 \end{matrix} \right]' \right) \quad (9.15)$$

Розв'язавши задачу на екстремум, знайдемо c . Тоді найкращий лінійний незміщений прогноз за моделлю, оцінки якої знайдені за методом Ейткена, визначатиметься зі співвідношення

$$Y_{np} = X_0 \hat{A} + W V^{-1} u, \quad (9.16)$$

де $u = \left[\begin{matrix} Y_{np} \\ -Y_0 \end{matrix} \right] - X A$ – вектор залишків, який відповідає оцінці параметрів моделі на основі методу найменших квадратів.

МЕТОДИКА ПЕРЕВІРКА ЕКОНОМЕТРИЧНОЇ МОДЕЛІ НА НАЯВНІСТЬ ГЕТЕРОСКЕДАСТИЧНОСТІ

Задача. Перевірити наявність гетероскедастичності при побудові економетричної моделі, яка описуватиме залежність між доходом і рівнем

заощаджень, використовуючи: μ –критерій; тести Гольдфреда-Квандта, Глейзера, Спірмена.

№ спостереження	Дохід (тис.грн.) x	Заощадження (тис.грн.) y
1	8,1	0,36
2	9,7	0,25
3	10,0	0,1
4	10,1	0,22
5	11,0	0,1
6	11,9	0,12
7	12,1	0,41
8	13,8	0,55
9	14,8	0,43
10	15,1	0,59
11	16,9	0,9
12	17,5	0,95
13	18,1	0,82
14	19,5	1,04
15	21,1	1,53
16	23,0	1,94
17	23,9	1,75
18	23,5	1,99

Розв'язання.

Перевіримо наявність гетероскедастичності згідно з μ -критерієм.

Розіб'ємо дані на три групи, по шість спостережень в кожній.

№	x	y	№	x	y	№	x	y
1	8,1	0,36	7	12,1	0,41	13	18,1	0,82
2	9,7	0,25	8	13,8	0,55	14	19,5	1,04
3	10,0	0,1	9	14,8	0,43	15	21,1	1,53
4	10,1	0,22	10	15,1	0,59	16	23,0	1,94
5	11,0	0,1	11	16,9	0,9	17	23,9	1,75
6	11,9	0,12	12	17,5	0,95	18	23,5	1,99
Сума		1,15			3,83			9,07
Середн.		0,192			0,638			1,512

Обчислимо суму квадратів відхилень індивідуальних значень кожної групи від свого середнього значення:

$$\bar{y}_1 = 0,192; \quad \bar{y}_2 = 0,638; \quad \bar{y}_3 = 1,512$$

$$S_1 = \sum_{i=1}^6 (y_{1i} - \bar{y}_1)^2 = 0,054; \quad S_2 = \sum_{i=1}^6 (y_{2i} - \bar{y}_2)^2 = 0,271;$$

$$S_3 = \sum_{i=1}^6 (z_{3i} - \bar{y}_3)^2 = 1,170.$$

Знайдемо суму квадратів відхилень за всіма трьома групами:

$$\sum_{r=1}^3 S_r = S_1 + S_2 + S_3 = 0,054 + 0,271 + 1,170 = 1,495.$$

Обчислимо критерій

$$\mu = -2 \ln \left[\frac{\binom{6}{2}^{-1} \cdot 0,054 \binom{6}{2}^{-1} \cdot 0,271 \binom{6}{2}^{-1} \cdot 1,17 \binom{6}{2}^{-1}}{\frac{1}{18^9} \cdot 1,495^9} \right] = -2 \ln 0,0027 = 11,82.$$

Табличне значення χ^2 при рівні довіри 0,95 і ступеня вільності $k-1=2$ (k – кількість груп) дорівнює $\chi^2_{\text{табл.}}=5,99$. Оскільки $\mu > \chi^2_{\text{табл.}}$, то дисперсія може змінюватися, тобто гетероскедастичність може бути наявна.

Перевіримо наявність гетероскедастичності використовуючи параметричний *тест Гольдфреда-Квандта*.

Впорядкуємо масив x від найменшого до найбільшого значення. У нашій початковій таблиці вже впорядковано. Розбиваємо цей масив на три частини і виключаємо $c = \frac{4}{15}n = \frac{4}{15} \cdot 18 \approx 4$ значень, які містяться всередині масиву.

Поділимо сукупність спостережень на дві частини, кожна з яких містить $(n-c)/2 = (18-4)/2 = 7$ спостережень:

№ спостереження	Дохід (тис.грн.) x	Заощадження (тис.грн.) y
1	8,1	0,36
2	9,7	0,25
3	10,0	0,1
4	10,1	0,22
5	11,0	0,1
6	11,9	0,12
7	12,1	0,41
12	17,5	0,95
13	18,1	0,82
14	19,5	1,04
15	21,1	1,53
16	23,0	1,94
17	23,9	1,75
18	23,5	1,99

Побудуємо економетричну модель за першою сукупністю семи спостережень у вигляді $y = a_0 + a_1x + u$. Використовуючи в MS Excel функцію ЛИНЕЙН, обчислимо невідомі параметри: $a_0=0,398364$ і $a_1=-0,01685$. Тоді $\hat{y} = 0,39836 - 0,01685x$. Обчислимо суму квадратів залишків за першою моделлю: $\sum u_i^2 = 0,092069$.

№	x	y	\hat{y}	u	u^2
1	8,1	0,36	0,2618587	0,0981413	0,0096317
2	9,7	0,25	0,2348947	0,0151053	0,0002282
3	10,0	0,1	0,2298389	-0,129839	0,0168581
4	10,1	0,22	0,2281537	-0,008154	6,648E-05
5	11,0	0,1	0,2129864	-0,112986	0,0127659
6	11,9	0,12	0,1978191	-0,077819	0,0060558
7	12,1	0,41	0,1944486	0,2155514	0,0464624
					0,0920686

Для другої сукупності спостережень отримаємо $\hat{y} = -1,84035 + 0,154122x$.

Сума квадратів залишків за другою моделлю: $\sum u_i^2 = 0,129155$.

№	x	y	\hat{y}	u	u^2
12	17,5	0,95	0,85677516	0,09322484	0,00869087
13	18,1	0,82	0,94924812	-0,1292481	0,01670508
14	19,5	1,04	1,16501837	-0,1250184	0,01562959
15	21,1	1,53	1,41161294	0,11838706	0,0140155
16	23,0	1,94	1,70444398	0,23555602	0,05548664
17	23,9	1,75	1,84315343	-0,0931534	0,00867756
18	23,5	1,99	2,08974799	-0,099748	0,00994966
					0,1291549

Обчислимо критерій

$$F^* = \frac{\sum u_2^2}{\sum u_1^2} = \frac{0,129155}{0,092069} = 1,420811$$

Для рівня значущості $\alpha=0,05$ і ступенів вільності $k_1=k_2=(n-c-2m)/2=(18-4-2\cdot 2)/2=5$, (n -загальна кількість спостережень; c -кількість центральних неврахованих спостережень; m -кількість оцінюваних параметрів у кожній моделі) табличне значення F -критерію дорівнює $F_{\text{табл.}}=5,05$. оскільки $F^* < F_{\text{табл.}}$, то у вихідних даних гетероскедастичність відсутня.

Тест Глейзера.

Будуємо економетричну модель $y = a_0 + a_1x + u$ для початкових даних та знаходимо відхилення u . Отримуємо: $\hat{y} = 1,05538 + 0,11798x$, та

№	x	y	\hat{y}	$u = y - \hat{y}$	$ u $
1	8,1	0,36	-0,0997	0,45972	0,459724
2	9,7	0,25	0,08905	0,16095	0,160952
3	10	0,1	0,12444	-0,0244	0,024442
4	10,1	0,22	0,13624	0,08376	0,08376
5	11	0,1	0,24242	-0,1424	0,142425
6	11,9	0,12	0,34861	-0,2286	0,228609
7	12,1	0,41	0,37221	0,03779	0,037795
8	13,8	0,55	0,57278	-0,0228	0,022775
9	14,8	0,43	0,69076	-0,2608	0,260758
10	15,1	0,59	0,72615	-0,1362	0,136153
11	16,9	0,9	0,93852	-0,0385	0,038521
12	17,5	0,95	1,00931	-0,0593	0,05931
13	18,1	0,82	1,0801	-0,2601	0,2601
14	19,5	1,04	1,24528	-0,2053	0,205275
15	21,1	1,53	1,43405	0,09595	0,095953
16	23	1,94	1,65821	0,28179	0,281786
17	23,9	1,75	1,7644	-0,0144	0,014398
18	23,5	1,99	1,7172	0,2728	0,272795

Розглянемо першу залежність $|u| = a_0 + a_1 x_j$.

Використовуючи в MS Excel функцію ЛИНЕЙН, обчислимо невідомі параметри: $a_0=0,166471$ і $a_1=-0,000753144$. Тоді $|u| = 0,166471 - 0,00075x$.

-0,000753144	0,166471
0,005943324	0,097111
0,001002633	0,125648
0,016058228	16
0,000253518	0,252599

$$S_{a_0} = 0,097111, S_{a_1} = 0,005943324$$

Визначаємо достовірність параметрів:

$$t_{a_0} = \frac{0,166471}{0,097111} = 1,714 \quad t_{a_1} = \frac{0,000753144}{0,005943324} = 0,12672$$

Для рівня значущості $\gamma=0,05$ і ступенів вільності $k=18-1-1=16$ табличне значення критерію Стьюдента $t_{табл.}=2,120$. Оскільки $t_{a_0} < t_{табл.}$, і $t_{a_1} < t_{табл.}$ – гіпотезу про наявність гетероскедастичності відкидаємо.

Тест рангової кореляції Спірмена.

Нами побудована модель $\hat{y} = 1,05538 + 0,11798x$, та знайдені відхилення

u .

Здійснимо ранжування змінних $|u|$ та x . При заповненні сьомого стовпця вибирається найменше число в п'ятому стовпці і поруч з ним записується 1. найменшому з тих чисел, що залишилися, відповідає 2, і т.д.

Знаходимо $d = \text{ранг}|u| - \text{ранг}x$, а також d^2 .

Розрахунки проведемо в таблиці.

№	x	y	\hat{y}	$u = y - \hat{y}$	$ u $	Ранг x	Ранг $ u $	d	d^2
	1	2	3	4	5	6	7	8	9
1	8,1	0,36	-0,0997	0,45972	0,459724	1	18	17	289
2	9,7	0,25	0,08905	0,16095	0,160952	2	11	9	81
3	10	0,1	0,12444	-0,0244	0,024442	3	3	0	0
4	10,1	0,22	0,13624	0,08376	0,08376	4	7	3	9
5	11	0,1	0,24242	-0,1424	0,142425	5	10	5	25
6	11,9	0,12	0,34861	-0,2286	0,228609	6	13	7	49
7	12,1	0,41	0,37221	0,03779	0,037795	7	4	-3	9
8	13,8	0,55	0,57278	-0,0228	0,022775	8	2	-6	36
9	14,8	0,43	0,69076	-0,2608	0,260758	9	15	6	36
10	15,1	0,59	0,72615	-0,1362	0,136153	10	9	-1	1
11	16,9	0,9	0,93852	-0,0385	0,038521	11	5	-6	36
12	17,5	0,95	1,00931	-0,0593	0,05931	12	6	-6	36
13	18,1	0,82	1,0801	-0,2601	0,2601	13	14	1	1
14	19,5	1,04	1,24528	-0,2053	0,205275	14	12	-2	4
15	21,1	1,53	1,43405	0,09595	0,095953	15	8	-7	49
16	23	1,94	1,65821	0,28179	0,281786	16	17	1	1
17	23,9	1,75	1,7644	-0,0144	0,014398	17	1	-16	256
18	23,5	1,99	1,7172	0,2728	0,272795	18	16	-2	4
									922

Обчислимо коефіцієнт рангової кореляції Спірмена:

$$r_s = 1 - \frac{6 \cdot 922}{18^3 - 18} = 0,0485$$

Перевіримо значущість отриманого коефіцієнта рангової кореляції за – критерієм Стюдента. Знаходимо розрахункове значення t -критерія:

$$t = \frac{0,0485 \cdot \sqrt{18-2}}{\sqrt{1-0,0485^2}} = 0,194$$

Знаходимо табличне значення t-критерія для $k=18-2=16$ ступенів свободи і рівня надійності $\gamma=0,05$ ($t_{\text{табл.}}=2,120$.)

Оскільки $t_{\text{розрах.}} < t_{\text{табл.}}$, то гіпотеза про гетероскедастичність не підтверджується.

АВТОКОРЕЛЯЦІЯ

10.1 ПРИЧИНИ ВИНИКНЕННЯ АВТОКОРЕЛЯЦІЇ В ЕКОНОМЕТРИЧНИХ МОДЕЛЯХ. ПЕРЕВІРКА НАЯВНОСТІ АВТОКОРЕЛЯЦІЇ.

В економетричних дослідженнях часто виникають і такі ситуації, коли дисперсія залишків стала, але спостерігається їх коваріація. Це явище називають ще автокореляцією залишків.

Загалом *автокореляція* - це взаємозв'язок послідовних елементів часового або просторового ряду даних.

Розрізняють поняття автокореляції і серійної кореляції. Автокореляція має місце, коли існує залежність між значеннями однієї вибірки із запізненням на лаг. Наприклад, якщо між значеннями однієї вибірки u_1, u_2, \dots, u_t та u_2, u_3, \dots, u_{t+1} є залежність, то маємо справу з автокореляцією, якщо така залежність є між значеннями двох різних вибірок u_1, u_2, \dots, u_t та v_2, v_3, \dots, v_{t+1} то це свідчить про наявність серійної кореляції. Автокореляція може виникнути у зв'язку з інерційністю та циклічністю багатьох економічних процесів. Автокореляція може бути також наслідком помилкової специфікації економетричної моделі.

Автокореляція залишків найчастіше спостерігається тоді, коли економетрична модель будується на основі часових рядів. Якщо існує кореляція між послідовними значеннями деякої незалежної змінної, то спостерігатиметься і кореляція послідовних значень залишків.

При автокореляції небажана оцінка параметрів методом найменших квадратів, бо вона призводить до неефективних оцінок і неможливості застосування статистичних критеріїв t - та F -статистики.

Критерій Дарбіна-Уотсона

Найбільш відомим і поширеним тестом перевірки моделі на наявність кореляції між залишками є критерій Дарбіна-Уотсона:

$$d = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2} \quad (10.1)$$

Теоретично доведено, що значення d критерію Дарбіна-Уотсона знаходиться в проміжку $[0; 4]$.

Далі за таблицею Дарбіна-Уотсона при заданому рівні значущості α , кількості факторів m та кількості спостережень n знаходимо два значення d_1 та d_2 . Якщо розраховане значення d критерію Дарбіна-Уотсона знаходиться в проміжку $(0; d_1)$, то це свідчить про наявність позитивної автокореляції. Якщо значення d потрапляє в проміжок $[d_1; d_2]$ або $[4-d_2; 4-d_1]$, то не можна зробити висновки ні про наявність, ні про відсутність автокореляції. Якщо d належить $(4-d_1; 4)$, то маємо негативну автокореляцію. Якщо d належить $(d_2; 4-d_2)$, то автокореляції немає.

Критерій фон Неймана. Для виявлення автокореляції залишків використовується також критерій фон Неймана:

$$Q = \frac{\frac{1}{n-1} \sum_{t=2}^n (u_t - u_{t-1})^2}{\frac{1}{n} \sum_{t=1}^n u_t^2} \quad (10.2)$$

Звідси $Q = d \frac{n}{n-1}$. При $n \rightarrow \infty$ $Q = d$. Фактичне значення критерію фон

Неймана порівнюється з табличним для вибраного рівня значущості і заданого числа спостережень. Якщо $Q_{\text{факт}} < Q_{\text{табл}}$, то автокореляція існує.

Нециклічний коефіцієнт автокореляції

Нециклічний коефіцієнт автокореляції відображає ступінь взаємозв'язку залишків кожного наступного значення з попереднім, а саме:

1 ряд u_1, u_2, \dots, u_{n-1} ;

2 ряд u_2, u_3, \dots, u_n .

Він обчислюється за формулою:

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

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

Коефіцієнт r^* може набувати значень в інтервалі $(-1;1)$. Від'ємні значення його свідчать про від'ємну автокореляцію, додатні – про додатну. Значення, що містяться в деякій критичній області біля нуля, свідчать про відсутність автокореляції, тобто стверджують нульову гіпотезу про відсутність автокореляції залишків. Оскільки ймовірнісний розподіл r^* встановити важко, то на практиці замість r^* обчислюють циклічний коефіцієнт автокореляції r^0 .

Циклічний коефіцієнт автокореляції

Циклічний коефіцієнт автокореляції відображає ступінь взаємозв'язку рядів:

1 ряд $u_1, u_2, \dots, u_{n-1}, u_n$;

2 ряд $u_2, u_3, \dots, u_n, u_1$.

Циклічний коефіцієнт обчислюється за формулою:

$$r^0 = \frac{\sum_{t=2}^n u_t u_{t-1} + u_n u_1 - \frac{1}{n} \left(\sum_{t=1}^n u_t \right)^2}{\sum_{t=1}^n u_t^2 - \frac{1}{n} \left(\sum_{t=1}^n u_t \right)^2} \quad (10.4)$$

Для досить довгих рядів вплив циклічних членів на величину коефіцієнта r^0 незначний, тому можна вважати, що ймовірнісний розподіл r^* наближається до розподілу r . Якщо останній член ряду дорівнює першому, тобто $u_1 = u_n$ то нециклічний коефіцієнт автокореляції дорівнює циклічному. Очевидно, що коли залишки не містять тренду, то припущення про рівність $u_1 = u_n$ недалеко від реальності і циклічний коефіцієнт автокореляції наближається до нециклічного.

Фактично обчислене значення циклічного коефіцієнта автокореляції порівнюється з табличним (таблиця критерію Дарбіна-Уотсона) для вибраного рівня значущості і довжини ряду n . Якщо $r_{факт} \geq r_{табл}$ то існує автокореляція.

Припускаючи, що $\sum_{t=1}^n u_t \approx \sum_{t=2}^n u_{t-1} \approx 0$, циклічний коефіцієнт автокореляції можна записати у вигляді

$$r = \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2} \quad (10.5)$$

На практиці часто замість (10.5) обчислюють

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

10.2 ОЦІНКА ПАРАМЕТРІВ МОДЕЛІ З АВТОКОРЕЛЬОВАНИМИ ЗАЛИШКАМИ.

Метод Ейткена

Розглянемо економетричну модель

$$y_t = a_0 + a_1 x_t + u_t \quad (10.7)$$

де припускається, що між випадковими величинами є лінійна залежність:

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1 \quad (10.8)$$

ε_t – нормально розподілені випадкові залишки. Тоді, щоб усунути автокореляцію залишків u_t , треба перетворити основну модель так, щоб вона мала залишки ε_t . Оскільки $\varepsilon_t = u_t - \rho u_{t-1}$ то для такого перетворення треба записати модель для попереднього періоду $y_{t-1} = a_0 + a_1 x_{t-1} + u_{t-1}$ помножити її ліву і праву частини на ρ та відняти від моделі для періоду t .

У результаті дістанемо таку економетричну модель:

$$y_t - \rho y_{t-1} = a_0 (1 - \rho) + a_1 (x_t - \rho x_{t-1}) + \varepsilon_t - \rho \varepsilon_{t-1} \quad (10.9)$$

Звідси очевидно, що коли вихідні дані перетворені, а саме $y_t - \rho y_{t-1}$, $x_t - \rho x_{t-1}$, то для оцінювання параметрів можна застосувати метод найменших

квадратів. Причому для перетворення можна використати перші різниці $y_t - y_{t-1}$, і $x_t - x_{t-1}$ коли ρ наближається до одиниці. Зауважимо, що коли $\rho=1$, у перетвореній моделі відсутній вільний член (як виняток може бути ситуація, коли вихідна модель містить часовий тренд). Якщо залишки вихідної моделі характеризувались додатною автокореляцією, використання перших різниць спричинюється до від'ємної автокореляції.

Параметр ρ наближено можна знайти на основі залишків, якщо обчислити циклічний коефіцієнт кореляції r . На практиці, як правило, $\rho \approx r$, але r коригується на величину зміщення:

$$\rho \approx \frac{n}{n-1} \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2} + \frac{m+1}{n} \quad (10.10)$$

Для оцінювання параметрів економетричної моделі, що має автокореляцію залишків, можна застосувати узагальнений метод найменших квадратів (метод Ейткена), який базується на скоригованій вихідній інформації з урахуванням коваріації залишків.

На основі методу Ейткена маємо:

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

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

де \hat{A} – вектор оцінок параметрів економетричної моделі; X – матриця незалежних змінних; X' – матриця, транспонована до матриці X ; S^{-1} матриця, обернена до матриці кореляції залишків; V^{-1} – матриця, обернена до матриці V , де $V = \sigma_u^2 S$, а σ_u^2 – залишкова дисперсія; Y – вектор залежних змінних. Матриця S має вигляд

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

У цій симетричній матриці ρ^s виражає коефіцієнт автокореляції s -го порядку для залишків u_t . Очевидно, що коефіцієнт автокореляції нульового порядку дорівнює 1.

Дисперсія залишків з урахуванням зміщення обчислюється так:

$$\sigma_u^2 = \frac{u'u}{n-m-1} \left(n - \frac{1+\lambda\rho}{1-\lambda\rho} \right) \quad (10.11)$$

де λ можна обчислити методом найменших квадратів за допомогою авторегресійного рівняння $x_t = \lambda x_{t-1} + \varepsilon_t$:

$$\lambda = \frac{\sum_{t=2}^n x_t x_{t-1}}{\sum_{t=2}^n x_{t-1}^2} \quad (10.12)$$

Тут x_t взято як відхилення від свого середнього значення.

Метод перетворення вихідної інформації

Випадок, коли залишки задовольняють авторегресійну модель першого порядку, допускає альтернативний підхід до пошуку оцінок параметрів моделі за допомогою двокрокової процедури:

- перетворення вихідної інформації при застосуванні для цього параметра ρ ;
- застосування методу найменших квадратів для оцінки параметрів на основі перетворених даних.

Для цього треба знайти матрицю перетворення T , щоб модель

$$TY = TXA + Tu \quad (10.13)$$

мала скалярну дисперсійну матрицю

$M(uu'T') = \sigma_u^2 E$. Розглянемо матрицю T_1 розміром $n \times m$:

$$T_1 = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} \quad (10.14)$$

Легко переконатись, що $M(uu'T')^{-1} = \sigma_u^2 E$. А це означає, що можна застосувати метод найменших квадратів до перетворених даних $T_1 Y$ і $T_1 X$.

Іноді для перетворення вихідної інформації використовується матриця T_2 розміром $(n-1) \times n$, яка отримується з матриці T_1 внаслідок викреслювання першого рядка

$$T_2 = \begin{pmatrix} -\rho & 1 & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} \quad (10.15)$$

Застосування методу найменших квадратів до даних $T_1 Y$ і $T_1 X$ дає таку саму оцінку параметрів моделі, як і метод Ейткена, а для даних $T_2 Y$ і $T_2 X$ – забезпечує добру апроксимацію.

У загальному випадку, коли ми не маємо інформації ні про порядок авторегресійної моделі, ні про значення параметрів у ній, а через це не можемо застосувати ні метод Ейткена, ні метод перетворення вихідної інформації, в економетричній літературі пропонуються наближені методи Кочрена-Оркатта і Дарбіна.

Метод Кочрена-Оркатта

Нехай задано економетричну модель

$$y_t = a_0 + a_1 x_t + u_t; \quad (10.16)$$

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

Перетворивши вихідну інформацію за допомогою ρ , отримаємо:

$$y_t - \rho y_{t-1} = a_0 (-\rho) + a_1 (x_t - \rho x_{t-1}) + (u_t - \rho u_{t-1}) + \varepsilon_t \quad (10.17)$$

У цій моделі залишки ε_t мають скалярну дисперсійну матрицю. Сума квадратів залишків на основі (3.26) визначатиметься співвідношенням

$$\sum_{t=2}^n \varepsilon_t^2 = \sum_{t=2}^n \left[y_t - \rho y_{t-1} - a_0 (-\rho) - a_1 (x_t - \rho x_{t-1}) - (u_t - \rho u_{t-1}) \right]^2. \quad (10.18)$$

Безпосередня мінімізація функції (10.18) приводить до системи нелінійних рівнянь, тому аналітичний вираз оцінок параметрів \hat{a}_0 , \hat{a}_1 , ρ дістати важко.

Метод наближеного пошуку параметрів \hat{a}_0 , \hat{a}_1 , ρ , які мінімізують суму квадратів (10.18), дає ітеративний метод, запропонований Кочреном і Оркаттом і названий на їхню честь.

Алгоритм

Крок 1. Довільно вибирають значення параметра ρ , наприклад $\rho=r_1$. Підставивши його в (10.18), обчислюють \hat{a}_0 і \hat{a}_1 .

Крок 2. Поклавши $\hat{a}_0 = \hat{a}_0^{(1)}$ і $\hat{a}_1 = \hat{a}_1^{(1)}$, підставимо їх у (10.18) і обчислимо $\rho=r_2$.

Крок 3. Підставивши $\hat{a}_0^{(1)}$ у співвідношення (10.18) значення $\rho=r_2$, знайдемо $\hat{a}_0^{(2)}$, $\hat{a}_1^{(2)}$.

Крок 4. Використовуємо $a_0 = \hat{a}_0^{(2)}$ і $a_1 = \hat{a}_1^{(2)}$ для мінімізації суми квадратів залишків (10.18) за невідомим параметром $\rho=r_3$. Процедура триває доти, доки наступні значення параметрів \hat{a}_0 , \hat{a}_1 , ρ , не будуть відрізнятись менш як на задану величину.

Цей ітеративний метод, як і інші подібні процедури, має дві проблеми: а) збіжності; б) характеру знайденого мінімуму – локальний чи глобальний.

Проведені дослідження за цими двома проблемами показали, що в результаті застосування методу Кочрена-Оркатта завжди знаходимо глобальний екстремум і алгоритм забезпечує порівняно добру збіжність.

Часто пропонується **альтернативний підхід** до використання цього ітеративного методу. На відміну від попереднього, у ньому подальші ітерації припиняються тоді, коли на основі критерію Дарбіна-Уотсона робиться висновок про відсутність автокореляції залишків.

Алгоритм

Крок 1. Приймається гіпотеза $r_1=0$ і мінімізується на основі методу найменших квадратів сума квадратів:
$$\sum_{t=1}^n (y_t - a_0 - a_1 x_t)^2.$$

Отже, так само й далі обчислюються параметри як і для лінійної моделі.

Крок 2. Знаходяться залишки і на основі критерію Дарбіна-Уотсона перевіряється нульова гіпотеза відносно автокореляції залишків. Якщо гіпотеза відхиляється, то переходять до наступного кроку.

Крок 3. Наданому кроці мінімізується сума квадратів відхилень:

$$\sum_{t=1}^n (y_t - a_0 - a_1 x_t)^2 + r (y_{t-1} - a_0 - a_1 x_{t-1})^2$$

де $a_0^{(1)}$ і $a_1^{(1)}$ – оцінки параметрів, знайдені на першому кроці методу найменших квадратів. У результаті параметр r_2 визначається як коефіцієнт регресії залишків, знайдених методом найменших квадратів, на їх лагові змінні, які стосуються минулого періоду.

Крок 4. Використовуючи значення оцінки параметра r_2 , визначають оцінки параметрів $a_0^{(2)}$ і $a_1^{(2)}$ на основі методу найменших квадратів, який застосовується до перетворених даних $(y_t - r_2 y_{t-1}), (x_t - r_2 x_{t-1})$.

Крок 5. Визначаються залишки і перевіряються на наявність автокореляції. Якщо гіпотеза про наявність автокореляції відхиляється, то ітеративний процес припиняється. У протилежному випадку переходимо до кроку 3, де використовуються знайдені оцінки параметрів $a_0^{(2)}$ і $a_1^{(2)}$.

Коли ітеративний процес припиняється, то виконується перевірка значущості параметрів за допомогою останньої економетричної моделі. У такому разі звичайні формули дадуть обґрунтовані оцінки дисперсій залишків.

Метод Дарбіна.

Дарбін запропонував просту двокрокову процедуру, яка також дає оцінки параметрів, вони асимптотично мають той самий вектор середніх і ту саму матрицю дисперсій, що й оцінки методу найменших квадратів.

Крок 1. Підставимо значення залишків, яке підпорядковане авторегресійній моделі першого порядку $u_t = \rho u_{t-1} + \varepsilon_t$, до економетричної моделі $y_t = a_0 + a_1 x_t + u_t$, Тоді отримаємо $y_t = a_0 + a_1 x_t + \rho u_{t-1} + \varepsilon_t$ де $u_{t-1} = y_{t-1} - a_0 + a_1 x_{t-1}$.

Звідси

$$y_t = a_0 (-\rho) + a_1 x_t + \rho y_{t-1} + a_1 x_t - a_1 \rho x_{t-1} + \varepsilon_t$$

Де ε_t , має скалярну матрицю дисперсій.

Згідно з методом найменших квадратів, визначаються параметри цієї моделі, куди входить і коефіцієнт ρ . У результаті обчислень маємо $\rho=r$.

Крок 2. Значення $\rho=r$ використовується для перетворення змінних $(y_t - r y_{t-1})$, $(x_t - r x_{t-1})$, а метод найменших квадратів застосовується до перетворених даних. Коефіцієнт при $(x_t - r x_{t-1})$ є оцінкою параметра a_1 , а вільний член, поділений на $-r$, оцінює параметр a_0 .

Метод Дарбіна дуже просто поширюється на випадок кількох незалежних змінних і для автокореляції вищих порядків. Нехай задано модель

$$y_t = a_0 + a_1 x_{1t} + a_2 x_{2t} + \dots + a_m x_{mt} + u_t \quad (10.19)$$

Де $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$.

Підставивши значення u_t в (10.19), одержимо:

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + a_1 x_{1t} + a_2 x_{2t} + \dots + a_m x_{mt} - \rho_1 a_1 x_{1,t-1} - \dots - \rho_1 a_m x_{m,t-1} - \rho_2 a_1 x_{1,t-2} - \dots - \rho_2 a_m x_{m,t-2} + \varepsilon_t$$

Застосувавши метод найменших квадратів, обчислимо параметри цієї моделі. Коефіцієнти $\rho_1 = r_1$ і $\rho_2 = r_2$ використаємо для перетворення даних:

$$(y_t - r_1 y_{t-1} - r_2 y_{t-2}), (x_{jt} - r_1 x_{j,t-1} - r_2 x_{j,t-2}), j=1,2,\dots,m; t=1,2,\dots,n.$$

Знову застосуємо метод найменших квадратів для цих перетворених даних і знайдемо оцінки параметрів моделі a_0, a_j ($j= 1,2,\dots, m$).

Описаний щойно ітеративний метод Кочрена-Оркатта і розглянута двокрокова процедура Дарбіна за наявності автокореляції залишків асимптотично ефективніші, ніж метод найменших квадратів.

Гриліхес і Рао за допомогою методу Монте-Карло отримали такі результати:

- Метод найменших квадратів дає менш ефективні оцінки порівняно з іншими методами, якщо сукупність спостережень $n=20$ одиниць, а $|p|>0,3$.
- Якщо $|p|<0,3$, то зниження ефективності оцінок методу найменших квадратів порівняно зі складнішими процедурами невелике.
- Метод Дарбіна забезпечує найкращу оцінку для ширшого кола параметрів порівняно з іншими методами.
- Нелінійний метод оцінювання параметрів не дає відчутних переваг порівняно з двокроковою процедурою Дарбіна.

Прогноз

Нехай маємо модель: $Y = XA + u$, де $M(u) = 0$ і $M(uu') = \sigma_u^2 S = V$, яка побудована для n спостережень.

Використаємо цю модель для визначення прогнозу залежної змінної Y_{n+1} для періоду $n+1$, коли для цього періоду задано незалежну змінну x_{n+1} . Формула дає найкращий незміщений прогноз:

$$y_{n+1} = x_{n+1}A + W'V^{-1}u, \quad (10.20)$$

де A – оцінка параметрів моделі методом Ейткена,

$$u = Y - XA,$$

$$W = \begin{pmatrix} M(u_1 u_{n+1}) \\ M(u_2 u_{n+1}) \\ \dots \\ M(u_n u_{n+1}) \end{pmatrix}$$

Якщо залишки описуються авторегресійною моделлю першого порядку, то з урахуванням рівності $M(u_t u_{t-s}) = \rho^s \sigma_u^2$ можна записати:

$$W = \rho^s \sigma_u^2 \begin{pmatrix} \rho^{n-1} & & & \\ & \rho^{n-2} & & \\ & & \dots & \\ & & & 1 \end{pmatrix}$$

Отже, вектор W можна дістати, помноживши p на останній стовпець матриці V . Але оскільки $VV^{-1}=E$, то добуток $W'V^{-1}$ являє собою останній рядок матриці E , помножений на p .

Звідси $W'V^{-1}u_n = pu_n$.

Формула прогнозу має вигляд:

$$y_{n+1} = x_{n+1}A + pu_n, \quad (10.21)$$

де y_{n+1} – прогнозний рівень залежної змінної; x_{n+1} – прогнозне значення незалежної змінної.

МЕТОДИКА ПЕРЕВІРКИ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ НА НАЯВНІСТЬ АВТОКОРЕЛЯЦІЇ

Задача. Побудувати економетричну модель, яка описує залежність роздрібного товарообігу від доходів населення. Перевірити наявність автокореляції.

у - роздрібний товарообіг	34,1	35,1	36,8	38,1	39,9	41,7	44,9	49,2	54,5
х - дохід	38,2	39,3	40,4	42,4	45,1	46,8	39,9	54,3	61,1

Розв'язання.

Використовуючи в MS Excel функцію **ЛИНЕЙН**, обчислимо невідомі параметри моделі $y = a_0 + a_1x + u$: $a_0 = 0,398364$ і $a_1 = -0,01685$. Тоді $\hat{y} = 0,39836 - 0,01685x$.

Обчислимо розрахункові значення роздрібного товарообігу на основі отриманої моделі і визначимо залишки u_t .

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1			x	y	\hat{y}_t	u_t	u_t^2	$u_t - u_{t-1}$	$(u_t - u_{t-1})^2$	$u_t u_{t-1}$		оцінки параметрів моделі		
2		1	30,2	24,1	24,74396	-0,64396	0,414682							
3		1	31,3	25,1	25,89251	-0,79251	0,628078	-0,14856	0,022069	0,510346		1,044142	-6,78913	
4		1	32,4	26,8	27,04107	-0,24107	0,058115	0,551444	0,30409	0,191052		0,039364	1,466808	
5		1	33,4	28,1	28,08521	0,014788	0,000219	0,255858	0,065463	-0,00356		0,990149	0,7233	
6		1	34,1	29,9	28,81611	1,083888	1,174814	1,069101	1,142976	0,016028		703,5804	7	
7		1	35,8	31,7	30,59115	1,108847	1,229541	0,024959	0,000623	1,201866		368,0868	3,662136	
8		1	39,9	34,9	34,87214	0,027864	0,000776	-1,08098	1,168523	0,030897				
9		1	44,3	39,2	39,46636	-0,26636	0,070948	-0,29422	0,086568	-0,00742				
10		1	49,4	44,5	44,79148	-0,29148	0,084963	-0,02512	0,000631	0,07764				
11			330,8	284,3	259,556		3,662136		2,790944	2,016842				
12														
13				d=		0,762108								
14														
15														
16														

Рис.10.1 Результати досліджень

Обчислимо величину критерію Дарбіна-Уотсона, отримаємо:

$$d = \frac{2,7909}{3,6621} = 0,7621$$

За таблицею Дарбіна-Уотсона при рівні значущості $\alpha=0,05$, кількості факторів $m=1$ та кількості спостережень $n=9$ знаходимо значення $d_1=0,824$ та $d_2=1,32$. оскільки $d < d_1$, то можна стверджувати, що залишки u_t мають додатну автокореляцію.

Для оцінювання параметрів економетричної моделі з автокорельованими залишками застосуємо *узагальнений метод найменших квадратів* (метод Ейткена). Згідно з методом Ейткена, маємо:

$$\hat{A} = (X^T S^{-1} X)^{-1} X^T S^{-1} Y$$

де матриця S має вигляд:

$$S = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 & \rho^7 & \rho^8 \\ \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 & \rho^7 \\ \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 & \rho^6 \\ \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 & \rho^5 \\ \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 & \rho^4 \\ \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 & \rho^3 \\ \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho & \rho^2 \\ \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 & \rho \\ \rho^8 & \rho^7 & \rho^6 & \rho^5 & \rho^4 & \rho^3 & \rho^2 & \rho & 1 \end{pmatrix}$$

Припустимо, що залишки описуються авто кореляційною моделлю першого порядку $u_t = \rho u_{t-1} + \varepsilon_t$, де

$$\rho \approx \frac{n \sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2} + \frac{m+1}{n} = \frac{9}{8} \cdot \frac{2,0168}{3,6621} + \frac{2}{9} = 0,8418$$

Отже матриця S має вигляд:

$$S = \begin{pmatrix} 1 & 0,824 & 0,709 & 0,597 & 0,502 & 0,423 & 0,356 & 0,3 & 0,252 \\ 0,824 & 1 & 0,824 & 0,709 & 0,597 & 0,502 & 0,423 & 0,356 & 0,3 \\ 0,709 & 0,824 & 1 & 0,824 & 0,709 & 0,597 & 0,502 & 0,423 & 0,356 \\ 0,597 & 0,709 & 0,824 & 1 & 0,824 & 0,709 & 0,597 & 0,502 & 0,423 \\ 0,502 & 0,597 & 0,709 & 0,824 & 1 & 0,824 & 0,709 & 0,597 & 0,502 \\ 0,423 & 0,502 & 0,597 & 0,709 & 0,824 & 1 & 0,824 & 0,709 & 0,597 \\ 0,356 & 0,423 & 0,502 & 0,597 & 0,709 & 0,824 & 1 & 0,824 & 0,709 \\ 0,3 & 0,356 & 0,423 & 0,502 & 0,597 & 0,709 & 0,824 & 1 & 0,824 \\ 0,252 & 0,3 & 0,356 & 0,423 & 0,502 & 0,597 & 0,709 & 0,824 & 1 \end{pmatrix}$$

Обчисливши значення параметрів моделі методом Ейткена, отримаємо:

$$\hat{A} = \begin{pmatrix} X^T S^{-1} X \end{pmatrix}^{-1} X^T S^{-1} Y = \begin{pmatrix} -5,627 \\ 1,007 \end{pmatrix}, \text{ а модель має вигляд: } \hat{y} = -5,627 + 1,007x.$$

Обчислимо для одержаної моделі критерій Дарбіна-Уотсона: $d=0,575$.

Оскільки $d < d_1$, то можна стверджувати, що залишки u_t мають додатну

автокореляцію. Отже припущення, що залишки описуються авто кореляційною моделлю першого порядку, є не правильним. Для моделей, в яких залишки описуються авто регресійною схемою вищого порядку, слід оцінювати параметри наближеними методами.

Критерій фон Неймана.

$$Q_{\text{факт}} = 0,7621 \cdot \frac{9}{8} = 0,857$$

Знаходимо табличне значення данного критерію $Q_{\text{табл}}=1,256$ при заданому рівні $\gamma=0,05$ та числі спостережень $n=9$.

Оскільки $Q_{\text{факт}} < Q_{\text{табл}}$, то існує додатна автокореляція.

Циклічний коефіцієнт автокореляції

Знайдемо коефіцієнт автокореляції

$$r = \frac{9}{9-8} \cdot \frac{2,016842}{3,662136} = 0,62$$

Знайдемо табличне значення коефіцієнта автокореляції $r_{\text{табл.}}=0,366$ при $\gamma=0,05$, та числі спостережень $n=9$. Оскільки $r_{\text{факт.}} < r_{\text{табл.}}$, то існує автокореляція.

АВТОРЕГРЕСІЙНІ І ДИСТРИБУТИВНО-ЛАГОВІ МОДЕЛІ.

11.1 ПРИРОДА АВТОРЕГРЕСИВНИХ МОДЕЛЕЙ.

При розгляді зв'язків економічних явищ часто доводиться на даний момент часу враховувати рівень економічного явища за попередній період. Якщо значення показника y змінюється через невеликий проміжок часу після зміни значення фактора x , то в рівнянні регресії повинні бути присутні відповідні фактори з відповідним *лагом* (запізненням). Модель, яка включає поточні і попередні значення незалежних змінних, називається *розподілено-лаговою* моделлю.

Нехай фактор x впливає на показник y із запізненням τ -періодів. Тоді рівняння регресії у випадку лінійної залежності матиме вигляд

$$y_t = a_0 + a_1 x_{t-\tau} + u_t.$$

Якщо характер впливу залишається незмінним у часі, то значення показника y може бути виражене через τ попередніх значень факторної змінної

$$y_t = a + a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \dots + a_\tau x_{t-\tau} + u_t \quad (11.1)$$

де a – параметри моделі при лагових змінних; x_{t-j} – незалежна лагова змінна; τ – часовий лаг; u_t – залишки, що розподілені нормально, тобто мають нульове математичне сподівання і сталу дисперсію.

Якщо в (11.1) τ – обмежене число, то така модель називається *скінченною розподілено-лаговою моделлю*. Якщо τ – необмежене число, то маємо нескінченну модель, для якої повинні виконуватися умови:

а) $a_j \geq 0, j=1,2,3,\dots;$

б) $\sum_{j=0}^{\infty} a_j = w < \infty;$

в) $\sum_{j=0}^{\infty} \frac{a_j}{w} = 1$

При побудові скінченної розподілено-лагової моделі виникає проблема з'ясування довжини максимального лагу, який повинен бути включений в

регресію. Це можна вяснити способом послідовного оцінювання: спочатку побудувати регресію y_t за x_t та оцінити невідомі параметри, потім – регресію y_t за x_t і за x_{t-1} потім за x_t , x_{t-1} і x_{t-2} і так далі. Ця послідовна процедура припиняється, коли параметри при лагових змінних x_t починають бути статистично незначимими або коефіцієнт хоча б однієї змінної змінює свій знак. Така методика призводить до двох статистичних труднощів. Одна трудність пов'язана з оцінкою параметрів, яким відповідає мале число ступенів вільності, друга трудність – з високою кореляцією між різними лаговими значеннями фактора. У зв'язку з цим розроблені методи оцінки параметрів регресії з лаговими значеннями факторів.

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 узагальненої моделі розподіленого лагу, яка містить не лише лагові змінні, а й інші фактори – змінні z_1, z_2, \dots, z_m , значення яких характеризують поточні умови функціонування економічних систем у період t .

Теоретично побудову розподілено-лагової моделі 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 тісний зв'язок кожного елемента вектора залежної змінної y_t з елементом вектора незалежної x_t , зсунутим один відносно одного на часовий лаг τ :

$$r_{(\tau)} = \frac{\left(\sum_{t=1}^{n-\tau} y_t x_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} x_{t+\tau} \right)}{\sqrt{\left[\sum_{t=1}^{n-\tau} y_t^2 - \left(\sum_{t=1}^{n-\tau} y_t \right)^2 \right] \left[\sum_{t=1}^{n-\tau} x_t^2 - \left(\sum_{t=1}^{n-\tau} x_t \right)^2 \right]}} \quad (11.2)$$

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 $(-1;1)$. Найбільше значення $r_{(\tau)}$ за модулем (найближче до одиниці) визначає зрушення, або часовий лаг. Якщо серед множини значень $r_{(\tau)}$ є кілька таких, що 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.

Метод Койка

Метод Койка використовується в тих випадках, коли факторна змінна має нескінченну лагову структуру і лагові параметри регресії володіють однаковим законом зміни. Зокрема, Койк припустив, що вони змінюються в геометричній прогресії: $a_k = a_0 \lambda^k$, $k=1,2,\dots$, де λ такі, що $0 < \lambda < 1$ – темп зменшення розподіленого лагу, а $(1-\lambda)$ – швидкість пристосування. Тоді модель можна записати у вигляді:

$$y_t = a + a_0 x_t + a_0 \lambda x_{t-1} + a_0 \lambda^2 x_{t-2} + \dots + u_t. \quad (11.3)$$

Далі, згідно з підходом Койка, вводиться затримка на один період:

$$y_{t-1} = a + a_0 x_{t-1} + a_0 \lambda x_{t-2} + a_0 \lambda^2 x_{t-3} + \dots + u_{t-1} \quad (11.4)$$

Віднявши від (11.3) величину (11.4) помножену λ отримаємо

$$y_t = a(1 - \lambda) a_0 x_t + \lambda y_{t-1} + u_t - \lambda u_{t-1}. \quad (11.5)$$

Тепер для оцінки невідомих параметрів використовуємо метод найменших квадратів.

Перетворенням Койка розподілено-лагова модель (11.1) зведена до авторегресивної моделі (11.4), тобто моделі, в яку включене одне або більше попередніх значень залежної змінної y .

Поява у моделі (11.4) величини y_{t-1} може спричинити ряд статистичних проблем: y_{t-1} , як y_t – стохастична; це означає, що в модель вводимо стохастичну змінну. Для застосування методу найменших квадратів необхідно, щоб стохастична змінна y_{t-1} була розподілена незалежно від випадкової величини u_t .

Можна показати, що якщо залишки u_t , серійно некорельовані, то $(u_t - \lambda u_{t-1})$ серійно корельовані. Отже, можемо зіткнутися з проблемою серійної кореляції щодо стохастичної змінної y_{t-1} .

Модель часткових пристосувань

Існують і інші моделі, у правій частині яких фігурують лагові значення залежної змінної. Добре відомими моделями такого типу є модель часткових пристосувань і модель адаптивних очікувань.

Коли відсутнє повне уявлення про об'єкт, його інерційність, тобто швидкість зміни його показників з часом, то застосовується метод часткових пристосувань.

Розглянемо модель вигляду

$$y_t^* = \alpha + \beta x_t + u_t \quad (11.6)$$

У цьому рівнянні y_t^* розглядається як оптимальне значення y_t , яке відповідає x_t . Так, наприклад, y_t^* – оптимальна в довгостроковому періоді кількість капіталу, необхідна для того, щоб виробляти певну кількість продукції при даному рівні технології, відсотковій ставці тощо.

Нехай величина випуску x_t різко змінюється (збільшується чи зменшується). При цьому рівень капіталу y_t може не змінитись адекватно зміні випуску з різних причин: певна інерційність, недостатня інформація, договірні умови тощо. Оскільки бажаний рівень не можна спостерігати явно, Нерлоу запропонував гіпотезу:

$$y_t - y_{t-1} = \gamma (y_t^* - y_{t-1}) \quad 0 \leq \gamma \leq 1 \quad (11.7)$$

де γ – коефіцієнт пристосувань; $y_t - y_{t-1}$ – фактична змінна, $y_t^* - y_{t-1}$ – бажана змінна.

Оскільки $y_t - y_{t-1}$ зміна запасу капіталу між двома періодами є інвестуванням, то (11.6) можна переписати у вигляді:

$$I_t = \gamma (y_t^* - y_{t-1}) \quad (11.8)$$

де I_t – інвестування за період t .

Рівняння (11.7) показує, що фактична зміна в запасі капіталу (інвестиціях) у будь-який період часу t є певною часткою γ від бажаної зміни за цей період.

Значення величини γ міститься між двома екстремальними значеннями $\gamma=1$ (фактична зміна обсягу капіталу дорівнює бажаній) і $\gamma=0$ (рівень капіталу такий самий, як і в попередній період). Насправді перехід до бажаного рівня капіталу найчастіше буває неповним через інертність, контрактівні зобов'язання тощо. Звідси зрозуміла назва моделі, що описує подібні явища – модель часткових пристосувань.

Якщо підставити (11.6) в (11.5), то одержимо:

$$y_t = \alpha\gamma + \beta\gamma x_t + (1-\gamma)y_{t-1} + \eta_t \quad (11.9)$$

Запис (11.9) має назву *модель часткових пристосувань*.

Модель (11.5) відображає довгостроковий або рівноважний попит на обсяг необхідного капіталу, модель у вигляді (11.8) представляє *короткостроковий* попит на капітал, в короткостроковому періоді наявний обсяг капіталу не обов'язково відповідає обсягу капіталу в довгостроковому періоді.

Модель адаптивних очікувань

Модель адаптивних очікувань, характеризує зв'язок змінної y з очікуваним рівнем змінної x . Позначимо x_t^* , - очікуване значення x_t , яке сформоване в поточний момент часу. Розглянемо модель

$$y_t = \alpha + \beta x_t^* + u_t \quad (11.10)$$

де y – попит на гроші; x^* – точка рівноваги, оптимум, очікувана довгострокова або звичайна відсоткова ставка; u – випадкова величина.

Рівняння (11.10) показує, що попит на гроші є функцією від очікуваної відсоткової ставки.

Оскільки очікувану змінну x^* не можна спостерігати безпосередньо, то вводиться гіпотеза формування очікування:

$$x_t^* - x_{t-1}^* = \delta (x_t - x_{t-1}^*), \quad 0 \leq \delta \leq 1 \quad (11.11)$$

де δ – коефіцієнт очікування.

Гіпотеза (11.11) передбачає, що "сили, задіяні в економіці, пристосовують свої очікування до попереднього досвіду, і зокрема вчать на своїх попередніх помилках", тобто (11.11) стверджує, що очікування кожного періоду коригується на частку δ від розриву між поточним значенням змінної та її попереднім очікуваним значенням. Зокрема для розглядуваного випадку це означає, що очікування щодо відсоткової ставки кожного проміжку часу коригуються на частку δ від розходження між відсотковою ставкою у поточному періоді і її очікуваним значенням у попередньому періоді.

Запишемо (11.11) у вигляді:

$$x_t^* = \delta x_t + (1 - \delta) x_{t-1}^* \quad (11.12)$$

Це показує, що очікуване значення відсоткової ставки в момент часу t є зваженим середнім фактичного значення відсоткової ставки в момент t і його очікуваного значення в попередній період. Якщо $\delta=1$, то очікування справджуються негайно і повністю, тобто в поточний період часу. Якщо $\delta=0$, очікування незмінні, тобто "умови, що є сьогодні, зберігатимуться й у наступні періоди".

Підставляючи (11.12) у (11.11), маємо:

$$y_t = \alpha + \beta \delta x_t + \beta (-\delta \bar{x}_{t-1}^* + u_t) \quad (11.13)$$

Затримаємо (11.10) на один період, помноживши його на $(1-\delta)$ і віднімемо добуток від (11.11). В результаті отримаємо:

$$y_t = \delta \alpha + \delta \beta x_t + (-\delta \bar{y}_{t-1} + u_t) - (-\delta \bar{y}_{t-1}) \quad (11.14)$$

Модель (11.14) має назву *модель адаптивних очікувань*.

Незважаючи на зовнішню схожість, модель адаптивних очікувань і модель часткових пристосувань відрізняються. Перша базується на невизначеності (майбутні ціни, ставки відсотків тощо), а остання залежить від технічних і інституціональних обмежень, інерції, вартості обміну тощо.

11.2 МЕТОДИ ОЦІНЮВАННЯ ПАРАМЕТРІВ ЛАГОВОЇ МОДЕЛІ.

Коли схема формування вагових коефіцієнтів задовольняє припущення Койка, модель часткового коригування або модель адаптивних сподівань, то у правій частині економетричної моделі виникає і лагове значення залежної змінної Y . Це зумовлює певні проблеми при оцінюванні параметрів такої моделі. Розглянемо ці проблеми.

Нехай економетрична модель має вигляд

$$y_t = a_1 y_{t-1} + a_2 x_t + v_t \quad (11.15)$$

Як ми вже переконалися, методи оцінювання параметрів моделі залежать від гіпотез, які будуть прийняті щодо залишків v_t .

Гіпотеза 1. Залишки – випадкові величини і розподіляються нормально.

Гіпотеза 2. Залишки виражені через параметр λ , тобто:

$$v_t = u_t - \lambda u_{t-1}, \quad 0 \leq \lambda \leq 1:$$

а) v_t – нормально розподілені залишки;

б) $u_t = \rho u_{t-1} + \varepsilon_t$, $|\rho| < 1$, ε_t – нормально розподілені залишки.

Гіпотеза 3. Залишки $v_t = \rho v_{t-1} + \varepsilon_t$, $|\rho| < 1$, ε_t – нормально розподілені залишки.

Перша гіпотеза найпростіша, а тому єдина складність в оцінюванні параметрів моделі пов'язується з наявністю в правій частині лагової змінної y_{t-1} .

Друга гіпотеза відповідає схемі Койка і моделі адаптивних сподівань. При цьому розглядаються два варіанти:

- а) залишки u_t незалежні;
- б) залишки u_t описуються авторегресійною моделлю першого порядку.

Третя гіпотеза не пов'язується ні зі схемою Койка, ні з моделлю адаптивних сподівань. Згідно з цією гіпотезою, величина залишків v_t описується авторегресійною схемою першого порядку (найпростіший випадок).

Розглянемо особливості оцінки параметрів моделі при різних гіпотезах відносно залишків.

Гіпотеза 1а. Оскільки залишки не корельовані між собою, то оцінка параметрів може бути зроблена методом найменших квадратів. Цей метод дасть зміщену оцінку, бо залишки не можна вважати незалежними від лагової змінної y_{t-1} .

Щоб знайти величину зміщення розглянемо таку модель:

$$y_t = ay_{t-1} + u_t$$

де $|a| < 1$ і послідовні значення u_t некорельовані. Для такої моделі оцінка параметрів a на основі методу найменших квадратів дає

$$\hat{a} = \frac{\sum_{t=2}^n y_t y_{t-1}}{\sum_{t=1}^n y_t^2} \quad (11.16)$$

В економетричній літературі доведено, що в такому разі зміщення параметра a

$$M \left(\frac{\hat{a} - a}{n} \right) \approx \frac{-2a}{n} \quad (11.16')$$

Альтернативною оцінкою параметра a може слугувати коефіцієнт автокореляції першого порядку для y , тобто

$$r = \frac{\frac{1}{1-n} \sum_{t=2}^n y_t y_{t-1} - \frac{1}{n} \left(\sum_{t=2}^n y_t \right) \left(\sum_{t=2}^n y_{t-1} \right)}{\frac{1}{n} \sum_{t=1}^n y_t^2 - \frac{1}{n^2} \left(\sum_{t=2}^n y_t \right)^2} \quad (11.17)$$

Зміщення тоді визначатиметься так:

$$M(\hat{r}) - a \approx n^{-1} (3a) \quad (11.18)$$

тобто обидві оцінки мають тенденцію до завищення параметра a , причому рівень зміщення параметра r більший, ніж параметра a .

За допомогою методу Монте-Карло було досліджено оцінки параметрів a у моделі $y_t = ay_{t-1} + u_t$ із застосуванням таких прийомів:

- а) визначення параметра r ;
- б) використання параметра r , скоригованого на величину зміщення в (11.18);
- в) застосування методу найменших квадратів.

При цьому виявилось, що оцінка параметрів a на основі методу найменших квадратів має найменшу середньоквадратичну помилку.

Гіпотеза 2а. Коли в економетричній моделі серед незалежних змінних є лагове значення залежної змінної, застосування критерію Дарбіна-Уотсона для виявлення серійної кореляції залишків призводить до зміщення його оцінок. Тому Дарбін розробив методи перевірки автокореляції залишків, які можна застосувати і для моделей з лаговими змінними, що побудовані на базі великих сукупностей спостережень (n). Цей критерій має вигляд:

$$h = \hat{\rho} \cdot \sqrt{\frac{n}{1 - n \cdot D(\hat{a}_1)}} \quad (11.19)$$

де $\hat{\rho}$ – оцінка параметра в автокореляційній моделі першого порядку

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

$D(\hat{a}_1)$ – оцінка вибіркової дисперсії параметра a_1 який знаходиться при лаговій змінній y_{t-1} . Оцінку параметра ρ можна дістати з такого співвідношення:

$$\hat{\rho} \approx r = \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=2}^n u_{t-1}^2} \quad (11.20)$$

Для перевірки нульової гіпотези, обчислені величини h порівнюються з критичними значеннями (односторонній критерій) нормального розподілу (χ^2) при вибраному рівні значущості. З формули для цього критерію видно, що коли $D(a_1) \geq 1$, то його використовувати не можна. Для критерію h виконується така сама перевірка, як і в разі стандартного нормального відхилення, тобто коли при рівні значущості $\alpha=0,05$ $h > 1,645$, то гіпотеза про нульову автокореляцію відхиляється.

Розглянемо особливості оцінки параметрів, коли залишки мають форму для моделі адаптивних сподівань і схеми Койка, тобто $v_t = u_t - \lambda u_{t-1}$, $0 \leq \lambda \leq 1$, v_t – нормально розподілені залишки.

Тоді математичне сподівання залишків дорівнюватиме нулю $M \epsilon_t = 0$ для всіх t , а дисперсія матиме вигляд: $D \epsilon_t = \sigma_u^2 (1 + \lambda^2)$ для всіх t . А це означає, що для оцінювання параметрів моделі в даному разі можна використати узагальнений метод найменших квадратів:

$$A = (X' V^{-1} X)^{-1} X' V^{-1} Y$$

Оскільки дисперсія залишків пропорційна до величини $1 + \lambda^2$, тоді як коваріація для $\tau = \pm 1$ дорівнює $-\lambda \sigma_u^2$, а для $|\tau| \geq 2$ вона дорівнює нулю, то матриця V має вигляд

$$V = \sigma_u^2 \begin{pmatrix} 1 + \lambda^2 & -\lambda & 0 & \dots & 0 \\ -\lambda & 1 + \lambda^2 & -\lambda & \dots & 0 \\ \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & 0 & \dots & 1 + \lambda^2 \end{pmatrix} \quad (11.21)$$

а матриця X містить лагову змінну u_{t-1} .

Врахувавши, що параметр a при u_{t-1} дорівнює λ , дійдемо висновку: коли λ відома, модель спрощується і має вигляд

$$y_t - \lambda y_{t-1} = a_0 + a_2 x_t + v_t \quad (11.22)$$

Тоді матриця X складатиметься лише з двох стовпців, перший з яких утворюється одиницями, а другий – спостереженнями над X . Вектор Y в такому разі складається з перетворених даних $y_t - \lambda y_{t-1}$. Як бачимо, проблема оцінювання параметрів у цьому випадку зводиться до знаходження параметра λ .

Зельнер і Гейсел запропонували вибирати значення параметра з інтервалу $0 < \lambda < 1$. Це означає, що довільно вибирається параметр λ , на основі якого формується матриця V (11.21). Ця матриця в свою чергу дає змогу знайти оцінки параметрів узагальненим методом найменших квадратів. Вибирається те значення параметра λ , яке дає змогу мінімізувати суму квадратів залишків $v'V^{-1}v$ а звідси і стандартну помилку параметрів. Тобто використовується поступовий перебір значень λ на певному інтервалі, доки не буде знайдено той параметр, який забезпечує найкращий розв'язок.

Гіпотеза 2 б. Згідно з цією гіпотезою, залишки мають вигляд:

$$v_t = u_t - \lambda u_{t-1}, \quad 0 \leq \lambda \leq 1$$

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1, \quad \varepsilon_t - \text{нормально розподілені залишки.}$$

Зельнер і Гейсел запропонували процедуру пошуку параметрів λ для цієї моделі.

Запишемо економетричну модель (11.22) у вигляді

$$y_t = a_0 + \lambda y_{t-1} + a_2 x_t + u_t - \lambda u_{t-1}$$

Визначимо $w_t = y_t - u_t$. Отже,

$$w_t - \rho w_{t-1} = y_t - y_{t-1} - u_t - \rho u_{t-1}$$

Перепишемо це рівняння так:

$$w_t \ominus \rho w_{t-1} \stackrel{\text{def}}{=} y_t \ominus u_t \ominus \rho w_{t-1}$$

де $w_t \ominus \rho w_{t-1} \stackrel{\text{def}}{=} w_t - \rho w_{t-1}$.

Оскільки

$$y_t \ominus \rho w_{t-1} \stackrel{\text{def}}{=} a_0 \ominus \rho w_{t-1} \ominus \lambda y_{t-1} \ominus a_2 x_t \ominus u_t \ominus \rho w_{t-1}$$

то

$$w_t \stackrel{\sim}{=} a_0 (-\rho) \mp \lambda w_{t-1} \mp a_2 x_t \stackrel{\sim}{\cdot}$$

Шляхом послідовних підстановок можна записати $y_t \stackrel{\sim}{\cdot}$

$$y_t \stackrel{\sim}{=} \lambda w_0 \mp a_0 (-\rho) \left(1 + \lambda + \lambda^2 + \lambda^3 + \dots + \lambda^{t-1} \right) \mp a_2 \left[x_t \mp \lambda x_{t-1} \mp \dots + \lambda^{t-1} x_1 \right] \mp \varepsilon_t \quad (11.23)$$

Якщо λ і ρ відомі, то (11.23) визначає Y лінійну функцію від трьох невідомих параметрів $w_0 \stackrel{\sim}{\cdot}$, $a_0 (-\rho) \stackrel{\sim}{\cdot}$, і a_2 плюс випадкове відхилення. Тоді ці параметри можна відшукати на основі методу найменших квадратів. Матриця вихідних даних $X(\rho)$ матиме вигляд:

$$X \stackrel{\sim}{=} \begin{pmatrix} \lambda & 1 & x_1 \stackrel{\sim}{\cdot} \\ \lambda^2 & 1 + \lambda & x_2 \stackrel{\sim}{\cdot} \mp \lambda x_1 \stackrel{\sim}{\cdot} \\ \lambda^2 & 1 + \lambda + \lambda^2 & x_3 \stackrel{\sim}{\cdot} \mp \lambda x_2 \stackrel{\sim}{\cdot} \mp \lambda^2 x_1 \stackrel{\sim}{\cdot} \\ \dots & \dots & \dots \\ \lambda^n & 1 + \lambda + \lambda^2 \dots + \lambda^{n-1} & x_n \stackrel{\sim}{\cdot} \mp \lambda x_{n-1} \stackrel{\sim}{\cdot} \mp \dots + \lambda^{n-1} x_1 \stackrel{\sim}{\cdot} \end{pmatrix}$$

Для випадку, коли λ і ρ невідомі, Зельнер і Гейсел запропонували вибирати значення λ і ρ довільно на проміжку $0 < \lambda < 1$; $-1 < \rho < 1$. Для кожної пари λ і ρ послідовно обчислюються значення $y_t(\rho)$ і залишки. У кінці процедури вибираються ті значення λ і ρ , які забезпечують мінімальну суму квадратів відхилень.

Як бачимо, процедура оцінювання параметрів при гіпотезах **2а** і **2б** досить громіздка. Тому використовувати її слід лише тоді, коли є впевненість, що залишки мають ту специфікацію, яка визначає особливості прийнятої гіпотези.

Гіпотеза 3а. Згідно з цією гіпотезою, специфікується модель:

$$y_t = a_0 + a_1 y_{t-1} + a_2 x_t + v_t$$

де $v_t = \rho v_{t-1} + \varepsilon_t$, $|\rho| < 1$ ε_t – нормально розподілені залишки.

Ця гіпотеза не пов'язується ні зі схемою Койка, ні з моделлю адаптивних сподівань. Ідеться про оцінку параметрів моделі, яка має серед незалежних

змінних лагове значення залежної змінної і одночасно має автокорельовані залишки.

Метод Ейткена

Якщо ρ відоме, то можна записати матрицю

$$M \left(V' \right) = \sigma_u^2 S = \sigma_v^2 \begin{pmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{pmatrix}$$

і оцінити параметри моделі за методом Ейткена:

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

$$X = \begin{pmatrix} 1 & y_0 & x_1 \\ 1 & y_1 & x_2 \\ \dots & \dots & \dots \\ 1 & y_{n-1} & x_n \end{pmatrix}$$

Така процедура наближено еквівалентна застосуванню методу найменших квадратів до моделі

$$y_t - \rho y_{t-1} = a_0 \left(-\rho \right) + a_1 \left(y_{t-1} - \rho y_{t-2} \right) + a_2 \left(x_t - \rho x_{t-1} \right) + \varepsilon_t$$

відносно перетворених даних. У результаті дістаємо обґрунтовані і асимптотично ефективні оцінки параметрів, але через присутність лагового значення залежної змінної в правій частині, вони будуть зміщеними для кінцевих вибірок.

Якщо значення параметра ρ невідоме, то можна скористатись процедурою пошуку, запропонованою для гіпотези **2**.

Ітеративний метод

Як альтернативу вище наведеному методу можна запропонувати ітеративний метод. Розглянемо його.

Перепишемо останнє рівняння попереднього розділу у вигляді:

$$y_t = a_0 \left(-\rho \right) + \left(1 + \rho \right) y_{t-1} - a_1 \rho y_{t-2} + a_2 x_t - a_2 \rho x_{t-1} + \varepsilon_t \quad (11.24)$$

Щоб безпосередньо оцінити всі чотири параметри мінімізацією суми квадратів відхилень для (11.24), треба розв'язувати нелінійні рівняння відносно параметрів. Якщо розбити параметри на дві множини, внівши до однієї a_0, a_1, a_2 , а до іншої – ρ , то можна знайти умовний мінімум суми квадратів залишків для рівняння (11.24) по чергово відносно кожної множини параметрів. У такому разі оцінюватимуться лінійні рівняння.

Алгоритм

Крок 1. Вибирається деяке початкове значення $\rho = \hat{\rho}$, воно підставляється в рівняння (11.24), яке відповідно спрощується.

Крок 2. Мінімізується сума квадратів залишків рівняння (11.24) при фіксованому ρ , в результаті одержуються оцінки $\hat{a}_0, \hat{a}_1, \hat{a}_2$.

Крок 3. Підставимо значення параметрів $a_0 = \hat{a}_0, a_1 = \hat{a}_1, a_2 = \hat{a}_2$ в модель (11.24) і визначимо параметр $\hat{\rho}$ тобто застосовується метод найменших квадратів до рівняння $v_t = \rho v_{t-1} + \varepsilon_t$, що і дозволяє знайти $\hat{\rho}$.

Крок 4. Задавши $\hat{\rho}$ в моделі (11.24), знайдемо на основі методу найменших квадратів оцінку параметрів $a_0 = \hat{a}_0, a_1 = \hat{a}_1, a_2 = \hat{a}_2$.

Процес продовжується доти, поки не буде досягнуто збіжності оцінок параметрів моделі на двох останніх кроках з вибраною точністю.

Інструментальні змінні. Алгоритм Уоліса

Застосовується також процедура, що використовує так звані інструментальні змінні, коли y_t залежить від v_t а y_t залежить від y_{t-1} .

Одна зі складнощів моделі – це існування кореляції y_{t-1} і з v_t . Але, враховуючи зроблене припущення, коли незалежні змінні ймовірніше всього не корелюють з y_{t-1} , оцінку параметрів моделі

$$y_{t-1} = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + \dots + \varepsilon_t \quad (11.25)$$

можна знайти за допомогою методу найменших квадратів. Кількість лагових значень X , які включаються в цю модель, можна вибрати залежно від обсягу вибірки і від їх здатності пояснити поведження залежної змінної y_t . Якщо

значення змінної X має високу автокореляцію, то навряд чи потрібно брати більше ніж два її лагових значення. Записане вище співвідношення змістимо на один період назад, щоб дістати y_{t-1} і підставимо вираз y_{t-1} у праву частину моделі (11.25). Після цього застосовується метод найменших квадратів для оцінки параметрів a . Ці оцінки будуть обґрунтованими, бо всі незалежні змінні гранично не корельовані із залишками, але вони будуть не ефективними, оскільки при оцінюванні параметрів не була врахована автокореляція залишків.

Алгоритм Уоліса. Уоліс запропонував складніший трикроковий метод оцінювання.

Крок 1. Оцінюються параметри моделі

$$y_t = a_0 + a_1 y_{t-1} + a_2 x_t + \dots + v_t$$

де x_{t-1} використовується як інструментальна змінна для y_{t-1} . Таким чином, визначають:

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

$$Z = \begin{pmatrix} 1 & x_0 & x_1 \\ 1 & x_1 & x_2 \\ \dots & \dots & \dots \\ 1 & x_{n-1} & x_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & y_0 & x_1 \\ 1 & y_1 & x_2 \\ \dots & \dots & \dots \\ 1 & y_{n-1} & x_n \end{pmatrix}, \quad Y = \begin{pmatrix} y_0 \\ y_1 \\ \dots \\ y_{n-1} \end{pmatrix}$$

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

$$r = \frac{\sum_{t=2}^n \hat{v}_t \hat{v}_{t-1}}{\sum_{t=1}^n \hat{v}_t^2} + \frac{m+1}{n}, \quad r \approx \rho \quad (11.26)$$

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

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

і обчислюють оцінку вектора A узагальненим методом найменших квадратів:

$$\hat{A} = (X'S^{-1}X)^{-1} X'S^{-1}Y \quad (11.27)$$

Проведені Уолісоном дослідження показали, що його метод оцінювання призводить до значно менших величин зміщення і до меншої суми квадратів залишків, ніж застосування методу Ейткена безпосередньо до моделі (11.25).

Метод інструментальних змінних.

Метод інструментальних змінних є альтернативним до методу найменших квадратів у випадку, коли кореляція між факторними (незалежними) змінними та залишками в регресії суттєва і, отже, існують перешкоди до застосування останнього. Як основний інструмент методу використовуються спеціальним чином підібрані допоміжні змінні, які називають інструментальними.

В економетричній моделі $Y = XA + u$ незалежні змінні X можуть бути як детермінованими, що набувають своїх значень з певної множини фіксованих чисел, так і стохастичними, які набувають своїх значень з певним рівнем ймовірності.

Якщо незалежні змінні X — стохастичні і функція розподілу не пов'язана з параметрами a і σ_u^2 та змінні X розподілені незалежно від залишків u , то переважна більшість висновків про перевірку значущості моделі та її параметрів, побудову довірчих інтервалів виконуватиметься і для тих економетричних моделей, які мають стохастичну матрицю незалежних змінних X .

У загальній лінійній моделі зі стохастичними незалежними змінними асимптотичні властивості оцінок методом найменших квадратів мають вигляд:

$$\text{a) } M(u/x) = M(u) = 0; \quad (11.28)'$$

$$\text{б) } M(y/x) = XA + M(u/x) = XA; \quad (11.28)''$$

$$\text{в) } M(u'/x) = \sigma_u^2 E. \quad (11.28)'''$$

Крім того, припускається, що виконуються такі рівності:

$$\text{а) } Plim(u^{-1}u'u) = \sigma_u^2; \quad (11.29)'$$

$$\text{б) } Plim(u^{-1}x'x) = \sum xx; \quad (11.29)''$$

$$\text{в) } Plim(u^{-1}x'u) = 0, \quad (11.29)'''$$

де n - розмір вибіркової сукупності, взятої для оцінки параметрів.

Припущення (11.29)' означає: дисперсія стала для всіх залишків. Припущення (11.29)'' стверджує існування границі за ймовірністю для дисперсій (других моментів) змінної X , які утворюють матрицю $\sum xx$. Припущення (11.29)''' вказує, що границя за ймовірністю коваріацій між змінними X і залишками u дорівнює нулю.

У такому випадку метод найменших квадратів забезпечує обґрунтовану оцінку асимптотичних дисперсій і коваріацій помилок, коли в моделі незалежні змінні є стохастичними.

Якщо не виконується припущення $Plim(u^{-1}x'u) = 0$, то оцінки методом найменших квадратів для скінченних вибіркової сукупностей можуть мати зміщення.

Кореляція між змінними X і залишками u може виникнути з різних причин, але основними вважаються три:

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

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

Метод інструментальних змінних часто використовують, коли при вимірюванні значень змінних, які входять до економетричної моделі,

припускаються помилок. Тоді оцінка параметрів методом найменших квадратів матиме зміщення, яке визначається за формулою:

$$Plim(\hat{A} - A) = -Plim(X'X)^{-1} \cdot Plim(X'X \Delta).$$

Оператор оцінювання вектора A за допомогою інструментальних змінних має вигляд:

$$\hat{A} = (Z'X)^{-1} Z'Y. \quad (11.30)$$

Отриманий вектор забезпечує визначення обґрунтованої оцінки параметрів моделі.

Припускається, що інструментальні змінні, що задаються певною матрицею Z , гранично некорельовані із залишками u , а їх перехресні моменти із змінними X не всі дорівнюють нулю і створюють невироджену матрицю. Тобто матриця Z порядку $n \times m$ має наступні властивості:

$$a) Plim(Z'u) = \sigma_u^2; \quad (11.30)'$$

$$б) Plim(Z'X) = \sum xx; \quad (11.30)''$$

$$в) Plim(Z'X) = 0, \quad (11.30)'''$$

Асимптотична матриця коваріацій матиме вигляд

$$asy\ var(\hat{A}) = \sigma_u^2 (Z'X)^{-1} (Z'Z)^{-1} X'Z \quad (11.31)$$

де

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

В економетричних дослідженнях пропонуються три методи визначення інструментальних змінних, на основі яких знайдені оператори оцінювання параметрів моделі:

- 1) оператор оцінювання Вальда;
- 2) оператор оцінювання Бартлета;
- 3) оператор оцінювання Дарбіна.

Оператори оцінювання Вальда і Бартлета застосовуються тоді, коли економетрична модель характеризує зв'язок двох змінних, тобто модель має вигляд $Y_t = a_0 + a_1 X_t + u_t$, а оператор оцінювання Дарбіна може бути

застосованим і тоді, коли економетрична модель має більш ніж одну незалежну змінну.

Оператор оцінювання Вальда. Нехай економетрична модель має вигляд

$$Y_t = a_0 + a_1 X_t + u_t. \quad (11.32)$$

У такому разі, якщо вибірка 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 незалежних 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 незалежної 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 ці інструментальні змінні в операторі оцінювання $\hat{A} = (Z'X)^{-1}Z'Y$ отримаємо оцінки Вальда:

$$\hat{a}_1 = \frac{\bar{Y}_2 - \bar{Y}_1}{\bar{X}_2 - \bar{X}_1}, \quad \hat{a}_0 = \bar{Y} - \hat{a}_1 \bar{X}$$

де \bar{X}_2, \bar{X}_1 характеризують середні відхилення значень X відповідно вгору і вниз від медіани, а \bar{Y}_2, \bar{Y}_1 – середні значення залежної змінної, які відповідають середнім \bar{X}_2, \bar{X}_1 .

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 вибіркова дисперсія може бути досить великою, тобто оцінка є неефективною.

Особливості оцінювання методом Бартлета. В операторі оцінювання Бартлета інструментальні змінні визначаються як і в операторі Вальда. Проте 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 параметрів:

$$\hat{a}_1 = \frac{\bar{Y}_3 - \bar{Y}_1}{\bar{X}_3 - \bar{X}_1}, \quad \hat{a}_0 = \bar{Y} - \hat{a}_1 \bar{X}$$

де $\bar{X}_3, \bar{X}_1, \bar{Y}_3, \bar{Y}_1$ – середні значення змінних для спостережень першої та третьої груп.

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. В операторі оцінювання Дарбіна інструментальні змінні визначаються так: значення вектора X упорядковуються в порядку зростання і за інструментальну змінну береться порядковий номер (ранг), тобто числа 1, 2, 3, 4, ..., n . Дарбін показав, що для великих вибірових сукупностей ефективність застосування такого методу оцінювання досягає майже 96% ефективності оцінок методу найменших квадратів, 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:

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

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

де x_i характеризують відхилення від середнього значення x які впорядковуються за зростанням. Оператор оцінювання має вигляд

$$\hat{a}_1 = \frac{\sum_{k=1}^n k Y_k}{\sum_{k=1}^n k X_k}, \quad \hat{a}_0 = \bar{Y} - \hat{a}_1 \bar{X}$$

де k – порядковий номер (інструментальна змінна).

Дисперсія оцінок параметрів

$$\text{asy var} \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \end{pmatrix} = \sigma_u^2 \begin{pmatrix} 1 & \bar{X} \\ \bar{X} & \sum_{k=1}^n k^2 \end{pmatrix}^{-1}$$

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

МЕТОДИКА РОЗВ'ЯЗАННЯ ЗАДАЧ

Задача. На основі взаємопов'язаних часових рядів, які характеризують чисту продукцію та капітальні вкладення деякої республіка за 20 років, побудувати взаємну кореляційну функцію, використавши такі дані:

Рік	Капітальні вкладення млн.гр.од.	Чиста продукція млн.гр.од.	Рік	Капітальні вкладення млн.гр.од.	Чиста продукція млн.гр.од.
1	3857	24334	11	17006	72165
2	4686	28678	12	17352	78743
3	5515	33021	13	17838	80381
4	5209	32432	14	18878	82204
5	7522	40325	15	19090	77833
6	10390	49334	16	20016	81413
7	13678	54717	17	17736	77484
8	15976	53818	18	11951	75443
9	13880	55968	19	11469	85038
10	13949	61517	20	9068	75809

Розв'язання.

У результаті розрахунків для різних значень τ , отримано значення коефіцієнтів кореляції.

τ	1	2	3	4	5	6	7	8	9	10
r_{τ}	0,89	0,86	0,92	0,92	0,92	0,85	0,75	0,64	0,4	0,55

З таблиці видно, що найбільше значення коефіцієнта кореляції $r_{\tau}=0,92$. Він відповідає значенням $\tau=(3,4,5)$. Це означає, що найбільшого впливу капітальних вкладень на обсяг чистої продукції слід очікувати впродовж третього четвертого та п'ятого років.

Динамічна модель розподіленого лага в такому разі має вигляд

$$y_t = a_0 x_t + a_1 x_{t-3} + a_2 x_{t-4} + a_3 x_{t-5} + u_t$$

де y_t – чиста продукція в період t ; $a_j=0,1,2,3$ – вагові коефіцієнти лагових змінних; $x_{t-\tau}$ – капітальні вкладення в період $t-\tau$.

DUMMY-ЗМІННІ.

12.1 ПРИРОДА DUMMY-ЗМІННИХ.

Досі ми розглядали рівняння багатofакторної регресії, де фактори набували кількісних значень. Звичайно, економічні явища набагато різноманітніші. На залежну змінну поряд з кількісними факторами впливають і якісні: якість продукції, стать, релігія, страйки, війни, зміни в економічній політиці і т. ін.

Потрібно вміти вводити якісні дані в багатofакторні регресійні моделі, оцінювати параметри і аналізувати отримані результати. **Часто якісні змінні є бінарними: вони отримують "значення 1" при наявності певної якості і "значення 0" при її відсутності. Такі змінні називаються dummy-змінними.** Звичайно dummy-змінні не обов'язково приймають значення (0,1). Пара (0,1) може легко трансформуватись у будь-яку іншу пару лінійним перетворення $y = a + bz$, $b \neq 0$, де a та b константи, а $z=1$ або нулю. Наприклад, коли $z=1$, $y=a+b$, а коли $z=0$, $y=a$.

Dummy-змінні можуть використовуватись у регресійних моделях поряд з кількісними змінними, а можуть утворювати регресійні моделі, в яких усі фактори є dummy-змінними. Такі моделі називають AOV-моделями (analysis-of-variance).

Розглянемо як приклад найпростішу лінійну регресійну модель залежності рейтингу студентів ТНЕУ від успішного закінчення навчання в школі:

$$y_i = a_0 + a_1 d_i + \varepsilon_i \quad (12.1)$$

де y_i – триместровий середній рейтинг успішності студента 1 курсу;

d_i – (dummy-змінна) = 1, якщо студент відмінник у школі, або його середній бал >4.8; =0, у протилежному випадку.

Модель (12.1) встановлює залежність успішного навчання в університеті від успішного навчання в школі. Як і у разі класичної лінійної регресійної моделі, ми отримаємо:

$$\text{середній рейтинг відмінників } E(y_i | d_i = 1) \approx a_0 + a_1. \quad (12.2)$$

$$\text{середній рейтинг не відмінників } E(y_i | d_i = 0) \approx a_0$$

У цьому разі перетин a_0 показує середній рейтинг невідмінників; $(a_0 + a_1)$ – середній рейтинг відмінників; а нахил a_1 – наскільки середній рейтинг відмінників відрізняється від середнього рейтингу невідмінників.

Базуючись на реальних даних рейтингу студентів першого курсу ТНЕУ, проведемо розрахунки за моделлю (12.1). Графічно отримані результати подано на рис.12.1.

Як бачимо з графіка, вхідні дані розподіляються на дві групи: відмінники та невідмінники. Сама функція є ступінчастою функцією – середній рейтинг студентів невідмінників та середній рейтинг відмінників.

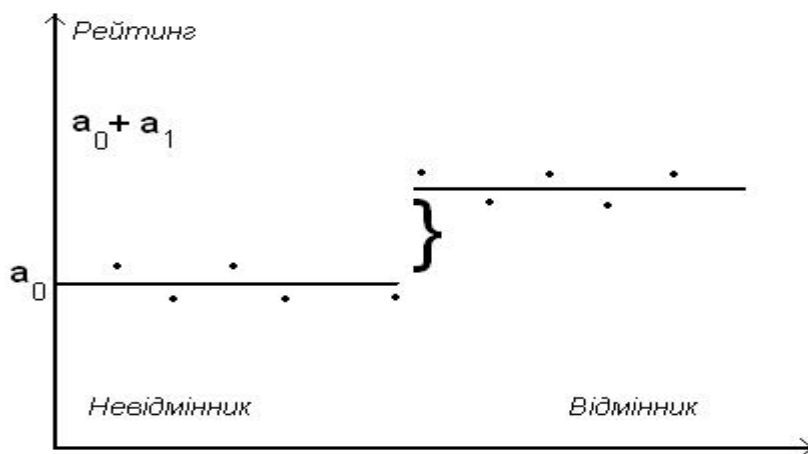


Рис. 12.1 Функція залежності рейтингу

12.2 РЕГРЕСІЯ ОДНІЄЇ КІЛЬКІСНОЇ ТА ОДНІЄЇ ЯКІСНОЇ ЗМІННОЇ ДВОХ КЛАСІВ, АБО КАТЕГОРІЙ.

Звичайно, більш поширеним випадком в економічних дослідженнях є випадок змішаних факторів – якісних та кількісних. Моделі такого типу називаються АСОВ-моделями (analysis of covariance). Розглянемо як приклад найпростішу АСОВ-модель, яка вміщує одну якісну та одну кількісну змінну.

Ускладнимо попередньо розглянуту модель і додамо ще один фактор – бал вступного іспиту; отримаємо таку АCOV-модель:

$$y_i = a_0 + a_1 d_i + b_1 x_{1i} + \varepsilon_i \quad (12.3)$$

де y_i – середній триместровий рейтинг студента;

$d_i=1$, якщо студент у минулому відмінник;

$=0$, у протилежному випадку;

x_1 – бал на вступному іспиті;

ε – випадкова величина.

Припускаючи, як завжди, $E(\varepsilon_i)=0$, отримаємо, що середній рейтинг не відмінника

$$E(y_i | d_{1i} = 0) = a_0 + b_1 x_{1i} \quad (12.4)$$

середній рейтинг відмінника

$$E(y_i | d_{1i} = 1) = a_0 + a_1 + b_1 x_{1i} \quad (12.5)$$

Геометрично АCOV-модель зображено на рис. 12.2 ($a_0 > 0$).

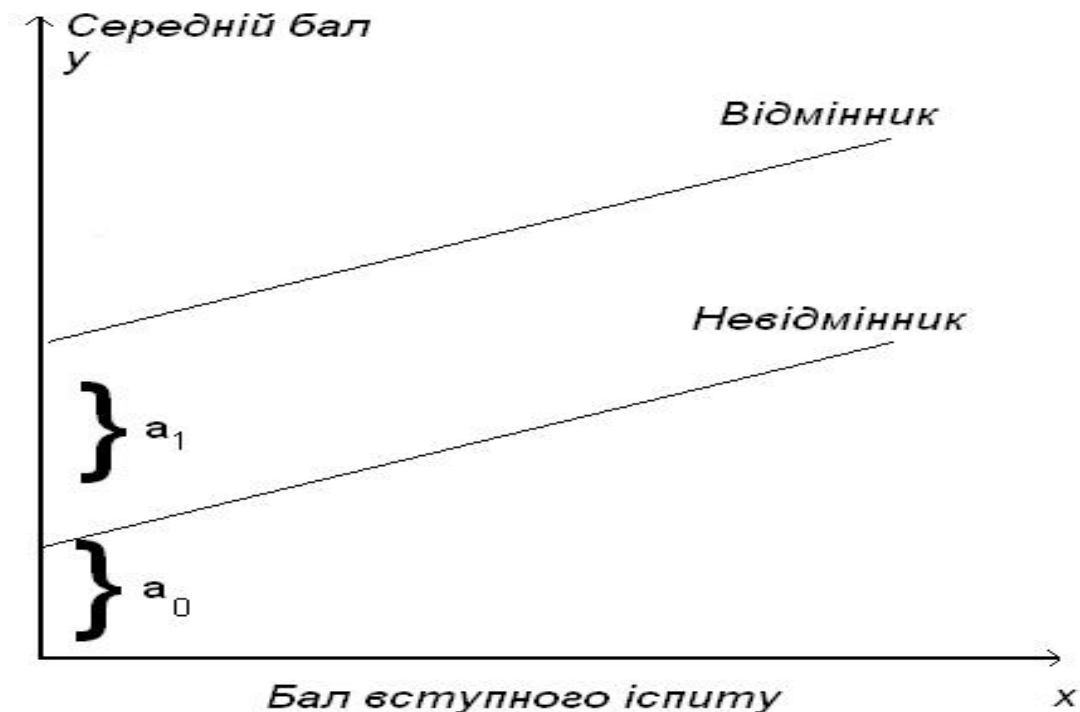


Рис. 12.2 АCOV-модель

Як бачимо, модель (12.3) залежно від відмінників і не відмінників розпадається на дві функції з однаковим нахилом, але різним перетином.

Іншими словами, рівень середнього триместрового рейтингу студентів відмінників відрізняється від рейтингу не відмінників на a_1 , але величина зміни середнього триместрового рейтингу студентів щодо балу, отриманого на вступному тестуванні, однакова для відмінників та не відмінників.

При умові, що нахил статистично значимий, зробимо перевірку гіпотези, за якою дві регресії (12.4) і (12.5) мають однаковий перетин (тобто немає впливу відмінної успішності в школі). Це можна легко здійснити при розгляді регресії (12.3) і визначити статистичну значимість оціненого коефіцієнта a_1 на основі традиційного t -тесту. Якщо t -тест показує, що a_1 статистично значимий, нуль-гіпотезу про те, що дві регресії мають однаковий перетин – відхиляємо.

Задамо запитання, що додають до регресії *dummy*-змінні, які особливості виникають при введенні *dummy*-змінних у модель?

По-перше, *dummy*-змінні відокремлюють різні класи або різні категорії. У нашому прикладі завдяки введенню *dummy*-змінних ми розрізняли дві категорії студентів: відмінників та не відмінників.

Отже, одна *dummy*-змінна може розрізняти дві категорії. Припустимо, що ми хочемо ускладнити наш приклад, поділивши ще детальніше категорії. Що при цьому виникає? Розглянемо таку модель:

$$y_i = a_0 + a_1 d_{1i} + a_{2i} d_{2i} + b_1 x_{1i} + \varepsilon_i \quad (12.6)$$

де y_i і x_i визначені попередньо;

$d_1=1$, якщо студент відмінник;

$=0$, в усіх інших випадках;

$d_2=0$, якщо студент не відмінник;

$=1$ в усіх інших випадках.

Модель (12.6) не може бути оцінена через повну колінеарність між d_1 і d_2 . Для пересвідчення, припустимо, що ми маємо модель, яка складається з трьох студентів відмінників та двох не відмінників. Інформаційна матриця при цьому виглядатиме таким чином:

			d_1	d_2	x	
відмінник	y_1		1	1	0	x_1
відмінник	y_2		1	1	0	x_2
невідмінник	y_3	=	1	0	1	x_3
відмінник	y_4		1	1	0	x_4
невідмінник	y_5		1	0	1	x_5

Перша колонка праворуч відображає перетин a_0 . Тепер можна побачити, що $d_1=1-d_2$, або $d_2=1-d_1$, тобто d_1 і d_2 – колінеарні. У разі досконалої мультиколінеарності звичайне оцінювання методом найменших квадратів неможливе. Найпростіший спосіб вирішення цієї проблеми – визначити *dummy*-змінні способом, описаним у моделі (12.3), скажімо, використати тільки одну *dummy*-змінну, якщо є тільки дві категорії або два класи якісної змінної. В даному випадку, попередня матриця інформації не матиме колонки d_2 , і проблема мультиколінеарності знімається. Тому в загальному випадку для уникнення проблеми мультиколінеарності потрібно дотримуватися такого правила: якщо якісна змінна має m категорій, у модель вводимо лише $m-1$ *dummy*-змінну. В нашому прикладі кінцевий результат навчання в школі ми підрозділяли на дві категорії – відмінник та не відмінник, тому ми ввели у модель лише одну *dummy*-змінну. Якщо це правило не виконано, ми попадаємо в так звану *dummy*-пастку, тобто ситуацію досконалої мультиколінеарності.

По-друге, припустимо, що ми ввели інші значення 1 і 0 для того, щоб відрізнити дві категорії. Нехай тепер $D=1$ для не відмінників і $D=0$ для відмінників. Тоді дві регресії, отримані з (12.3), будуть:

$$\text{студенти не відмінники: } E(y_i | d_{1i} = 1) = a_0 + a_1 + b_1 x_{1i} \quad (12.7)$$

$$\text{студенти відмінники: } E(y_i | d_{1i} = 0) = a_0 + b_1 x_{1i} \quad (12.8)$$

Якщо порівняти ці рівняння з рівняннями (12.4) та (12.5) у попередніх моделях, то a_1 показує, наскільки середній рейтинг студента-відмінника в школі відрізняється від середнього рейтингу не відмінника в школі. При наявності залежності a_1 буде негативним, тоді як попередньо очікувалось, що воно матиме позитивне значення. Тому під час інтерпретації результатів моделей, які

використовуюють *dummy*-змінні, важливо знати, які саме категорії позначались **1**, а які **0**.

По-третьє, клас, або категорія, позначена 0 (нулем), часто розглядається як базова, вихідна, контрольна категорія. Вона є базовою в тому розумінні, що порівняння робляться на основі саме цієї категорії. Таким чином, у моделі (5.33) студент-невідмінник у школі — базова категорія. Перетин a_0 є перетином базової категорії в тому сенсі, що при розгляді регресії стосовно студентів-невідмінників $D=0$ перетин дорівнюватиме a_0 .

По-четверте, коефіцієнт біля *dummy*-змінної D називається диференціальним коефіцієнтом перетину, тому що він показує, наскільки значення перетину першої категорії відрізняється від значення перетину базової категорії.

12.3 РЕГРЕСІЯ КІЛЬКІСНОЇ ЗМІННОЇ ТА ОДНІЄЇ ЯКІСНОЇ ЗМІННОЇ З БІЛЬШ НІЖ ДВОМА КЛАСАМИ.

Припустимо, що базуючись на міжгалузевій інформації, ми хочемо описати регресійний зв'язок щорічних оздоровчих витрат людини залежно від освіти та доходу. Оскільки освіта є якісною змінною, припустимо, що розглядаємо три взаємовиключні рівні освіти: середня, середня спеціальна і вища. Тепер, на відміну від попереднього випадку, маємо більш ніж дві категорії якісних змінних. Застосовуючи правило, що кількість змінних має бути на одиницю менша від кількості категорій, нам потрібно подати дві *dummy*-змінні, щоб простежити за двома рівнями освіти. Припустимо, три рівні освіти мають однаковий нахил, але різні перетини.

Модель щорічних оздоровчих витрат залежно від доходу матиме вигляд:

$$y_i = a_0 + a_1 D_{1i} + a_2 D_{2i} + b_1 x_i + \varepsilon_i \quad (12.9)$$

де y_i – щорічні оздоровчі витрати;

x_i – щорічний дохід;

$D_1=1$ – середня освіта;

=0 – у всіх інших випадках;

$D_2=1$ – вища освіта;

=0 – у всіх інших випадках.

Аналогічно з попереднім завданням ми трактуємо середню спеціальну освіту як базову категорію. Тому a_0 відображатиме перетин саме цієї категорії. Диференційні a_1 і a_2 , показують, наскільки перетини двох інших категорій відрізняються від перетину базової категорії. Це можна перевірити таким чином: припустимо, $E(\varepsilon_i)=0$, тоді з (5.38) отримаємо:

$$E(y_i / D_1 = 0, D_2 = 0, x_i) = a_0 + b_1 x_{1i} \quad (12.10)$$

$$E(y_i / D_1 = 1, D_2 = 0, x_i) = a_0 + a_1 + b_1 x_{1i} \quad (12.11)$$

$$E(y_i / D_1 = 0, D_2 = 1, x_i) = a_0 + a_2 + b_1 x_{1i} \quad (12.12)$$

які є функціями середніх оздоровчих витрат для трьох рівнів освіти: середньої, середньої спеціальної та вищої.

Після розгляду регресії (12.9) можна встановити, чи диференційні перетини a_1 і a_2 , статистично значимі, тобто чи відрізняються від базової групи. Гіпотезу $a_1=a_2=0$ можна перевірити за допомогою AOV або F -тесту.

Інтерпретація регресії (12.9) зміниться, якщо ми візьмемо іншу схему визначення думми-змінних. Тобто, якщо припустити, що $D_1=1$ – "середня освіта", і $D_2=1$ – "категорія середньої спеціальної освіти", то базовою категорією буде "вища освіта" і всі порівняння будуть пов'язані з цією категорією.

Регресія однієї кількісної і двох якісних змінних

Техніку думми-змінних можна розширити, маючи справу більше ніж з однією якісною змінною. Повернемось до регресії середнього рейтингу студентів (5.5.3), але тепер до навчання в школі та вступного балу додамо стать – чоловічу або жіночу. Тепер (5.5.3) можна виразити як

$$y_i = a_0 + a_1 D_{1i} + a_2 D_{2i} + b_1 x_{1i} + \varepsilon_i \quad (12.13)$$

де y_i та x_i – середній триместровий рейтинг студента та бал вступного іспиту;

$D_1=1$ – якщо студент-відмінник;

=0 – в іншому випадку;

$D_2=1$ – якщо студент-чоловік;

=0 – в іншому випадку.

Зазначимо, що кожна з двох якісних змінних, навчання в школі та стать, мають дві категорії і тому потребують *dummy*-змінну для кожної. Базовою категорією буде "невідмінник-дівчина". Припускаємо, що $E(\varepsilon_i)=0$, і з (12.13) отримуємо такі регресії:

Середній рейтинг невідмінниці-дівчини:

$$E(y_i / D_1 = 0, D_2 = 0, x_i) = a_0 + b_1 x_{1i} \quad (12.14)$$

Середній рейтинг невідмінника-хлопця:

$$E(y_i / D_1 = 1, D_2 = 0, x_i) = a_0 + a_1 + b_1 x_{1i} \quad (12.15)$$

Середній рейтинг відмінниці-дівчини:

$$E(y_i / D_1 = 0, D_2 = 1, x_i) = a_0 + a_2 + b_1 x_{1i} \quad (12.16)$$

Середній рейтинг відмінника-хлопця:

$$E(y_i / D_1 = 1, D_2 = 1, x_i) = a_0 + a_1 + a_2 + b_1 x_{1i} \quad (12.17)$$

Нагадаємо, що попередні регресії відрізняються лише коефіцієнтами перетину, а не нахилу.

За допомогою оцінювання методом найменших квадратів неможливо перевірити всі гіпотези. Але якщо a_2 статистично значимий, це означатиме, що стать людини впливатиме на середній триместровий рейтинг студентів. Аналогічно, якщо a_1 буде статистично значимим, це означатиме що навчання в школі також впливатиме на середній рейтинг студентів. А якщо обидва диференційні перетини статистично значимі, це означатиме, що навчання в школі та стать є важливими детермінантами середнього рейтингу. Отже, ми можемо розширити нашу модель, включивши більше ніж одну кількісну змінну та більше ніж одну якісну змінну. Єдине застереження: кількість *dummy*-змінних для кожної якісної змінної має бути на одиницю меншою від категорії цієї змінної. Наведемо приклад.

МЕТОДИКА РОЗВ'ЯЗАННЯ ЗАДАЧ

Задача. Додаткові заробітки

Людина, яка працює на кількох роботах, відома як заробітчанин. Шіско і Росткер зацікавилися факторами, які визначають зарплати заробітчан. Базуючись на вибірці щодо 318 заробітчан, вони отримали таку регресію:

$$w_m = 37,07 + 0,403w_0$$

(0.062)

$$90,06D_1 + 75,51D_2 + 47,33D_3 + 113,64D_4 + 2,26x_1$$

(12.18)

$$\left[\begin{matrix} 4,47 \\ 1,60 \\ 3,42 \\ 7,62 \\ 0,94 \end{matrix} \right]$$

$$R^2=0,34 \quad df=311,$$

де w_m – зарплата заробітчанина (центів на годину);

w_0 – початкова зарплата (центів на годину);

D_1 – раса =0, якщо білий, =1, якщо не білий;

D_2 – місто =0, якщо не міський, =1, якщо міський;

D_3 – область =0, якщо не західний район, =1, якщо західний район;

D_4 — освіта =0 – без освіти, =1 має освіту;

x_1 – вік людини.

У моделі (12.18) є дві кількісні пояснюючі змінні, w_0 та x_1 а також чотири якісні змінні. Коефіцієнти всіх цих змінних статистично значимі з рівнем похибки 5%. Цікаво, що всі якісні змінні суттєво впливають на зарплату заробітчан. Наприклад, за інших рівних умов рівень погодинної оплати праці на 47% вищий у осіб з освітою, ніж у тих, хто не має освіти.

З регресії (12.18) можна виділити декілька індивідуальних регресій, дві з яких такі: середня зарплата білих, неміських, не із західних регіонів заробітчан, які не мають вищої освіти (коли всі dummy-змінні дорівнюють нулю):

$$w_m = 37,07 + 0,403w_0 + 2,26x_1 \quad (12.19)$$

Середня зарплата чорношкірого, міського, з західного регіону заробітчанина з вищою освітою (коли всі dummy-змінні рівні одиниці):

$$w_m = 183,49 + 0,403w_0 + 2,26x_1 \quad (12.20)$$

Порівняння двох регресійних моделей

Дотепер у розглянутих моделях ми припускали, що якісні змінні впливають лише на перетин, але не на нахил. А що, коли нахили також різні? Якщо нахили відрізняються, то перевірка відмінностей у перетинах фактично мало важлива. Тому потрібно розвинути загальну методологію знаходження відмінностей у двох (або більше) регресіях. Відмінність може бути в перетинах, нахилах або в обох випадках. Щоб простежити, як це відбувається, розглянемо приклад.

Задача. Заощадження та дохід. У таблиці наведено інформацію про доходи та заощадження.

Інформація в таблиці поділена на два періоди, 1946 – 1954 рр. (одразу після Другої світової війни, період реконструкції) і 1955 – 1963 рр. (період після реконструкції). Припустимо, що ми хочемо дослідити, чи змінився взаємозв'язок агрегованих доходів та заощаджень між двома періодами.

Період 1	Заощадження	Доходи	Період 2	Заощадження	Доходи
1946	0,36	8,8	1955	0,59	15,5
1947	0,21	9,4	1956	0,90	16,7
1948	0,08	10,0	1957	0,95	17,7
1949	0,20	10,6	1958	0,82	18,6
1950	0,10	11,0	1959	1,004	19,7
1951	0,12	11,9	1960	1,53	21,1
1952	0,41	12,7	1961	1,94	22,8
1953	0,50	13,5	1962	1,75	23,9
1954	0,43	14,3	1963	1,99	25,3

$$\text{Період реконструкції: } y_i = a_0 + a_1x_i + \varepsilon_{1i}, i=1,2,\dots,N_1 \quad (12.21)$$

$$\text{Після періоду реконструкції: } y_i = \gamma_0 + \gamma_1x_i + \varepsilon_{2i}, i=1,2,\dots,N_2 \quad (12.22)$$

де y – заощадження (мільйони фунтів);

x – доходи (мільйони фунтів);

$\varepsilon_{1i}, \varepsilon_{2i}$ – відхилення в двох регресіях.

Звичайно потрібно, щоб кількість спостережень N_1 та N_2 в обох групах (періодах) була однаковою.

Між регресіями (12.21) і (12.22) можливі такі чотири варіанти співвідношень.

- 1) $a_0 = \gamma_0$ і $a_1 = \gamma_1$, тобто дві регресії ідентичні.
- 2) $a_0 \neq \gamma_0$ але $a_1 = \gamma_1$, тобто дві регресії відрізняються лише за розміщенням (перетинами).
- 3) $a_0 = \gamma_0$ і $a_1 \neq \gamma_1$, дві регресії мають однакові перетини, але різні нахили.
- 4) $a_0 \neq \gamma_0$ і $a_1 \neq \gamma_1$, обидві регресії абсолютно різні.

Використовуючи дані табл., можна записати дві окремі регресії (12.21) і (12.22), а потім застосувати відповідні статистичні засоби для перевірки всіх попередніх варіантів. Одним з таких засобів є Chow-тест, іншим – dummy-змінні. Ми обговоримо це в наступних двох частинах.

Chow-тест для порівняння регресійних моделей

Одним з відомих методів тестування відмінностей між двома (і більше) регресіями є Chow-тест. Цей тест базується на таких припущеннях.

1. Випадкові величини в обох регресійних моделях – нормально розподілені випадкові величини з математичним сподіванням нуль та постійною дисперсією:

$$\varepsilon_{1i} \approx N(0, \sigma^2)$$

$$\varepsilon_{2i} \approx N(0, \sigma^2)$$

2. Випадкові величини ε_{1i} , ε_{2i} незалежно розподілені. Враховуючи ці припущення, використаємо інформацію табл. 5.10 про доходи та заощадження для застосування Chow-тесту.

Крок 1. Об'єднаємо всі спостереження N_1 та N_2 в обох періодах і розглянемо одну "збірну" регресію {примітка: в даному випадку $N_1 = N_2 = 9$ }.

$$y_t = a_0 + a_1 x_t + \varepsilon_t \quad (12.23)$$

З даної регресії отримаємо суму квадратів залишків (SSE) з $df=N_1-N_2-k$, де k означає кількість оцінених параметрів. Для нашого прикладу $k=2$.

Крок 2. Розглянемо окремо дві регресії (12.21) і (12.22). Отримаємо їх SSE, скажімо, S_2 і S_3 , $df=N_1-k$ і $df=N_2-k$, відповідно; k у даному випадку дорівнює 2. Додамо дві SS, $S_4=S_2+ S_3$ з $df=N_1+N_2-2k$.

Крок 3. Отримаємо $S_5=S_1-S_4$.

Крок 4. Застосуємо F-тест:

$$F = \frac{S_5 / k}{S_4 / (N_1 + N_2 - 2k)} \quad (12.24)$$

де $df=(k; N_1+N_2-2k)$. Якщо обчислене F перевищує критичне F , відхиляємо гіпотезу про те, що дві регресії однакові.

Застосуємо ці кроки до нашого прикладу доходів та заощаджень.

Крок 1.

$$\begin{aligned} \hat{y}_t &= -1,0821 + 0,1178x_t \\ &\quad (0,1451) \quad (0,0088) \\ t &= (-7,4576) \quad (3,3864) \quad R_2 = 0,9185 \quad S_1 = 0,5722 \quad df = 16 \end{aligned} \quad (12.25)$$

Крок 2.

$$\begin{aligned} \hat{y}_t &= -1,266 + 0,0470x_t \\ &\quad (0,3053) \quad (0,0266) \\ t &= (-0,8719) \quad (7,669) \quad R_2 = 0,3092 \quad S_2 = 0,1396 \quad df = 7 \end{aligned} \quad (12.26)$$

Період після реконструкції:

$$\begin{aligned} \hat{y}_t &= -1,7502 + 0,1504x_t \\ &\quad (0,3576) \quad (0,0175) \\ t &= (-4,88943) \quad (3,5943) \quad R_2 = 0,9131 \quad S_3 = 0,1931 \quad df = 7 \end{aligned} \quad (12.27)$$

$$S_4 = S_2 + S_3 = 0,3327$$

Крок 3

$$S_5 = S_1 - S_4 = 0,2395$$

Крок 4.

$$F = \frac{0,2395/2}{0,3327/14} = 5,04$$

Тоді $F_{2,14}$ з рівнем значимості 5% дорівнює 3,74. Тому обчислене $F=5,04$ статистично значиме. Можна зробити висновок, що дві регресії різні. Chow-тест є спеціальним застосуванням методу обмежених найменших квадратів: коли ми оцінюємо (12.23) замість (12.21) і (12.22), ми накладаємо обмеження $a_0 = \gamma_0 = a$ і $a_1 = \gamma_1 = \beta$.

Якщо не виконана умова гомоскедастичності, то Chow-тест не можна застосовувати. Тому необхідно спочатку перевірити умову гомоскедастичності $\sigma_1^2 = \sigma_2^2 = \sigma^2$, застосовуючи один (і більше) методів, обговорюваних раніше. Також має бути достатньо ступенів вільності, щоб розглядати регресії окремо. У винятковому випадку, якщо в прикладі доходів та заощаджень є тільки два спостереження для останнього періоду, буде нуль ступенів вільності $(N-2)=(2-2)=0$. В таких ситуаціях F-тест у (12.24) потрібно модифікувати.

Приклад, який ми обговорювали, мав дві групи, але Chow-тест можна легко узагальнити і для більше, ніж двох груп.

Порівняння двох регресій: підхід з використанням dummy-змінної

Процедуру Chow-тесту можна скоротити, використовуючи dummy-змінні. Хоча висновки, які ми отримуємо з Chow-тесту та тесту dummy-змінних однакові, застосування dummy-змінних дає деякі переваги.

Давайте об'єднаємо всі N_1 та N_2 спостереження і оцінимо таку регресію:

$$y_i = a_1 + a_2 D_i + b_1 x_i + b_2 (D_i x_i) + \varepsilon_i \quad (12.28)$$

де y_i , та x_i – відповідно заощадження і дохід, $D_i=1$ для спостережень у першому періоді та періоді реконструкції і $D_i=0$ для спостережень у періоді після реконструкції.

Припускаючи, що $E(\varepsilon_i)=0$, отримаємо:

$$E(y_i / D_i = 0, x_i) = a_1 + b_1 x_i \quad (12.29)$$

$$E \left(\frac{y_i}{D_i} \mid D_i = 1, x_i \right) = \alpha_1 + a_2 + \alpha_1 + b_2 x_i \quad (12.30)$$

які є відповідно функціями середніх заощаджень для другого (після реконструкції) та першого (реконструкції) періодів і є такими ж як (12.22) і (12.21) з $\gamma_0 = a_1$, $\gamma_1 = b_1$, $a_0 = (a_1 + a_2)$ і $a_1 = (b_1 + b_2)$. Тому оцінювання (12.25) є таким самим, як і оцінювання окремих функцій заощаджень (12.21) та (12.22).

У (12.28) a_2 є диференційним перетином, а b_2 є диференційним нахилом, який показує, наскільки нахил функції збережень першого періоду відрізняється від нахилу функції збережень другого періоду. Якщо перемножити D і x , то неможливо зрозуміти відмінність між нахилами двох періодів.

Повертаючись до інформації про доходи та заощадження з табл., емпіричну частину (12.26) можемо записати таким чином:

$$\hat{y}_t = -1,7502 + 1,4839D_t + 0,1504x_t - 0,1034D_t x_t \quad (12.31)$$

$$\begin{matrix} (0,3319) & (0,4704) & (0,0163) & (0,0332) \end{matrix}$$

$$t = (-5,2733) \quad (1,1545) \quad (0,2270) \quad (-3,1144) \quad \bar{R}^2 = 0,94425$$

Як показує регресія, диференційні коефіцієнти перетину та нахилу статистично значимі, вони чітко показують, що регресії для двох періодів різні. Тоді, розглядаючи (12.29) та (12.30), можемо отримати дві регресії (примітка: $D=1$ для першого періоду).

Період реконструкції:

$$\hat{y}_t = (-1,77502 + 1,4839) + (0,1504 - 0,1034)x_t = -0,2663 + 0,00470x_t \quad (12.32).$$

Після періоду реконструкції:

$$\hat{y}_t = -1,7502 + 0,1504x_t \quad (12.33)$$

Ці регресії такі ж самі, як і регресії, отримані за допомогою Chow-процедури, відповідно (12.26) і (12.27).

Тепер можемо продемонструвати переваги техніки dummy-змінних (оцінювання (12.28) над Chow-тестом (оцінювання трьох регресій: (12.21), (12.22) і "зібраної" регресії окремо).

1. Нам потрібно розглядати тільки одну регресію, тому що окремі регресії

можна легко вивести з неї, як показано в рівняннях (12.29) і (12.30).

2. Для множини гіпотез можна застосувати одну регресію. Однак, якщо диференційний коефіцієнт перетину a_2 статистично незначимий, то можемо припустити гіпотезу, що дві регресії мають однаковий перетин. Відповідно, якщо диференційний нахил b_2 статистично незначимий, але a_2 — значимий, щонайменше ми можемо не відхилити гіпотезу про те, що дві регресії мають однаковий нахил, тобто лінії двох регресій паралельні. Перевірку стабільності регресії ($a_2=b_2=0$) можна здійснити звичайним F-тестом. Якщо гіпотеза прийнята, лінія регресії збігається.

3. Chow-тест чітко не показує, який коефіцієнт — перетину чи нахилу — відмінний або (як у даному прикладі) обидва відмінні в двох періодах. Тобто можна отримати значимий Chow-тест, оскільки відмінний лише нахил або лише перетин, або обидва. Іншими словами, Chow-тест не може дати відповідь на запитання, який з чотирьох варіантів є в даному випадку. В цьому відношенні підхід *dummy*-змінних має чітку перевагу, тут він вказує не лише на відмінність регресій, а й на причини цієї відмінності — чи вона стосується перетину або нахилу, чи відразу обох, що дуже важливо на практиці.

4. Оскільки "складання" збільшує ступені вільності, можна довести відносну точність оцінених параметрів.

Використання *dummy*-змінних у сезонному аналізі

Економічні процеси дуже часто підпорядковані сезонним коливанням. Прикладом можуть бути різдвяний розпродаж товарів, попит на гроші домогосподарств під час свят, попит на морозиво і напої влітку, сільськогосподарське виробництво. Крім того, багато первісної інформації (індекси споживчих, оптових цін, цін виробника) підпорядковані також впливу сезонності.

В економічному аналізі інколи виникає проблема вилучення сезонних коливань з метою виявлення тенденцій. Розглянемо один з методів вилучення сезонних коливань. Для прикладу розглянемо квартальну динаміку прибутків

деяких приватних фірм України. Необхідні умовні дані містяться в наведеній нижче табл. 5.11. У ній також міститься інформація щодо способу підготовки матриці для об'єднання *dummy*-змінних.

Візуальний аналіз свідчить про наявність сезонності – обсяги прибутку і продажу найбільші в другому кварталі щороку. Використовуючи дані, побудуємо модель:

$$y_i = a_0 + a_1 D_{1t} + a_2 D_{2t} + b_1 x_{1t} + \varepsilon_t \quad (12.34)$$

де $D_1=1$ для другого кварталу, $D_1=0$ – в усіх інших випадках;

$D_2=1$ для третього кварталу, $D_2=0$ – в усіх інших випадках;

$D_3=1$ для четвертого кварталу, $D_3=0$ – в усіх інших випадках.

Припустимо, що змінна "сезон" має чотири класи (чотири квартали). Отже, необхідно використовувати три *dummy*-змінні. За наявності сезонного впливу він буде відображений оцінками диференційованого перетину a_1 a_2 та a_3 , якщо вони будуть статистично значимі. Можливий випадок статистичної значимості лише декількох диференційованих перетинів.

Використовуючи дані з таблиці, отримаємо такі результати (млн. грн.):

$$y_t = 6688,3789 + 1322,8938 D_{1t} - 217,8037 D_{2t} + 183,8597 D_{3t} + 0,038 x_t.$$

$$\sigma_{at} \quad (1711,377) \quad (638,47) \quad (632,25) \quad (654,29) \quad (0,0115)$$

$$t \quad (3,9082) \quad (2,0720) \quad (-0,3445) \quad (0,2810) \quad (3,3313)$$

$$R^2 = 0,5255. \quad (12.35)$$

Результати свідчать, що лише коефіцієнт продаж і диференційний перетин другого кварталу статистично значимі з рівнем помилки 5% . Звідси випливає, що сезонний фактор присутній у другому кварталі щороку. Коефіцієнт продажу 0,0383 показує, що після врахування впливу сезонних коливань збільшення продаж на 1 млн. грн. приведе до підвищення прибутків на 0,04 млн. грн. Середній рівень прибутків у базовому першому кварталі становив 6688 млн. грн., а в другому підвищився на 1323 млн. грн., тобто дорівнював 8011 млн.грн.

Рік, квартал	Прибуток, млн.грн.	Продаж, млн.грн.	D_1	D_2	D_3
1992 – I	10.53	114.9	0	0	0
- II	12.09	124.0	1	0	0
- III	10.84	121.46	0	1	0
- IV	12.20	131.92	0	0	1
1993 - I	12.25	129.91	0	0	0
- II	14.00	140.98	1	0	0
- III	12.21	137.83	0	1	0
- IV	12.82	145.47	0	0	1
1994 - I	11.34	136.99	0	0	0
- II	12.61	145.13	1	0	0
- III	11.01	141.54	0	1	0
- IV	12.73	151.78	0	0	1
1995 - I	12.54	148.86	0	0	0
- II	14.85	158.91	1	0	0
- III	13.20	155.72	0	1	0
- IV	14.95	168.41	0	0	1
1996 - I	14.15	162.78	0	0	0
- II	15.95	176.06	1	0	0
- III	14.02	172.42	0	1	0
- IV	14.31	183.32	0	0	1

Оскільки тільки оцінки другого кварталу статистично значимі, то можна модифікувати (12.34), використовуючи лише одну *dummy*-змінну, для того, щоб його виділити:

$$\hat{y}_i = 6515,6 + 1331,4D_1 + 0,0393x_i$$

$$t = \begin{matrix} \begin{matrix} \text{[623,1]} \\ \text{[93,02]} \\ \text{[0,0106]} \end{matrix} \\ \begin{matrix} \text{[0,0143]} \\ \text{[0,7004]} \\ \text{[0,7173]} \end{matrix} \end{matrix} \quad (12.36)$$

$$R^2=0,5155$$

де $D_1=1$ для спостережень у другому кварталі. У всіх інших випадках $D_1=0$.

Зрозуміло, що (12.36) – обмежена версія (12.34); обмеження для першого, третього і четвертого кварталів однакові. Судячи з результатів (12.34), можна сподіватися, що ці обмеження дійсні. Тому висновок залишається таким самим: сезонні коливання мають найбільший вплив тільки у другому кварталі.

У формуванні моделі (12.34) припускалося, що лише перетин відрізняється між кварталами, нахил змінної продаж залишається таким самим. Однак, це припущення можна перевірити.

Зауваження.

Треба бути уважним, застосовуючи техніку dummy-змінних. По-перше, якщо модель регресії містить константу, то кількість dummy-змінних має бути на одиницю меншою за кількість категорій кожної якісної змінної. По-друге, коефіцієнти dummy-змінних завжди мають інтерпретуватись за відношенням до базової групи, яка набуває значення нуль. Нарешті, якщо модель має кілька якісних змінних з кількома категоріями, введення dummy-змінних може призвести до великої кількості ступенів вільності. Тому завжди потрібно зважувати кількість dummy-змінних для введення у модель із загальною кількістю спостережень

ЛІТЕРАТУРА

1. Айвазян С. А. Прикладная статистика и основы эконометрики / С. А. Айвазян, В. С. Мхатарян. – М. : ЮНИТИ, 1998. – 1022 с.
2. Березька К. М. Економетрія: основи теорії та комп'ютерний практикум / К. М. Березька. – Тернопіль : Тайп, 2007. – 137 с.
3. Доугерти Кристофер Введение в эконометрику: пер. с англ. / Доугерти Кристофер. — М.: ИНФРА-МД997.— С. 402.
4. Економіко-математичне моделювання: навчальний посібник / за ред. О. Т. Іващука. – Тернопіль : ТНЕУ Економічна думка, 2008. – 704 с.
5. Єлейко В. Основи економетрії : навчальний посібник / В. Єлейко. — Львів: Марка Лтд, 1995.— 235 с
6. Єрмоєнко В. О. Економетрія (економетрика) : навчальний посібник для студентів заочної форми навчання економічних спеціальностей / В. О. Єрмоєнко, А. М. Алілуйко, О. М. Мартинюк, С. Ю. Попіна. – Тернопіль : Підручники і посібники, 2011. – 114 с.
7. Іващук О. Т. Економетричні методи та моделі : навч. посібник / О. Т. Іващук. – Тернопіль : ТАНГ Економічна думка, 2002. – 348 с.
8. Кейн З. Экономическая статистика и эконометрия. Введение в количественный экономический анализ / З. Кейн. — М.: Статистика, 1997.— 254 с.
9. Корольов О. А. Практикум з економетрії: завдання з практичними рекомендаціями, алгоритмами та прикладом їх наскрізного виконання / О. А. Корольов, В. В. Рязанцева. – Київ: Вид-во Європ. Ун-ту, 2002. – 250 с.
10. Корольов О. А. Практикум з економетрії: завдання з практичними рекомендаціями, алгоритмами та прикладом їх наскрізного виконання / О. А. Корольов, В. В.Рязанцева. — Київ: Вид-во Європ. ун-ту, 2002 — 250 с.
11. Кулинич О. Економетрія навчальний посібник / О. Кулинич — Хмельницький: Поділля, 1997.— 275 с

12. Лещинський О. Л. Економетрія : навчальний посібник для студ. вищ. навч. закл. / О. Л. Лещинський, В. В. Рязанцева, О. О. Юнькова. – К. : МАУП, 2003. – 208 с.
13. Личковський Е. І. Вища математика. Теорія наукових досліджень у фармації та медицині : підручник / Е.І. Личковський, П. Л. Свердан. – К. : Знання, 2012. – 476 с.
14. Лоднар С. І. Економетрія засобами MS Excel : навчальний посібник / С. І. Лоднар, Р. В. Юринець – К. : Вид-во Європ. Ун-ту, 2004. – 242 с.
15. Лукьяненко І. Г. Економетрика : підручник / І. Г. Лукьяненко, Л. І. Краснікова – К. : Товариство «Знання», КОО, 1998. – 494 с.
16. Магнус Я.Р. Эконометрика Начальный курс. 3-е изд. / Я. Р. Магнус, П. К. Катыхев, А. А. Пересецкий. — М.: Дело, 1999.— 248 с.
17. Назаренко О. М. Основи економетрики : підручник вид. 2-ге, допов. та перероб. / О. М. Назаренко. – Київ : Центр навчальної літератури, 2005. – 395 с.
18. Наконечний С. І. Економетрія : підручник вид. 2-ге, допов. та перероб. / С. І. Наконечний, Т. О. Терещенко, Т. П. Романюк. — К. : КНЕУ, 2000. – 296 С.
19. Толбатов Ю.А. Економетрика / Ю.А. Толбатов Ю.А.— К.: Четверта хвиля, 1997.— 320 с
20. Четыркин Е.М. Статистические методы прогнозирования / Е. М. Четыркин. – М. : Финансы и статистика, 1977. –220 с.

ДОДАТКИ

ДОДАТОК 1

Алгоритм побудови вбудованої діаграми

1. Виділяємо область, яку потрібно зобразити на діаграмі (у випадку необхідності відмічаємо окремі блоки тримаємо клавішу Ctrl).
2. Курсор на Майстер діаграм ЛК.
3. Після активізації Майстра діаграм (курсор перетворюється на маленький хрестик) переносимо його в поле робочого аркуша, ЛК і не відпускаючи її розтягуємо прямокутник до бажаної величини графіка.
4. На екрані появляється діалогове вікно Майстра діаграм – крок 1. **Далее.**
5. На екрані діалогове вікно Майстра діаграм – крок 2. Вибираємо вид діаграми. **Далее.**
6. На екрані діалогове вікно Майстра діаграм – крок 3. Вибираємо вид формату. **Далее.**
7. На екрані діалогове вікно Майстра діаграм – крок 4. Вибираємо варіант представлення рядів даних.в цьому вікні демонструється зовнішній вигляд графіка. Назначаємо число рядків і стовбців, дані яких є мітками по осях. Якщо отриманий результат не задовольняє, можна повернутися до попереднього діалогу, натиснувши кнопку < **Назад.** **Далее.**
8. На екрані діалогове вікно Майстра діаграм – крок 5. Водимо назву діаграми, осей, вводимо легенду. Текст, який вводиться зразу демонструється на екрані. **Готово.**

При необхідності розмір одержаної діаграми можна змінити з допомогою маркерів, а також переміщувати по екрану, копіювати, коректувати вихідні дані.

Алгоритм дій з вбудованою діаграмою з маркерами

1. Зміна розміру:

- ◇ курсор наставляємо на маркер;
- ◇ після того як він набуде вигляду стрілочки, ЛН;
- ◇ відпускаємо коли досягнуто потрібних розмірів.

2. Переміщення по екрану:

- ◇ курсор наставляємо на поле діаграми;
- ◇ ЛН;
- ◇ Переміщуємо в потрібне місце.

3. Копіювання:

- ◇ курсор наставляємо на поле діаграми;
- ◇ <Ctrl> + ЛН;
- ◇ Переміщуємо в місце розміщення копії.

4. Коректування вихідних даних:

- ◇ курсор в комірку, де необхідно змінити дані;
- ◇ ввести нові дані;
- ◇ <Enter1>.

На побудованій діаграмі виникають зміни.

Алгоритм переходу від вбудованої діаграми до виділеної

1. Є вбудована діаграма з маркерами (без маркерів).
2. Наставляємо курсор всередині діаграми.
3. ЛК2.
4. Одержуємо виділену діаграму.

Навпаки

1. Є виділена діаграма.
2. ЛК1 (ЛК2).
3. Одержуємо вбудовану діаграму з маркерами (без маркерів).

Алгоритм форматування діаграми

1. Від вбудованої діаграми до виділеної (попередній алгоритм).
2. Заберемо рамку і фон:

- ◇ курсор на область побудови діаграми;
- ◇ М2;

На екрані: Форматування області побудови

- ◇ ввести:
 - рамка: невидимая
 - заливка: нет заливки

- ◇ ОК.;
- ◇ курсор поза виділеною областю

3. Вводимо мітки:

- ◇ курсор на будь-який стовбчик;
- ◇ ЛК;

На стовбцях появляються маркери

- ◇ **Вставка, Метки значений...**;
- ◇ Значення елементів ряду;
- ◇ ОК.;
- ◇ Курсор на кнопку Напівжирний
- ◇ ЛК
- ◇ Курсор в поле діаграми
- ◇ Забираємо виділення міток: ЛК
- ◇ Курсор поза діаграмою
- ◇ ЛК.

Алгоритм знаходження оберненої матриці

1. Відмічаємо поле, де буде знаходитись результат.

2. Входимо в майстер функцій. У категоріях вибираємо математику, а в функціях МОБР. Вводимо адресу матриці для якої знаходимо обернену. Натискаємо клавішу Готово.

3. Для того щоб отримати на екрані значення всієї оберненої матриці, натискаємо клавіші F2 Ctrl+Shift+Enter.

Алгоритм множення матриць

1. Відмічаємо поле, де буде знаходитись результат.

2. Входимо в майстер функцій. У категоріях вибираємо математику, а в функціях МУМНОЖ. Вводимо адресу 1-ої матриці в перший масив і 2-ої – в другий. Натискаємо клавішу Готово.

3. Для того щоб отримати на екрані значення всієї оберненої матриці, натискаємо клавіші F2 Ctrl+Shift+Enter.

Порядок знаходження оцінок параметрів економетричної моделі з використанням функції ЛИНЕЙН:

1. Відмічаємо блок, де мають знаходитись розрахункові дані. Висота цього блоку завжди дорівнює 5-ти рядкам, а ширина – числу оцінюваних параметрів.

2. В діалоговому вікні Майстер функцій, вибираємо категорію СТАТЕСТИЧЕСКИЕ функцію ЛИНЕЙН і натискаємо на кнопку Далее>.

3. В наступному діалоговому вікні вводимо: в перший рядок блок даних показника; в другий рядок блок даних факторів; в третій рядок вводиться слово ИСТИНА – якщо a_1 не дорівнює нулю, і слово ЛОЖЬ – якщо a_0 дорівнює нулю; в четвертий рядок вводиться слово ИСТИНА, якщо необхідно знайти не лише параметри регресії, а й додаткову регресійну статистику, і слово ЛОЖЬ якщо необхідно знайти лише параметри регресії. Після цього натискаємо кнопку ГОТОВО.

4. Для того щоб отримати на екрані значення всіх розрахункових даних, натискаємо клавіші F2 Ctrl+Shift+Enter.

Таблиця розрахункових значень має вигляд (для трьох оцінюваних параметрів):

Рядок 1	a_1	a_2	a_3
Рядок 2	σ_{a_1}	σ_{a_2}	σ_{a_3}
Рядок 3	r^2	σ_{xy}	#НД
Рядок 4	F	k	#НД
Рядок 5	$\sum (x - \bar{x})^2$	$\sum (y - \bar{y})^2$	#НД

В першому рядку знаходяться оцінки економетричної моделі.

В другому рядку знаходяться середні квадратичні відхилення оцінок параметрів.

В третьому рядку в першій комірці знаходиться коефіцієнт детермінації, а в другій – середнє квадратичне відхилення показника (стандартна похибка оцінки за рівнянням економетричної моделі).

В четвертому рядку в першій комірці знаходиться розрахункове значення F -статистики, в другій – k – число ступенів вільності.

В п'ятому рядку в першій комірці знаходиться сума квадратів відхилень розрахункових значень показника від його середнього значення, в другій комірці – сума квадратів відхилень розрахункових значень показника від статистичних.

Використання стандартної програми РЕГРЕССИЯ

1. Заходимо в меню Сервис, відкриваємо розділ Анализ данных, вибираємо функцію РЕГРЕССИЯ, Enter.
2. У діалоговому вікні вводимо: в першу комірку блок даних результативного показника Y , вказуючи діапазон комірок або ім'я блоку даних; в другу – блок факторів X або ім'я цього блоку.
3. Якщо відмічаємо вихідний інтервал, то вводимо діапазон вихідного блоку.
4. Якщо Новый рабочий лист, результати виводяться на окремому аркуші

Примітка: Результати розрахунку за цією програмою дають найбільшу кількість характеристик взаємозв'язку.

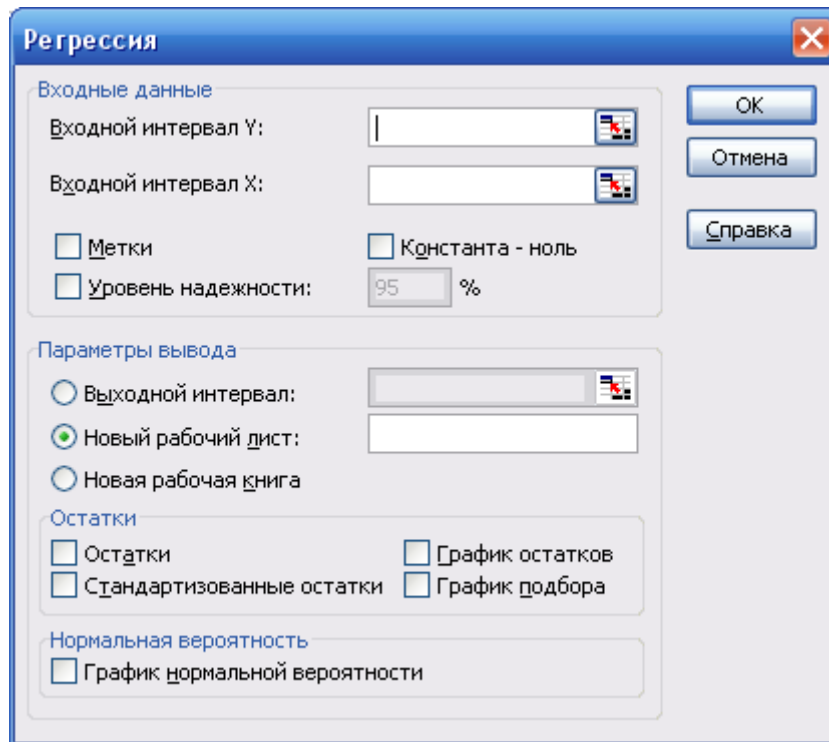


Рис.1. Диалогове вікно РЕГРЕССИЯ

Регресійна статистика:

«Множественный R» – коефіцієнт кореляції.

«R-квадрат» – коефіцієнт детермінації без врахування ступенів вільності.

«Нормированный R-квадрат» –

«Стандартная ошибка» – коефіцієнт детермінації з врахуванням ступенів вільності.

«Наблюдения» – кількість спостережень

Дисперсійний аналіз:

«df» – ступені вільності.

«SS» – суми квадратів відхилень: регресії, залишків та залежної змінної відповідно.

«MS» – виправлені дисперсії.

«Значимость F» – розрахункове значення статистики критерію Фішера.

Коефіцієнти:

Цей блок результатів містить 9 стовпчиків: 1-2, це назва та рівень оцінок параметрів моделі; 3-й – стандартні похибки оцінок параметрів моделі; 4-й –

критерії; 5-й – рівень значущості; інші стовпці з імовірність 0,95 визначають верхні та нижні границі оцінок параметрів моделі, в яких вони існують.

Вывод Итогов

Множеств	0,995062
R-квадрат	0,990149
Нормиров	0,988742
Стандартн	0,7233
Наблюден	9

Дисперсионный анализ

	df	SS	MS	F	значимость F
Регрессия	1	368,0868	368,0868	703,5804	2,77E-08
Остаток	7	3,662136	0,523162		
Итого	8	371,7489			

Коэффициентная статистика

	df	SS	MS	F	значимость F	верхние 95%	нижние 95%	
Y-пересеч		1,466808	-4,62851	0,002402	-10,2576	-3,32068	-10,2576	-3,32068
Переменн		0,039364	26,52509	2,77E-08	0,95106	1,137224	0,95106	1,137224

Рис.2 Диалогове вікно Вивід результатів

ДОДАТОК 2
ДОВІДКОВІ ТАБЛИЦІ

Розподіл Кочрена

Критичні точки критерію Кочрена при $\alpha = 0,05$.

$v \backslash k$	3	4	5	6	7	8	9	10
1	0,9669	0,9065	0,8412	0,7808	0,7271	0,6798	0,6385	0,6020
2	0,8709	0,7679	0,6338	0,6161	0,5612	0,5157	0,4775	0,4450
3	0,7977	0,6841	0,5981	0,5321	0,4800	0,4377	0,4027	0,3733
4	0,7457	0,6287	0,5440	0,4803	0,4307	0,3910	0,3584	0,3311
5	0,7071	0,5895	0,5063	0,4447	0,3974	0,3595	0,3286	0,3029
6	0,6771	0,5598	0,4783	0,4184	0,3726	0,3362	0,3067	0,2823
7	0,6530	0,5365	0,4564	0,3980	0,3535	0,3185	0,2901	0,2666
8	0,6333	0,5175	0,4387	0,3817	0,3384	0,3043	0,2768	0,2541
9	0,6167	0,5017	0,4241	0,3682	0,3259	0,2926	0,2659	0,2439
10	0,6025	0,4884	0,4118	0,3568	0,3154	0,2829	0,2568	0,2353
16	0,5466	0,4366	0,3645	0,3135	0,2756	0,2462	0,2226	0,2032
36	0,4748	0,3720	0,3066	0,2612	0,2278	0,2022	0,1820	0,1655
144	0,4031	0,3093	0,2013	0,2119	0,1833	0,1616	0,1446	0,1308
∞	0,3333	0,2500	0,2000	0,1667	0,1429	0,1250	0,1111	0,1000

Розподіл Кочрена

Критичні точки критерію Кочрена при $\alpha = 0,01$.

$v \backslash k$	3	4	5	6	7	8	9	10
1	0,9933	0,9676	0,9279	0,8828	0,8376	0,7945	0,7544	0,7175
2	0,9423	0,8643	0,7885	0,7218	0,6644	0,6152	0,5727	0,5358
3	0,8831	0,7814	0,6957	0,6258	0,5685	0,5209	0,4810	0,4469
4	0,8335	0,7212	0,6329	0,5635	0,5080	0,4627	0,4251	0,3934
5	0,7933	0,6761	0,5875	0,5195	0,4659	0,4226	0,3870	0,3572
6	0,7606	0,6410	0,5531	0,4866	0,4347	0,3932	0,3592	0,3308
7	0,7335	0,6129	0,5259	0,4608	0,4105	0,3704	0,3378	0,3106
8	0,7107	0,5897	0,5037	0,4401	0,3911	0,3522	0,3207	0,2945
9	0,6912	0,5702	0,4854	0,4229	0,3751	0,3373	0,3067	0,2813
10	0,6743	0,5536	0,4697	0,4084	0,3616	0,3248	0,2950	0,2704
16	0,6059	0,4884	0,4094	0,3529	0,3105	0,2779	0,2514	0,2297
36	0,5153	0,4057	0,3351	0,2858	0,2494	0,2214	0,1992	0,1811
144	0,4230	0,3251	0,2644	0,2229	0,1929	0,1700	0,1521	0,1376
∞	0,3333	0,2500	0,2000	0,1667	0,1429	0,1250	0,1111	0,1000

Приклад. Якщо кожна з п'яти вибірок ($k=5$) має по десять об'єктів ($v=9$), тоді на рівні значущості $\alpha=0,01$ маємо: $C^*=C(\alpha=0,01; v=9; k=5)=0,485$

Таблиця t – розподілу Стюдента

Ступені вільно- сті	Довірча ймовірність						
	0,5	0,8	0,9	0,95	0,98	0,99	0,998
1	1,000	3,078	6,314	12,706	31,821	63,657	318,31
2	0,861	1,886	2,920	4,303	6,965	9,925	22,327
3	0,765	1,638	2,353	3,182	4,541	5,841	10,214
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893
6	0,718	1,440	1,943	2,227	3,143	3,707	3,208
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646
18	0,688	1,330	1,434	2,101	2,552	2,878	3,610
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408
29	0,683	1,311	1,699	2,045	2,462	2,756	3,396
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307
60	0,679	1,296	1,671	2,000	2,390	2,660	3,232
120	0,677	1,289	1,658	1,980	2,358	2,167	3,160
∞	0,674	1,282	1,645	1,960	2,326	2,576	3,090

Таблиця χ^2 - розподілу ($\alpha=0,05$)

Ступені вільності	0,95	Ступені вільності	0,95
1	3,84	11	19,68
2	5,99	12	21,03
3	7,82	13	22,36
4	9,49	14	23,69
5	11,07	15	25,00
6	12,59	16	26,30
7	14,07	17	27,59
8	15,51	18	28,57
9	16,92	19	30,14
10	18,31	20	31,41

Таблиця критерію Дарбіна-Уотсона
(m - кількість незалежних змінних, $\alpha = 0,05$)

n	$m = 1$		$m = 2$		$m = 3$		$m = 4$	
	d_1	d_2	d_1	d_2	d_1	d_2	d_1	d_2
6	0.610	1.400						
7	0.700	1.356	0.467	1.896				
8	0.763	1.332	0.559	1.777	0.368	2.287		
9	0.824	1.320	0.629	1.699	0.455	2.128	0.296	2.588
10	0.879	1.320	0.697	1.641	0.525	2.016	0.376	2.414
11	0.927	1.324	0.658	1.604	0.595	1.928	0.444	2.283
12	0.971	1.311	0.812	1.579	0.658	1.864	0.512	2.177
13	1.010	1.340	0.861	1.562	0.715	1.816	0.574	2.094
14	1.045	1.350	0.905	1.551	0.767	1.779	0.632	2.030
15	1.077	1.361	0.946	1.543	0.814	1.750	0.685	1.977
16	1.106	1.371	0.982	1.539	0.857	1.728	0.734	1.935
17	1.133	1.381	1.015	1.536	0.897	1.710	0.779	1.900
18	1.158	1.391	1.046	1.535	0.933	1.696	0.820	.872
19	1.180	1.401	1.074	1.536	0.967	1.685	0.859	.848
20	1.201	1.411	1.100	1.537	0.998	1.676	0.894	.828
21	1.221	1.420	1.125	1.538	1.026	1.669	0.927	.812
22	1.239	1.429	1.147	1.541	1.053	1.664	0.958	.797
23	1.257	1.437	1.168	1.543	1.078	1.660	0.986	.785
24	1.273	1.446	1.188	1.546	1.101	1.656	1.013	.775
25	1.288	1.454	1.206	1.550	1.123	1.654	1.038	1.767
26	1.302	1.461	1.224	1.553	1.143	1.652	1.062	1.759
27	1.316	1.469	1.240	1.556	1.162	1.651	1.084	1.753
28	1.328	1.476	1.225	1.560	1.181	1.650	1.104	1.747
29	1.341	1.483	1.270	1.563	1.198	1.650	1.124	1.743
30	1.352	1.489	1.284	1.567	1.214	1.650	1.143	1.739
31	1.363	1.496	1.297	1.570	1.229	1.650	1.160	1.735

Критичні значення коефіцієнта кореляції

Критичні значення $r^* = r(\alpha; i)$ визначені при $t^*_2 = t(p = 1 - \frac{\alpha}{2}; v = n-2)$ за

формулою $r^* = t^*_2 / \sqrt{n-2 + (t^*_2)^2}$

Об'єм вибірки	Рівень значущості					
	0,1	0,05	0,02	0,01	0,005	0,001
4	0,9000	0,9500	0,9800	0,9900	0,9950	0,9990
5	0,8054	0,8783	0,9343	0,9587	0,9740	0,9911
6	0,7293	0,8114	0,8822	0,9172	0,9417	0,9741
7	0,6694	0,7545	0,8329	0,8745	0,9056	0,9509
8	0,6215	0,7067	0,7887	0,8343	0,8697	0,9249
9	0,5822	0,6664	0,7498	0,7977	0,8359	0,8983
10	0,5494	0,6319	0,7155	0,7646	0,8046	0,8721
11	0,5214	0,6021	0,6851	0,7348	0,7759	0,8470
12	0,4973	0,5760	0,6581	0,7079	0,7496	0,8233
13	0,4762	0,5529	0,6339	0,6835	0,7255	0,8010
14	0,4575	0,5324	0,6120	0,6614	0,7034	0,7800
15	0,4409	0,5140	0,5923	0,6411	0,6831	0,7604
16	0,4259	0,4973	0,5742	0,6226	0,6643	0,7419
17	0,4124	0,4821	0,5577	0,6055	0,6470	0,7247
18	0,4000	0,4683	0,5425	0,5897	0,6308	0,7084
19	0,3887	0,4555	0,5285	0,5751	0,6158	0,6932
20	0,3783	0,4438	0,5155	0,5614	0,6018	0,6788
21	0,3687	0,4329	0,5034	0,5487	0,5886	0,6652
22	0,3598	0,4227	0,4921	0,5368	0,5763	0,6524
27	0,3233	0,3809	0,4451	0,4869	0,5243	0,5974
32	0,2960	0,3494	0,4093	0,4487	0,4840	0,5541
37	0,2746	0,3246	0,3810	0,4182	0,4518	0,5189
42	0,2573	0,3044	0,3578	0,3932	0,4252	0,4896
47	0,2429	0,2876	0,3384	0,3721	0,4028	0,4647
52	0,2306	0,2732	0,3218	0,3542	0,3836	0,4432
62	0,2108	0,2500	0,2948	0,3248	0,3522	0,4079
72	0,1954	0,2319	0,2737	0,3017	0,3274	0,3798
82	0,1829	0,2172	0,2565	0,2830	0,3072	0,3568
92	0,1726	0,2050	0,2422	0,2673	0,2903	0,3375
102	0,1638	0,1946	0,2301	0,2540	0,2759	0,3211
122	0,1496	0,1779	0,2104	0,2324	0,2526	0,2943
142	0,1386	0,1648	0,1951	0,2155	0,2343	0,2733

Приклад. Для $\alpha = 0,01; n = 11$ маємо: $r^* = r(\alpha = 0,01; n = 11) = 0,735$.

Критичні значення коефіцієнта автокореляції

Об'єм вибірки	Додатна автокореляція		Від'ємна автокореляція	
	$\alpha = 0,05$	$\alpha = 0,01$	$\alpha = 0,05$	$\alpha = 0,01$
5	0,253	0,297	-0,753	-0,798
6	0,345	0,447	-0,708	-0,863
7	0,370	0,510	-0,674	-0,799
8	0,371	0,531	-0,625	-0,764
9	0,366	0,533	-0,593	-0,737
10	0,360	0,525	-0,564	-0,705
11	0,353	0,515	-0,539	-0,679
12	0,348	0,505	-0,516	-0,655
13	0,341	0,495	-0,497	-0,634
14	0,335	0,485	-0,479	-0,615
15	0,328	0,475	-0,462	-0,597
20	0,299	0,432	-0,399	-0,524
25	0,276	0,398	-0,356	-0,473
30	0,257	0,370	-0,324	-0,433
35	0,242	0,347	-0,300	-0,401
40	0,229	0,329	-0,279	-0,376
45	0,218	0,313	-0,262	-0,356
50	0,208	0,301	-0,248	-0,339
60	0,191	0,278	-0,225	-0,310
70	0,178	0,259	-0,207	-0,287
80	0,170	0,246	-0,201	-0,271
90	0,161	0,233	-0,184	-0,255
100	0,154	0,221	-0,174	-0,242

Приклад. Для об'єму вибірки $n = 15$ та рівня значущості $\alpha = 0,05$ при додатній автокореляції маємо: $r^* = r(n = 15; \alpha = 0,05) = 0,335$.

Таблиця F – розподілу Фішера (k_1, k_2 - ступені вільності, $\alpha=0,05$)

$k_2 \backslash k_1$	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,45	19,46	19,47	19,48	19,49	19,50
3	10,13	9,55	9,28	9,1	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,931
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,711
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,29	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,64	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00

Навчальне видання

Руська Руслана Василівна

ЕКОНОМЕТРИКА

Навчальний посібник