О.В. Козьменко О.В. Кузьменко

ЕКОНОМІКО-МАТЕМАТИЧНІ МЕТОДИ ТА МОДЕЛІ (ЕКОНОМЕТРИКА)

Навчальний посібник

Рекомендовано Міністерством освіти і науки України
як навчальний посібник

Суми
Університетська книга
2014

УДК 330.43(075.8)
ББК 65в631я73
К 59

Рекомендовано вченою радою Державного вищого навчального закладу „Українська академія банківської справи Національного банку України." Протокол № 8 від 12 квітня 2013 р.

Рецензенти:
Клебанова Т.С., доктор економічних наук, професор, завідувач кафедри економічної кібернетики Харківського національного економічного університету імені Семена Кузнеця;
Чаплига В.М., доктор технічних наук, професор, завідувач кафедри економічної кібернетики Львівського інституту банківської справи Університету банківської справи НБУ, заслужений працівник освіти України.
Черняк О.І., доктор економічних наук, професор, завідувач кафедри економічної кібернетики Київського національного університету імені Тараса Шевченка

Гриф надано Міністерством освіти і науки України.
Лист № 1/11-12833 від 12.08.2014 р.

Козьменко О. В.
К 59 Економіко-математичні методи та моделі (економетрика) : навчальний посібник / О. В. Козьменко, О. В. Кузьменко. – Суми : Університетська книга, 2014. – 406 с.

ISBN 978-966-680-722-2

У навчальному посібнику комплексно викладено теоретико-методичні засади та практичні аспекти статистичного вивчення явищ і процесів фінансового сектору з урахуванням міжнародного й вітчизняного досвіду. Розглянуто методи оцінювання параметрів класичної економетричної моделі. Висвітлено методи оцінювання параметрів узагальненої моделі та динамічних економетричних моделей. Наведено підходи до визначення параметрів економетричних моделей, які побудовані на основі системи одночасних структурних рівнянь.

Для студентів вищих навчальних закладів економічного спрямування, аспірантів, викладачів, науковців, фахівців-практиків та читачів, які цікавляться застосуванням економіко-математичного інструментарію для вирішення сучасних проблем в економіці.

УДК 330.43(075.8)
ББК 65в631я73

ISBN 978-966-680-722-2

Зміст

Вступ

У сучасних умовах розвитку вітчизняних економіки важливого значення набувають дослідження проблем ринку фінансових послуг, соціальної політики, ціноутворення, попиту та пропозиції тощо, де взаємозв'язки релевантних показників часто неочевидні та мінливі. Крім того, прийняття адекватних управлінських рішень в економічній діяльності, а також розроблення повноцінної інформаційної бази для отримання нових знань з економіки неможливі без поєднання економічної теорії та математико-статистичного інструментарію. З огляду на це, слід зауважити, що важливу роль у реалізації зазначених вище завдань у процесі професійної підготовки майбутніх фахівців (бакалаврів) з економічних спеціальностей відіграє дисципліна «Економіко-математичні методи та моделі (економетрика)».

Методи економетрії є найсучаснішими засобами аналізу та дослідження різних соціально-економічних систем. За допомогою економетричних методів можна відхилити деякі економічні гіпотези або показати неможливість застосовування їх у конкретних умовах. Оволодіння базовим інструментарієм економетрії дозволяє обґрунтовано прогнозувати розвиток економічних систем, оцінювати вплив управлінських рішень чи урядових постанов щодо зміни цін, податків тощо на фінансовий стан будь-якого суб'єкта господарювання, розробляти шляхи ефективного управління ним та формувати ефективні стратегічні програми розвитку.

У процесі професійної підготовки фахівців галузі «Економіка і підприємництво» важливим є навчитися аналізувати інформаційні потоки в соціально-економічних системах, прогнозувати їх поведінку, оцінювати та будувати економетричні моделі різного рівня. Зазначені вміння є підґрунтям формування професійної компетентності майбутніх фахівців з фінансів, обліку та аудиту, міжнародної економіки та економічної кібернетики.

Застосування різноманітних економетричних моделей на різних рівнях економічної діяльності дає можливість розв'язувати фінансові 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 бюджетно-податкова політика, політика у сфері страхування та грошова-кредитна політика. Узгодженість усіх складових фінансової 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 впливу інших деструктивних чинників.

На мікрорівні економетричні дослідження передбачають наукове обґрунтування управлінських рішень, що приймаються на підприємствах різних форм власності й мають ураховувати постійний вплив зміни зовнішнього та внутрішнього середовища. Моделі можуть використовуватися для аналізу економічних і соціально-економічних показників, що характеризують відповідну економічну систему для прогнозування їх подальшої трансформації або для імітації можливих сценаріїв соціально-економічного розвитку досліджуваної системи за умови, що деякі показники можна змінювати цілеспрямовано. Зазначені аспекти наголошують важливу роль дисципліни «Економіко-математичні методи та моделі (економетрика)» у формуванні бакалаврів з економічних спеціальностей.

Навчальний посібник «Економіко-математичні методи та моделі (економетрика)» складається із десяти тем, кожна з яких містить опис методів кількісного вимірювання взаємозв'язків між економічними показниками, а також їх застосування в економічних дослідженнях і практичній діяльності суб'єктів господарювання та державних органів влади.

Так, перша тема «Предмет, методи і завдання дисципліни «економетрія»» присвячена висвітленню сукупності теоретичних підходів, 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 розвитку економіки.

Питання формального опису найважливіших та найсуттєвіших зв'язків економічних змінних і об'єктів та адекватного відображення тверджень економічної теорії, формулювання її понять та висновків висвітлені в таких трьох темах, як «Методи побудови загальної лінійної економетричної моделі», «Мультиколінеарність та її вплив на оцінки параметрів моделі» та «Узагальнені економетричні моделі».

П'ята («Економетричні моделі динаміки»), шоста («Економетричні методи кількісного аналізу на основі статистичних рівнянь») та сьома («Побудова економетричної моделі з автокореольованими залишками») теми містять підходи, що дозволяють прогнозувати поведінку економічних систем.

Застосування різноманітних економетричних моделей на різних рівнях економічної діяльності дає можливість розв'язувати економічні проблеми різного рівня складності шляхом використання інструментарію, наведеного в таких темах, як «Методи інструментальних змінних», «Моделі розподіленого лагу», «Економетричні моделі на основі системи структурних рівнянь».

Так, структуру розробленого навчального посібника складають зміст, вступ, навчальна програма комплексу, основний текст за темами та практичні заняття, запитання для самоконтролю, тести, питання до іспиту, словник термінів та список рекомендованої літератури. Слід зазначити, що практичні заняття розроблені до всіх тем та містять завдання для самостійної роботи, методику розв'язку типових задач. Виклад матеріалу в навчальному посібнику характеризується об'єктивністю, науковістю і чіткою логічною послідовністю. Крім того, переважна більшість практичних завдань побудована авторами на основі реальної статистичної інформації, що дозволить майбутнім фахівцям економіки при їх розв'язанні ознайомитися з фактичним станом справ у фінансовому секторі України, банківській системі та міжнародній економіці, виявити основні проблеми та визначити основні тенденції подальшого розвитку.

Навчальна програма курсу

Тема 1. Предмет, методи і завдання дисципліни «Економетрія»

Основні завдання економетрії. Роль економетричних досліджень в економіці. Предмет, цілі, завдання та структура курсу. Місце і значення курсу серед дисциплін фундаментальної підготовки фахівців з економіки. Зв'язки курсу із суміжними дисциплінами. Особливості економетричних моделей. Вибір змінних і структура зв'язків. Роль і місце економетричних моделей у математичному моделюванні. Приклади економетричних моделей: виробнича функція Кобба – Дугласа; моделі пропозиції і попиту на конкурентному ринку, модель Кейнса; модель споживання.

Тема 2. Методи побудови загальної лінійної економетричної моделі

Загальна лінійна економетрична модель. Емпірична модель множинної лінійної регресії. Зведення нелінійних економетричних моделей до лінійного вигляду. Метод найменших квадратів. Оператор оцінювання 1МНК. Передумови застосування методу найменших квадратів (1МНК) – умови Гауса – Маркова. Верифікація моделі. Перевірка значущості та довірчі інтервали. Етапи дослідження загальної лінійної моделі множинної регресії. Прогнозування за лінійною моделлю. Методи побудови багатофакторної регресійної моделі.

Тема 3. Мультиколінеарність та її вплив на оцінки параметрів моделі

Поняття мультиколінеарності. Її вплив на оцінки параметрів моделі. Ознаки мультиколінеарності. Алгоритм Фаррара – Глобера. Методи усунення мультиколінеарності. Метод головних компонент.

Тема 4. Узагальнені економетричні моделі

Моделі з порушенням передумов використання звичайного методу найменших квадратів. Узагальнений метод найменших квадратів. Суть гетероскедастичності. Гетероскедастичність і зважений метод найменших квадратів.

Тема 5. Економетричні моделі динаміки

Основні поняття і попередній аналіз рядів динаміки. Поняття часового ряду. Основні характеристики динаміки часового ряду. Систематичні та випадкові компоненти часового ряду. Перевірка гіпотези про існування тренда. Методи фільтрації сезонної компоненти. Проблеми аналізу сезонності (та/або циклічності). Фільтрація сезонної компоненти за допомогою індексу сезонності. Метод декомпозиції часового ряду. Методи прогнозування часових рядів. Методи соціально-економічного прогнозування. Прогнозування тенденцій часового ряду за середніми характеристиками. Прогнозування тенденцій часового ряду за механічними методами. Прогнозування тенденцій часового ряду за аналітичними методами.

Тема 6. Економетричні методи кількісного аналізу на основі статистичних рівнянь

Вибір форми моделі. Ознаки "гарної" моделі. Види помилок специфікації. Виявлення та коригування помилок специфікації. Дослідження залишкового члена моделі. Проблеми специфікації. Сутність, типологізація та прикладна спрямованість завдань класифікації об'єктів. Класифікація з навчанням.

Тема 7. Побудова економетричної моделі з автокореольованими залишками

Природа і наслідки автокореляції. Методи визначення автокореляції. Критерій Дарбіна – Уотсона. Критерій фон Неймана. Коефіцієнти автокореляції та їх застосування. Моделі з автокорельованими залишками. Метод оцінювання параметрів Ейткена. Метод Кочрена – Оркатта. Метод перетворення вихідної інформації. Метод Дарбіна.

Тема 8. Методи інструментальних змінних

Сутність методу інструментальних змінних. Оператор оцінювання Вальда. Особливості оцінювання методом Бартлета. Оператор оцінювання Дарбіна. Побудова економетричної моделі на основі даних, що містять помилки вимірювання змінних. Оцінка параметрів моделі за допомогою алгоритму Уоліса.

Тема 9. Моделі розподіленого лага

Поняття лага і лагових змінних. Моделі розподіленого лага. Взаємна кореляційна функція. Оцінка параметрів моделей з лагами в незалежних змінних: метод послідовного збільшення кількості лагів, перетворення Койка (метод геометричної прогресії). Оцінювання параметрів авторегресійних моделей. Виявлення автокореляції залишків в авторегресійних моделях. Авторегресійне перетворення. Перетворення методом ковзного середнього. Перетворення ARMA і ARIMA.

Тема 10. Економетричні моделі на основі системи структурних рівнянь

Системи одночасних структурних рівнянь. Структурна і приведена форми моделі. Проблеми ідентифікації. Ідентифікована, неідентифікована і надідентифікована системи рівнянь. Проблема оцінювання параметрів системи, загальна характеристика методів. Непрямий метод оцінювання параметрів ідентифікованої системи рівнянь. Двокроковий метод найменших квадратів оцінювання параметрів надідентифікованих систем одночасних рівнянь (2МНК-оцінка). Трикроковий метод найменших квадратів. Рекурсивні системи одночасних рівнянь, їх характеристика, можливість застосування МНК-оцінки для розрахунку параметрів рекурсивних систем. Прогноз і загальні довірчі інтервали.

ТЕМА 1

Предмет, методи і завдання дисципліни «Економетрія»

Основні завдання економетрії ● Роль економетричних досліджень в економіці ● Предмет, цілі, завдання та структура курсу ● Місце і значення курсу серед дисциплін фундаментальної підготовки фахівців з економіки ● Взаємозв'язки курсу із суміжними дисциплінами ● Особливості економетричних моделей ● Вибір змінних і структура зв'язків ● Роль і місце економетричних моделей в математичному моделюванні ● Приклади економетричних моделей: виробнича функція Кобба – Дугласа; моделі пропозиції і попиту на конкурентному ринку; модель Кейнса; модель споживання.

1.1. Основні завдання економетрії

Термін «економетрика» (економетрія) був уведений у словесний обіг ще в 1910 р. П. Чомпою в праці «Нарис економетрії і природничої бухгалтерії, основаної на політичній економії», яка була надрукована в м. Львові.

У подальшому цьому термінові значну увагу приділяв норвезький вчений Р. Фріш (1895–1973), який наголосив:

> ***Економетрія*** є синтезом економічної теорії, математики і статистики.

Офіційною датою народження нового напряму економічних досліджень вважають 1931 р., коли було створено «Міжнародне товариство розвитку економічної теорії в її зв'язку зі статистикою і математикою». У 1933 р. це товариство почало видавати журнал «Економетрика».

Поняття «економетрика» було загальноприйнятим терміном тільки в період його зародження. У подальшому набули активного розвитку окремі напрями економетрики (рис. 1.1).

Рис. 1.1. Напрями історичного розвитку економетрики

Економетрія пов'язана з науковою діяльністю таких видатних вчених, лауреатів Нобелівської премії, як Р. Фішер, Я. Тімберген, В. Леонтьєв, Т. Кумпанс. Варто також нагадати про внесок в економетрію вчених: В.К. Оштрієва, В.І. Борткевича, Н.А. Столярова, Н.Н. Шапошникова, Е.Е. Слуцького, Л.В. Канторовича.

Отже, розглянемо поширені тлумачення терміна «економетрія» в сучасній економічній літературі, працях та доробках вітчизняних і закордонних теоретиків і практиків (табл. 1.1).

Таблиця 1.1. Тлумачення терміна «економетрія» в сучасній економічній літературі

Визначення	Автор	Джерело
Економетрія (**економетрика**) – наука, яка вивчає кількісні та якісні економічні взаємозв'язки з використанням математичних і статистичних методів та моделей.	–	Большая советская энциклопедия. – 3-е изд. – М. : Сов. энциклопедия, 1978. – Т. 28. Чаган-Экс-ле-Бен. – 640 с.
Економетрія (інший варіант терміна – економетрика) – це інструментальна наука, що дозволяє вивчати кількісні взаємозв'язки економічних об'єктів і процесів за допомогою математичних і статистичних методів і моделей. Дослівно цей термін означає «економічний вимір».	В.І. Суслов	Эконометрия / В. И. Суслов, Н. М. Ибрагимов, Л. П. Талышева, А. А. Цыплаков. – Новосибирск : Изд-во «Новосибирский государственный университет», 2005. – 742 с.

Продовження табл. 1.1

Визначення	Автор	Джерело
Економетрика є одночасно нашим телескопом і нашим мікроскопом для вивчення навколишнього економічного світу.	З. Грилліхес	Эконометрика : учебник / И. И. Елисеева, С. В. Курышева, Т. В. Костеева и др. ; под ред. И. И. Елисеевой. – 2-е изд., перераб. и доп. – М. : Финансы и статистика, 2007. – 576 с.
Економетрика дозволяє проводити кількісний аналіз реальних економічних явищ, базуючись на сучасному розвитку теорії і спостереженнях, пов'язаних з методами отримання висновків.	П. Самуэльсон	
Основне **завдання економетрики** – наповнити емпіричним змістом апріорні економічні міркування».	В. Клейн	
Мета економетрики – емпіричний висновок економічних законів. Економетрика доповнює теорію, використовуючи реальні дані для перевірки і уточнення постульованих відносин.	Є. Маленво	Маленво Э. Статистические методы эконометрии / Э. Маленво. – М. : Статистика, 1975–1976. – Т. 1, 2.
Курс **економетрики** переслідує дві основні мети. Перша – необхідно показати, як різні кількісні методи можуть бути використані для моделювання статистичних даних. Друга – виробити розуміння статистичних властивостей цих моделей, а також того, чому вони працюють в одних випадках, але не працюють в інших	К. Доугерті	Доугерти К. Введение в эконометрику. – М. : Инфра-М; 1997 (перепечатка в 1999).
Економетрика – набір статистичних методів, що використовуються для спостереження за ходом розвитку економіки, її аналізу і прогнозів.	Я.Р. Магнус, П.К.Катишев, А.А. Пересецький	Магнус Я. Р. Эконометрика : начальный курс / Я. Р. Магнус, П. К. Катишев, А. А. Пересецкий. – М. : Дело, 1998 (второе расширенное издание 2000).
Економетрика – економіко-математична наукова дисципліна, що розробляє та використовує методи, моделі, прийоми, що дозволяють надати конкретне кількісне вираження загальним (якісним) закономірностям економічної теорії на базі економічної статистики і з використанням математико-статистичного інструментарію.	С.А. Айвазян, В.С. Мхитарян	Айвазян С. А. Прикладная статистика и основы эконометрики / Айвазян С. А., Мхитарян В. С. – М. : Юнити, 1998.

Продовження табл. 1.1

Визначення	Автор	Джерело
Визначення економетрики як галузі економіки, що стосується емпіричної оцінки економічних відносин, моделей, разом з даними, які є основними елементами будь-якого економетричного дослідження. Як правило, теорія розглянутих явищ розвивається і в подальшому доопрацьовується в економетричній моделі. Ця модель потім оцінюється на основі даних, що відносяться до досліджуваних явищ за допомогою економетричних методів. Розрахункова модель може бути використана для різних цілей, у тому числі структурного аналізу, прогнозування та оцінки політики.	Kennedy P.	Kennedy P. A Guide to Econometrics. – / P. Kennedy. – Third edition. – M.I.T. Press : Cambridge, MA, 1992.
Економетрика надає емпіричного змісту економічній теорії. Головною суттю математичної економіки є економічна теорія в математичній формі без урахування одиниць вимірювання або емпіричної перевірки теорії. Економетрика є, в основному, зацікавлена в емпіричній верифікації економічної теорії. В економетриці часто використовуються математичні рівняння, але ці рівняння представляються в такій формі, щоб вони піддавались емпіричній перевірці. І це перетворення математичних рівнянь в економетричні вимагає великої винахідливості і практичних навичок.	Damodar N. Gujarati	Damodar N. Gujarati. Basic Econometrics / Damodar N. Gujarati. – McGrow-Hill, 1995.
Економетрика, як результат певного погляду на роль економіки, являє собою застосування математичної статистики в економічних даних з метою емпіричного обґрунтування моделей, побудованих за допомогою економіко-математичних методів і отримання чисельних результатів.	Gerhard Tintner	Gerhard Tintner. Methodology of Mathematical Economics and Econometrics / Gerhard Tintner. – Chicago : The University of Chicago Press, 1968. – P. 74.
Економетрика може бути визначена як кількісний аналіз реальних економічних явищ на основі поступального розвитку теорії і спостережень, пов'язаних за допомогою належних методів логічного виводу.	P. A. Samu-elson	Samuelson P. A. Report of the Evaluative Committee for Econometrica / P. A. Samuelson, T. C. Koopmans, and J. R. N. Stone // Econometrica. – 1954. – Vol. 22, no. 2. – P. 141–146.

Продовження табл. 1.1

Визначення	Автор	Джерело
Економетрика може бути визначена як соціальна наука, в якій інструменти економічної теорії, математичний і статистичний висновок застосовуються для аналізу економічних явищ.	Arthur S.	Arthur S. Goldberger, Econometric Theory / S. Arthur. – New York : John Wiley & Sons, 1964. – P. 1.
Економетрика стосується емпіричного визначення економічних законів.	H. Theil	Theil H. Principles of Econometrics / H. Theil. – New York : John Wiley & Sons, 1971. – P. 1.
Мистецтво **економетристів** перш за все полягає у визначенні набору припущень, які є достатньо визначеними і достатньо реалістичними, щоб побудувати модель на основі доступних статистичних даних.	E. Malinvaud	Malinvaud E., Statistical Methods of Econometrics / E. Malinvaud. – Chicago : Rand McNally, 1966. – P. 514.
Econometricians are a positive help in trying to dispel the poor public image of economics (quantitative or otherwise) as a subject in which empty boxes are opened by assuming the existence of can-openers to reveal contents which any ten economists will interpret in 11 ways. **Економетрика** – позитивна допомога у спробі розвіяти непрестижність економіки (кількісно або іншим шляхом).	Adrian C. Darnell, J. Lynne Evans	Darnell Adrian C. The Limits of Econometrics / Adrian C. Darnell and J. Lynne Evans. – Hants, England : Edward Elgar Publishing, 1990. – P. 54.
Метод **економетричного дослідження** спрямований, по суті, на поєднання економічної теорії і реальних вимірювань, використовуючи теорії і методики статистичного виводу.	T. Haavelmo	Haavelmo T. The Probability Approach in Econometrics / T. Haavelmo // Supplement to Econometrica. – Vol. 12. 1944 (preface p. Iii).

Грунтований аналіз табл.1.1 надає можливість систематизувати існуючі підходи до визначення економетрії та сформулювати такі висновки.

Економетрія в широкому розумінні є сукупністю різноманітних економічних досліджень, що здійснюються з використанням математичних методів. Поряд з економіко-математичними дослідженнями економетрія охоплює також усі сфери застосування математичних методів для розв’язання прикладних економічних завдань.

Економетрія у вузькому розумінні – це використання статистичних методів в економічних дослідженнях, а саме: побудова математико-статистичних моделей економічних процесів, оцінка параметрів моделей.

Якщо результати економічної теорії мають якісний зміст, то економетрія привносить у них емпіричну суть. Якщо математична економіка виражає економічні закони у вигляді математичних співвідношень, то економетрія здійснює статистичну перевірку цих законів, використовуючи емпіричну інформацію. Одержані математичними методами і виражені мовою математики результати лише тоді мають цінність, якщо їх можна інтерпретувати мовою економіки.

Економетрія використовує традиційні математико-статистичні та спеціально розроблені методи для виявлення кількісних взаємозв'язків між економічними показниками. Економетрія має економічну та математичну складові, причому економічній складовій надається перевага.

Важливою проблемою економіки є правильне прогнозування певної реальної ситуації, в якій відбувається досліджуваний економічний процес, а також знаходження таких важелів впливу на цей процес, за допомогою яких він розвивався б необхідним чином. На практиці часто виникають ситуації, коли розв'язуючи одну й ту саму економічну проблему, її дослідники можуть пропонувати різні, часом навіть протилежні методи її вирішення. Тому досвідчений політик, керівник підрозділу, галузі чи підприємства спирається, як правило, на інтуїцію, обираючи певну стратегію розв'язання проблеми, щоб одержати бажаний результат. Поганий чи гарний вибір керівника – перевірити практично неможливо, оскільки економічна ситуація є функціональною залежністю від часу, тобто вона є динамічною, і ніколи не може повторюватися. Отже, експериментально неможливо перевірити кінцеві результати розвитку подій за кількома запропонованими стратегіями в одних і тих самих умовах.

Розглянемо основні завдання економетрії, наведені в табл.1.2.

Отже***, першим з основних завдань*** економетрії є дослідження розвитку економічних процесів і прогнозування їх динаміки. Вдалим чи невдалим буде цей прогноз – залежатиме від того, чи поталанить дослідникові виявити рушійні чинники, які впливають на ці процеси і які не завжди можна визначити. Урахування цих факторів у математичних моделях дає можливість раціонально керувати ними, а отже, досягти окресленої мети.

Таблиця 1.2. Основні завдання економетрії

№ пор.	Сутність	Характеристика
1.	Дослідження розвитку економічних процесів і прогнозування їх динаміки	Якість прогнозу; виявлення ключових чинників, що впливають на досліджувані процеси; раціональне керування даними факторами в математичних моделях з метою досягнення визначеної мети
2.	Правильний вибір факторів при побудові математико-статистичних моделей	Вибір основних факторів здійснюється, щоб питома вага решти факторів, які будуть неврахованими в моделі, була несуттєвою; виконання вимоги незначних відхилень поведінки модельованої системи (процесу) порівняно з реальною; регулювання економічних параметрів, щоб у майбутньому прийняти правильне рішення на мікро-, мезорівні чи в масштабах усієї країни
3.	Вибір та побудова математико-статистичної моделі, здійснення низки модельних експериментів, аналіз одержаних результатів і перенесення їх на реальну економічну систему (процес) як основу для прийняття належних управлінських рішень	Врахування невідповідності між показниками економічних величин; аналіз маловивчених і нестабільних зв'язків; використання апарату математичних моделей, ймовірнісні та статистичні методи аналізу параметрів цих моделей

Будь-який економічний процес характеризується певними економічними показниками (параметрами), значення яких залежать від великої кількості факторів, які впливають на них і які практично врахувати всі неможливо. Проте в кожній конкретній ситуації з усієї множини цих факторів суттєвий вплив на економічні параметри процесу має лише певна обмежена кількість факторів.

Тому ***другим завданням*** економетрії є правильний вибір факторів при побудові математико-статистичних моделей. Бажано, щоб питома вага решти факторів, які будуть неврахованими в моделі, була настільки несуттєвою, що ігнорування їх у процесі побудови моделі не призводило до значних відхилень поведінки модельованої системи (процесу) порівняно з реальною.

На сьогодні економічна теорія дослідила та вивчила значну кількість стабільних зв'язків між показниками економічних систем (процесів). Так, наприклад, добре вивчені такі зв'язки: між попитом споживача на продукцію та сумою коштів, що він може витрачати на неї;

між рівнем безробіття та інфляцією; між обсягом виробництва та низкою показників, таких, як основні фонди, термін їх експлуатації, кількість оборотних коштів, професійний рівень персоналу; між продуктивністю праці та рівнем механізації виробничих процесів, технологією виробництва.

Дослідження таких зв'язків і їм подібним забезпечує можливість спрямовувати економічні процеси в потрібному напрямку, тобто реалізовувати бажану економічну політику не лише на мікрорівні, а й на рівні економіки держави. Для ефективної реалізації економічної політики необхідно здійснювати регулювання певних економічних параметрів, а для цього слід володіти достеменною інформацією про зв'язок їх з іншими, ключовими величинами, щоб у майбутньому прийняти правильне рішення на мікро-, мезорівні чи в масштабах всієї країни.

Наприклад, у ринковій економіці не можна безпосередньо регулювати темп інфляції, але можна вплинути шляхом фіскального (бюджетно-податкового) важеля або монетарної (кредитно-грошової) політики. У цьому разі необхідно досконало вивчити залежність між пропозицією грошей та рівнем цін, які діють у державі.

Однак у реальних умовах, навіть у стабільних залежностях між показниками економічних величин, завжди проявляється певна невідповідність. Особливо виникають трудноці під час аналізу маловивчених і нестабільних зв'язків. Тому в сучасній економічній теорії в дослідженнях використовують апарат математичних моделей, імовірнісні та статистичні методи аналізу параметрів цих моделей.

Отже, ***третім важливим завданням*** економетрії є вибір та побудова математико-статистичної моделі, здійснення низки модельних експериментів, аналіз одержаних результатів і перенесення їх на реальну економічну систему (процес) як основу для прийняття належних управлінських рішень.

Якщо в природничих науках значною мірою мають справу з функціональними залежностями між змінними, то в економіці такі залежності, як правило, бувають відсутні. Наприклад, не може існувати жорсткої функціональної залежності між доходами громадян і їх витратами на споживання, між процентною ставкою на депозити (кредити) і попитом на нього; продуктивністю праці та стажем роботи працівників підприємства та ін.

Відсутність жорсткої функціональної залежності між змінними у сфері економіки пов'язана з низкою причин. Так, при аналізі впливу однієї змінної на іншу можуть бути не враховані фактори, що впливають або на кожну з змінних окремо, або на всі одночасно. Цей вплив може бути як безпосереднім, так і через цілий ланцюг інших факторів, урахувати які практично неможливо, оскільки вони мають випадкове походження. Тому в економічних дослідженнях, як правило, мають справу не з функціональною, а зі статистичною або кореляційною залежністю, вивченням якої займається кореляційний та регресійний аналіз (табл. 1.3).

Для прикладу розглянемо дві змінні *Y* та *X*, між якими можуть існувати три форми зв'язку – кореляційний, функціональний та регресійний.

У практичній діяльності з погляду економічних досліджень найбільш вживаними є кореляційний та регресійний зв'язки, які розглянемо більш докладно.

За наявності кореляційного зв'язку між *Y* та *X* ці змінні вважають рівноправними в тому розумінні, що їх не поділяють на залежну та незалежну. У цьому разі вирішується лише питання про наявність між цими змінними зв'язку, про який нас інформує кореляційний (коваріаційний) момент $K_{xy}(\text{cov}(x, y))$[1]. Якщо $K_{xy} \neq 0$ $(\text{cov}(x, y) \neq 0)$, цей зв'язок існує. У протилежному випадку ($K_{xy} = 0$ (cov $(x, y) = 0$) зв'язок відсутній. Суттєвість цього зв'язку (тісноту) вимірюють коефіцієнтом кореляції rxy ($|rxy| \leq 1$ або $-1 \leq rxy \leq 1$). Цей зв'язок не має напрямного характеру. Серед змінних *Y* та *X* немає залежної і незалежної.

Таблиця 1.3. Сутність кореляційного, функціонального та регресійного зв'язку між економічними показниками

Кореляційний зв'язок	Функціональний зв'язок	Регресійний зв'язок
Статистичний взаємозв'язок двох або кількох випадкових величин. При цьому зміна значень однієї чи кількох із цих величин відповідає систематичній зміні значень іншої/інших величин	Така залежність явищ, за якої зміна одного явища супроводжується зміною іншого	Функція, що описує відношення (залежність) між випадковими змінними величинами

[1] ***Коваріація*** (англ. *Covariance*) – у теорії ймовірностей та математичній статистиці числова характеристика залежності випадкових величин. Сутність коваріації полягає в тому, що вона виникає внаслідок невизначеності результату перемножування двох сукупностей чисел.

Коваріація двох випадкових величин *X, Y* позначається як Cov*(X, Y)* і має вигляд:

$$\text{Cov}(X, Y) = \text{E}[(X - \mu_x)(Y - \mu_y)] = \text{E}(XY) - \mu_x\mu_y = \mu_{xy} - \mu_x\mu_y,$$

де E – оператор математичного сподівання;

μ_x – середнє значення величини *X;*

μ_y – середнє значення величини *Y;*

E(*XY*) – математичне сподівання добутку величин *X, Y;*

μ_{xy} – це середнє значення добутку цих величин.

Регресійний зв'язок між змінними Y та X є таким, що коли одна з них, наприклад X, вибирається як незалежна змінна, то її називають пояснювальною змінною (регресором), а другу змінну Y – залежною (пояснювальною, регресандом). У цьому разі пояснювальна змінна X (регресор) є причиною зміни залежної змінної Y (регресанда). Так, збільшення доходу населення викликає збільшення витрат на споживання; збільшення ціни на ресурси призводить до зменшення попиту на них; зниження відсоткової ставки збільшує кількість інвестицій.

Однак залежності, наведені в прикладах, на практиці не будуть функціональними та взаємно однозначними, а є стохастичними, коли одному значенню x може відповідати множина значень Y. Отже, пояснювальна змінна x впливає на Y у сукупності вибірки, а не в середньому.

1.2. Роль економетричних досліджень в економіці

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

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

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

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

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

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

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

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

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

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

1.3. Предмет, цілі, завдання та структура курсу. Місце і значення курсу серед дисциплін фундаментальної підготовки фахівців з економіки. Зв'язки курсу із суміжними дисциплінами

Економетрика (рис. 1.2) посідає важливе місце серед дисциплін фундаментальної підготовки фахівців з економіки. Значення цього курсу обумовлене тим, що економетрика дозволяє провести узагальнення закономіростей, обґрунтованих економічною теорією, на основі

Об'єкт економетрії	економічні системи та простори різного рівня складності: від окремого підприємства чи фірми до економіки галузей, регіонів, держави й світу загалом
Предмет економетрії	методи побудови та дослідження математико-статистичних моделей економіки, кількісні дослідження економічних явищ, пояснення та прогнозування розвитку економічних процесів
Мета економетричного дослідження	аналіз реальних економічних систем і процесів, що в них відбуваються, за допомогою економетричних методів і моделей, їх застосування при прийнятті науково обгрунтованих управлінських рішень
Основне завдання економетрії	оцінити параметри моделей з урахуванням особливостей вхідної економічної інформації, перевірити відповідність моделей досліджуваному явищу і спрогнозувати розвиток економічних процесів

Рис. 1.2. Об'єкт, предмет, мета та основні завдання економетрії

інстурментарію математики та теорії імовірностей. Крім того, економетрика дозволяє на базі отриманих знань з дисциплін економічного спрямування здійснювати грунтовний економічний аналіз та приймати належні управлінські рішення.

Процес економетричного моделювання складається з таких кроків:

1. Вибір конкретної форми аналітичної залежності між економічними показниками (специфікація моделі) на підставі відповідної економічної теорії.
2. Збирання та підготовка статистичної інформації.
3. Оцінювання параметрів моделей.
4. Перевірка адекватності моделі та достовірності її параметрів.
5. Застосування моделі для прогнозування розвитку економічних процесів з метою подальшого керування ними.

Застосування різноманітних економетричних моделей на різних рівнях економічної діяльності дає можливість розв'язувати економічні проблеми різного рівня складності (табл. 1.4).

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 розрізняють: макрорівень (країна загалом), мезорівень (регіони, галузі, корпорації) та мікрорівень (сім'я, підприємство, фірма).

Таблиця 1.4. Напрями дослідження економетричних закономірностей на різних рівнях економіки

Рівень макроекономіки	Рівень мікроекономіки
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. 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	Економетричні дослідження передбачають наукове обгрунтування управлінських рішень, що приймаються на підприємствах різних форм власності й маютьураховувати постійний вплив зовнішнього середовища

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 можливість 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.

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

Методи економетрії є найсучаснішими засобами аналізу та дослідження різних соціально-економічних систем. За допомогою економетричних методів можна відхилити деякі економічні гіпотези або показати неможливість застосовування їх у конкретних умовах.

Хоча засоби економетрії не дають можливість довести теоретичні твердження, але з допомогою її методів можна показати, що те чи інше твердження не суперечить даним спостережень. Оволодівши елементарним інструментарієм економетрії, можна обґрунтовано прогнозувати розвиток цих систем, оцінювати вплив рішень чи урядових постанов щодо зміни цін, податків тощо на стан справ будь-якого підприємства, розробляти шляхи ефективного керування ними, приймати ефективні управлінські рішення.

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

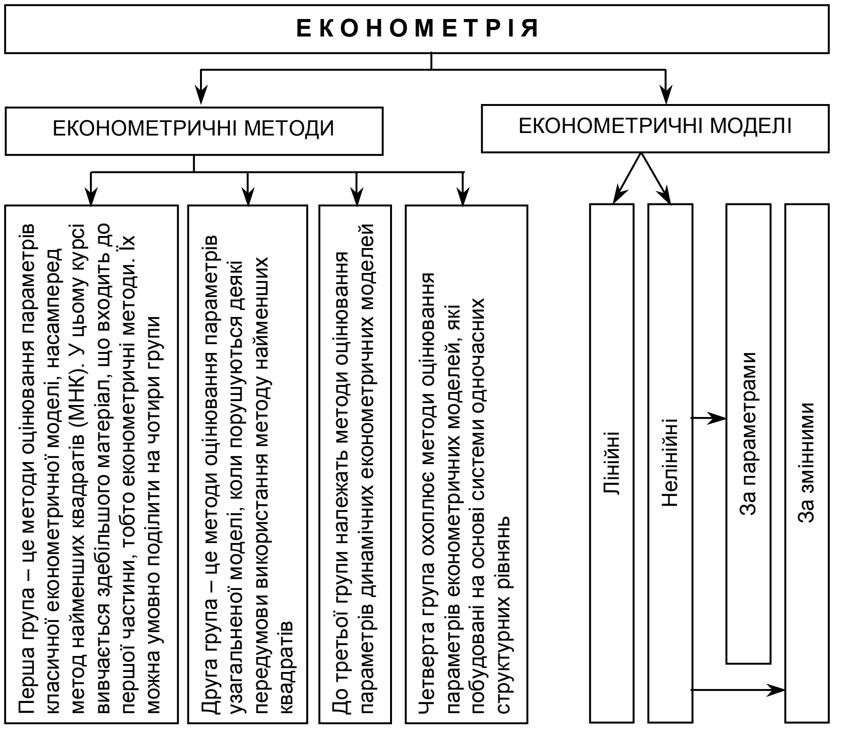


Рис. 1.3. Поділ економетрії на дві складові

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

Економетрія поділяється на дві частини – *економетричні методи* та *економетричні моделі* економічних процесів і явищ (рис. 1.3).

1.4. Особливості економетричних моделей

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

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

Статистична залежність – це залежність, коли зі змінюванням однієї випадкової величини змінюється закон розподілу ймовірностей іншої.

Кореляційна залежність – статистична залежність, яка виявляється в тому, що зі змінюванням однієї величини змінюється середнє значення іншої

Наприклад, у землеробстві з однакових за площею ділянок землі при рівних кількостях внесених добрив збирають різний врожай. Звичайно, немає строгої функціональної залежності між прибутковістю банківської установи та розміром ставки за кредитами (депозитами). Це пояснюється впливом випадкових факторів (пора року, розташування банку тощо). Водночас, як показує досвід, зазначені показники, напевне, пов'язані кореляційною залежністю.

Можна зазначити два типи взаємозв'язку змінних. В одному випадку невідомо, яка зі змінних незалежна, а яка – залежна, тобто вони рівноправні й зв'язок можна розглядати як в один, так і в інший бік. У другому випадку змінні нерівноправні, тобто зміна лише однієї з них впливає на змінювання іншої, а не навпаки. У цьому разі при розгляді зв'язку між двома змінними величинами важливо встановити на основі логічного міркування, яка з ознак є причиною, а яка – наслідком. Наприклад, прибутковість банківської установи залежить від розміру ставки за кредитами (депозитами), а не навпаки, тобто економічна оцінка процентної ставки є незалежною змінною, а прибутковість – залежною.

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

1.5. Вибір змінних і структура зв'язків

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

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

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

де $M(Y|x)=f(x)$ – умовне математичне сподівання.

Функція регресії Y на X	функція *f{x)*
Незалежна (пояснювальна) змінна (регресор)	*X*
Залежна (пояснювана) змінна (регресанд)	*Y*

$$M(Y\mid x_1,x_2,...,x_m)=F(x_1,x_2,...,x_m) \qquad (1.2)$$

Термін «регресія» (рух назад, повернення до попереднього стану) запропонував Френсіс Галтон наприкінці XIX ст., проаналізувавши залежність між зростом батьків і зростом дітей. Він помітив, що зріст дітей у дуже високих батьків у середньому менший, ніж середній зріст батьків. У дуже низьких батьків, навпаки, середній зріст дітей вищий. В обох випадках середній зріст дітей прямує (повертається) до середнього зросту людей у даному регіоні. Звідси й вибір терміна, що відбиває таку залежність (рис. 1.4).

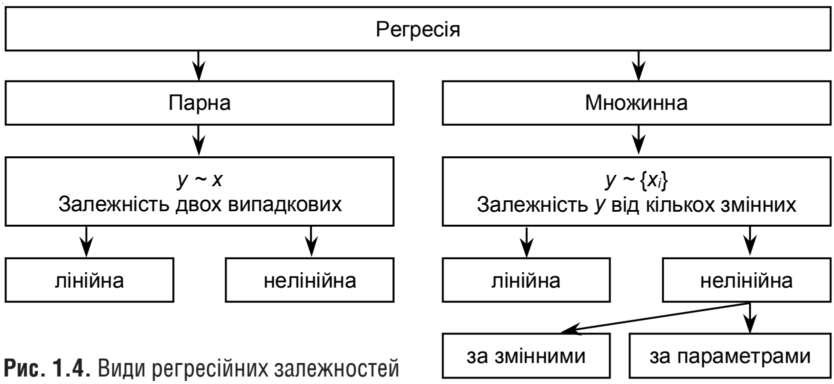


Рис. 1.4. Види регресійних залежностей

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

Регресійні рівняння (моделі) – зв'язки між залежною та незалежною (незалежними) змінними, що описуються співвідношеннями.

$$y = f(x) + \varepsilon, \tag{1.3}$$

$$y = F(x_1, x_2,..., x_m) + \varepsilon, \tag{1.4}$$

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

1. ***Введення до моделі не всіх пояснювальних змінних.*** Будь-яка регресійна (зокрема, економетрична) модель – це спрощення реальної ситуації. Остання завжди є складною композицією різних факторів, багато з яких у моделі не враховуються, що призводить до відхилення реальних значень залежної змінної від її модельних значень. Наприклад, попит на кредитні ресурси визначається процентною ставкою, прибутком споживачів, тощо. Безперечно, навести всі пояснювальні змінні практично неможливо. Зокрема, неможливо врахувати такі фактори, як традиції, національні чи релігійні особливості, географічне положення району, погоду та багато інших, вплив яких призводить до деяких відхилень реальних спостережень від модельних. Ці відхилення можуть бути описані як випадкова складова моделі.У деяких випадках заздалегідь невідомо, які фактори за умов, що склалися, насправді є визначальними, а якими можна знехтувати. Крім того, інколи безпосередньо врахувати якийсь фактор неможливо через відсутність статистичних даних. Наприклад, обсяг заощаджень домогосподарств може визначатися не лише прибутками їх членів, а й станом здоров'я останніх, інформація про яке в цивілізованих країнах становить лікарську таємницю. У деяких ситуаціях низка факторів має принципово випадковий характер, що додає неоднозначності певним моделям, наприклад порі року в моделях, що прогнозують прибутковість банків.

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

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

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

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

5. ***Непередбачуваність людського фактора.*** Ця причина може «зіпсувати» найякіснішу модель. Дійсно, за правильного вибору форми моделі, скрупульозного добору пояснювальних змінних неможливо спрогнозувати поведінку кожного індивідуума.

6. Кореляційно-регресійний аналі Сукупність методів, за допомогою яких досліджуються та узагальнюються взаємозв’язки кореляційно пов’язаних змінних.

Здебільшого процедура аналізу зв’язку між змінними дає можливість встановити його природу, тобто визначити форму залежності між змінними (рис. 1.5).

Побудова якісного рівняння регресії, що відповідає емпіричним даним і цілям досліджень, є досить складним процесом. Його можна поділити на три етапи (рис. 1.6).

Специфікація моделі регресії – вибір форми зв’язку змінних.

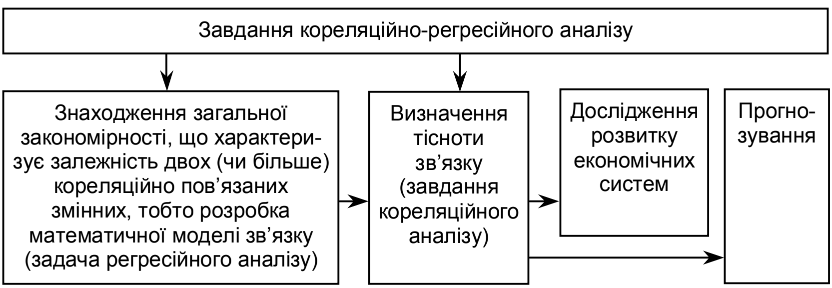


Рис. 1.5. Задачі кореляційно-регресійного аналізу

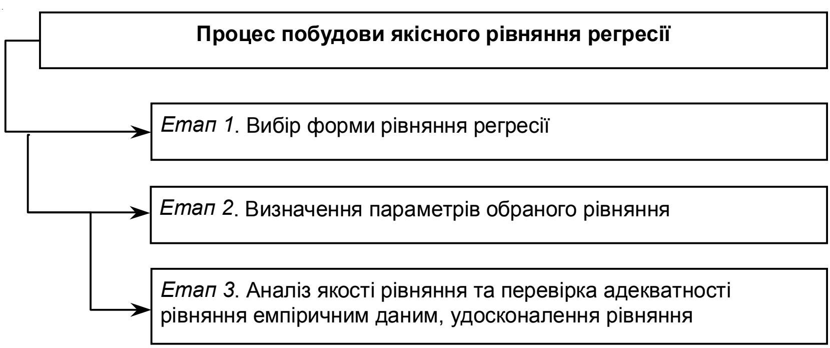


Рис. 1.6. Процес побудови якісного рівня регресії

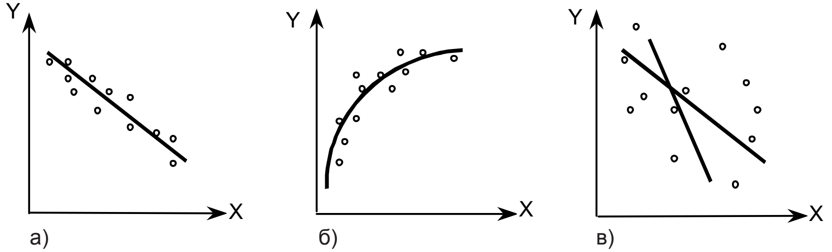


Рис.1.7. Кореляційне поле

Кореляційне поле (діаграма розсіювання) – за парної регресії вибір формули звичайно здійснюється за графічним зображенням реальних статистичних даних у вигляді точок у декартовій системі координат (рис.1.7).

На рис. 1.7*а* взаємозв'язок між X і Y близький до лінійного, і пряма 1 досить добре узгоджується з емпіричними точками. Тому, щоб описати залежність між X і Y, доцільно вибрати лінійну функцію $Y = b_0 + b_1X$.

На рис. 1.7*б* реальний взаємозв'язок між X і Y, найімовірніше, описується логарифмічною функцією $Y = a*ln(bX)$.

На рис. 1.7*в* явний взаємозв'язок між X і Y відсутній. Тому щоб краще вибрати форму зв'язку, необхідно, можливо, збільшити кількість спостережень – точок кореляційного поля або скористатися іншими способами вимірювання показників.

У разі множинної регресії визначити форми залежності ще складніше.

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

Наприклад, Кейнс запропонував лінійну формулу залежності індивідуального споживання C від доходу Y: $C = c0 + bY$, де $c0 > 0$ – величина автономного споживання; b – гранична схильність до споживання, $0 < b < 1$.

Однак поки не обчислено кількісні значення коефіцієнтів $c0$ і b та не перевірено надійність отриманих результатів, зазначена формула залишається лише гіпотезою.

1.6. Роль і місце економетричних моделей у математичному моделюванні

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

Модель має форму рівняння чи системи рівнянь, що характеризують визначені 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 (табл. 1.5).

Отже, сформулюємо таке визначення економетричної моделі.

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 з цих показників розглядаються як залежні змінні, а інші – як незалежні.

Таблиця 1.5. Класифікація економетричних моделей

Класифікаційна ознака	Класифікація
За рівнем агрегування змінних	Мікро- та макроекономічні показники
За способом відображення змінних	У постійних та поточних цінах, в абсолютних значеннях та приростах показників
За кількістю змінних	Одно- чи багатофакторні моделі
За кількістю рівнянь	Одне та кілька
За часом спостережень	Річні, квартальні та місячні дані
За призначенням та метою використання	Аналітичні, імітаційні, прогностичні

У загальному випадку рівняння в економетричній моделі має вигляд

$$Y = f(x_1, x_2, .., x_m, \varepsilon), \tag{1.5}$$

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

Процес опису явища чи процессу, тобто вибір аналітичної форми моделі, називається специфікацією моделі. Інакше кажучи, ***специфікація моделі*** – це аналітична форма залежності між економічними показниками.

Екзогенні змінні (регресори) – незалежні змінні $x_1, x_2, .., x_m$, що задані заздалегідь чи за межами моделі.

Ендогенна змінна (регресанд) – залежна змінна Y, що визначається як розв'язок рівняння.

Функція f у кожному конкретному випадку, крім змінних $x_1, x_2, ..., x_m$ і u, містить ще щонайменше деякі коефіцієнти, які поєднують змінні у певних співвідношеннях і визначають структуру рівняння. Ці коефіцієнти називаються ***параметрами моделі***.

Параметризація рівняння регресії, або ***оцінювання параметрів,*** – визначення значень коефіцієнтів (параметрів) обраної форми статистичного зв'язку змінних на підставі відповідних статистичних даних.

Існує відмінність між змінними та параметрами моделі.

Змінні – економічні величини, що можуть набувати певних значень з деякої множини допустимих величин.

Параметри – 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, безперечно, 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.

Статичні моделі – моделі, у яких зв'язки розглядаються у фіксований момент часу і часові зміни в них ролі не відіграють.

Динамічні моделі – моделі, у яких взаємозв'язки вивчаються в розвитку й час є необхідним фактором змін.

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 обгрунтування 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.

Перехресні дані – 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: існують певні закономірності у відхиленнях від загальної тенденції розвитку або 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.

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

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

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

В означеній ситуації найприйнятнішими методами дослідження є методи математичної статистики, адаптовані до економічних явищ. Саме ці методи дають можливість 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.

1.7. Приклади економетричних моделей

1.7.1. Виробнича функція Кобба – Дугласа

На сучасному етапі розвитку економетрії застосовується модель, що характеризує та кількісно описує зв'язок основних результативних показників, які виникають у процесі здійснення виробничо-господарської діяльності разом з множиною факторів, що визначають дані показники. Основними показниками є дохід, прибуток, рентабельність, продуктивність праці, собівартість та ін.

Поняття виробничої функції походить із взаємозв'язку математичного моделювання технологічної залежності між обсягом продукції, що випускається, та кількісними характеристиками витрат ресурсів. Ще у 30-х роках XX ст. її вперше побудували такі дослідники, як Кобб і Дуглас у Сполучених Штатах Америки, використовуючи дані про функціонування обробної промисловості, які охоплювали період у двадцять років, що є класичним прикладом економетричного моделювання.

Функція Кобба – Дугласа (CDPF) набула всеохопного використання в процесі економічних досліджень як на макро- так і на мікрорівні, також вона є найвідомішою виробничою функцію. Функція Кобба – Дугласа у класичному розумінні має такий вигляд:

$$Y = aK^{\alpha}L^{1-\alpha}, \tag{1.6}$$

де Y – обсяг продукції;
K – основний капітал;
L – робоча сила.

Параметри a, α і $1 - \alpha$ у даній функції є невід'ємними. Це доводиться шляхом вилучення з виробничої функції одного з факторів. Поділивши ліву та праву частини залежності $Y = f(K,L)$ на L, отримуємо функцію двох змінних

$$W = f(V), \tag{1.7}$$

де $W = \dfrac{Y}{L}$ – продуктивність праці;
$V = \dfrac{K}{L}$ – фондоозброєність праці.

Нехай залежність між W і V має вигляд степеневої функції, тобто

$$W = aV^{\alpha}.$$

Підставивши в цю функцію $W = \dfrac{Y}{L}$ і $V = \dfrac{K}{L}$, дістанемо:

$$\frac{Y}{L} = a\left(\frac{K}{L}\right)^{\alpha}, \text{ або } Y = aK^{\alpha}L^{1-\alpha}.$$

Сума параметрів або ступінь однорідності, класичної функції Кобба – Дугласа дорівнює одиниці, у свою чергу це означає, що в разі зростання обох виробничих ресурсів на одиницю збільшення обсягу продукції відбудеться на одиницю. Саме тому ефективність ресурсів у такому випадку стала.

На практиці було виявлено, що припущення про лінійну однорідність на практиці виконується рідко. Отже, було запропоновано виробничу функцію більш загального вигляду

$$Y = aK^{\alpha}L^{\beta}. \tag{1.8}$$

На відміну від попереднього випадку сума параметрів (a + b) може бути як меншою, так і більшою від одиниці. За умови $(\alpha + \beta) > 1$ темпи зростання обсягу продукції вищі за темпи зростання виробничих ресурсів, а якщо $(a + b) < 1$, то, навпаки, темпи збільшення продукції нижчі за темпи збільшення ресурсів.

Якщо зробити припущення, що рівень кожного виробничого ресурсу буде збільшено на r %, тоді величини їх дорівнюватимуть $K(1+\frac{i}{100})$ і $L(1+\frac{i}{100})$.

Обсяг продукції на основі виробничої функції матиме такий вигляд:

$$Y = a\left[K\left(1+\frac{i}{100}\right)\right]^{\alpha}\left[L\left(1+\frac{i}{100}\right)\right]^{\beta} = aK^{\alpha}L^{\beta}\left(1+\frac{i}{100}\right)^{\alpha}\left(1+\frac{i}{100}\right)^{\beta} = aK^{\alpha}L^{\beta}\left(1+\frac{i}{100}\right)^{\alpha+\beta}. \qquad (1.9)$$

Отже, якщо $\alpha + \beta > 1$, обсяг продукції зростатиме більш ніж на r%; якщо $\alpha + \beta < 1$ – менш ніж на r %; при $\alpha + \beta = 1$ продукція збільшиться на r%. Визначивши ряд коефіцієнтів еластичності[1] для виробничої функції Кобба – Дугласа, матимемо:

---

[1] (Точковою) еластичністю (коефіцієнтом еластичності) змінної y до x називається величина $E_{y/x} = \frac{\partial \ln y}{\partial \ln x} = \frac{dy/y}{dx/x} = \frac{\delta y}{\delta x}$. Це значення визначає еластичність в конкретній точці. Еластичність постійна тільки в рамках логарифмічною (або степеневої) моделі залежності. У багатьох випадках (у тому числі й для лінійної моделі залежності) еластичність в різних точках відрізняється. Тому розраховують також середню (дугову) еластичність як відношення відсоткових змін y та x. $E_{y/x} = \frac{\delta y}{\delta x} = \frac{\Delta y / y_1}{\Delta x / x_1}$. Іноді замість x_1 та y_1 використовують середню точку в інтервалі зміни їх значень. $E_{y/x} = \frac{\delta y}{\delta x} = \frac{\Delta y / \bar{y}}{\Delta x / \bar{x}}$, де $\bar{y} = \frac{(y_1 + y_2)}{2}$, $\bar{x} = \frac{(x_1 + x_2)}{2}$. Перевагою останнього способу є симетричність щодо знака зміни фактора.

Якщо коефіцієнт еластичності за модулем менше від одиниці, то говорять про нееластичність змінної y за x. Якщо коефіцієнт еластичності більше ніж 1, то кажуть, що y еластична за x, оскільки кожен відсоток зміни фактора призводить до ще більшого зміни y. Якщо коефіцієнт еластичності дорівнює 1, то говорять про одиничну еластичність. У граничному випадку, коли коефіцієнт еластичності дорівнює нескінченності, говорять про досконалу еластичності. Відповідно, за нульового коефіцієнта еластичності – про досконалу нееластичність.

$$Y_K = \frac{\partial Y}{\partial K} = \alpha \frac{Y'}{K}; \tag{1.10}$$

$$Y_L = \frac{\partial Y}{\partial L} = \beta \frac{Y'}{K}. \tag{1.11}$$

Як результат ми визначили, що граничний приріст продукції шляхом зростання кожного ресурсу визначається як добуток двох показників – коефіцієнта еластичності та середньої ефективності ресурсів. Параметр *a* у функції Кобба – Дугласа має залежність від обраних одиниць вимірювання *Y*, *K*, *L*; в той самий час абсолютне значення даного параметра визначається ефективністю виробничого процесу. Це можна довести шляхом порівняння двох виробничих функцій, які відрізняються одна від одної лише значенням параметра *a*.

У разі з фіксованими значеннями *K* і *L* тій функції, у якій більше числове значення параметра *a*, відповідає більше значення *Y*. Через те і виробничий процес, який описується даною функцією, буде ефективнішим. Другі похідні функції Кобба – Дугласа мають такий вигляд:

$$\frac{\partial Y}{\partial K \partial K} = \frac{\alpha(\alpha - 1)Y}{F^2}; \tag{1.12}$$

$$\frac{\partial Y}{\partial L \partial L} = \frac{\beta(\beta - 1)Y}{L^2}. \tag{1.13}$$

Враховуючи, що $0 < \alpha < 1$ і $0 < \beta < 1$, $Y_{KK} < 0$ і $Y_{LL} < 0$, зробимо висновок: у разі збільшення ресурсів граничний приріст обсягу продукції зменшуватиметься. Вважаючи обсяг продукції у функції Кобба – Дугласа сталим (або таким, що дорівнює const), можна обчислити граничні норми заміщення ресурсів:

$$h = \frac{Y_K}{Y_L} = \frac{\alpha \frac{Y}{K}}{\beta \frac{Y}{L}} = \frac{\alpha}{\beta} \frac{L}{K}. \tag{1.13}$$

Результатом чого є гранична норма заміщення ресурсів у функції Кобба – Дугласа, що визначається як добуток співвідношень обсягів ресурсів та коефіцієнтів їх еластичності.

Обчислення швидкості зміни норми заміщення ресурсів через зміну їх величини здійснюється в такий спосіб:

$$\frac{\partial h}{\partial K} = \frac{\beta}{\alpha L}; \quad \frac{\partial h}{\partial L} = -\frac{\beta K}{\alpha L^2}. \tag{1.15}$$

Відношенням зміни величини ресурсів до зміни величини h визначається еластичність заміщення ресурсів K і L, що є мірою швидкості зміни h:

$$h_{K/L} = \frac{{}^{L}\!/_{K}\,\partial\, {}^{L}\!/_{K}}{\partial h /_{h}} = \frac{h(Lh+K)}{KL(h\frac{\partial h}{\partial K} - \frac{\partial h}{\partial L})} = 1 \tag{1.15}$$

Саме тому еластичність заміщення в кожній точці кривої, що описує виробничу функцію Кобба – Дугласа, дорівнює одиниці.

Також необхідно розглянути поведінку функції при зміні масштабу виробництва. Зробимо припущення, що витрати кожного ресурсу виробництва збільшилися в λ раз, тоді нове значення Y визначатиметься за формулою:

$$Y = a(\lambda K)^{\alpha}(\lambda\, L)^{\beta} = \lambda^{\alpha+\beta} Y. \tag{1.17}$$

Ступінь однорідності цієї функції дорівнює $\alpha + \beta$. Якщо $\alpha + \beta = 1$, рівень ефективності ресурсів не залежить від масштабів виробництва. Якщо $\alpha + \beta < 1$, то з розширенням масштабів виробництва середні витрати в розрахунку на одиницю продукції зменшуються, а якщо $\alpha + \beta > 1$ – збільшуються. Причому ці властивості не залежать від числових значень K і L і зберігають силу в кожній точці виробничої функції.

Якщо припустити, що мета господарської діяльності – максимізація прибутку, то можна проілюструвати інші властивості виробничої функції. Запишемо функцію прибутку:

$$\Pi = bY^{r+1} - wL - rK + \lambda\,[f(K,L) - Y]. \tag{1.18}$$

Суб’єкт господарської діяльності обирає такі значення Y, L, K, які будуть максимізувати прибуток при обмеженнях, що накладаються виробничою функцією. Величини b, w, r – параметри функції прибутку,

λ – множник Лагранжа[1]. Якщо виробничий процес у даному співвідношенні описується функцією Кобба – Дугласа, то можна записати умови максимізації прибутку:

$$w=\frac{\lambda\beta Y}{L};\quad r=\frac{\lambda\alpha Y}{K};\quad \frac{w}{r}=\frac{\beta F}{\alpha L},\tag{1.19}$$

$\lambda=(r+1)P$ при $r\neq -1$, де $P=bY^r$.

Звідси обсяги ресурсів такі:

$$L=\frac{(r+1)P\beta Y}{w};\quad K=\frac{(r+1)P\alpha Y}{r}.\tag{1.20}$$

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

$$Y=aK^{\alpha}L^{\beta}=a\left[\frac{(r+1)P\alpha Y}{r}\right]^{\alpha}\left[\frac{(r+1)P\beta Y}{w}\right]^{\beta}.\tag{1.21}$$

Якщо $r=1$, згідно із записаними щойно умовами максимізації матимемо:

$$K=\frac{w\alpha L}{\beta r};\quad Y=a\left(\frac{w\alpha}{\beta r}\right)^{\alpha}L^{\alpha+\beta}.\tag{1.22}$$

Отже, необхідні умови для забезпечення максимізації прибутку дають можливість визначити відповідні витрати робочої сили і основного капіталу. Поступове розширення масштабів промислового виробництва призводить до зниження ефективності витрат ресурсів, що характеризується максимізацією прибутку в умовах досконалої конкуренції. Наведена виробнича функція показує, що дана економетрична модель допомогає досить повно проаналізувати виробничу діяль-

---

Метод множників Лагранжа – метод знаходження умовного екстремуму функції $f(x)$, де $x\in R^n$, щодо m обмежень $\varphi_i(x)=0$, де i змінюється від одиниці до m.

- Складемо функцію Лагранжа у вигляді лінійної комбінації функцій f і функцій φ_i, узятих з коефіцієнтами, які називаються множниками Лагранжа, – λ_i:

$$L(x,\lambda)=f(x)+\sum_{i=1}^{m}\lambda_i\varphi_i(x),\ \text{де } \lambda=(\lambda_1,\ ..,\ \lambda_m).$$

- Складемо систему $n+m$ рівнянь, прирівнявши до нуля частинні похідні функції Лагранжа $L\ (x,\lambda)$ за x_j та λ_i.
- Якщо отримана система має рішення щодо параметрів x_j' і λ_i', тоді точка x' може бути умовним екстремумом, і є рішенням вихідної задачі. Зауважимо, що це умова носить необхідний, але не достатній характер.

ність, а також допомагає підвищити її ефективність шляхом її вдосконалення. Від вірогідності економетричної моделі, а саме, від того, наскільки вона адекватна реальному процесу. залежить обґрунтованість такого аналізу.

Побудови виробничої функції або інших технологічних взаємозалежностей у виробництві – класична проблема економетрії, яка висвітлюється далі.

1.7.2. Моделі пропозиції і попиту на конкурентному ринку

Шляхом взаємодії попиту та пропозиції на конкурентному ринку встановлюється рівновага обміну. Припустимо, що g_1 і g_2 – кількість попиту і пропозиції деякого продукту у встановлений день на певному ринку; p – ціна, за якою реалізується продукція. Величини g_1 і g_2 залежать від p, через ціну, яка не є врівноваженою, кількість проданого товару зменшується. Це дає нам можливість записати такі функції:

$D = f(p, u)$ – функція попиту;

$S = \mathrm{Y}(p, \mathrm{e})$ – функція пропозиції.

За умови, що відома ціна p, можна розрахувати величини попиту і пропозиції. Існування рівноваги на ринку обумовлюється виконанням рівності. Модель має такий вигляд:

$$\begin{aligned} &D = S; \\ &D = f(p, u); \\ &S = \psi(p, \varepsilon). \end{aligned} \tag{1.23}$$

Вона містить функції, що описують залежність попиту і пропозиції від ціни, а також їх тотожність.

Попит і пропозиція певного товару у реальних умовах залежать не лише від його ціни, а й від цін на товари, субститути на комплементи. Також попит залежить від доходу покупців, а пропозиція – від виробничих умов і т.ін. Тоді модель (1.23) матиме такий вигляд:

$$\begin{aligned} &D_t = f(p_t, X_{1t}, X_{2t}, \ldots X_{mt}, u_t); \\ &S_t = \psi(p_{t-1}, X_{1t}, X_{2t}, \ldots X_{mt}, \varepsilon_t); \\ &D_t = S_t. \end{aligned} \tag{1.24}$$

На відміну від попередньої, ця модель формування попиту в періоді t залежить від ціни в цьому самому періоді, в той час як пропозиція в періоді t залежить від ціни попереднього періоду ($t - 1$).

Нехай залежність попиту і пропозиції від факторів, що впливають на них, лінійна. Тоді економетрична модель матиме такий вигляд:

$$D_t = a_0 + a_1 p_t + a_2 X_{1t} + a_3 X_{2t} + \ldots + a_m X_{mt} + u_t;$$
$$S_t = b_0 + b_1 p_{t-1} + b_2 X_{1t} + b_3 X_{2t} + \ldots + b_m X_{mt} + \varepsilon_t; \qquad (1.25)$$
$$D_t = S_t.$$

Щоб оцінити параметри даної моделі, необхідно застосувати один з численних економетричних методів, які розглядаються далі.

1.7.3. Модель Кейнса

Фази безробіття докладно не досліджувалися класичною економічною теорією. Вони розглядалися як тимчасові випадковості, що відбувалися в короткострокових періодах, які не розглядалися так детально на противагу проблемам рівноваги та зростанняву довгострокових періодах. І в той самий час протягом 1930–1940 рр. майже в усіх розвинених країнах спостерігалось тривале безробіття. Для передбачення розвитку економіки та вживання певних заходів впливу на економічний розвиток, потрібно було знати, як у даний момент фіксувати рівень випуску продукції та зайнятості і чому остання не буває ні дуже високою, ні дуже низькою.

Ідеї Джона Мейнарда Кейнса були, у першу чергу, спрямовані на розв'язання зазначеної проблеми. Його числені дослідження були спрямовані на пояснення рівня виробництва в період неповної завантаженості робочої сили та обладнання. Згодом численні дослідники висвітлювали це питання, намагаючись пояснити нечіткі позиції в теорії Кейнса або запропонувати власні розв'язання. Ці спроби дозволили дійти висновку про значну роль, яку капіталовкладення відіграють у кон'юнктурній еволюції, з двох причин:

1) інвестиційні рішення здебільшого є автономними, їх вплив характеризується зростанням обсягів виробництва у таких секторах — предметів споживання та засобів виробництва;
2) зростаючий обсяг виробництва впливає на збільшення доходів, що зумовлює зростання обсягів виробництва предметів споживання.

Наведені тези запишемо у вигляді математичної моделі.

Нехай P — загальний обсяг продукції; C — виробництво предметів споживання; I — виробництво засобів виробництва (що дорівнює капіталовкладенням); R – доходи, які розподіляються. Тоді модель матиме вигляд:

$$P = C + I;$$
$$C = F(R, u); \qquad (1.26)$$
$$R = P.$$

У наведеній моделі I задається автономно, а F є функція, що визначає відповідність між споживанням і розподіленими доходами.

Дана модель є спрощеною і не відтворює ні ідей Кейнса, ні складності фактів, яка присутня на ринку в повному обсязі. Водночас вона добре пояснює досягнутий рівень виробництва. Адже три раніше записані рівняння можна трансформувати в одне, що матиме вигляд:

$$P - F(R, u) = I. \qquad (1.27)$$

Розв'язавши його відносно P, знайдемо рівень виробництва, який пов'язаний з рівнем капіталовкладень. Так, наприклад, якщо $F(R)$ є лінійна функція

$$C = F(R) = \alpha R + \beta = \alpha P + \beta, \qquad (1.28)$$

то рівняння набуває вигляду

$$(1 - \alpha)P - \beta = I \qquad (1.29)$$

Звідси

$$P = \frac{I}{1-\alpha} + \frac{\beta}{1-\alpha}. \qquad (1.30)$$

Рівняння (1.30) визначає залежність обсягу виробництва P від обсягу капіталовкладень I, які задаються автономно. Коефіцієнти α і β в цьому рівнянні залежать від функції споживання (1.29), тобто від зв'язку між R і C. Зокрема, ця функція вимірює збільшення споживання α, яке пов'язане зі збільшенням доходу на одиницю і називається «граничною схильністю до споживання». Значення α, як правило, менше за одиницю. Зокрема, у моделі Кейнса $\alpha = 0{,}6$. Залежність (1.30) показує при цьому, що збільшення капіталовкладень на одиницю зумовлює зростання обсягу виробництва на $1/(1 - \alpha)$ – коефіцієнт, який завжди перевищує одиницю (якщо $\alpha = 0{,}6$, маємо $1/1 - \alpha = 2{,}5$). Цей коефіцієнт вимірює ефект взаємозв'язку між автономним зростанням капіталовкладень та обсягом виробництва і називається мультиплікатором.

Модель (1.28) формалізує теорію Кейнса в її найпростішому вигляді. Але цінність згаданої моделі виходить поза ці межі, бо вона дає можливість вивчати різні конкретні питання економічної кон'юнктури в країні, для якої було б знайдено адекватну форму функції $F(R, u)$. Для забезпечення надійності результатів необхідно, щоб модель з потрібним ступенем точності відповідала дійсності, але досягти цього за такої вельми віддаленої схематизації не можна. Кон'юнктурні моделі, застосовувані для короткострокового прогнозування, використовують набагато більше змінних і рівнянь, але їхня логічна природа досить близька до природи моделі (1.28).

1.7.4. Модель споживання

Головною передумовою існування виробничих систем є створення матеріальних благ, які споживаються в подальшому після їх виробництва або надходять у запаси. Це дає поштовх до моделювання системи споживання матеріальних благ, це питання посідає провідне місце серед проблем математичного моделювання виробничо-технічного рівня економічних систем. Споживання, або використання матеріальних благ, поділяється на два види: виробниче і невиробниче споживання. Суть виробничого споживання полягає у використанні матеріальних благ під час виробництва у вигляді сировини, основних фондів і т.ін. Невиробниче споживання – це безпосереднє задоволення потреб людей, як правило, до нього належать товари народного споживання. Потреба в них значною мірою визначає структуру та обсяг виробництва в цілому.

Метою дослідження обсягу споживання є пошук відповідних умов зміни споживання товару або групи товарів залежно від їх ціни, доходів та інших істотних параметрів. Спостереження дали можливість виявити закономірність зміни споживання, наприклад, дослідивши споживання окремих сімей протягом деякого часу, визначають зміну споживання того чи іншого товару при загальному підвищенні доходів. Ці дослідження використовують деякі гіпотези щодо стабільності залежностей між споживанням і факторами, які його визначають. Виникає запитання: чи можна кореляцію, що спостерігається для однієї обмеженої вибірки, інтерпретувати як доказ існування залежності в більш загальному випадку? При цьому гіпотези, які є основою для вивчення споживання, можна зобразити формально з допомогою моделі.

Нехай C_i – споживання деякого продукту i-ю сім'єю, дохід якої дорівнює Y_i. Припустимо, що для даного періоду відомі значення C_i і Y_i для невеликої кількості сімей. Як вивести звідси закономірність, на підставі якої можна визначити споживання даного продукту кожною сім'єю і в кожний період?

Найпоширеніший підхід полягає в доказі існування деякого точного функціонального зв'язку між C_i і Y_i, який не залежить від часу або від окремих характеристик кожної сім'ї. Тоді модель можна подати у вигляді

$$C_i = f(Y_i). \tag{1.31}$$

У той самий час можна стверджувати про неадекватність поданої гіпотези і моделі. Вони припускають, що дві сім'ї з одним і тим самим доходом мають однакове споживання, що не є абсолютним твердженням, тому від моделі (1.31) потрібно відмовитися.

Перше узагальнення може полягати в тому, щоб, крім доходу, розглянути й інші незалежні змінні: ціну, склад сім'ї, величину наявних коштів і т.ін. Тоді можна повністю описати споживання, але суто функціональний зв'язок лишиться недосяжним навіть за наявності п'яти і більше незалежних змінних. Дві сім'ї з однаковими доходами, структурним складом, заощадженнями тощо, все одно щодо споживання тих чи інших товарів поводитимуться по-різному.

Це означає, що в попередніх гіпотезах завжди має місце така фактична ситуація: споживання частково визначається невідомими нам факторами, які ми не можемо врахувати в моделі. Такі фактори є випадковими, і необхідно оцінити їх випадковий вплив. Для цього потрібно змінити модель (1.31), ввівши до неї випадкову складову:

$$C_i = f(Y_i) + u_i \tag{1.32}$$

У моделі споживання випадкова складова містить у собі вплив усіх випадкових факторів, а також факторів, які не належать моделі. Ця складова називається помилкою, або залишком. Ці терміни, використовуватимемо далі під час викладення матеріалу.

Загальний вигляд моделі споживання залежно від доходу сім'ї такий:

$$C = f(Y) + u. \tag{1.33}$$

Якщо сукупність спостережень (кількість досліджуваних сімей) буде достатньою, щоб забезпечити вірогідність зв'язку, який визначається згідно з моделлю (1.33), то характеристики взаємозв'язку можуть бути поширені на певну групу населення країни. При цьому слід пам'ятати, що специфікація та методи оцінювання параметрів моделі також впливають на вірогідність зв'язку, що визначається економетричною моделлю.

Завдання для самоконтролю

1. Сформулюйте сутність основних завдання економетрії.
2. Охарактеризуйте предмет, об'єкт та цілі курсу.
3. Яка модель належить до категорії економетричних?
4. Які особливості має економетрична модель?
5. Як записується економетрична модель в загальному вигляді? Які змінні у моделі є ендогенними, а які – екзогенними?
6. Дайте тлумачення випадкової складової економетричної моделі.
7. З яких причин у модель фактичних даних вводиться випадкова складова u?
8. Які етапи побудови економетричної моделі?
9. Що називається специфікацією економетричної моделі?
10. Сутність кореляційного та регресійного аналізу.
11. Охарактеризуйте прикладні аспекти використання економетричних моделей.

Практичне заняття 1

Збір статистичної інформації. Побудова та дослідження моделі парної лінійної регресії з використанням засобів MS EXCEL

Постановка завдання

1 етап	**Побудова інформаційного масиву дослідження шляхом збору статистичних даних та розроблення моделі парної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), аналіз моделі**
2 етап	Побудова множинної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), дослідження моделі
3 етап	Виявлення, дослідження та усунення мультиколінеарності факторів у моделі множинної лінійної регресії методом Фаррара – Глобера
4 етап	Виявлення ефекту гетероскедастичності на основі застосування параметричного тесту Гольдфельда – Квандта та усунення даного ефекту шляхом використання узагальненого методу найменших квадратів (методу Ейткена)
5 етап	Побудова адитивної та мультиплікативної моделі часового ряду на основі проведення декомпозиційного аналізу: визначення структури ряду (тенденції та сезонних (циклічних) коливань), здійснення автокореляційного аналізу, визначення якості моделі, розрахунок прогнозних значень
6 етап	Групування елементів економічної системи шляхом кластерного аналізу, побудова деревоподібної дендограми, графічне подання елементів кластеризації
7 етап	Перевірка наявності автокореляції залишків за допомогою критерію Дарбіна – Уотсона, оцінка параметрів рівняння множинної лінійної регресії методом Ейткена
8 етап	Побудова економіко-математичної моделі методом інструментальних змінних; побудова економетричної моделі на основі даних, що містять помилки вимірювання змінних, оцінка параметрів моделі за допомогою алгоритма Уоліса
9 етап	Побудова моделі розподіленого лагу, оцінка параметрів моделей з лагами в незалежних змінних, оцінювання параметрів авторегресійних моделей; перетворення ARMA і ARIMA
10 етап	Структурне моделювання (моделювання на основі системи структурних рівнянь), оцінка невідомих параметрів моделі, визначення адекватності моделі щодо опису структури вхідних даних

Мета роботи: набуття навичок побудови інформаційного масиву дослідження та розроблення моделі парної лінійної регресії, оцінки її параметрів методом найменших квадратів (1 МНК), аналізу моделі.

Початкові дані

Досліджується залежність результативної ознаки від рівня факторної (самостійно зібрана інформація (на оцінку 5 балів) або

див. на с. 54 табл. 1.6, відкориговану згідно свого варіанту на число *k* (на оцінку 3 бали).

Необхідно:

1. Визначити факторну і результативну ознаки, навести статистичні дані.
2. Побудувати лінійне рівняння парної регресії y від x.
3. Розрахувати лінійний коефіцієнт парної кореляції і середню помилку апроксимації. Зробити висновок про зв'язок між x і y (пряма або зворотна, тісна чи ні).
4. Оцінити статистичну значущість параметрів регресії і кореляції з допомогою F-критерію Фішера і t-критерію Стьюдента для рівня надійності $p = 0{,}95$.
5. Виконати прогноз рівня результативної ознаки y при прогнозному значенні факторної x, що дорівнює 107% середнього рівня.
6. Оцінити точність прогнозу, розрахувавши помилку прогнозу і його довірчий інтервал.
7. Побудувати поле кореляції і графік лінії регресії.
8. Перевірити результати розрахунків за допомогою надбудови MS Excel *Аналіз даних*.

Порядок виконання роботи

1. Провести збір статистичної інформації з погляду дослідження питань, які розкривають проблемні аспекти дослідження залежно від спеціальності, на якій навчається студент, – на оцінку 5 балів. За номером свого варіанта оберіть умову задачі, *до кожного значення діапазону вхідних даних додайте число k, яке призначається кожному студенту індивідуально,* – на оцінку 3 бали.
2. Побудуйте модель парної лінійної регресії, використовуючи метод найменших квадратів (1 МНК).
3. Дослідіть модель за допомогою класичних економетричних формул та перевірте результати розрахунків через надбудову MS Excel *«Аналіз даних»*.
4. Виконання окремих пунктів завдання супроводжуйте коментарями (формули, результати обчислень, висновки).

Для розв'язання задачі використати: інструктивні матеріали, надбудову *«Аналіз даних»* в MS Excel.

Вимоги до оформлення звіту

Звіт про проведення даної лабораторної роботи оформлюється разом з іншими звітами цього курсу у письмовому вигляді і містить:

- назву, тему, завдання, мету лабораторної роботи;

- вхідні дані варіанту;
- хід розв'язання поставлених завдань з використанням формул, що супроводжуються коментарями (див. приклад);
- електронні таблиці з результатами;
- електронні таблиці з електронними формулами;
- результати інструменту *Регресія* надбудови «*Аналіз даних*» в MS Excel.

Контрольні питання

1. Теоретичне та емпіричне рівняння лінійної регресії. Сутність параметрів рівняння регресії.
2. Які методи застосовуються для оцінювання параметрів класичної регресійної моделі?
3. Що називається специфікацією економетричної моделі?
4. Сутність методу найменших квадратів (1 МНК). Система нормальних рівнянь для знаходження оцінок парної лінійної регресії. Навести формули.
5. Як можна інтерпретувати параметри простої економетричної моделі?
6. За яких умов неможливе використання 1 МНК?
7. Запишіть оператор оцінювання 1 МНК.
8. Що таке гомоскедастичність? При порушенні якої умови застосування 1 МНК має місце гомоскедастичність?
9. Сутність кореляційного та регресійного аналізу.
10. Побудова точкового та інтервального прогнозу залежної змінної в моделі парної лінійної регресії. Навести відповідні формули.
11. Коефіцієнт детермінації: формули для обчислення та сутність.
12. Коефіцієнт кореляції: формули для обчислення та сутність.
13. Критерій Стьюдента: сутність та формула розрахунку.
14. У яких межах має знаходитися похибка апроксимації, щоб можна було зробити висновок про «гарну» якість моделі?
15. Алгоритм побудови довірчих інтервалів для параметрів a_0 та a_1 та функції регресії. Навести відповідні формули.

Приклад. На підставі квартальних даних за період 2009–2012 роки наводяться статистичні дані характеристики (тис. грн) банківської системи України (табл. 1.6).

Таблиця 1.6. Вхідні статистичні дані характеристики банківської системи України

Дата	ВК (x), тис. грн.	Всього активів (y), тис. грн.
01.01.2009	119263048	926086498
01.04.2009	117081585	870633535
01.07.2009	112597492	864694968
01.10.2009	117968018	889958533
01.01.2010	120207619	873449574
01.04.2010	126646323	874964709
01.07.2010	127162304	885255711
01.10.2010	132802031	917497465
01.01.2011	137725113	942083994
01.04.2011	138434527	995033185
01.07.2011	147816550	1019811043
01.10.2011	151866259	1029162518
01.01.2012	155486926	1054272287
01.04.2012	162236166,3	1082473105
01.07.2012	163775906,4	1104395259
01.10.2012	165810129,5	1117445882
01.01.2013	170196261,8	1127179379

Розв'язання

Для розрахунку параметрів рівняння лінійної регресії будуємо розрахункову таблицю 1.7.

Таблиця 1.7

	x	y	yx	x^2	y^2	$\hat{y}_x$	$y-\hat{y}_x$	A_i
1	119	926	110448	14224	857636	879	48	5,13
2	117	871	101935	13708	758003	868	3	0,30
3	113	865	97362	12678	747697	846	18	2,11
4	118	890	104987	13916	792026	872	18	1,98
5	120	873	104995	14450	762914	883	–10	1,11
6	127	875	110811	16039	765563	914	–39	4,48

Продовження табл. 1.7

	x	y	yx	x^2	y^2	$\hat{y}_x$	$y-\hat{y}_x$	A_i
7	127	885	112571	16170	783678	917	–31	3,55
8	133	917	121846	17636	841802	944	–26	2,88
9	138	942	129749	18968	887522	968	–26	2,71
10	138	995	137747	19164	990091	971	24	2,41
11	148	1020	150745	21850	1040015	1016	3	0,34
12	152	1029	156295	23063	1059175	1036	–7	0,65
13	155	1054	163926	24176	1111490	1053	1	0,09
14	162	1082	175616	26321	1171748	1086	-3	0,32
15	164	1104	180873	26823	1219689	1093	11	1,00
16	166	1117	185284	27493	1248685	1103	14	1,28
17	170	1127	191842	28967	1270533	1124	3	0,25
Усього	2367	16574	2337032	335647	16308268	16574	0	31
Среднє значення	139	975	137472	19744	959310	–	–	2
σ	19	96	–	–	–	–	–	–
σ^2	378	9301	–	–	–	–	–	–

$$b = \frac{\overline{y \cdot x} - \overline{y} \cdot \overline{x}}{\overline{x^2} - \overline{x}^2} = \frac{137472 - 139 \cdot 975}{19744 - 139^2} = 4{,}82,$$

$$a = \overline{y} - b \cdot \overline{x} = 975 - 4{,}82 \cdot 139 = 303{,}18.$$

Одержане рівняння регресії:

$y = 303{,}18 + 4{,}82 \cdot x.$

Отже, зі збільшенням власного капіталу банків на 1 млрд грн, активи зростають в середньому на 4,82 млрд грн.

Тісноту лінійного зв’язку оцінимо за допомогою коефіцієнта кореляції:

$$r_{xy} = b \cdot \frac{\sigma_x}{\sigma_y} = 4{,}82 \cdot \frac{19}{96} = 0{,}97;$$

$$r_{xy}{}^2 = 0{,}95.$$

Це означає, що 94,7% варіації активів банків (y) пояснюється варіацією фактора x – величиною власного капіталу.

Якість моделі визначає середня помилка апроксимації:

$$\overline{A} = \frac{1}{n}\sum A_i = \frac{31}{17} = 1,8\%.$$

Якість побудованої моделі оцінюється як висока, оскільки $\bar{A}$ не перевищує 8–10%.

Оцінку значущості рівняння регресії в цілому проведемо за допомогою F-критерію Фішера. Фактичне значення F-критерію дорівнює

$$F_{факт} = \frac{r_{xy}^2}{1 - r_{xy}^2} \cdot (n-2) = \frac{0,95}{1-0,95} \cdot 15 = 269,03.$$

Табличне значення критерію при п'ятипроцентному рівні значущості і ступенях свободи $k_1 = 1$ і $k_2 = 12 - 2 = 10$ становить $F_{табл.} = 4,96$. Оскільки $F_{факт} = 269,03 > F_{табл.} = 4,96$, то рівняння регресії вважається статистично значущим.

Оцінку статистичної значущості параметрів регресії проведемо за допомогою t-статистики Стьюдента і шляхом розрахунку довірчого інтервалу кожного з показників.

Табличне значення t-критерію для кількості ступенів свободи $df = n - 2 = 12 - 2 = 10$ і $\alpha = 0,05$ становитиме $t_{табл.} = 2.23$.

Визначимо випадкові помилки m_a, m_b, $m_{r_{xy}}$:

$$m_a = S_{ост} \cdot \frac{\sqrt{\sum x^2}}{n \cdot \sigma_x} = 22,89 \cdot \frac{\sqrt{335647}}{17 \cdot 19} = 40,10;$$

$$m_b = \frac{S_{ост}}{\sigma_x \cdot \sqrt{n}} = \frac{22,89}{19 \cdot \sqrt{17}} = 0,29;$$

$$m_{r_{xy}} = \sqrt{\frac{1 - r_{xy}^2}{n-2}} = \sqrt{\frac{1-0,973}{17-2}} = 0,06.$$

Тоді

$$t_a = \frac{a}{m_a} = \frac{303,18}{40,10} = 7,56;$$

$$t_b = \frac{b}{m_b} = \frac{4,82}{0,29} = 16,62;$$

$$t_{r_{xy}} = \frac{r_{xy}}{m_{r_{xy}}} = \frac{0,973}{0,059} = 16,49.$$

Фактичні значення t-статистики перевершують табличне значення:

$$t_a = 7,56 > t_{табл.} = 2,3;$$

$$t_b = 16,62 > t_{табл.} = 2,3;$$

$$t_{r_{xy}} = 16,49 > t_{табл.} = 2,3,$$

тому параметри a, b і r_{xy} не випадково відрізняються від нуля, а є статистично значущими.

Розрахуємо довірчі інтервали для параметрів регресії a і b. Для цього визначимо граничну помилку для кожного показника:

$$\Delta_a = t_{табл.} \cdot m_a = 2,23 \cdot 40,10 = 89,42;$$

$$\Delta_b = t_{табл.} \cdot m_b = 2,23 \cdot 0,29 = 0,65.$$

Довірчі інтервали

$$\gamma_a = a \pm \Delta_a = 303,18 \pm 89,42;$$

$$\gamma_{a\min} = 303,18 - 89,42 = 253,36;$$

$$\gamma_{a\max} = 303,18 + 89,42 = 353,00;$$

$$\gamma_b = b \pm \Delta_b = 4,82 \pm 0,65;$$

$$\gamma_{b\min} = 4,82 - 0,65 = 4,17;$$

$$\gamma_{b\max} = 4,82 + 0,65 = 5,15.$$

Аналіз верхньої і нижньої меж довірчих інтервалів надає можливість дійти висновку про те, що з імовірністю $p = 1 - \alpha = 0,95$ параметри a і b, знаходячись у вказаних межах, не приймають нульових значень, тобто не є статистично незначущими й істотно відмінні від нуля.

Одержані оцінки рівняння регресії дозволяють застосувати його для прогнозу. Якщо прогнозне значення власного капіталу банків становитиме:

$$x_p = \bar{x} \cdot 1,07 = 139 \cdot 1,07 = 148,73 \text{ млрд грн},$$

тоді прогнозне значення заробітної плати становитиме:

$y_p = 303{,}18 + 4{,}82 \cdot 148{,}73 = 1020{,}06$ млрд грн.

Помилка прогнозу становитиме:

$$m_{y_p} = S_{ост} \cdot \sqrt{1 + \frac{1}{\mathrm{n}} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2}} = 22{,}89 \cdot \sqrt{1 + \frac{1}{17} + \frac{(148{,}73 - 139)^2}{6056{,}50}} = 1{,}04.$$

Гранична помилка прогнозу, яка в 95% випадків не буде перевищена, становитиме:

$$\Delta_{y_p} = t_{табл.} \cdot \mathrm{m}_{y_p} = 2{,}23 \cdot 1.04 = 2{,}32.$$

Довірчий інтервал прогнозу:

$\gamma_{y_p} = y_p \pm \Delta_{y_p}$;

$\gamma_{y_{p\min}} = 1020{,}06 - 2{,}32 = 1017{,}74$ млрд грн;

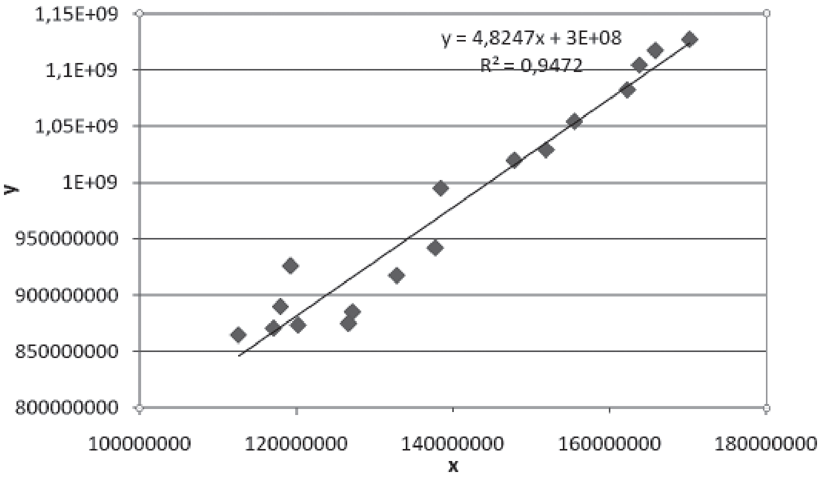


Рис.1.8. Кореляційне поле точок та парне лінійне рівняння регресії залежності активів від власного капіталу банків у цілому за показниками характеристики банківської системи України

$$\gamma_{\hat{y}_{p_{\max}}} = 1020{,}06 + 2{,}32 = 1022{,}38 \text{ млрд грн;}$$

Виконаний прогноз активів банків є надійним ($p = 1 - a = 1 - 0{,}05 = 0{,}95$) і знаходиться в межах від 1017,74 млрд грн до 1022,38 млрд грн.

Побудуємо на одному графіку початкові дані і теоретичну пряму (рис. 1.8).

Тести

1. Виберіть правильну відповідь з правого стовпчика.

1. Об'єктом економетрії є:	б) економічні системи та простори різного рівня складності – від окремого підприємства чи фірми до економіки галузей, регіонів, держави й світу загалом;
2. Предмет економетрії – ...	в) методи побудови та дослідження математико-статистичних моделей економіки, здійснення кількісних досліджень економічних явищ, пояснення та прогнозування розвитку економічних процесів;
3. Метою економетричного дослідження є:	г) аналіз реальних економічних систем і процесів, що в них відбуваються, за допомогою економетричних методів і моделей, їх застосування при прийнятті науково обгрунтованих управлінських рішень;
4. Основне завдання економетрії – ...	а) оцінити параметри моделей з урахуванням особливостей вхідної економічної інформації, перевірити відповідність моделей досліджуваному явищу і спрогнозувати розвиток економічних процесів.

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

3. Які з наведених особливостей характеризують економетричну модель:
   1) є логічним (звичайно математичним) описом того, що економічна теорія вважає особливо важливим при дослідженні певної проблеми;

2) має форму рівняння чи системи рівнянь;
3) поєднує не лише теоретичний, якісний аналіз взаємозв’язків, а й емпіричну інформацію;
4) завжди наявні стохастичні залишки;
5) описує кореляційно-регресійний зв’язок між економічними показниками.

4. Поняття “економетрика” охоплює такі напрями:
   а) теоретичні дослідження;
   б) абстрактно-теоретичні дослідження;
   в) дослідження суто емпірично-статистичного спрямування;
   г) дослідження математичних моделей економіки, які не використовують емпіричних даних;
   д) дослідження, що грунтувалися на використанні математики й статистики;
   е) -.

5. Виберіть правильну відповідь з правого стовпчика:

1. Економетрія в широкому розумінні є ...	а) сукупністю різноманітнихекономічних досліджень, що здійснюються з використанням математичних методів, а також усі сфери застосування математичних методів для розв'язання прикладних економічних завдань;
2. Економетрія у вузькому розумінні є ...	б) використанням статистичних методів в економічних дослідженнях, а саме побудова математико-статистичних моделей економічних процесів, оцінка параметрів моделей

6. Визначте основні завдання економетрії:
   1) дослідження розвитку економічних процесів і прогнозування їх динаміки;
   2) правильний вибір факторів при побудові математико-статистичних моделей;
   3) вибір та побудова математико-статистичної моделі;
   4) здійснення низки модельних експериментів;
   5) аналіз одержаних результатів і перенесення їх на реальну економічну систему (процес).

7. Вставте пропущене слово:
   А. Економетрія – це самостійна наукова дисципліна, яка об’єднує сукупність теоретичних результатів, засобів, прийомів, методів і моделей, призначених для того, щоб на базі економічної теорії, еко-

номічної статистики та ... інструментарію надавати конкретних кількісних значень загальним (якісним) закономірностям, обґрунтованим економічною теорією.

Б. Економетрія – це прикладна ... дисципліна, яка вивчає методи ... вимірювання взаємозв'язків між ... показниками та напрями їх застосування в економічних дослідженнях і ... економічній діяльності.

1) економіко-математичний;
2) кількісний;
3) економічний;
4) практичній.

8. Зв'язки між залежною та незалежною (незалежними) змінними, що описуються співвідношеннями $y = f(x) + u$, $y = F(x_1, x_2, ..., x_m) + u$ (1.4), називають:

1) регресійними рівняннями (моделями);
2) множинною регресією;
3) кореляційною залежністю;
4) кореляційно-регресійною залежністю;
5) статистичною залежністю.

9. Визначте причини обов'язкової наявності в регресійних моделях випадкового фактора:

1) введення в модель всіх пояснювальних змінних;
2) неправильний вибір функціональної форми моделі;
3) агрегування змінних;
4) помилки вимірювань;
5) помилки обчислень;
6) обмеженість статистичних даних;
7) непередбачуваність людського фактора;
8) обмеженість часу проведення дослідження.

10. Виберіть правильну відповідь з правого стовпчика:

1. Завдання регресійного аналізу	а) знаходження загальної закономірності, що характеризує залежність двох (чи більше) кореляційно пов'язаних змінних, тобто розроблення математичної моделі зв'язку
2. Завдання кореляційного аналізу	б) визначення міцності зв'язку

11. Побудова якісного рівняння регресії передбачає етапи:
    1) визначення параметрів обраного рівняння;
    2) вибір форми рівняння регресії;
    3) аналіз якості рівняння та перевірка адекватності рівняння емпіричним даним, удосконалення рівняння.

12. Залежну змінну *Y* називають:
    1) ендогенною змінною;
    2) регресандом;
    3) результативною ознакою;
    4) факторною ознакою;
    5) екзогенною змінною;
    6) регресором.

13. Моделі класифікують за призначенням та метою використання:
    1) аналітичні;
    2) імітаційні;
    3) прогностичні;
    4) макроекономічні;
    5) мікроекономічні;
    6) однофакторні;
    7) багатофакторні.

14. Виберіть правильну відповідь з правого стовпчика:

1. Перехресні дані ...	а) характеризують один і той самий об'єкт, але в різні моменти часу;
2. Часові ряди ...	б) дані за якимось економічним показником, отримані для різних однотипних об'єктів (фірм, регіонів). Причому дані отримано в один і той самий момент часу, або часова належність є несуттєвою

ТЕМА 2

Методи побудови загальної лінійної економетричної моделі

Загальна лінійна економетрична модель ● Емпірична модель множинної лінійної регресії ● Зведення нелінійних економетричних моделей до лінійного вигляду ● Метод найменших квадратів ● Оператор оцінювання МНК 1 ● Передумови застосування методу найменших квадратів (1 МНК) – умови Гауса – Маркова ● Верифікація моделі ● Перевірка значущості та довірчі інтервали ● Етапи дослідження загальної лінійної моделі множинної регресії ● Прогнозування за лінійною моделлю ● Методи побудови багатофакторної регресійної моделі

2.1. Загальна лінійна економетрична модель

На будь-який економічний показник Y, як правило, впливають не один, а декілька факторів (регресорів) $X_1, X_2, ..., X_m$. Так, наприклад, попит населення на певний банківський продукт або послугу буде визначатися не тільки вартістю її надання, а й вартістю замінників, доходами клієнтів банку та іншими факторами. У низці досліджень аналізується зв'язок доходу працівника банку з його рівнем освіти, віком, стажем роботи в цій галузі.

У подібних випадках маємо справу з множинною лінійною моделлю (регресією), яка описує взаємний зв'язок між залежною змінною Y та регресорами $X_1, X_2, ..., X_m$ і яку можна подати у такому вигляді:

$$\mathrm{M}\left(\mathrm{Y}/X_1, X_2, .., X_m\right) = \alpha\left(X_1, X_2, .., X_m\right). \tag{2.1}$$

Цей математичний запис інформує про функціональну залежність умовного математичного сподівання залежної змінної Y від m регресорів (незалежних, пояснювальних) змінних $X = (X_1, X_2, ..., X_m)$.

Отже, постає завдання виявлення статистичного взаємозв'язку між Y та X.

Загальний запис теоретичної лінійної множинної регресії можна подати в такому вигляді:

$$
\begin{aligned}
y_1 &= a_0 + a_1 x_{11} + a_2 x_{12} + \ldots + a_j x_{1j} + \ldots + a_m x_{1m}, \\
y_2 &= a_0 + a_1 x_{12} + a_2 x_{22} + \ldots + a_j x_{2j} + \ldots + a_m x_{2m}, \\
&\vdots \\
y_i &= a_0 + a_1 x_{i1} + a_2 x_{i2} + \ldots + a_j x_{ij} + \ldots + a_m x_{im}, \\
&\vdots \\
y_n &= a_0 + a_1 x_{n1} + a_2 x_{n2} + \ldots + a_j x_{nj} + \ldots + a_m x_{nm},
\end{aligned} \tag{2.2}
$$

де $a_j (j = \overline{1, m})$ – теоретичні коефіцієнти регресії (часткові коефіцієнти), або параметри теоретичної регресії, які характеризують реакцію залежної змінної $y_i (i = \overline{1, n})$ на зміну кожного регресора $X_j (j = \overline{1, m})$;
a_0 – вільний член, який визначає значення y_i за умови, коли значення регресорів дорівнюють нулеві;
$x_{ij} (i = \overline{1, n},\ j = \overline{1, m})$ – значення X_j-го регресора при i-му спостереженні;
ε_i – випадковий збудник при i-му спостереженні.

Для однозначного визначення параметрів a_j моделі (2.2) необхідно, щоб виконувалася нерівність:

$$n \geq m + 1, \tag{2.3}$$

де n – кількість спостережень;
m – кількість регресорів у моделі.

У векторно-матричній формі теоретичну модель (2.2) можна подати так:

$$\vec{Y} = X \cdot \vec{\alpha}, \tag{2.4}$$

де
$$
\vec{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}, \quad
X = \begin{pmatrix}
1 & x_{11} & x_{12} & \ldots & x_{1j} & \ldots & x_{1m} \\
1 & x_{21} & x_{22} & \ldots & x_{2j} & \ldots & x_{2m} \\
\vdots & & \vdots & & \vdots & & \vdots \\
1 & x_{i1} & x_{i2} & \ldots & x_{ij} & \ldots & x_{im} \\
\vdots & & \vdots & & \vdots & & \vdots \\
1 & x_{n1} & x_{n2} & \ldots & x_{nj} & \ldots & x_{nm}
\end{pmatrix}, \quad
\vec{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_j \\ \vdots \\ \alpha_m \end{pmatrix}.
$$

Компоненти a_j вектора $\vec{a}$ є величинами сталими ($a_j = const$), але невідомими. Їх необхідно оцінити шляхом обробки вибірки, а тому надалі будемо мати справу з емпіричною моделлю, яка є прообразом теоретичної (2.2), (2.4):

$$\vec{y} = X \cdot \vec{\alpha} + \vec{\varepsilon}, \tag{2.5}$$

де $\vec{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix}$, $X = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1m} \\ 1 & x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2m} \\ \vdots & & \vdots & & \vdots & & \vdots \\ 1 & x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{im} \\ \vdots & & \vdots & & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{nm} \end{pmatrix}$, $\vec{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_j \\ \vdots \\ \alpha_m \end{pmatrix}$, $\vec{\varepsilon}^{\,*} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_i \\ \vdots \\ \varepsilon_n \end{pmatrix}$.

Тут вектор $\vec{\alpha}$ є статистичною оцінкою теоретичного вектора $\vec{a}$ лінійної множинної регресії (2.4).

Вектор похибок $\vec{\varepsilon}$ присутній для емпіричної моделі.

2.2. Емпірична модель множинної лінійної регресії

Емпірична модель становить собою статистичний аналог теоретичної моделі (2.2). За її допомогою визначаються статистичні оцінки параметрів α_j. При цьому використовується статистична обробка вибірки.

У загальному вигляді емпірична модель записується як

$$\begin{aligned} y_1 &= \alpha_0 + \alpha_1 x_{11} + \alpha_2 x_{12} + \alpha_3 x_{13} + \dots + \alpha_m x_{1m} + \varepsilon_1, \\ y_2 &= \alpha_0 + \alpha_1 x_{21} + \alpha_2 x_{22} + \alpha_3 x_{23} + \dots + \alpha_m x_{2m} + \varepsilon_2, \\ y_3 &= \alpha_0 + \alpha_1 x_{31} + \alpha_2 x_{32} + \alpha_3 x_{33} + \dots + \alpha_m x_{3m} + \varepsilon_3, \\ &\vdots \\ y_n &= \alpha_0 + \alpha_1 x_{n1} + \alpha_2 x_{n2} + \alpha_3 x_{n3} + \dots + \alpha_m x_{nm} + \varepsilon_n. \end{aligned} \tag{2.6}$$

У векторно-матричній формі система (2.6) має вигляд:

$$\vec{y} = X \cdot \vec{\alpha} + \vec{\varepsilon}, \tag{2.7}$$

де $\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$, $X = \begin{pmatrix} 1 & x_{11} & x_{12} & & x_{1m} \\ 1 & x_{21} & x_{22} & \cdots & x_{2m} \\ & \vdots & & & \\ & & & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix}$, $\vec{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_m \end{pmatrix}$, $\vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \vdots \\ \varepsilon_n \end{pmatrix}$.

Компоненти α_j вектора $\vec{\alpha}$ є статистичними оцінками компонент α_j теоретичного вектора $\vec{\alpha}$ лінійної множинної регресії (2.4), а компоненти ε_i вектора похибок $\vec{\varepsilon}$ – статистичні оцінки випадкових збудників ε_i вектора $\vec{\varepsilon}$.

Якщо теоретичний вектор $\vec{a}$ є величиною сталою і нам невідомою, то емпіричний вектор $\vec{\alpha}$ ми можемо визначити шляхом обробки статистичної інформації вибірки обсягом n. Враховуючи те, що вибірка становить лише незначну частину генеральної сукупності $(n \leq N)$, то інформація, яку одержимо при статистичній обробці, про регресори X_j моделі буде неповною, і для кожної іншої вибірки буде потерпати певних змін. Отже, компоненти α_j емпіричного вектора $\vec{\alpha}$ будуть містити елемент випадковості. Таким чином, α_j, як і сам вектор $\vec{\alpha}$, буде випадковою величиною, що має певні закони розподілу ймовірностей із відповідними числовими характеристиками.

На підставі наведеного можна тепер стверджувати, що $\vec{\alpha}$ є статистичною оцінкою для теоретичного вектора $\vec{a}$. А тому постають питання математичної статистики: зміщена чи незміщена ця статистична оцінка; у якому довірчому інтервалі із заданою надійністю γ можуть перебувати теоретичні компоненти (параметри) a_j і сама функція регресії; як здійснити перевірку на статистичну значущість теоретичних параметрів a_j за заданим рівнем значущості α.

Для вирішення цих питань слід визначити числові характеристики для параметрів α_j $(j = 0, 1, 2, .., m)$ і для самої функції регресії, використовуючи при цьому елементи матричної алгебри як інструментарію, застосовуючи який ми можемо без громіздких викладок отримати необхідні результати.

2.3. Зведення нелінійних економетричних моделей до лінійного вигляду

Дія багатьох чинників на результативну змінну може бути описана лінійною моделлю:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m + \varepsilon, \tag{2.8}$$

де y – досліджувана (залежна, пояснювана) змінна, або регресанд;
$x_1, x_2, .., x_m$ – незалежні, пояснювальні змінні, або регресори;
$\alpha_1, \alpha_2, .., \alpha_m$ – параметри моделі;
ε – випадкова складова регресійного рівняння.

Функція (2.8) є лінійною відносно незалежних змінних і параметрів моделі, але саме лінійність за параметрами є більш суттєвою, оскільки це пов'язано з методами оцінювання параметрів. Випадкова складова ε є результативною дією всіх неконтрольованих випадкових факторів, що зумовлюють відхилення реальних значень досліджуваного показника *y* від аналітичних (обчислених на підставі обраної регресійної залежності).

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

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

$$y = a_0 x_1^{a_1} x_2^{a_2} \dots x_m^{a_m} \qquad (2.9)$$

після логарифмування набуває вигляду

$$\ln y = \ln a_0 + x_1 \ln a_1 + x_2 \ln a_2 + \dots + x_m \ln a_m, \qquad (2.10)$$

і після заміни $\ln a_i = a_i^*$, i=1, 2,..., m є лінійною відносно параметрів a_1^*, a_2^*, .., a_m^*. Показникова функція

$$y = a_0 a_1^{x_1} a_2^{x_2} \dots a_m^{x_m} \qquad (2.11)$$

після логарифмування набуває вигляду

$$\ln y = \ln a_0 + a_1 \ln x_1 + a_2 \ln x_2 + \dots + a_m \ln x_m \qquad (2.12)$$

і після заміни $\ln x_i = \beta_i$, i =1, 2, ..., m є лінійною відносно параметрів β_i.

Гіперболічна

$$y = a_0 + \frac{a_1}{x_1} + \frac{a_2}{x_2} + \dots + \frac{a_m}{x_m} \qquad (2.13)$$

і квадратична

$$y = a_0 + a_1 x_1^2 + a_2 x_2^2 + \ldots + a_m x_m^2 \qquad (2.14)$$

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

$$y = a_0 + a_1 z_1 + a_2 z_2 + \ldots + a_m z_m. \qquad (2.15)$$

Таким чином, переходи від змінних нелінійних функцій до нових змінних лінеаризованих економетричних моделей можна подати у формі таблиці (табл. 2.1).

Приклад

У реальних ситуаціях показник характеристики розглянутої економічної системи визначається, як правило, не одним, а багатьма факторами, тому частіше за все застосовують багаторесурсні або багатофакторні виробничі функції. Найпоширенішою серед них є виробнича функція Кобба – Дугласа, яка описує залежність між обсягом виробленої продукції Y і витратами праці L та капіталу F:

$$Y = aF^{\alpha} L^{\beta}.$$

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

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

Зазначена функція має такі властивості:

- коефіцієнт α показує, на скільки відсотків зміниться обсяг випуску продукції, якщо витрати праці зміняться на 1%, а витрати капіталу залишаться незмінними. Такий показник називається коефіцієнтом еластичності випуску за витратами праці;
- коефіцієнт β є коефіцієнтом еластичності випуску за витратами капіталу;
- сума параметрів $\alpha + \beta$ описує масштаб виробництва.

Якщо ця сума дорівнює одиниці, маємо постійний масштаб виробництва. А це означає, що зі збільшенням обох виробничих ресурсів на одиницю обсяг продукції також зросте на одиницю. Якщо сума менша від одиниці, то масштаб виробництва спадний, тобто

Таблиця 2.1. Зведення нелінійних економетричних моделей до лінійного вигляду

№	Вигляд функції	Форма заміни змінної
1	$y = a_0 + a_1 x$	–
2	$y = a_0 + a_1 \ln x$	$x^* = \ln x$
3	$y = a_0 + a_1 \dfrac{1}{x}$	$x^* = \dfrac{1}{x}$
4	$y = a_0 \cdot a_1^{\,x}$	$y^* = \ln y$ $a_0^* = \ln a_0$ $a_0^* = \ln a_1$
5	$y = a_0 \cdot x^{a_1}$	$y^* = \ln y$ $a_0^* = \ln a_0$
6	$\mathrm{y} = \mathrm{e}^{\mathrm{a}_0 + \frac{a_1}{x}}$	$y^* = \ln y$ $x^* = \dfrac{1}{x}$
7	$y = \dfrac{1}{a_0 + a_1 x}$	$y^* = \dfrac{1}{y}$
8	$y = \dfrac{1}{a_0 + \ln x}$	$y^* = \dfrac{1}{y}$ $x^* = \ln x$
9	$y = \dfrac{x}{a_0 + a_1 x}$	$y^* = \dfrac{1}{y}$ $x^* = \dfrac{1}{x}$

темпи зростання обсягу продукції нижчі за темпи зростання обсягу ресурсів. Якщо сума перевищує одиницю, маємо зростаючий масштаб: темпи зростання обсягу продукції перевищують темпи зростання обсягу виробничих ресурсів.

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

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

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

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

Отже, предметом наших досліджень буде узагальнена багатофакторна лінійна регресійна модель (2.8).

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

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

2.4. Метод найменших квадратів

В економічних дослідженнях найбільш широке використання знайшли моделі лінійної регресії, хоча це і є спрощений засіб у моделюванні реальних економічних процесів. Ґрунтовне вивчення і застосування методики побудови лінійних моделей надає необхідну теоретичну базу для створення більш складних, нелінійних моделей, які більшою мірою відповідають реальним економічним процесам.

Прості лінійні регресійні моделі встановлюють лінійну залежність між двома змінними.

Залежна та незалежна змінні – одна із змінних вважається залежною змінною (y) та розглядається як функція від незалежної змінної (x) (табл. 2.2).

Таблиця 2.2. Приклади визначення залежної та незалежної змінних лінійної регресійної моделі

№ прикладу	Залежна змінна	Незалежна змінна
1	обсяги резервів банку	склад кредитного портфелю
2	обсяги витрат банку	обсяги депозитів
3	зміни рейтингу банку	фактор часу

Якщо в рівнянні є лише одна пояснювальна змінна, то одержуємо теоретичну модель, яка дістала назву парної лінійної регресії:

$$y_i = a_0 + a_1 x_i. \tag{2.16}$$

Теоретичну модель для парної лінійної регресії можна записати в такому вигляді:

$$\begin{aligned} &y_1 = a_0 + a_1 x_1, \\ &y_2 = a_0 + a_1 x_2, \\ &\dots\dots\dots\dots\dots\dots \\ &y_i = a_0 + a_1 x_i, \\ &\dots\dots\dots\dots\dots\dots \\ &y_n = a_0 + a_1 x_n, \end{aligned} \tag{2.17}$$

або у векторно-матричній формі, співвідношення (2.17) буде мати вигляд

$$\vec{y} = X\vec{a}, \tag{2.18}$$

де $\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{pmatrix}$, $\vec{a} = \begin{pmatrix} a_0 \\ a_1 \end{pmatrix}$, $X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}$.

Для визначення теоретичних коефіцієнтів a_0, a_1 необхідно буде використати всі значення (x_i, y_i) $(i = \overline{1,n})$ змінних Y і X генеральної сукупності, що практично здійснити неможливо. Тому переходимо до побудови так званого емпіричного рівняння на базі інформації, одержаної зі статистичної вибірки.

Емпіричне рівняння регресії має вигляд:

$$
\begin{aligned}
&y_1 = \alpha_0 + \alpha_1 x_1 + \varepsilon_1, \\
&y_2 = \alpha_0 + \alpha_1 x_2 + \varepsilon_2, \\
&\dots\dots\dots\dots\dots\dots \\
&y_i = \alpha_0 + \alpha_1 x_i + \varepsilon_i, \\
&\dots\dots\dots\dots\dots\dots \\
&y_n = \alpha_0 + \alpha_1 x_n + \varepsilon_n,
\end{aligned}
\tag{2.19}
$$

який, аналогічно з теоретичною моделлю, запишемо у векторно-матричній формі:

$$\vec{y} = X\vec{\alpha} + \vec{\varepsilon}. \tag{2.20}$$

$$\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_n \end{pmatrix}, \quad \vec{\alpha} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \vec{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Метод найменших квадратів

У загальному вигляді проста вибіркова регресійна модель запишеться так:

$$y = \alpha_0 + \alpha_1 x + \varepsilon, \tag{2.21}$$

де y — вектор спостережень за залежною змінною, $y = \{y_1, y_2, \dots, y_n\}$;
x— вектор спостережень за незалежною змінною, $x = \{x_1, x_2, \dots, x_n\}$;
α_0, α_1 — невідомі параметри регресійної моделі;
ε — вектор випадкових величин (помилок); $\varepsilon = \{\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n\}$.

Лінійна регресійна модель. Регресійна модель називається ***лінійною***, якщо вона лінійна за своїми параметрами. Отже, модель (2.21) є лінійною регресійною моделлю, її ще можна трактувати і як пряму на площині, де α_0 — перетин з віссю ординат, а α_1 — нахил (звичайно, якщо абстрагуватися від випадкової величини ε).

Щоб мати явний вигляд залежності, необхідно знайти (оцінити) невідомі параметри α_0, α_1 цієї моделі. Як це зробити? Яким критерієм краще користуватися? Щоб відповісти на ці питання, розглянемо спочатку приклад.

Приклад регресійної моделі

Відділ економічного аналізу комерційного банку оцінює ефективність кредитного відділу. Для такої оцінки вони мають досвід праці у п'яти географічних зонах з майже однаковими умовами (потенційні клієнти, ставлення до даного банку і т. ін.). У цих зонах вони зафіксували протягом однакового періоду обсяги наданих кредитів та витрати банку, пов'язані з рекламною кампанією (млн грн). Дані наведені в табл. 2.2. [Козьменко О. В. Статистика: банківський досвід : навчальний посібник : у 2 ч. / О. В. Козьменко, О. В. Меренкова. – Суми : ДВНЗ «УАБС НБУ», 2009. – Ч. 2 – С. 5–136].

Таблиця 2.2. Вхідні дані моделювання

i	y_i	x_i
1	25	5
2	30	6
3	35	9
4	45	12
5	65	18

Реальні спостереження y зобразимо точками в системі координат (X, Y) (рис. 2.1).

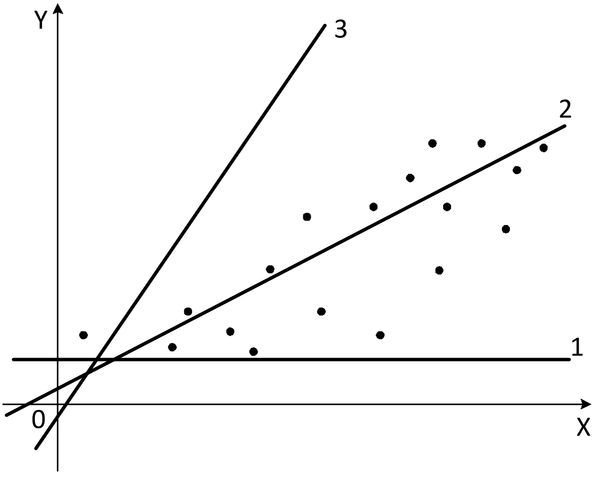


Рис. 2.1. Залежність між обсягами наданих банком кредитів та витратами на рекламу

Яка пряма більше підходить? Інтуїтивно ми обираємо 2. Запишемо це в математичних термінах. Що таке математичні терміни? Це координати точок.

Візуально можна припустити, що між даними є лінійна залежність, тобто їх можна апроксимувати прямою лінією.

Узагалі існує необмежена кількість прямих $y = \alpha_0 + \alpha_1 x$, які можна провести через множину спостережуваних точок. Яку ж із них вибрати?

Щоб це визначити, потрібно мати в розпорядженні певний критерій, що дозволяв би вибрати з множини можливих прямих «найкращу» за даним критерієм. Найпоширенішим є критерій мінімізації суми квадратів відхилень. На рис. 2.1, наприклад, пряма (1), як і інші, розташована таким чином, що деякі точки знаходяться вище, деякі нижче цієї прямої, на основі чого можна встановити відхилення (помилки) відносно цієї прямої:

$$\varepsilon_i = y_{istat} - y_{iteor} = y_{istat} - a_o - a_1 x, i = \overline{1, n}, \tag{2.22}$$

де y_{iteor} – i-та точка на прямій, яка відповідає значенню x_i (див. рис. 2.2), теоретичний рівень результативної ознаки;
y_{istat} – емпіричний (статистичний) теоретичний рівень результативної ознаки;

Відхилення, або помилки, ще інколи називають залишками. Логічно, що треба проводити пряму так, щоб сума квадратів помилок була мінімальною. У цьому й полягає критерій найменших квадратів: невідомі параметри a_0 та a_1 визначаються так, щоб мінімізувати $\sum\limits_{i=1}^{n} \varepsilon_i^2$. Справді, за критерієм маємо

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_{istat} - a_0 - a_1 x_i)^2 = f(a_0, a_1) \to \min. \tag{2.23}$$

Це функція двох змінних a_0 та a_1.

Знайдемо мінімум функції двох змінних.

Визначимо значення a_0 та a_1, які мінімізують вираз (2.23). Мінімум функції (2.22) досягається за необхідних умов, коли перші похідні дорівнюють нулю (рис. 2.2).

$$\frac{\partial(\sum\limits_{i=1}^{n} \varepsilon_i^2)}{\partial a_0} = \frac{\partial f(a_0, a_1)}{\partial a_0} = 0; \frac{\partial(\sum\limits_{i=1}^{n} (y_{istat} - a_0 - a_1 x_i)^2)}{\partial a_1} = \frac{\partial f(a_0, a_1)}{\partial a_1} = 0, \tag{2.24}$$

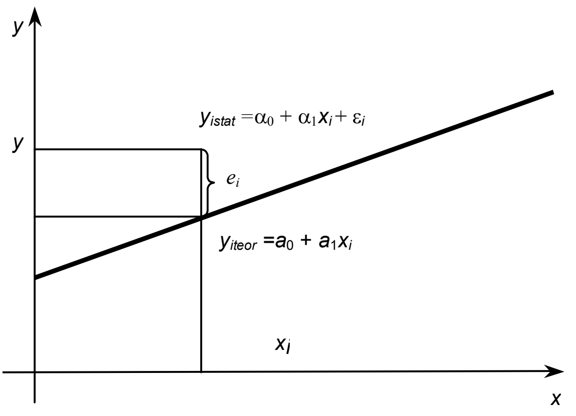


Рис. 2.2. Відхилення теоретичних значень від фактичних

$$\begin{cases} \dfrac{\partial(\sum\limits_{i=1}^{n}\varepsilon_i^2)}{\partial a_0} = -2\sum\limits_{i=1}^{n}(y_{istat} - a_0 - a_1 x_i) = 0 \\ \dfrac{\partial(\sum\limits_{i=1}^{n}\varepsilon_i^2)}{\partial a_1} = -2\sum\limits_{i=1}^{n}x_i(y_i - a_0 - a_1 x_i) = 0. \end{cases} \quad (2.25)$$

Вирішимо систему лінійних алгебраїчних рівнянь за допомогою теореми Кронекера – Капеллі[1].

Отримаємо систему лінійних рівнянь

$$\begin{cases} \sum\limits_{i=1}^{n} y_i = n a_0 + a_1 \sum\limits_{i=1}^{n} x_i; \\ \sum\limits_{i=1}^{n} y_i x_i = a_0 \sum\limits_{i=1}^{n} x_i + a_1 \sum\limits_{i=1}^{n} x_i^2, \end{cases} \quad (2.26)$$

яка називається нормальною.

[1] Теорема Кронекера – Капеллі – критерій сумісності системи лінійних алгебраїчних рівнянь. Система лінійних алгебраїчних рівнянь сумісна тоді і тільки тоді, коли ранг основної матриці дорівнює рангу її розширеної матриці, причому система має єдине рішення, якщо ранг дорівнює числу невідомих, і нескінченну безліч рішень, якщо ранг менше числа невідомих.

Розв'язок (2.26) відносно нахилу прямої (невідома a_1) дає

$$a_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}. \tag{2.27}$$

З метою спрощення виразу для a_1 чисельник та знаменник виразу (2.27) помножимо на $1/n$. Отримаємо:

$$a_1 = \frac{\frac{1}{n}\left(\sum_{i=1}^{n} x_i y_i\right) - \bar{x}\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2}, \tag{2.28}$$

де $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i.$

Отже, кут нахилу прямої регресії можна встановити за формулою (2.28).

Для визначення параметра a_0 повернемося до (2.27). Маємо:

$$\frac{\partial\left(\sum_{i=1}^{n} \varepsilon_i^2\right)}{\partial a_0} = \sum_{i=1}^{n} \frac{\partial}{\partial a_0}\left[\left(y_{istat} - a_0 - a_1 x_i\right)^2\right] = \sum_{i=1}^{n}\left(y_{istat} - a_0 - a_1 x_i\right) = 0. \tag{2.29}$$

Вираз (2.29) дає нам, по-перше, підтвердження того, що сума помилок дорівнює нулеві. Справді,

$$y_{istat} - a_0 - a_1 x_i = \varepsilon_i \Rightarrow \sum_{i=1}^{n} \varepsilon_i. \tag{2.30}$$

По-друге, розділивши (2.30) на n, маємо вираз для визначення a_0:

$$\frac{1}{n}\sum_{i=1}^{n} y_{istat} - a_0 - a_1 \frac{1}{n}\sum_{i=1}^{n} x_i = 0 \Rightarrow a_0 = \bar{y} - a_1 \bar{x}. \tag{2.31}$$

Таким чином, ми знайшли формули для визначення невідомих параметрів a_0 та a_1 і можемо записати у явному вигляді регресію y від x, у якій параметри обчислені за методом найменших квадратів, її інколи називають регресією найменших квадратів y від x. Маємо:

$$y_{teor} = a_0 + a_1 x \tag{2.32}$$

або

$$y_{stat} = y_{teor} + \varepsilon = \alpha_0 + \alpha_1 x + \varepsilon. \tag{2.33}$$

Приклад ілюстрації побудови рівняння регресії

Для ілюстрації цих викладок повернемося до нашого прикладу про дослідження ефективності витрат на рекламу. Проведені попередні розрахунки подамо у вигляді табл. 2.3.

Для обчислення невідомих параметрів a_0 , a_1 необхідно послідовно здійснити такі розрахунки:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{50}{5} = 10; \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{200}{5} = 40.$$

Таблиця 2.3. Дослідження ефективності витрат на рекламу

i	y_i	x_i	x_i^2	$x_i\,y_i$
1	25	5	25	125
2	30	6	36	180
3	35	9	81	315
4	45	12	144	540
5	65	18	324	1170
Сума	200	50	610	2330
Середнє значення	40	10	122	466

$$\mathrm{var}(x) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{610}{5};$$

$$\mathrm{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{xy} = \frac{2330}{5} - 400 = 66;$$

$$a_1 = \frac{\mathrm{cov}(x,y)}{\mathrm{var}(x)} = \frac{66}{22} = 3;$$

$$a_0 = \bar{y} - a_1\bar{x} = 40 - 3 * 10 = 10.$$

Таблиця 2.4. Систематизація причин появи випадкових збудників в регресійних рівняннях

Причина появи випадкових збудників	Сутнісна характеристика
Регресійна модель є певною мірою спрощенням реальної ситуації, яка на практиці є складним переплетінням різних факторів, багато з яких практично неможливо врахувати в моделі	Попит на банківські послуги буде визначатися як процентною ставкою, так і процентною ставкою на ті послуги, які можуть їх заміняти, доходами споживачів і та ін. Однак у цьому переліку не враховуються традиції (як релігійні, так і національні), особливості кліматичних умов і багато інших факторів. При цьому ще виникає проблема визначення факторів, які за певних умов будуть домінуючими, а якими можна знехтувати. У ряді випадків існують фактори, які не можна використати в моделі тому, що для них проблематично одержати необхідні статистичні дані. Наприклад, величина заощаджень родини визначається не лише доходами її членів, а й їхнім здоров'ям, інформацію про що в цивілізованих країнах тримають у таємниці. Крім того, багато факторів має випадковий характер (погода, стихійні лиха), які посилюють неоднозначність
Неправильно вибрана функціональна залежність	Може трапитися внаслідок недостатнього дослідження процесу, який підлягає моделюванню. Так, виробнича функція, яка описує залежність *Y* від одного фактора *X,* може бути виражена лінійним співвідношенням: $Y = \alpha_0 + \alpha_1 X,$ хоча, насправді, при більш ретельному дослідженні, стане відомо, що співвідношення між Y та X матиме нелінійний характер, наприклад: $Y = \alpha_0 X^{\alpha_1}.$ Вибір форм функціональної залежності між змінними називають специфікацією моделі
Неправильно вибрані пояснювальні змінні	-
Залежність між факторами має складну форму зв'язку між цілими комплексами подібних величин	При дослідженні залежності попиту на банківські послуги *Y* як пояснювальну вибирають змінну, яка уособлює складну комбінацію індивідуальних запитів, що мають на неї певний вплив поряд із факторами, які враховані в моделі. Здійснюється так зване агрегування пояснювальних змінних, що може бути однією з причин появи в моделі випадкового збудника ε_i
Помилки при аналізі та обробці статистичних даних	Як правило, це помилки механічного характеру
Обмеженість статистичної інформації	Більшість моделей описується неперервними функціями, але при цьому використовуються вибіркові дані, які мають дискретну структуру

Продовження табл. 2.4

Причина появи випадкових збудників	Сутнісна характеристика
Наявність людського фактора, який тією чи іншою мірою обов'язково є в будь-якому економічному процесі, але врахувати який у моделі поки що практично неможливо	У певних ситуаціях цей фактор може навіть якісну модель деформувати до примітивного рівня

Знаючи параметри a_0, a_1, отриману пряму запишемо у вигляді:

$\hat{y} = 3x + 10.$

Деяка інформація про випадкові збудники $(i = \overline{1,n})$

Причини, які спонукають появу випадкових збудників ε_i в рівняннях (2.17), можуть бути такими (табл. 2.4).

2.5. Оператор оцінювання 1 МНК

Нехай відомо n спостережень незалежних змінних $x_1, x_2, ..., x_m$ і n спостережень залежної змінної y. Необхідно за МНК оцінити параметри $\alpha_1, \alpha_2, ..., \alpha_m$ лінійної моделі (2.18).

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

1. Незалежні змінні записати у вигляді матриці

$$X = \{x_0, x_1, x_2, ..., x_m)\}, \tag{2.34}$$

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

2. Обчислити матрицю X^TX і вектор X^Ty, де X^T – транспонована матриця X, y – вектор спостережень залежної змінної.
3. Обчислити обернену матрицю $(X^TX)^{-1}$.
4. Обчислити параметри моделі за формулою

$$\alpha = \left(X^T X\right)^{-1}\left(X^T y\right). \tag{2.34}$$

де α – вектор параметрів, дорівнює $(\alpha_0, \alpha_1, \alpha_2, ..., \alpha_m)^T$.

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

$$\left(X^T X\right)a = X^T y. \tag{2.36}$$

2.6. Передумови застосування методу найменших квадратів (1 МНК) – умови Гауса – Маркова

Для визначення емпіричного вектора $\vec{\alpha}^*$ необхідно використати метод найменших квадратів (МНК), а для цього потрібно, щоб виконувалися певні умови, які називаються умовами Гаусса – Маркова, а саме:

1. *Математичне сподівання випадкових відхилень ε_i повинно дорівнювати нулеві:*

$$M(\vec{\varepsilon}) = 0. \tag{2.37}$$

Ця умова вимагає, щоб випадкові відхилення в середньому не впливали на залежну змінну Y, тобто в кожному конкретному спостереженні відхилення ε_i може набувати додатні або від'ємні значення, але не повинно спостерігатися систематичне зміщення відхилень здебільшого в бік одного знаку.

З урахуванням наведеного вище, використовуючи рівняння (2.32), будемо мати:

$$M\left(\vec{y}\right) = M\left(X\vec{\alpha} + \vec{\varepsilon}\right) = X\vec{\alpha} + M\left(\vec{\varepsilon}\right) = X\vec{\alpha}. \tag{2.38}$$

2. *Дисперсія випадкових відхилень ε_i повинна бути сталою величиною*

$$D(\varepsilon_i) = \sigma_\varepsilon^2(\varepsilon_i) = \sigma_\varepsilon^2 = const, \quad (i = \overline{1,n}). \tag{2.30}$$

Ця вимога передбачає, що незважаючи на те, що при кожному конкретному спостереженні випадкове відхилення може виявитися порівняно великим або малим, це не повинно складати основу для апріорної причини, тобто причини, що не базується на досвіді, яка спонукала б велику похибку.

3. *Випадкові відхилення ε_i та ε_j, $i \neq j$ повинні бути незалежними випадковими величинами.*

Виконання цієї умови передбачає, що між будь-якими випадковими відхиленнями відсутній систематичний зв'язок, тобто величина та знак будь-якого випадкового відхилення не буде причиною величини та знаку будь-якого іншого випадкового відхилення. Цю умову можна записати так:

$$\mathrm{co\upsilon}\left(\varepsilon_i,\varepsilon_j\right)=\begin{cases}0, & i\neq j,\\ \sigma_\varepsilon^2=\mathrm{cons}, & i=j.\end{cases} \tag{2.40}$$

Тут $co\upsilon(\varepsilon_i,\varepsilon_j)=M(\varepsilon_i\cdot\varepsilon_j)-M(\varepsilon_i)\cdot M(\varepsilon_i)=M(\varepsilon_i\cdot\varepsilon_j)=0$ є математичний запис коваріаційного (кореляційного) моменту.

4. *Випадковий вектор відхилень $\vec{\varepsilon}$ повинен бути незалежним від регресорів $\vec{X}_j$ $(j=\overline{1,m})$ матриці X.*

   Ця умова виконується автоматично, коли пояснювальні змінні X_j $(j=\overline{1,m})$ не є стохастичними величинами в заданій моделі:

$$\begin{aligned} co\upsilon(\vec{\varepsilon},\vec{X}_j)&=M[(\vec{\varepsilon}-M(\vec{\varepsilon}))\cdot(\vec{X}_j-M(\vec{X}_j)'] =\\ &=M[\vec{\varepsilon}(\vec{X}_j-M(\vec{X}_j))]=M(\vec{\varepsilon}\vec{X}_j)-M(\vec{\varepsilon}\vec{X}_j)=0, \end{aligned} \tag{2.41}$$

бо $M(\vec{\varepsilon})=0$, а $M(\vec{X}_j)=\vec{X}_j$ ($\vec{X}_j$ не є випадковою величиною).

5. *Компоненти ε_i випадкового вектора $\vec{\varepsilon}$ повинні мати нормальний закон розподілу $\varepsilon_i\sim N(0,\sigma_\varepsilon)$.*

   Тоді випадковий вектор $\vec{\varepsilon}$ матиме нормальний закон розподілу вигляду $\vec{\varepsilon}\sim N(0,\sigma_\varepsilon I)$.

6. *Між регресорами $\vec{X}_j,\vec{X}_k$ $(k,j=\overline{1,m})$ матриці X повинна бути відсутня лінійна (кореляційна) залежність.* Для цього випадку повинна виконуватися умова

$$\det\left(X'X\right)\neq 0. \tag{2.42}$$

   Слід при цьому наголосити, що матриця $X'X$ є симетричною.

7. *Економетричні моделі повинні бути лінійними відносно своїх параметрів.*

Класичні економетричні моделі – економетричні моделі, для яких виконуються умови (1–6).

Гомоскедастичнімоделі – моделі, для яких виконується умова (2) (сталість дисперсії випадкових відхилень).

Гетероскедастичні моделі – моделі, для яких не виконується умова (2) ($D(\varepsilon_i) = \sigma_\varepsilon^2 \neq const,\ i = \overline{1,n}$).

Слід також зауважити, що ранг матриці X повинен бути

$$r(X) \leq m + 1. \qquad (2.43)$$

Виконання наведених умов дає нам право на використання МНК для визначення статистичних оцінок параметрів теоретичної лінійної множинної регресії, на перевірку статистичних гіпотез та побудову інтервальних статистичних оцінок.

2.7. Верифікація моделі

У класичному регресійному аналізі вважається, що функція регресії відома до оцінювання параметрів, тобто регресійна модель специфікована правильно. Однак в емпіричних економічних і соціальних дослідженнях не завжди відомо, скільки факторів має бути введено в модель і яка форма залежності краще описує реальні зв'язки. Щоб забезпечити найбільш адекватне відтворення досліджуваного явища чи процесу необхідно вибрати регресійну функцію серед багатьох варіантів, використовуючи спеціальні критерії якості моделі.

Верифікація моделі – пізньолат. *verificatia* – підтвердження; лат. *verus* – істинний, *facio* – роблю) – доказ того, що вірогідний факт або твердження є істиним. У науці: логіко-методологічна процедура встановлення істинності наукової гіпотези на підставі їх відповідності емпіричним даним або теоретичним положенням, що відповідають емпіричним даним.

Для перевірки коректності побудови моделі визначають насамперед:

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

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

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

$$S_\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^2. \qquad (2.44)$$

Поправка на кількість ступенів свободи дає незміщену оцінку дисперсії залишків:

$$\sigma_\varepsilon^2 = \frac{1}{n-m-1}\sum_{i=1}^{n}\varepsilon_i^2. \qquad (2.45)$$

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

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

$$R^2 = 1 - \frac{S_\varepsilon^2}{S_y^2}, \qquad (2.46)$$

де $S_y^2 = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2;$

$\bar{y}$ – середнє значення залежної змінної:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i. \qquad (2.47)$$

На значення коефіцієнта детермінації впливає кількість факторів, що врахована в моделі. Введення до моделі кожної нової змінної збільшує значення коефіцієнта детермінації. Тому, щоб запобігти невиправданому розширенню моделі й мати можливість порівнювати моделі з різною кількістю факторів, вводять спеціальний оцінений коефіцієнт детермінації

$$\overline{R}^2 = 1 - \frac{\sigma_\varepsilon^2}{\sigma_y^2}, \tag{2.48}$$

де σ_ε^2 – незміщена оцінка дисперсії залишків;

σ_y^2 – незміщена оцінка дисперсії залежної змінної, $\sigma_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(y_i - \overline{y}\right)^2$.

Неважко помітити, що обидва коефіцієнти пов'язані такою залежністю

$$\overline{R}^2 = 1 - \left(1 - R^2\right)\frac{n-1}{n-m-1}. \tag{2.49}$$

Обчислений у такий спосіб коефіцієнт детермінації називається скоригованим. Крім того, застосовують також наступний підхід до коригування, що виконується за формулою:

$$\overline{R}^2 = 1 - \left(1 - R^2\right)\frac{n+m}{n-m-1}. \tag{2.50}$$

Обидва коефіцієнти (2.49) та (2.50) враховують той факт, що введення в модель кожного нового регресора зменшує кількість ступенів свободи. А для застосування статистичних критеріїв перевірки якості отриманих результатів ступенів свободи бажано мати якомога більше.

Очевидно, для кожного зі скоригованих коефіцієнтів детермінації виконується нерівність $\overline{R^2} \le R^2$, тобто зі збільшенням кількості факторів моделі оцінені коефіцієнти детермінації зростають повільніше, ніж R^2. Крім того, якщо $R^2 = 1$, то і $\overline{R^2} = 1$. Якщо R^2 прямує до нуля, оцінені коефіцієнти стають від'ємними. Така властивість скоригованих коефіцієнтів детермінації дає можливість більш об'єктивно оцінювати якість моделей з різною кількістю факторів, причому в разі застосування скоригованого коефіцієнта $\overline{R^2}$ перевага однозначно надається рівнянню з меншою кількістю регресорів.

Коефіцієнт детермінації має ще два рівноцінні визначення. За першим коефіцієнт детермінації R^2 дорівнює квадрату емпіричного коефіцієнта кореляції між двома рядами спостережень (теоретичними значеннями регресанда y_i та його розрахунковими значеннями (y_{iteor}, $i = 1, 2, .., n$) і обчислюється за формулою:

$$R^2 = \frac{(\Sigma(y_i - \overline{y})(y_{iteor} - \overline{y}))^2}{\Sigma(y_i - \overline{y})^2 \sum(y_{iteor} - \overline{y})^2}. \quad (2.51)$$

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

$$R^2 = \frac{\Sigma(y_{iteor} - \overline{y})^2}{\Sigma(y_i - \overline{y})^2}. \quad (2.52)$$

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

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

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

2.8. Перевірка значущості та довірчі інтервали

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

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

У задачах регресійного аналізу важливе значення має припущення про нормальний розподіл випадкових величин[1], що задіяні в даній моделі. Певні перетворення нормально розподілених величин забезпечують їх розподіл за законом Стьюдента чи за законом Фішера: на підставі першого з них визначаються довірчі інтервали, а другий дає можливість оцінювати відношення двох випадкових величин.

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

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

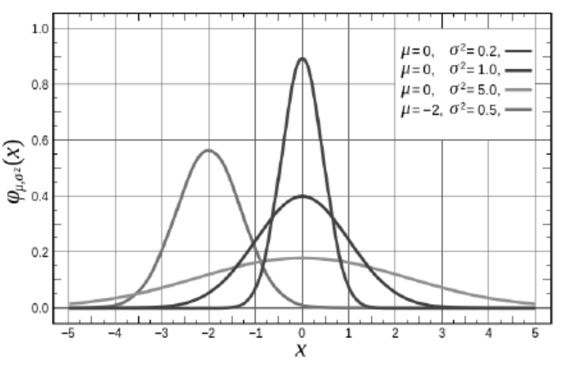
За заданим рівнем значущості множина допустимих значень розбивається на дві неперетинні множини: одна містить значення випадкової величини, імовірність досягнення яких перевищує заданий рі-

---

[1] Нормальний розподіл (розподіл Гаусса) – розподіл імовірностей випадкової величини, що характеризується щільністю ймовірності:

$$f(x,\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp(-\frac{(x-\mu)^2}{2\sigma^2}),$$

де μ – математичне сподівання, σ^2 – дисперсія випадкової величини. Параметр σ також відомий як стандартне відхилення. Розподіл із $\mu=0$ та $\sigma^2=1$ називають стандартним нормальним розподілом.



вень значущості, а інша – критична область – визначає ті значення, що досягаються рідко (імовірність потрапити до такої області нижча від заданого рівня), і розташована вона, як правило, на «хвостах розподілу».

Залежно від альтернативної гіпотези критична область може складатися з одного чи двох проміжків на числовій осі. Це буде один проміжок (правий чи лівий «хвіст» розподілу), якщо зазначається напрямок нерівності (більше або менше деякої величини), і два проміжки (обидва «хвости» розподілу), якщо встановлюється нерівність (не дорівнює певній величині).

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

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

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

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

$$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_n = 0. \tag{2.53}$$

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

Для перевірки цих гіпотез застосовують F-критерій Фішера з m і $n-m-1$ ступенями свободи. За отриманими в моделі значеннями коефіцієнта детермінації R^2 обчислюють експериментальне значення F-статистики:

$$F_{експ} = \frac{R^2}{1-R^2} \cdot \frac{n-m-1}{m}, \tag{2.54}$$

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

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

$$t_{експ} = \frac{R\sqrt{n-m-1}}{\sqrt{1-R^2}} \tag{2.55}$$

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

$$|t| > t_{табл}, \tag{2.56}$$

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

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

На підставі одного з найважливіших припущень МНК – припущення про нормальний розподіл випадкової складової рівняння з нульовим математичним сподіванням і сталою дисперсією – доведено, що кожний параметр лінійної регресії також має нормальний розподіл. Причому математичне сподівання параметра дорівнює значенню параметра узагальненої регресії, а дисперсія – незміщеній дисперсії випадкової складової рівняння, помноженій на відповідний діагональний елемент оберненої матриці $(X^TX)^{-1}$.

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

$$H_0 : \alpha_j = 0, \tag{2.57}$$

альтернативна –

$$H_A : \alpha_j \neq 0. \tag{2.58}$$

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

$$t_j = \frac{\alpha_j}{\sqrt{\sigma_\varepsilon^2 c_{jj}}} = \frac{\alpha_j}{S_{\alpha_j}}, \tag{2.59}$$

де c_{jj} – діагональний елемент матриці $(X^TX)^{-1}$;
S_{α_j} – стандартизована похибка оцінки параметра моделі,
$S_{\alpha_j} = \sigma_\varepsilon \sqrt{c_{jj}}$.

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

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

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

$$\left(\alpha_j - t_{табл}\sqrt{\sigma_\varepsilon^2 c_{jj}}; \alpha_j + t_{табл}\sqrt{\sigma_\varepsilon^2 c_{jj}}\right). \tag{2.60}$$

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

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

$$\Delta R_j^2 = \frac{(1-R^2)t_j^2}{n-m}. \tag{2.61}$$

Часткові коефіцієнти детермінації $\Delta\overline{R_T^2}$ і $\Delta\overline{R_A^2}$, обчислені за відповідними значеннями $\overline{R_T^2}$ і $\overline{R_A^2}$, можуть бути як додатними, так і від'ємними, що дає можливість більш об'єктивно оцінювати моделі з різною кількістю регресорів.

2.9. Етапи дослідження загальної лінійної моделі множинної регресії

Розглядається багатофакторна лінійна регресійна модель

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m, \tag{2.62}$$

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

Для дослідження зазначеної моделі слід виконати такі кроки.

1. За даними спостережень оцінити параметри $\alpha_1, \alpha_2, .., \alpha_m$.

2. Для перевірки адекватності отриманої моделі обчислити:
   а) залишки моделі – розбіжності між спостереженими та розрахунковими значеннями залежної змінної $\varepsilon_i = y_i - \hat{y}_i$;
   б) відносну похибку залишків та її середнє значення;
   в) залишкову дисперсію;
   г) коефіцієнт детермінації;
   д) вибірковий коефіцієнт множинної кореляції.
3. Перевірити статистичну значущість отриманих результатів:
   а) перевірити адекватність моделі загалом: за допомогою F-критерію Фішера перевірити гіпотезу
   $$H_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_m = 0 \tag{2.63}$$
   проти альтернативної H_A: існує хоча б один коефіцієнт $a_j \neq 0$;
   б) перевірити значущість коефіцієнта множинної кореляції, тобто розглянути гіпотезу $H_0 : R = 0$;
   в) перевірити істотність кожного коефіцієнта регресії: за допомогою t-критерію Стьюдента перевірити гіпотезу
   $$H_0 : \alpha_j = 0 = 1, 2, .., m \tag{2.64}$$
   проти відповідних альтернативних гіпотез
   $$H_A : \alpha_j \neq 0 = 1, 2, .., m; \tag{2.65}$$
   г) оцінити вплив кожного регресора на якість моделі, тобто обчислити часткові коефіцієнти детермінації ΔR_j^2, скоригувати їх за Тейлом і за Амемією та дати їх відповідну інтерпретацію;
   д) оцінити вплив окремих груп регресорів на змінювання регресанда, застосувавши F-критерій Фішера.
4. Обчислити та інтерпретувати коефіцієнти еластичності.
5. Визначити довірчі інтервали регресії при рівні значущості a.
6. Побудувати довірчі інтервали для параметрів регресії.
7. Обчислити прогнозні значення y_p за значеннями x_{1p}, x_{2p}, ..., x_{mp}, що перебувають за межами базового періоду, і знайти межі довірчих інтервалів індивідуальних прогнозованих значень і межі довірчих інтервалів середнього прогнозу.

Розглянемо етапи побудови та аналізу економетричних моделей на прикладі.

Практичне завдання 1

Дано: Статистичні дані – показники депозитів домашніх господарств $X(i)$ за місяцями 2007–2008 рр. залежно від строку (табл. 2.5):

x_1 – іпотечні кредити строком до одного року;

x_2 – іпотечні кредити строком від одного року до 5 років;

x_3 – іпотечні кредити на придбання, будівництво та реконструкцію нерухомості;

y – усього іпотечні кредити.

Знайти: 1. Обрати форму багатофакторної моделі.

2. Оцінити всі параметри моделі.

3. Визначити зони надійності при рівні значущості $a = 95$.

4. Оцінити коефіцієнти детермінації, автокореляції і перевірити показники на мультиколінеарність між факторами.

Хід виконання

1. Вивчення зв'язку між трьома і більше зв'язаними між собою ознаками має назву множинної (багатофакторної) регресії.

У нашому випадку залежна змінна – усього іпотечні кредити (тис. грн), позначимо y.

Пояснювальні змінні:

x_1 – іпотечні кредити строком до одного року;

x_2 – іпотечні кредити строком від одного року до 5 років;

x_3 – іпотечні кредити на придбання, будівництво та реконструкцію нерухомості;

Кількість спостережень $n = 9$.

Вимагається побудувати статистичну модель, що виражає залежність результативної ознаки від чотирьох пояснювальних змінних – факторів і проаналізувати побудовану модель.

Загальна множинна лінійна регресійна модель може бути записана у вигляді

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m + \varepsilon, \qquad (2.66)$$

де $x_1, x_2, .., x_m$ – незалежні змінні (фактори);

$a_0, a_1, .., a_m$ – параметри моделі;

ε – чисто випадкова величина.

У нашому випадку маємо модель множинної регресії з трьома факторами:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3. \qquad (2.67)$$

Таблиця 2.5. Кредити, надані домашнім господарствам, за цільовим спрямуванням і строками погашення за місяцями 2007–2008 рр. (залишки коштів на кінець періоду, млн грн)

Період	Усього	Кредити на поточні потреби				Іпотечні кредити									Інші кредити			
		усього	у тому числі за строками			усього	у тому числі за строками			з них на придбання, будівництво та реконструкцію нерухомості	у тому числі за строками				усього	у тому числі за строками		
			до одного року	від одного року до 5 років	більше 5 років		до одного року	від одного року до 5 років	більше 5 років		до одного року	від одного року до 5 років	від 5 року до 10 років	більше 10 років		до одного року	від одного року до 5 років	більше 5 років
2007																		
Січень	83 489	56 789	12 626	44 162	…	21 105	325	20 780	…	20 988	323	20 666	…	…	5 596	195	5 401	…
Лютий	87 181	59 633	12 887	29 496	17 251	25 744	376	2 360	23 008	21 586	342	1 985	4 728	14 532	1 803	183	719	901
Березень	93 262	63 695	13 503	30 522	19 670	27 607	433	4 524	22 650	23 172	392	2 967	4 872	14 941	1 960	217	831	913
Квітень	99 339	67 736	14 058	31 661	22 017	29 625	466	3 539	25 619	24 759	423	3 031	5 161	16 144	1 978	213	740	1 025
Травень	105 060	71 694	14 714	32 663	24 317	31 222	494	3 731	26 997	26 047	447	3 198	5 434	16 967	2 145	186	826	1 133
Червень	111 819	76 293	15 097	34 091	27 105	33 205	524	3 766	28 916	27 754	479	3 212	5 913	18 150	2 321	208	864	1 249
Липень	120 244	82 655	17 084	35 763	29 809	35 049	547	3 950	30 553	29 437	487	3 397	6 196	19 357	2 539	190	941	1 408
Серпень	127 244	87 501	17 539	37 425	32 536	36 916	519	4 142	32 254	31 229	488	3 522	6 461	20 758	2 827	332	983	1 512
Вересень	134 011	92 189	17 891	38 575	35 724	38 887	540	4 223	34 124	33 118	504	3 577	6 724	22 314	2 934	292	1 036	1 606
Жовтень	141 712	97 513	18 684	40 444	38 385	41 077	547	4 409	36 121	35 352	510	3 735	7 007	24 100	3 123	282	1 066	1 775
Листопад	150 108	103 171	19 435	42 008	41 729	43 586	538	4 501	38 547	37 821	499	3 749	7 425	26 149	3 351	235	1 128	1 988
Грудень	160 386	110 121	19 990	44 593	45 538	46 626	534	4 568	41 523	40 826	503	3 828	7 875	28 619	3 640	208	1 272	2 160
2008																		
Січень	164 775	113 355	20 884	45 332	47 138	47 445	573	4 526	42 345	41 889	524	3 808	8 034	29 523	3 976	206	1 324	2 446
Лютий	174 234	120 350	22 773	47 056	50 521	50 103	685	4 618	44 800	44 221	605	3 862	8 328	31 426	3 782	155	1 309	2 317
Березень	183 580	126 593	23 757	48 914	53 922	53 023	825	4 785	47 413	47 173	746	4 013	8 785	33 630	3 964	137	1 390	2 438
Квітень	191 899	132 109	24 826	50 123	57 160	55 294	833	4 775	49 686	49 894	738	3 952	9 284	35 920	4 497	139	1 409	2 949
Травень	193 546	133 933	25 894	50 077	57 962	55 157	845	4 615	49 698	50 011	749	3 791	9 120	36 352	4 455	148	1 377	2 930
Червень	198 650	137 291	25 429	51 063	60 800	56 746	841	4 581	51 323	51 658	763	3 779	9 376	37 739	4 613	149	1 401	3 063

Багатофакторна модель (2.67) може бути записана у вигляді

$$Y = XA + E, \tag{2.68}$$

де $y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ – векторний стовпець розмірності;

$A = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix}$ – вектор-стовбчик розмірності $((p + 1) \cdot 1)$ невідомих параметрів рівняння;

X – матриця спостережень розмірності $n \cdot (p + 1)$;

E – вектор – стовпець розмірності $n \cdot 1$ випадкових величин – помилок.

Вектор невідомих параметрів ми знаходимо методом найменших квадратів (МНК), мінімізуючи суму квадратів залишків:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_1 - \alpha_0 - \alpha_1 x_{1i} - \alpha_2 x_{2i} - \ldots - \alpha_p x_{pi})^2. \tag{2.69}$$

Після знаходження часних похідних і, прирівнявши їх до нуля, після відповідних перетворень ми одержимо систему з $(p + 1)$ невідомим:

$$\begin{cases} n\alpha_0 + \alpha_1 \sum_{i=1}^{n} x_{1i} + \alpha_2 \sum_{i=1}^{n} x_{2i} + \ldots + \alpha_p \sum_{i=1}^{n} x_{pi} = \sum_{i=1}^{n} y_1 \\ {}_0 \sum_{i=1}^{n} x_{1i} + \alpha_1 \sum_{i=1}^{n} x_{1i}^2 + \alpha_2 \sum_{i=1}^{n} x_{1i} x_{2i} + \ldots + \alpha_p \sum_{i=1}^{n} x_{1i} x_{pi} = \sum_{i=1}^{n} x_{1i} y_1 \\ \alpha_0 \sum_{i=1}^{n} x_{2i} + \alpha_1 \sum_{i=1}^{n} x_{2i} x_{1i} + \alpha_2 \sum_{i=1}^{n} x_{2i}^2 + \ldots + \alpha_p \sum_{i=1}^{n} x_{2i} x_{pi} = \sum_{i=1}^{n} x_{2i} y_1 \\ \alpha_0 \sum_{i=1}^{n} x_{pi} + \alpha_1 \sum_{i=1}^{n} x_{pi} x_{1i} + \alpha_2 \sum_{i=1}^{n} x_{pi} x_{2i} + \ldots + \alpha_p \sum_{i=1}^{n} x_{pi}^2 = \sum_{i=1}^{n} x_{pi} y_1 \end{cases} \cdot (2.70)$$

У нашому випадку ми маємо:

$$\begin{cases} na_0 + a_1\sum\limits_{i=1}^{n} x_1 + a_2\sum\limits_{i=1}^{n} x_2 + a_3\sum\limits_{i=1}^{n} x_3 = \sum\limits_{i=1}^{n} y \\ a_0\sum\limits_{i=1}^{n} x_1 + a_1\sum\limits_{i=1}^{n} x_1^2 + a_2\sum\limits_{i=1}^{n} x_1x_2 + a_3\sum\limits_{i=1}^{n} x_1x_3 = \sum\limits_{i=1}^{n} x_1y \\ a_0\sum\limits_{i=1}^{n} x_2 + a_1\sum\limits_{i=1}^{n} x_2x_1 + a_2\sum\limits_{i=1}^{n} x_2^2 + a_3\sum\limits_{i=1}^{n} x_2x_3 = \sum\limits_{i=1}^{n} x_2y \\ a_0\sum\limits_{i=1}^{n} x_3 + a_1\sum\limits_{i=1}^{n} x_3x_1 + a_2\sum\limits_{i=1}^{n} x_3x_2 + a_3\sum\limits_{i=1}^{n} x_3^2 = \sum\limits_{i=1}^{n} x_3y \end{cases}. \tag{2.71}$$

У матричному вигляді (т.к.р. = 3, $n = 9$):

$$\begin{pmatrix} n & \sum x_1 & \sum x_2 & \sum x_3 \\ \sum x_1 & \sum x_1^2 & \sum x_2x_1 & \sum x_3x_1 \\ \sum x_2 & \sum x_2x_1 & \sum x_2^2 & \sum x_2x_3 \\ \sum x_3 & \sum x_3x_1 & \sum x_3x_2 & \sum x_3^2 \end{pmatrix} \cdot \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & \dots & x_{19} \\ x_{21} & x_{22} & \dots & x_{29} \\ x_{31} & x_{32} & \dots & x_{39} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}. \tag{2.72}$$

У скороченому вигляді можна записати:

$$X'XA = X'Y. \tag{2.73}$$

Звідси одержуємо рівняння для знаходження невідомих параметрів у матричному вигляді:

$$A = \left(X'X\right)^{-1} X'Y. \tag{2.74}$$

де $\left(X'X\right) = \begin{pmatrix} n & \sum x_1 & \sum x_2 & \sum x_3 \\ \sum x_1 & \sum x_1^2 & \sum x_1x_2 & \sum x_1x_3 \\ \sum x_2 & \sum x_2x_1 & \sum x_2^2 & \sum x_2x_3 \\ \sum x_3 & \sum x_3x_1 & \sum x_{32} & \sum x_3^2 \end{pmatrix};$

$$X'Y = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & \dots & x_{19} \\ \dots & \dots & \dots & \dots \\ x_{31} & x_{32} & \dots & x_{39} \end{pmatrix} \cdot \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} \sum y \\ \sum x_1y \\ \sum x_2y \\ \sum x_3y \end{pmatrix}.$$

З табл. 2.6 одержуємо, що

$$X'X = \begin{pmatrix} 9,00 & 6221,74 & 41377,49 & 398845,72 \\ 6221,74 & 4467803,45 & 28709567,58 & 281992011,20 \\ 41377,49 & 28709567,58 & 190351154,94 & 1837782795,03 \\ 398845,72 & 281992011,20 & 1837782795,03 & 17941575238,10 \end{pmatrix};$$

$$X'Y = \begin{pmatrix} 449056,07 \\ 316489875,21 \\ 2068600037,56 \\ 20157138073,89 \end{pmatrix}.$$

Значення для матриць беремо з табл. 2.6 з рядка «Сума» і стовпців, а також стовпці YX_1, YX_2, YX_3 в, враховуючи, що $n = 9$.

Потім знаходимо зворотну матрицю $(X'X)^{-1}$ і множимо її на матрицю $(X'Y)$.

У результаті одержуємо вектор-стовпець зі знайденими параметрами регресійної моделі:

$$A = \begin{pmatrix} -1856,5591 \\ -0,3949 \\ 2,2695 \\ 0,9385 \end{pmatrix} = \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}.$$

Тепер можемо записати лінійну модель множинної регресії:

$$Y = -1856,5591 - 0,3949 \cdot x_1 + 2,2695 \cdot x_2 + 0,9385 \cdot x_3$$

2. Коефіцієнт детермінації R^2 – так називають квадрат R – коефіцієнта множинної кореляції. Його шукаємо за формулою

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - y_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2} = 1 - \frac{\sum_{i=1}^{n}\varepsilon_i^2}{\sum_{i=1}^{n}(y_i - \overline{y}\ )^2}; \qquad (2.75)$$

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

Використовуючи з табл. 2.6 рядок «Сума» стовпців e_i^2, i, $\left(y_i-\overline{y}\right)^{\wedge 2}$, знаходимо $R^2=0{,}99>0{,}8$, отже, рівняння регресії достовірне. Коефіцієнт детермінації показує, що 99,95% варіації залежної ознаки (Y) пояснюється включеними в модель факторами.

$R=0{,}99$ – чим ближче до 1 коефіцієнт множинної кореляції, тим більш міцний зв'язок між Y і безліччю X.

Для подальших розрахунків скористаємося табл. 2.6:

Таблиця 2.6. Проміжні розрахунки

№	y	x^1	x^2	x^3	yx^1	yx^2	yx^3	y^2
1	41 077	547	4409	35 352	22463982,6	181111030,7	1452146320	1687332006
2	43 586	538	4501	37 821	23467103,3	196178843,3	1648465111	1899747852
3	46 626	534	4568	40 826	24915046,8	212983288,5	1903517772	2173955527
4	47 445	573	4526	41 889	27195300,9	214756622,6	1987428130	2250999368
5	50 103	685	4618	44 221	34342781,6	231370327,6	2215611219	2510300488
6	53 023	825	4785	47 173	43722736	253701052,6	2501248282	2811393778
7	55 294	833	4775	49 894	46055389,5	264014155,4	2758861753	3057429311
8	55 157	845	4615	50 011	46590182,9	254544372,9	2758493249	3042338223
9	56 746	841	4581	51 658	47737351,6	259940343,9	2931366238	3220055742
Сума	**449 056**	**6 222**	**41 377**	**398 846**	**316 489 875**	**2 068 600 038**	**20 157 138 074**	**22 653 552 296**

№	x_1^2	x_2^2	x_3^2	x^1x^2	x^1x^3	x^2x^3
1	299070,1	19439686,63	1249741562	2411188,21	19332881,48	155867200,9
2	289883,2	20258550,9	1430419947	2423347,49	20363071,36	170229948,3
3	285543,8	20866057,57	1666722185	2440937,05	21815641,47	186488393,9
4	328558,2	20488858,24	1754718649	2594568,02	24010982,3	189610605,3
5	469834,8	21325028,12	1955516121	3165318,52	30311212,77	204209295,3
6	679975,1	22894062,23	2225317214	3945553,35	38899359,83	225713426,2
7	693752,4	22798065,67	2489450253	3976960,23	41557936,36	238232345,3
8	713479,2	21297052,81	2501130528	3898077,08	42243398,23	230795816,5
9	707706,6	20983792,77	2668558779	3853617,63	43457527,39	236635763,2
Сума	**4 467 803**	**190 351 155**	**17 941 575 238**	**28 709 568**	**281 992 011**	**1 837 782 795**

Продовження табл. 2.6

	$\hat{y}$	$y-\hat{y}$	$(y_i-\hat{y})(y_{i-1}-\hat{y}_{i-1})$	$(y-\hat{y})^2$
1	41111,32	–34	–	1167,917
2	43640,58	–54	1862,039	2968,696
3	46614,03	12	–635,821	136,177
4	47502,98	–58	–680,085	3396,432
5	49854,61	248	–14470	61647,28
6	52948,85	74	18304,92	5435,276
7	55476,54	–183	–13455,6	33310,65
8	55218,81	–61	11208,92	3771,764
9	56688,35	57	–3511,92	3269,975
Сума	**449 056**	**–0**	**–1 378**	**115 104**

3. Проведемо відсів неістотних факторів.
   Для цього розрахуємо коефіцієнти парної кореляції за формулами:

$$r_{yx_i}=\frac{\sum yx_i-\dfrac{\sum y\cdot\sum x_i}{n}}{\sqrt{\left(\sum y^2-\dfrac{(\sum y)^2}{n}\right)\left(\sum x_i^2-\dfrac{(\sum x_i)^2}{n}\right)}}. \tag{2.76}$$

Дані беремо з табл. 2.6 з рядка «Сума» відповідних стовпців:

$$r_{yx_1}=\frac{316\,489\,875-\dfrac{449\,056\cdot 6\,222}{9}}{\sqrt{\left(22\,653\,552\,296-\dfrac{449\,056^2}{9}\right)\left(4\,467\,803-\dfrac{6\,222^2}{9}\right)}}=0{,}9421.$$

Аналогічно:
$r_{yx_2}=0{,}751;$
$r_{yx_3}=0{,}9992.$

Набуті значення коефіцієнтів парної кореляції більше за модулем 0,3, отже, усі три фактори слід включити в модель, що розробляється.

4. Перевірка показників на мультиколінеарність.

Під ***мультиколінеарністю*** розуміють наявність висококореляційного зв'язку між двома факторами.

Для цього визначаємо парні коефіцієнти кореляції за формулами:

$$r_{x_1x_2}=\frac{\sum x_1x_2-\frac{\sum x_1\cdot\sum x_2}{n}}{\sqrt{\left(\sum x_1^2-\frac{(\sum x_1)^2}{n}\right)\left(\sum x_2^2-\frac{(\sum x_2)^2}{n}\right)}}; \tag{2.77}$$

$$r_{x_1x_3}=\frac{\sum x_1x_3-\frac{\sum x_1\cdot\sum x_3}{n}}{\sqrt{\left(\sum x_1^2-\frac{(\sum x_1)^2}{n}\right)\left(\sum x_3^2-\frac{(\sum x_3)^2}{n}\right)}}; \tag{2.78}$$

$$r_{x_2x_3}=\frac{\sum x_2x_3-\frac{\sum x_2\cdot\sum x_3}{n}}{\sqrt{\left(\sum x_2^2-\frac{(\sum x_2)^2}{n}\right)\left(\sum x_3^2-\frac{(\sum x_3)^2}{n}\right)}}. \tag{2.79}$$

Маємо

$$r_{x_1x_2}=\frac{28\,709\,568-\frac{6\,222\cdot 41\,377}{9}}{\sqrt{\left(4\,467\,803-\frac{6\,222^2}{9}\right)\left(190\,351\,155-\frac{41\,377^2}{9}\right)}}=0{,}748758.$$

Аналогічно:

$r_{x_1x_3}=0{,}940895$;

$r_{x_2x_3}=0{,}72899906$.

Теоретично вважається, що мультиколінеарність відсутня, якщо коефіцієнти парної кореляції менше ніж 0,85 за модулем.

У нас коефіцієнт $r_{x_1x_3}=0{,}940895$ більше 0,85, тому робимо висновок про наявність мультиколінеарності.

Оцінка будь-якої регресії страждатиме від мультиколінеарності певною мірою, якщо тільки всі незалежні змінні не виявляться абсолютно некорельованими.

У нашому випадку на наявність мультиколінеарності може впливати недостатня кількість спостережень. Можливо, збільшення кількості спостережень зменшить проблему мультиколінеарності.

5. Перевірка надійності впливу окремих факторів на результат.

Ця перевірка здійснюється за допомогою коефіцієнтів надійності:

$$m_{yx_i} = \frac{r_{yx_i} \cdot \sqrt{n}}{1-\left(r_{yx_i}\right)^2}, \qquad (2.80)$$

$$m_{yx_1} = \frac{0{,}9421 \cdot \sqrt{9}}{1-(0{,}9421)^2} = 25{,}13;$$

$$m_{yx_2} = \frac{0{,}751 \cdot \sqrt{9}}{1-(0{,}751)^2} = 5{,}17;$$

$$m_{yx_3} = \frac{0{,}9992 \cdot \sqrt{9}}{1-(0{,}9992)^2} = 1916{,}66.$$

Оскільки всі коефіцієнти надійності за модулем більше ніж 2,6, то всі дані фактори дійсно впливають на обсяг споживчих кредитів.

6. Розрахунок коефіцієнта кореляції.

Автокореляція – це кореляційна залежність між послідовними значеннями рівнів однієї і тієї самої ознаки.

Щоб виявити наявність автокореляції за часом, у помилках використовують таку ідею: якщо кореляція є в помилок E, то вона присутня і в залишках, одержуваних після застосування методу МНК.

Лінійний коефіцієнт автокореляції першого порядку розраховується за формулою:

$$r_1 = \frac{\sum \varepsilon_i \cdot \varepsilon_{i-1}}{\sqrt{\sum \varepsilon_i^2 \cdot \sum \varepsilon_{i-1}^2}}, \qquad (2.81)$$

де $\varepsilon_i = y_t - \hat{y}$ – залишки;

$\hat{y}$ – розрахункові значення y за рівнянням регресії;
ε_{i-1} – залишки, зсунуті на один крок.

Тоді (за даними табл. 2.6):

$$r_1 = \frac{-1378}{\sqrt{115\,104 \cdot 113\,936}} = -0{,}9998.$$

Якщо коефіцієнт автокореляції першого порядку менше ніж 0,5, то автокореляція відсутня, а якщо більше ніж 0,5, то автокореляція наявна.

У нашому випадку $|r_1| > 0{,}5$, тому ми можемо стверджувати, що автокореляція наявна.

Перевіряючи гіпотезу про існування лінійної автокореляції першого порядку, виконують перевірочну процедуру, засновану на обчисленні d-статистики Дарбіна – Уотсона:

$$d = DW = \frac{\sum(\varepsilon_i - \varepsilon_{i-1})^2}{\sum \varepsilon_i^2}, \tag{2.82}$$

де d – взвішена сума квадратів різниць послідовних залишків.

Маємо

$$d = \frac{228\,525}{115\,104} = 1{,}99.$$

За таблицею значень d-статистики Дарбіна – Уотсона рівень значущості $\alpha = 0{,}05\%$, кількості спостережень $n = 9$, кількості незалежних змінних у рівнянні регресії $R = 3$ знаходимо два критичні значення: нижнє $d_L = 0{,}455$ (межа для визнання позитивної автокореляції залишків) і верхнє $d_U = 2{,}128$ (межа визнання її відсутності).

Тоді
$4 - d_U = 1{,}872$;
$4 - d_L = 3{,}545$.

Обчислене значення $d = 1{,}98538$ потрапляє в інтервал $4 - d_U < d < 4 - d_L$.

За табл. 4 із завдання 1 одержуємо, що d потрапляє в область невизначеності (у разі передбачуваної негативної кореляції), тому гіпотеза про відсутність автокореляції не приймається і не відкидається.

7. Визначимо часткові коефіцієнти множинної кореляції для зміни тісноти зв'язку між y і факторами x_1, x_2, x_3 за формулами:

$$R_{y(x_1;x_2)} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2}}; \tag{2.83}$$

$$R_{y(x_1;x_3)} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_3}^2 - 2r_{yx_1} \cdot r_{yx_3} \cdot r_{x_1x_3}}{1 - r_{x_1x_3}^2}}; \tag{2.84}$$

$$R_{y(x_2;x_3)} = \sqrt{\frac{r_{yx_2}^2 + r_{yx_3}^2 - 2r_{yx_2} \cdot r_{yx_3} \cdot r_{x_2x_3}}{1 - r_{x_1x_3}^2}}. \tag{2.85}$$

Маємо

$$R_{y(x_1;x_2)} = \sqrt{\frac{0{,}9421^2 + 0{,}751^3 - 2 \cdot 0{,}9421 \cdot 0{,}751 \cdot 0{,}748758}{1-(0{,}748758)^2}} = 0{,}94.$$

Аналогічно:

$R_{y(x_1;\, x_3)} = 0{,}99;$

$R_{y(x_2;\, x_3)} = 0{,}99.$

8. Перевірка достовірності одержаної моделі здійснюється:
   1) за допомогою розрахунку теоретичних значень за одержаним рівнянням $\hat{y}$.

   Якщо $\left.\begin{array}{l}\sum \hat{y} = 449\,056 \\ \sum y = 449\,056\end{array}\right\}$,

то рівняння достовірне;

2) за допомогою розрахунку коефіцієнта детермінації.

$R^2 = 0{,}9995 > 0{,}8$, отже, рівняння достовірне і може бути використане для прогнозу.

Знайдемо скоректований коефіцієнт R^2 за формулою:

$$\overline{R}^2 = 1 - \left(1 - R^2\right) \cdot \frac{n-1}{n-m-1}, \tag{2.86}$$

де n – кількість спостережень;
m – кількість незалежних змінних.

Маємо

$$\overline{R}^2 = 1 - (1 - 0{,}9995) \cdot \frac{9-1}{9-3-1} = 0{,}9993.$$

За невеликої кількості спостережень величина, як правило, завищується.

9. Обчислимо стандартну помилку регресії за формулою

$$S = \sqrt{\frac{\sum \varepsilon_i^2}{n-2}}. \qquad (2.87)$$

$$S = \sqrt{\frac{115104}{9-2}} = 151{,}73.$$

Знайдемо відношення стандартної помилки до середнього значення залежної змінної:

$$V = \frac{S}{\overline{y}} = \frac{151{,}73}{49\,895} = 0{,}00304, \text{ або } 0{,}304\%.$$

Це відношення служить критерієм прогнозних якостей оціненої регресійної моделі.

Алгоритм використання інструменту КОРРЕЛЯЦИЯ

Інструмент аналізу **Корреляция** (який також поставляється разом з надбудовою **Пакет анализа**) використовується для оцінки ступеня залежності між двома наборами даних. Щоб застосувати інструмент **Корреляция**, виконайте низку дій.

4.1. Виберіть команду **Сервис – Анализ данных.**

4.2. У діалоговому вікні, що відкрилося, **Анализ данных** **в** списку **Инструменты анализа** виберіть пункт **Корреляция** і клацніть на кнопці «OK».

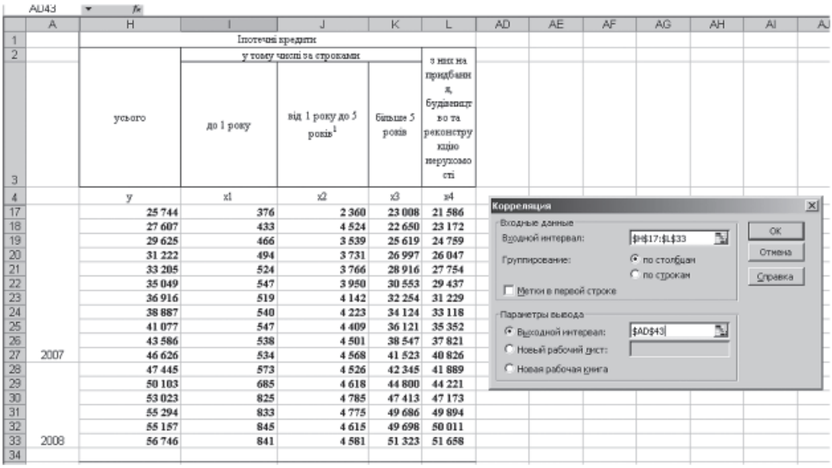
Excel відобразить на екрані діалогове вікно **Корреляция.**

4.3. Визначте значення X і У, які повинні бути проаналізовані.

У полі **Входной интервал** вкажіть посилання на комірки. Якщо цей діапазон містить підписи даних, встановіть прапорець **Метки в** першому рядку. Перевірте, чи правильно вибраний перемикач **Группирование,** що визначає спосіб організації даних у виділеному діапазоні комірок.

4.4. Вкажіть місце, куди повинні бути поміщені результати обчислень, що здійснюються.

Використовуйте перемикачі групи **Параметри выводов**, щоб вказати Excel, де повинен бути розміщений звіт про одержані результати. Наприклад, щоб помістити цей звіт на тому самому листі, де розташовані початкові дані, виберіть перемикач **Выходной интервал** і в полі праворуч від нього вкажіть адресу комірок, які повинні містити обчислені значення. Щоб помістити звіт в якомусь іншому місці, виберіть один із двох інших перемикачів.



	A	H	I	J	K	L
1		Іпотечні кредити				
2			у тому числі за строками			з них на придбання, будівництво та реконструкцію нерухомості
3		усього	до 1 року	від 1 року до 5 років[1]	більше 5 років	
4		y	x1	x2	x3	x4
17		25 744	376	2 360	23 008	21 586
18		27 607	433	4 524	22 650	23 172
19		29 625	466	3 539	25 619	24 759
20		31 222	494	3 731	26 997	26 047
21		33 205	524	3 766	28 916	27 754
22		35 049	547	3 950	30 553	29 437
23		36 916	519	4 142	32 254	31 229
24		38 887	540	4 223	34 124	33 118
25		41 077	547	4 409	36 121	35 352
26		43 586	538	4 501	38 547	37 821
27	2007	46 626	534	4 568	41 523	40 826
28		47 445	573	4 526	42 345	41 889
29		50 103	685	4 618	44 800	44 221
30		53 023	825	4 785	47 413	47 173
31		55 294	833	4 775	49 686	49 894
32		55 157	845	4 615	49 698	50 011
33	2008	56 746	841	4 581	51 323	51 658
34						

4.5. Клацніть на кнопці «ОК».

Excel обчислить коефіцієнт кореляції для вказаних вами даних і помістить його в задане місце. Значення цього коефіцієнта, наприклад, дорівнює числу 0,91802024, що означає, що майже на 92% обсяги депозитів залежать від кількості виходів реклами.

	Столбец 1	*Столбец 2*	*Столбец 3*	*Столбец 4*	*Столбец 5*
Столбец 1	1				
Столбец 2	0,91802024	1			
Столбец 3	0,76095654	0,66251471	1		
Столбец 4	0,99919175	0,91708731	0,73456426	1	
Столбец 5	0,99906638	0,9238193	0,74934722	0,99883724	1

Побудова моделі множинної регресії

Вхідні дані для побудови моделі множинної регресії лінійного вигляду та у вигляді функції Кобба – Дугласа наведені нижче.

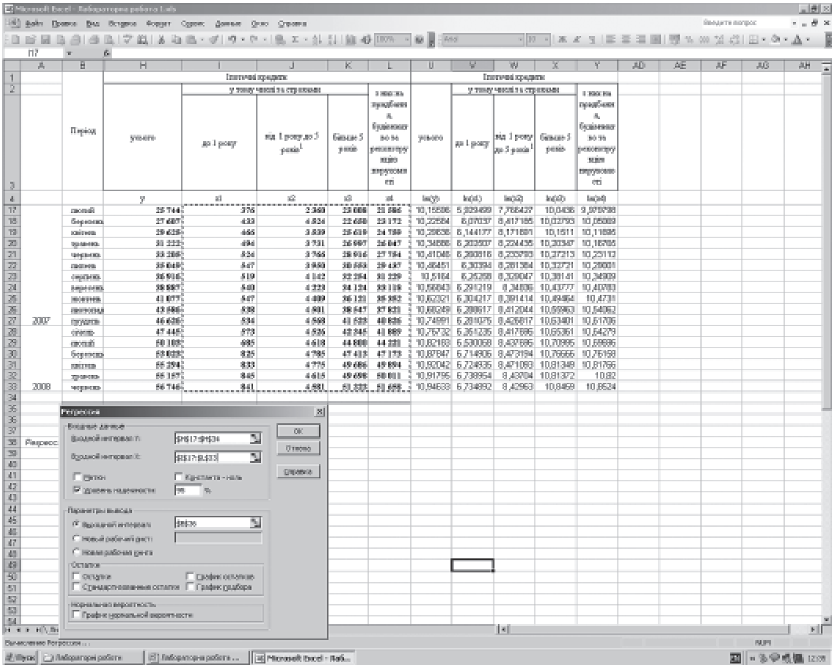
A	B	H	I	J	K	L	O	V	W	X	Y
	Період	Іпотечні кредити					Іпотечні кредити				
			у тому числі за строками			з них на придбання, будівництво та реконструкцію нерухомості		у тому числі за строками			з них на придбання, будівництво та реконструкцію нерухомості
		усього	до 1 року	від 1 року до 5 років[1]	більше 5 років		усього	до 1 року	від 1 року до 5 років[1]	більше 5 років	
		y	x1	x2	x3	x4	ln(y)	ln(x1)	ln(x2)	ln(x3)	ln(x4)
	лютий	25 744	376	2 360	23 008	21 586	10,15596	5,929499	7,766427	10,0436	9,979798
	березень	27 607	433	4 524	22 650	23 172	10,22584	6,07037	8,417186	10,02793	10,05069
	квітень	29 625	466	3 539	25 619	24 759	10,29636	6,144177	8,171691	10,1511	10,11696
	травень	31 222	494	3 731	26 997	26 047	10,34886	6,202507	8,224436	10,20347	10,16766
	червень	33 205	524	3 766	28 916	27 754	10,41046	6,260816	8,233793	10,27213	10,23112
	липень	35 049	547	3 950	30 553	29 437	10,46451	6,30394	8,281384	10,32721	10,29001
	серпень	36 916	519	4 142	32 254	31 229	10,5164	6,25258	8,329047	10,38141	10,34909
	вересень	38 887	540	4 223	34 124	33 118	10,56843	6,291219	8,34836	10,43777	10,40763
	жовтень	41 077	547	4 409	36 121	35 352	10,62321	6,304217	8,391414	10,49464	10,4731
	листопад	43 586	538	4 501	38 547	37 821	10,68249	6,288617	8,412044	10,55963	10,54062
2007	грудень	46 626	534	4 568	41 523	40 826	10,74991	6,281075	8,426817	10,63401	10,61706
	січень	47 445	573	4 526	42 345	41 889	10,76732	6,351236	8,417696	10,65361	10,64279
	лютий	50 103	685	4 618	44 800	44 221	10,82183	6,530068	8,437696	10,70995	10,69696
	березень	53 023	825	4 785	47 413	47 173	10,87847	6,714906	8,473194	10,76666	10,76158
	квітень	55 294	833	4 775	49 686	49 894	10,92042	6,724935	8,471093	10,81349	10,81766
	травень	55 157	845	4 615	49 698	50 011	10,91795	6,738954	8,43704	10,81372	10,82
2008	червень	56 746	841	4 581	51 323	51 658	10,94633	6,734892	8,42963	10,8459	10,8524

5.1. Вибираємо команду **Сервис – Анализ данных**.

5.2. У діалоговому вікні, що відкрилося, **Анализ данных** у списку **Инструменты анализа** вибираємо пункт **Регрессия** і клацаємо на кнопці «ОК». На екрані відображається діалогове вікно **Регрессия**.

5.3. Визначте значення *X* і *У*.

У полі **Входной интервал *У*** вкажіть посилання на діапазон комірок, у яких міститься набір залежних значень (обсяг продажів). Потім перейдіть до поля **Входной интервал *X***. Переконайтеся, наприклад, що дані про розміри торгової площі і чисельності персоналу розташовуються в сусідніх колонках. Позначте блок, що складається з цих обох колонок, як значення *X*. В цьому полягає єдина відмінність введення даних для лінійного і множинного регресійного аналізу. Потім аналіз здійснюється аналогічно попередньому, але, коли з'являються його результати, ми бачимо два коефіцієнти при *X* (див. рис. на с. 106).



Отримаємо такі результати:

– лінійна модель:

$$y = -0{,}00056 + 0{,}99999x_1 + x_2 + 0{,}99999x_3 + 1{,}7 \cdot 10^7 x_4$$

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	1
R-квадрат	1
Нормированный R-к	1
Стандартная ошибк	0,000641771
Наблюдения	17

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	4	1,71E+09	427034006,4	1,03682E+15	4,10784E-87
Остаток	12	4,94E-06	4,1187E-07		
Итого	16	1,71E+09			

	Коэффициенты	*Стандартная ош*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*Верхние 95,0%*
Y-пересечение	-0,000564043	0,00193	-0,292263766	0,775076503	-0,00476896	0,00364087	-0,004768956	0,00364087
Переменная X 1	0,999995474	3,04E-06	329094,1727	4,17405E-61	0,999988853	1,000002095	0,999988853	1,000002095
Переменная X 2	1,000000285	4,62E-07	2163996,453	6,38735E-71	0,999999278	1,000001292	0,999999278	1,000001292
Переменная X 3	0,999999883	4,14E-07	2413421,219	1,72505E-71	0,99999898	1,000000785	0,99999898	1,000000785
Переменная X 4	1,71214E-07	4,29E-07	0,398635525	0,697163904	-7,6458E-07	1,10701E-06	-7,64585E-07	1,10701E-06

— модель у вигляді функції Кобба – Дугласа:

$$y = e^{0,589} x_1^{\ 0,014} x_2^{\ 0,097} x_3^{\ 0,666} x_4^{\ 0,206}$$

ВЫВОД ИТОГОВ

Регрессионная статистика	
Множественный R	0,999997
R-квадрат	0,999994
Нормированный R-квадрат	0,999991
Стандартная ошибка	0,000755
Наблюдения	17

Дисперсионный анализ

	df	*SS*	*MS*	*F*	*Значимость F*
Регрессия	4	1,06244345	0,265610863	465886,7359	4,99027E-31
Остаток	12	6,84143E-06	5,70119E-07		
Итого	16	1,062450291			

	Коэффициент	*Стандартная ошибка*	*t-статистика*	*P-Значение*	*Нижние 95%*	*Верхние 95%*	*Нижние 95,0%*	*рхние 95,0%*
Y-пересечение	0,569353	0,015491	36,75378012	1,0553E-13	0,535600818	0,603104797	0,535600818	0,603105
Переменная X 1	0,014276	0,002037997	7,005003518	1,42434E-05	0,00983576	0,018716586	0,00983576	0,018717
Переменная X 2	0,097637	0,00221116	44,1566773	1,18467E-14	0,092819795	0,102455204	0,092819795	0,102455
Переменная X 3	0,665531	0,015065878	44,17474358	1,1769E-14	0,632705554	0,698357009	0,632705554	0,698357
Переменная X 4	0,206308	0,015569578	13,25071681	1,59044E-08	0,172384875	0,240231268	0,172384875	0,240231

6. Для оцінки суттєвості лінійного коефіцієнта кореляції використаємо критерій Стьюдента. Порівняємо розраховане значення критерію Стьюдента з його критичним значенням. Критичне значення знаходимо за таблицями розподілу Стьюдента для заданого рівня ймовірності та кількості ступенів свободи $v = n - 2$, де n – кількість спостережень.

За умови, якщо $t_{факт} > t_{кр}$ лінійний коефіцієнт кореляції можна вважати статистично значущим. Для заданих умов $t_{крит}(0,05; 8) = 2,307$.

Оцінка суттєвості рівняння регресії здійснюється за критерієм Фішера. Якщо розрахункове значення критерію Фішера перевищує його критичне значення, модель можна вважати адекватною: $F_{крит}(0,05; 1; 8) = 5,32$.

Оскільки коефіцієнт детермінації для обох рівнянь суттєво не відрізняється і приблизно дорівнює 1, F-критерій Фішера показує адекватність обох рівнянь, а t-критерій Стьюдента підтверджує значущість лише коефіцієнтів при x_2, x_3, x_4 для лінійної моделі і значущість усіх коефіцієнтів для моделі у вигляді функції Кобба – Дугласа, тому більш прийнятною вважається модель у вигляді функції Кобба – Дугласа.

2.10. Прогнозування за лінійною моделлю

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

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

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

Значення y_{pteor} для майбутнього періоду чи додаткового елемента обчислюють за формулою (2.5) за відомим вектором оцінених параметрів $\alpha = (\alpha_0, \alpha_1, \alpha_2,.., \alpha_m)$ і за вектором значень незалежних змінних $x_p = (1, x_{1p}, x_{2p},.., x_{mp},)$, що не належать до базового періоду. Розрізняють прогноз середній (оцінку математичного сподівання регресанда) та індивідуальний (оцінку певної реалізації регресанда y_{pteor}, що відповідає моменту p). Перша з них базується на передумові МНК про нульове математичне сподівання випадкової складової рівняння регресії, а друга застосовує оцінене значення $\hat{u}_p$. Оцінену дисперсію прогнозу розраховують за формулами

$$\sigma_e^2 = \sigma_u^2 x_p^T (X^T X)^{-1} x_p; \tag{2.88}$$

$$\sigma_{e(i)}^2 = \sigma_u^2 + \sigma_u^2 x_p^T (X^T X)^{-1} x_p. \tag{2.89}$$

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

$$(y_{pteor} - t_{a/2}\sqrt{\sigma_e^2};\ y_{pteor} + t_{a/2}\sqrt{\sigma_e^2}); \tag{2.90}$$

$$(y_{pteor} - t_{a/2}\sqrt{\sigma_{e(i)}^2};\ y_{pteor} + t_{a/2}\sqrt{\sigma_{e(i)}^2}), \tag{2.91}$$

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

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

2.11. Методи побудови багатофакторної регресійної моделі

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

Для вибору компромісного рішення не існує єдиної процедури.

Тому для побудови «найкращого» рівняння застосовують один із таких методів.

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

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

3. *Покроковий регресійний метод* діє у зворотному порядку порівняно з попереднім методом, тобто до моделі послідовно включаються фактори, що мають найбільший коефіцієнт кореляції із залежною змінною. Модель аналізується за значеннями коефіцієнта детермінації та частковими F-критеріями. Фактори, що не задовольняють критерії, з моделі вилучаються. Процес припиняється, якщо жоден з факторів рівняння вилучити не вдається, а новий претендент на включення не відповідає частковому F-критерію. На практиці цей метод найпоширеніший.

Завдання для самоконтролю

1. Схарактеризуйте стисло алгоритм покрокової регресії.
2. Чим відрізняються коефіцієнти парної та часткової кореляції?
3. Запишіть співвідношення між коефіцієнтами кореляції і детермінації.
4. Як визначаються дисперсія залишків, загальна дисперсія і дисперсія регресії?
5. Який між ними зв'язок?
6. Критерій Фішера. Перевірка лінійної регресійної моделі на адекватність.
7. Покажіть залежність між F-критерієм і R^2.
8. Як оцінити вірогідність коефіцієнта кореляції?
9. Як обчислюється t-критерій?
10. Теоретична та емпірична лінійна множинна модель та їх запис у векторно-матричній формі.
11. Що таке стандартна помилка оцінок параметрів моделі?
12. Описати алгоритм побудови довірчих інтервалів із заданою надійністю γ для параметрів β_i теоретичної множинної лінійної регресії. Навести відповідні формули.
13. Перевірка статистичної значущості параметрів β_i та перевірка загальної якості множинної регресії. Навести відповідні формули.

Практичне заняття 2

Побудова та дослідження моделі множинної лінійної регресії з використанням засобів MS Excel

Постановка завдання

1 етап	Побудова інформаційного масиву дослідження шляхом збору статистичних даних та розробка моделі парної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), анаіз моделі
2 етап	**Побудова множинної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), дослідження моделі**
3 етап	Виявлення, дослідження та усунення мультиколінеарності факторів у моделі множинної лінійної регресії методом Фаррара – Глобера
4 етап	Виявлення ефекту гетероскедастичності на основі застосування параметричного тесту Гольдфельда – Квандта та усунення даного ефекту шляхом використання узагальненого методу найменших квадратів (методу Ейткена)
5 етап	Побудова адитивної та мультиплікативної моделі часового ряду на основі проведення декомпозиційного аналізу: визначення структури ряду (тенденції та сезонних (циклічних) коливань), проведення авто кореляційного аналізу, визначення якіості моделі, розрахунок прогнозних значень
6 етап	Групування елементів економічної системи шляхом кластерного аналізу, побудова деревоподібної дендограми, графічне подання елементів кластеризації
7 етап	Перевірка наявності автокореляції залишків за допомогою критерію Дарбіна – Уотсона, оцінка параметрів рівняння множинної лінійної регресії методом Ейткена
8 етап	Побудова економіко-математичної моделі методом інструментальних змінних; побудова економетричної моделі на основі даних, що містять помилки вимірювання змінних, оцінка параметрів моделі за допомогою алгоритма Уоліса
9 етап	Побудова моделі розподіленого лагу, оцінка параметрів моделей з лагами в незалежних змінних, оцінювання параметрів авторегресійних моделей; перетворення ARMA і ARIMA
10 етап	Структурне моделювання (моделювання на основі системи структурних рівнянь), оцінка невідомих параметрів моделі, визначення адекватності моделі щодо опису структури вхідних даних

Мета роботи: набуття навичок побудови множинної лінійної регресії, оцінки її параметрів методом найменших квадратів (1 МНК). Дослідження моделі.

Початкові дані

За даними 17 рівнів часових рядів вивчається залежність величини активів банків Y (млрд грн) від власного капіталу X_1 млрд грн) і величини зобов’язань X_2 (млрд грн).

Необхідно:

1. Сформувати масив вхідних статистичних даних шляхом корегування даних розглянутого прикладу на номер варіанта.
2. Побудувати лінійну модель множинної регресії. Записати стандартизоване рівняння множинної регресії. На основі стандартизованих коефіцієнтів регресії і середніх коефіцієнтів еластичності ранжувати фактори за ступенем їх впливу на результат.
3. Знайти коефіцієнти парної, частинної та множинної кореляції. Проаналізувати їх.
4. Знайти скоригований коефіцієнт множинної детермінації. Порівняти його з нескоригованим (загальним) коефіцієнтом детермінації.
5. За допомогою F-критерію Фішера оцінити статистичну значущість рівняння регресії та коефіцієнта детермінації $R^2_{yx_1x_2}$.
6. За допомогою часткових F-критеріїв Фішера оцінити доцільність включення в рівняння множинної регресії фактора x_1 після фактора x_2 і фактора після фактора .
7. Скласти рівняння парної лінійної регресії, обравши тільки один значущий фактор.

Приклад виконання

За даними 17 рівнів часових рядів вивчається залежність величини активів банків Y (млрд грн) від власного капіталу X_1 млрд грн) і величини зобов'язань X_2 (млрд грн).

Таблиця 2.9. Вхідні дані для побудови моделі

Дата	Усього активів Y	ВК, X_1	Усього зобов'язань, X_2
01.01.2009	119263048	806823449	926086498
01.04.2009	117081585	753551950	870633535
01.07.2009	112597492	752097475	864694968
01.10.2009	117968018	771990516	889958533
01.01.2010	120207619	753241949	873449574
01.04.2010	126646323	748243365	874964709
01.07.2010	127162304	758093409	885255711
01.10.2010	132802031	784695450	917497465
01.01.2011	137725113	804358877	942083994
01.04.2011	138434527	856598661	995033185
01.07.2011	147816550	871994493	1019811043
01.10.2011	151866259	877296268	1029162518

Продовження табл. 2.9

Дата	Усього активів Y	ВК, X_1	Усього зобов'язань, X_2
01.01.2012	155486926	898785345	1054272287
01.04.2012	162236166,3	920236938,9	1082473105
01.07.2012	163775906,4	940619353,1	1104395259
01.10.2012	165810129,5	951635752,2	1117445882
01.01.2013	170196261,8	956983117,5	1127179379

Хід розв'язання

Таблиця 2.10. Проміжні розрахунки

№ пор.	y	x_1	x_2	yx_1	yx_2	$x_1\,x_2$	x_1^2	x_2^2	y^2
1	7,0	3,9	10,0	27,3	70,0	39,0	15,21	100,0	49,0
2	7,0	3,9	14,0	27,3	98,0	54,6	15,21	196,0	49,0
3	7,0	3,7	15,0	25,9	105,0	55,5	13,69	225,0	49,0
4	7,0	4,0	16,0	28,0	112,0	64,0	16,0	256,0	49,0
5	7,0	3,8	17,0	26,6	119,0	64,6	14,44	289,0	49,0
6	7,0	4,8	19,0	33,6	133,0	91,2	23,04	361,0	49,0
7	8,0	5,4	19,0	43,2	152,0	102,6	29,16	361,0	64,0
8	8,0	4,4	20,0	35,2	160,0	88,0	19,36	400,0	64,0
9	8,0	5,3	20,0	42,4	160,0	106,0	28,09	400,0	64,0
10	10,0	6,8	20,0	68,0	200,0	136,0	46,24	400,0	100,0
11	9,0	6,0	21,0	54,0	189,0	126,0	36,0	441,0	81,0
12	11,0	6,4	22,0	70,4	242,0	140,8	40,96	484,0	121,0
13	9,0	6,8	22,0	61,2	198,0	149,6	46,24	484,0	81,0
14	11,0	7,2	25,0	79,2	275,0	180,0	51,84	625,0	121,0
15	12,0	8,0	28,0	96,0	336,0	224,0	64,0	784,0	144,0
16	12,0	8,2	29,0	98,4	348,0	237,8	67,24	841,0	144,0
17	12,0	8,1	30,0	97,2	360,0	243,0	65,61	900,0	144,0
18	12,0	8,5	31,0	102,0	372,0	263,5	72,25	961,0	144,0
19	14,0	9,6	32,0	134,4	448,0	307,2	92,16	1024,0	196,0
20	14,0	9,0	36,0	126,0	504,0	324,0	81,0	1296,0	196,0
Сумма	192	123,8	446	1276,3	4581	2997,4	837,74	10828,0	1958,0
Сер. знач.	9,6	6,19	22,3	63,815	229,05	149,87	41,887	541,4	97,9

1. Побудувати лінійну модель множинної регресії. Записати стандартизоване рівняння множинної регресії. На основі стандартизованих коефіцієнтів регресії і середніх коефіцієнтів еластичності ранжувати фактори за ступенем їх впливу на результат.

Для знаходження параметрів лінійного рівняння множинної регресії $y = a + b_1x_1 + b_2x_2$ необхідно розв'язати таку систему лінійних рівнянь відносно невідомих параметрів a, b_1, b_2:

$$\begin{cases} na + b_1\sum x_1 + b_2\sum x_2 = \sum y; \\ a\sum x_1 + b_1\sum x_1^2 + b_2\sum x_1x_2 = \sum yx_1; \\ a\sum x_2 + b_1\sum x_1x_2 + b_2\sum x_2^2 = \sum yx_2 \end{cases}$$

або скористатися готовими формулами:

$$b_1 = \frac{\sigma_y}{\sigma_{x_1}} \cdot \frac{r_{yx_1} - r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2}; \quad b_2 = \frac{\sigma_y}{\sigma_{x_2}} \cdot \frac{r_{yx_2} - r_{yx_1}r_{x_1x_2}}{1 - r_{x_1x_2}^2};$$

$$a = \overline{y} - b_1\overline{x}_1 - b_2\overline{x}_2.$$

Оберемо другий шлях. Розрахуємо спочатку середньоквадратичні відхилення ознак:

$$\sigma_y = \sqrt{\overline{y^2} - \overline{y}^2} = \sqrt{97,9 - 9,6^2} = 2,396;$$

$$\sigma_{x_1} = \sqrt{\overline{x_1^2} - \overline{x}_1^2} = \sqrt{41,887 - 6,19^2} = 1,890;$$

$$\sigma_{x_2} = \sqrt{\overline{x_2^2} - \overline{x}_2^2} = \sqrt{541,4 - 22,3^2} = 6,642.$$

Знайдемо парні коефіцієнти кореляції:

$$r_{yx_1} = \frac{\operatorname{cov}(y, x_1)}{\sigma_y \cdot \sigma_{x_1}} = \frac{63,815 - 6,19 \cdot 9,6}{1,890 \cdot 2,396} = 0,970;$$

$$r_{yx_2} = \frac{\operatorname{cov}(y, x_2)}{\sigma_y \cdot \sigma_{x_2}} = \frac{229,05 - 22,3 \cdot 9,6}{6,642 \cdot 2,396} = 0,941;$$

$$r_{x_1x_2} = \frac{\operatorname{cov}(x_1, x_2)}{\sigma_{x_1} \cdot \sigma_{x_2}} = \frac{149,87 - 6,19 \cdot 22,3}{1,890 \cdot 6,642} = 0,943.$$

Матимемо:

$$b_1 = \frac{2,396}{1,890} \cdot \frac{0,970 - 0,941 \cdot 0,943}{1 - 0,943^2} = 0,946;$$

$$b_2 = \frac{2,396}{6,642} \cdot \frac{0,941 - 0,970 \cdot 0,943}{1 - 0,943^2} = 0,0856;$$

$a = 9,6 - 0946 \cdot 6,19 - 0,0856 \cdot 22,3 = 1,835.$

Таким чином, рівняння множинної регресії набуде вигляду:

$$\hat{y} = 1,835 + 0,946 \cdot x_1 + 0,0856 \cdot x_2.$$

Коефіцієнти β_1 і β_2 стандартизованого рівняння регресії $t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \varepsilon$ знаходять за формулами:

$$\beta_1 = b_1 \frac{\sigma_{x_1}}{\sigma_y} = 0,946 \cdot \frac{1,890}{2,396} = 0,746;$$

$$\beta_2 = b_2 \frac{\sigma_{x_2}}{\sigma_y} = 0,0856 \cdot \frac{6,642}{2,396} = 0,237.$$

Тобто рівняння набуде вигляду:

$$\hat{t}_y = 0,746 \cdot t_{x_1} + 0,237 \cdot t_{x_2}.$$

Оскільки стандартизовані коефіцієнти регресії можна порівнювати між собою, можна сказати, що ведення нових основних фондів здійснює більш значний вплив на випуск продукції, ніж питома вага робочих високої кваліфікації.

Порівнювати вплив факторів на результат можна також за допомогою коефіцієнтів еластичності:

$$\overline{E}_t = b_i \cdot \frac{\overline{x}_t}{y_{xt}}.$$

Обчислимо:

$$\overline{E}_1 = 0,946 \cdot \frac{6,19}{9,6} = 0,61; \quad \overline{E}_2 = 0,0856 \cdot \frac{22,3}{9,6} = 0,20.$$

Тобто збільшення тільки основних фондів (від свого середнього значення) або тільки питомої ваги робочих високої кваліфікації на 1%

збільшує в середньому випуск продукції на 0,61% або 0,20% відповідно. Таким чином, підтверджується більший вплив на результат y фактора x_1, ніж фактора x_2.

2. Знайти коефіцієнти парної, частинної та множинної кореляції. Проаналізувати їх.

Коефіцієнти парної кореляції ми вже знайшли:

$r_{yx_1} = 0{,}970;\quad r_{yx_2} = 0{,}941;\quad r_{x_1x_2} = 0{,}943;$

Вони показують досить міцний зв'язок кожного фактора з результатом, а також високу міжфакторну залежність (фактори x_1 і x_2 явно колінеарні, оскільки $r_{x_1x_2} = 0{,}943 > 0{,}7$). За умов такої міцної міжфакторної залежності рекомендується один із факторів виключити з розгляду.

Частинні коефіцієнти кореляції характеризуються щільністю зв'язку між результатом і відповідним фактором при елімінуванні (виключенні впливу) інших факторів, що включені в рівняння регресії.

При двох факторах часткові коефіцієнти кореляції розраховуються в такий спосіб:

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{\left(1 - r_{yx_1}^2\right) \cdot \left(1 - r_{x_1x_2}^2\right)}} = \frac{0{,}941 - 0{,}970 \cdot 0{,}943}{\sqrt{\left(1 - 0{,}970^2\right) \cdot \left(1 - 0{,}943^2\right)}} = 0{,}325.$$

Якщо порівнювати коефіцієнти парної і часткової кореляції, то можна побачити, що висока міжфакторна залежність дає підвищені оцінки міцності зв'язку. Саме з цієї причини рекомендується за наявності сильної колінеарності (взаємозв'язку) факторів виключати з дослідження той фактор, у якого тіснота парної залежності менше, ніж тіснота міжфакторного зв'язку.

Коефіцієнт множинної кореляції визначаємо через матрицю парних коефіцієнтів кореляції:

$$R_{yx_1x_2} = \sqrt{1 - \frac{\Delta_r}{\Delta_{r_{11}}}},$$

$$\Delta_r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{yx_1} & 1 & r_{x_1x_2} \\ r_{yx_2} & r_{x_2x_1} & 1 \end{vmatrix}$$ – визначник матриці парних коефіцієнтів кореляції;

$$\Delta_{r_{11}} = \begin{vmatrix} 1 & r_{x_1x_2} \\ r_{x_2x_1} & 1 \end{vmatrix}$$ — визначник матриці міжфакторної кореляції;

$$\Delta_r = \begin{vmatrix} 1 & 0,970 & 0,941 \\ 0,970 & 1 & 0,943 \\ 0,941 & 0,943 & 1 \end{vmatrix} = 1 + 0,8607 + 0,8607 -$$

$$-0,8855 - 0,8892 - 0,9409 = 0,0058;$$

$$\Delta_{r_{11}} = \begin{vmatrix} 1 & 0,943 \\ 0,943 & 1 \end{vmatrix} = 1 - 0,8892 = 0,1108.$$

Коефіцієнт множинної кореляції:

$$R_{yx_1x_2} = \sqrt{1 - \frac{0,0058}{0,1108}} = 0,973.$$

Аналогічний результат отримаємо при використанні інших формул:

$$R_{yx_1x_2} = \sqrt{1 - \frac{\sigma^2_{\text{ост}}}{\sigma^2_y}} = \sqrt{1 - \frac{0,305}{5,74}} = 0,973;$$

$$R_{yx_1x_2} = \sqrt{\sum \beta_i \cdot r_{yx_i}} = \sqrt{0,746 \cdot 0,970 + 0,237 \cdot 0,941} = 0,973;$$

$$R_{yx_1x_2\ldots x_m} = \sqrt{1 - \left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{yx_2 \cdot x_1}\right)} =$$

$$= \sqrt{1 - \left(1 - 0,970^2\right) \cdot \left(1 - 0,325^2\right)} = 0,973.$$

Коефіцієнт множинної кореляції показує досить міцний зв'язок усього набору факторів з результатом.

3. Нескоригований коефіцієнт множинної детермінації $R^2_{yx_1x_2} = 0,947$ оцінює частку варіації результату за рахунок поданих у рівнянні факторів у загальній варіації результату. У цьому прикладі ця частка становить 94,7% і вказує на вельми високий ступінь обумовленості варіації результату варіацією факторів, інакше кажучи, на досить міцний зв'язок факторів з результатом.

4. Знайти скоригований коефіцієнт множинної детермінації. Порівняти його з нескоригованим (загальним) коефіцієнтом детермінації.

Скоригований коефіцієнт множинної детермінації

$$\hat{R}^2 = 1-(1-R^2)\frac{(n-1)}{n-m-1} = 1-(1-0{,}947)\frac{20-1}{20-2-1} = 0{,}941$$

визначає щільність зв'язку з урахуванням ступенів свободи загальної та залишкової дисперсій. Він дає таку оцінку щільності зв'язку, яка не залежить від кількості факторів і тому може порівнюватися за різними моделями з різною кількістю факторів. У даному прикладі обидва коефіцієнти вказують на досить високу (більше ніж 94%) детермінованість результату y у моделі факторами x_1 и x_1.

5. За допомогою F-критерію Фішера оцінити статистичну значущість рівняння регресії та коефіцієнта детермінації $R^2_{yx_1x_2}$.

Оцінку надійності рівняння регресії в цілому і показника щільності зв'язку $R_{yx_1x_2}$ дає F-критерій Фішера:

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

У нашому випадку експериментальне значення F-критерію Фішера:

$$F_{\text{факт}} = \frac{0{,}973^2}{1-0{,}973^2}\cdot\frac{20-2-1}{2} = 151{,}88.$$

Отримали, що $F_{факт} > F_{табл} = 3{,}49$ (якщо $n = 20$), тобто ймовірність отримати таке значення F-критерію не перевищує припустимий рівень значущості 5%. Відповідно, отримане значення не випадкове, воно сформувалося під впливом суттєвих факторів, тобто підтверджується статистична значущість усього рівняння і показника щільності зв'язку $R^2_{yx_1x_2}$.

6. За допомогою часткових F-критеріїв Фішера оцінити доцільність включення в рівняння множинної регресії фактора x_1 після фактора x_2 і фактора x_2 після фактора x_1.

Часткові F-критерії Фішера розраховуються за формулами:

$$F_{\text{част},\, x_1} = \frac{R^2_{yx_1x_2} - R^2_{yx_2}}{1-R^2_{yx_1}}\cdot\frac{n-m-1}{m};$$

$$F_{\text{част},\, x_2} = \frac{R^2_{yx_1x_2} - R^2_{yx_1}}{1 - R^2_{yx_2}} \cdot \frac{n-m-1}{m}.$$

Знайдемо $R^2_{yx_1}$ і $R^2_{yx_2}$:

$$R^2_{yx_1} = r^2_{yx_1} = 0{,}970^2 = 0{,}941;$$

$$R^2_{yx_2} = r^2_{yx_2} = 0{,}941^2 = 0{,}885.$$

Маємо:

$$F_{\text{част},\, x_1} = \frac{0{,}947 - 0{,}885}{1 - 0{,}941} \cdot \frac{20-2-1}{2} = 8{,}9322;$$

$$F_{\text{част},\, x_2} = \frac{0{,}947 - 0{,}941}{1 - 0{,}885} \cdot \frac{20-2-1}{2} = 0{,}4435.$$

Отримали, що $F_{част,\, x_2} < F_{табл} = 3{,}49$. Отже, включення у модель фактора x_2 після того, як у модель включений фактор x_1, статистично недоцільне: приріст факторної дисперсії за рахунок додаткової ознаки x_2 є незначним, несуттєвим; фактор x_2 включати в рівняння після фактора x_1 не слід.

Якщо змінити початковий порядок включення факторів у модель і розглянути варіант включення x_1 після x_2, то результат розрахунку часткового F-критерію для x_1 буде іншим. $F_{част,\, x_1} > F_{табл} = 3{,}49$, тобто ймовірність його випадкового формування менше прийнятого стандарту $\alpha = 0{,}05$ (5%). Отже, значення часткового F-критерію для додатково включеного фактора x_1 не випадкове, є статистично значущим, надійним, достовірним: приріст факторної дисперсії за рахунок додаткового фактора x_1 є суттєвим. Фактор x_1 має бути в рівнянні, у тому числі у варіанті, коли він додатково включається після фактора x_2.

Загальний висновок: множинна модель з факторами x_1 і x_2 з $R^2_{yx_1x_2} = 0{,}947$ містить неінформативний фактор x_2.

7. Скласти рівняння парної лінійної регресії, обравши тільки один значущий фактор.

Якщо виключити фактор x_2, то можна обмежитися рівнянням парної регресії:

$\hat{y}_x = \alpha_0 + \alpha_1 \cdot x = 1{,}99 + 1{,}23 \cdot x,\ \ r_{yx}{}^2 = 0{,}941.$

Тести

1. Сутність 1 МНК полягає в:
   а) мінімізації суми залишкових величин;
   б) мінімізації дисперсії результативної ознаки;
   в) мінімізації суми квадратів залишкових величин.

2. Наочним видом вибору рівняння парної регресії є:
   а) аналітичний;
   б) графічний;
   в) експериментальний (табличний).

3. Суть коефіцієнта детермінації полягає в такому:
   а) оцінює якість моделі з відносних відхилень за кожним спостереженням;
   б) характеризує частку дисперсії результативної ознаки, що пояснюється регресією, у загальній дисперсії результативної ознаки;
   в) характеризує частку дисперсії, викликану впливом не врахованих у моделі чинників.

4. Якість моделі з відносних відхилень за кожним спостереженням оцінює:
   а) коефіцієнт детермінації;
   б) F-критерій Фішера;
   в) середня помилка апроксимації;
   г) стандартна похибка рівняння.

5. Коефіцієнт при незалежній змінній лінійного парного рівняння регресії:
   а) показує середню зміну результату зі зміною чинника на одну одиницю;
   б) оцінює статистичну значущість рівняння регресії;
   в) показує, на скільки відсотків зміниться в середньому результат, якщо чинник зміниться на 1%.

6. Класичний метод до оцінювання параметрів регресії заснований на:
   а) методі найменших квадратів;
   б) методі максимальної правдоподібності;
   в) кроковому регресійному аналізі.

7. Для оцінки значущості коефіцієнтів регресії розраховують:
   а) F-критерій Фішера;
   б) t-критерій Стьюдента;
   в) коефіцієнт детермінації.

8. Якщо регресія має R2 = 0,8, то регресійна лінія:
   а) пояснює 80% варіації змінної x;
   б) пояснює 80% варіації змінної y;
   в) не пояснює зв'язок між x і y.

9. Коефіцієнт кореляції може набувати значення:
   а) від –1 до 1;
   б) від 0 до 1;
   в) будь-які.

10. Кількість ступенів свободи для залишкової суми квадратів в лінійній моделі множинної регресії рівна:
   а) $n-1$;
   б) m;
   в) $n - m - 1$.

11. Залишкова сума квадратів відхилень дорівнює нулю:
   а) коли правильно підібрана регресійна модель;
   б) коли між ознаками існує точний функціональний зв'язок;
   в) ніколи.

12. Моделі, для яких виконується умова $M(\vec{y}) = M\left(X\vec{\alpha} + \vec{\varepsilon}\right) = X\vec{\alpha} + M(\vec{\varepsilon}) = X\vec{\alpha}$, називаються:
   а) гомоскедастичними;
   б) гетероскедастичними;
   в) економетричними.

13. Розташуйте в правильній послідовності виконання процедур МНК на конкретному прикладі:

а) $$X^T = \begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 28 & 33 & 19 & 20 & 22 & 31 & 30 & 33 & 18 & 24 & 26 & 33 & 34 & 25 & 26 & 28 & 40 & 37 & 41 & 33 \\ 7 & 14 & 8 & 14 & 8 & 12 & 8 & 16 & 7 & 12 & 18 & 12 & 14 & 15 & 11 & 16 & 22 & 15 & 18 & 25 \\ 18 & 26 & 24 & 20 & 23 & 8 & 19 & 21 & 27 & 23 & 25 & 24 & 29 & 28 & 25 & 26 & 32 & 28 & 27 & 33 \end{vmatrix}$$

б) $X = \begin{vmatrix} 1 & 28 & 7 & 18 \\ 1 & 33 & 14 & 26 \\ 1 & 19 & 8 & 24 \\ 1 & 20 & 14 & 20 \\ 1 & 22 & 8 & 23 \\ 1 & 31 & 12 & 18 \\ 1 & 30 & 8 & 19 \\ 1 & 33 & 16 & 21 \\ 1 & 18 & 7 & 27 \\ 1 & 24 & 12 & 23 \\ 1 & 26 & 18 & 25 \\ 1 & 33 & 12 & 24 \\ 1 & 34 & 14 & 29 \\ 1 & 25 & 15 & 28 \\ 1 & 26 & 11 & 25 \\ 1 & 28 & 16 & 26 \\ 1 & 40 & 22 & 32 \\ 1 & 37 & 15 & 28 \\ 1 & 41 & 18 & 27 \\ 1 & 33 & 25 & 33 \end{vmatrix}$;

в) $X^T Y = \begin{vmatrix} 1252 \\ 37040 \\ 17471 \\ 31317 \end{vmatrix}$

г) $\left(X^T X\right)^{-1} = \begin{vmatrix} 2{,}678451 & -0{,}03654 & 0{,}052314 & -0{,}09188 \\ -0{,}03654 & 0{,}001862 & -0{,}00161 & 0{,}000175 \\ 0{,}52314 & -0{,}00161 & 0{,}005336 & -0{,}00315 \\ -0{,}09188 & 0{,}000175 & -0{,}00315 & 0{,}005227 \end{vmatrix}$

д) $a = \begin{vmatrix} 36{,}79129 \\ 0{,}578703 \\ 0{,}448008 \\ 0{,}117116 \end{vmatrix}$

е) $X^T X = \begin{vmatrix} 20 & 581 & 272 & 496 \\ 581 & 17713 & 8267 & 14601 \\ 272 & 8267 & 4150 & 7005 \\ 496 & 14601 & 7005 & 12642 \end{vmatrix}$

ТЕМА 3

Мультиколінеарність та її вплив на оцінки параметрів моделі

Поняття мультиколінеарності. Її вплив на оцінки параметрів моделі • Ознаки мультиколінеарності • Алгоритм Фаррара – Глобера • Методи усунення мультиколінеарності • Метод головних компонент

3.1. Поняття мультиколінеарності. Її вплив на оцінки параметрів моделі

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

Суть мультиколінеарності: у багатофакторній регресійній моделі дві або більше незалежних змінних мають лінійну залежністю або, інакше кажучи, мають високий ступінь кореляції.

Математично сутнісну характеристику мультиколінеарності можна записати у вигляді співвідношення (модуль коефіцієнта кореляції прямує до одиниці):

$$\left|r_{x_i x_j}\right| \to 1, i \neq j, \tag{3.1}$$

де $\left|r_{x_i x_j}\right|$ – модуль числа $r_{x_i x_j}$.

Мультиколінеарність впливає на оцінки параметрів моделі. Визначник матриці спостережень $|X^T X|$ наближається до нуля, і оператор оцінювання за звичайним МНК стає надзвичайно чутливий до похибок вимірювань і похибок обчислень. При цьому МНК-оцінки можуть бути зміщеними відносно дійсних оцінок узагальненої моделі.

Розглянемо природу мультиколінеарності на прикладі залежності між ціною акції, дивідендами на акцію та отриманим прибутком на акцію. Дивіденди та отриманий прибуток на одну акцію мають висо-

Практичні наслідки мультиколінеарності

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

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

Рис. 3.1. Негативні наслідки присутності мультиколінеарності при розробенні економетричних моделей

кий ступінь кореляції. Виникає ситуація, коли два фактори змінюються в одному напрямку. Таким чином, неможливо оцінити вплив кожного з цих факторів на залежну змінну.

З'ясуємо, до яких наслідків може призвести мультиколінеарність. Це одне з найважливіших питань, яке потрібно зрозуміти при розробенні економетричних моделей (рис. 3.1).

Для ілюстрації розглянемо двофакторну регресійну модель:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \varepsilon \tag{3.1}$$

та її вибірковий аналог

$$y = a_0 + a_1 x_1 + a_2 x_2 . \tag{3.2}$$

Дисперсія оцінок параметрів a_1 і a_2 має вигляд:

$$D(a_1) = \frac{\sigma_\varepsilon^2}{\left(1 - r_{xy}{}^2\right)\sum_{i=1}^{n}\left(x_{1i} - \bar{x}_1\right)^2}; \tag{3.3}$$

$$D(a_2) = \frac{\sigma_\varepsilon^2}{\left(1 - r_{xy}{}^2\right)\sum_{i=1}^{n}\left(x_{2i} - \bar{x}_2\right)^2}; \tag{3.4}$$

$$\operatorname{cov}(a_1, a_2) = \frac{-r_{xy}\sigma_\varepsilon^2}{\left(1 - r^2\right)\sqrt{\sum_{i=1}^{n}\left(x_{1i} - \bar{x}_1\right)^2 \sum_{i=1}^{n}\left(x_{2i} - \bar{x}_2\right)^2}}. \tag{3.5}$$

де r_{xy} – коефіцієнт кореляції між x_1 і x_2.

З (3.4), (3.5) випливає, що якщо r зростає, то $D(a_1)$, $D(a_2)$ також зростають.

З (3.5) випливає, що якщо r_{xy} збільшується, $\text{cov}(a_1, a_2)$ зростає за абсолютною величиною. Причому при наближенні до граничного значення це збільшення має експоненціальний характер.

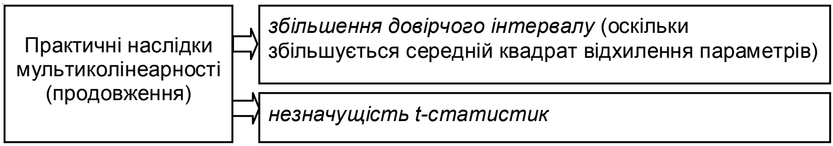


Рис. 3.2. Негативні наслідки присутності мультиколінеарності при розробенні економетричних моделей (продовження)

Оскільки значення t-статистики Стьюдента $t = \dfrac{a_1}{\sigma_{a_i}}$ то у випадку мультиколінеарності $\sigma_{a_i} \to \infty$, а отже, $t \to 0$ (рис. 3.2).

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

3.2. Ознаки мультиколінеарності

Єдиного способу визначення мультиколінеарності немає (рис. 3.3). Для визначення мультиколінеарності застосовують такі тести (рис. 3.4).

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

Тест, що застосовує характеристичні значення (власні числа матриці спостережень) та умовний індекс R (що обчислюється як відно-

Зовнішні ознаки наявності мультиколінеарності

велике значення R^2 і незначущість t-статистики

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

велике значення парних коефіцієнтів кореляції

Якщо значення хоча б одного коефіцієнта кореляції $r_{x_i x_j} > 0{,}8$, $i \neq j$, то мультиколінеарність є серйозною проблемою.
Зауважимо, що велике значення парних коефіцієнтів кореляції – достатня, але не необхідна умова наявності мультиколінеарності. Мультиколінеарність може мати місце навіть при відносно невеликих значеннях парних коефіцієнтах кореляції у більш ніж двофакторній регресійній моделі.

Рис. 3.3. Зовнішні ознаки присутності мультиколінеарності при розробенні економетричних моделей

Тести ідентифікації явища мультиколінеарності

F-тест, запропонований Глобером і Фарраром (побудова допоміжної регресії)

характеристичні значення та умовний індекс

Рис. 3.4. Тести ідентифікації присутності мультиколінеарності при розробенні економетричних моделей

шення максимального власного числа матриці до її мінімального власного числа), використовується в сучасних статистичних пакетах. Цим тестом розраховується не тільки умовне кількість R, а й умовний індекс $CI = \sqrt{R}$. Якщо $100 \le R \le 1000$, мультиколінеарність *помірна*, якщо $R > 1000$ – *висока*. Аналогічно, якщо $10 \le CI \le 33$, мультиколінеарність *помірна*, якщо $CI > 33$ – *висока*.

Ми розглянули основні методи тестування мультиколінеарності. Вони є універсальним, але мають один спільний недолік: жоден із них не проводить чіткої межі між тим, що треба вважати "суттєвою" мультиколінеарністю, яку необхідно враховувати, і тим, коли її можна не враховувати.

3.3. Алгоритм Фаррара – Глобера

Найповніше дослідити мультиколінеарність дає можливість алгоритм Фаррара – Глобера, який застосовує три види статистичних критеріїв для виявлення цього явища.

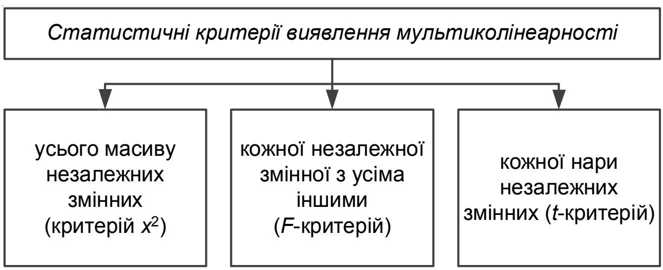


Рис. 3.5. Статистичні критерії ідентифікації присутності мультиколінеарності при розробленні економетричних моделей

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

Складемо покроковий ***алгоритм Фаррара – Глобера***.

1-й крок

нормалізувати змінні $x_1, x_2, \ldots, x_m$ економетричної моделі

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

де n – кількість спостережень ($i = 1, 2, \ldots, n$);

m – кількість незалежних змінних ($j = 1, m$);

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

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

2-й крок

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

$$R = X^{*tr}X^* = \begin{pmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{pmatrix}, \qquad (3.6)$$

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

Однак на основі залежності (3.5) не можна стверджувати, що отриманий зв'язок є явищем мультиколінеарності. Якщо діагональні елементи матриці R не дорівнюють одиниці, то на діагоналі цієї матриці ми проставляємо одиниці, а до решти елементів додаємо різницю між одиницею й значенням діагонального елемента.

3-й крок

на основі матриці X^*, елементами якої є нормалізовані незалежні змінні $x_{ij}{}^*$, обчислити кореляційну матрицю (матрицю моментів нормалізованої системи нормальних рівнянь) визначити $|R|$ – визначник кореляційної матриці R; обчислити критерій χ^2 порівняти значення χ^2 з табличним при $\frac{1}{2}m(m-1)$ ступенях свободи і рівні значущості *б* (якщо $\chi^2 > \chi^2_{табл}$, то в масиві незалежних змінних існує мультиколінеарність)

$$\chi^2 = -\left[n - 1 - \frac{1}{6}(2m+5)\right] \cdot \ln|R|; \qquad (3.7)$$

4-й крок

визначити матрицю похибок

(3.8)

5-й крок

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

$$F_k = \frac{(c_{kk} - 1)(n - m)}{(m - 1)}, \tag{3.9}$$

де c_{kk} – діагональні елементи матриці C

розрахувати коефіцієнти детермінації для кожної змінної

$$R_k^2 = 1 - \frac{1}{c_{kk}}. \tag{3.10}$$

6-й крок

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

$$r_{kj} = \frac{-c_{kj}}{\sqrt{c_{kk} c_{jj}}}, \tag{3.11}$$

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

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

7-й крок

розрахувати t-критерії; значення критеріїв t_{kj} порівняти з табличними при $(n–m)$ ступенях свободи та рівні значущості a; якщо $t_{kj} > t_{табл}$, то між незалежними змінними x_k і x_j існує мультиколінеарність

$$t_{kj} = \left| r_{kj} \right| \cdot \frac{\sqrt{n-m}}{\sqrt{1-r_{kj}^2}}. \qquad (3.12)$$

3.4. Методи усунення мультиколінеарності

Виявлення мультиколінеарності є лише частиною справи. Інша частина – як її усунути. Безпомилкових і абсолютно правильних порад немає, оскільки мультиколінеарність є прикладною проблемою (рис. 3.6).

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

Методи усунення мультиколінеарності:

- використання додаткової або первинної інформації
- об'єднання інформації
- відкидання змінної з високою кореляцією
- перетворення даних (використання перших різниць)
- збільшення кількості спостережень

Рис. 3.6. Підходи до усунення мультиколінеарності при розробенні економетричних моделей

3.5. Метод головних компонент

Цей метод призначений для оцінювання моделей великого розміру, а також для оцінки параметрів моделі, якщо до неї входять мультиколінеарні змінні.

Існують різні модифікації методу головних компонентів, які відрізняються між собою залежно від того, що береться за основу при визначенні ортогональних змінних – коваріаційна чи кореляційна матриця незалежних змінних.

Нехай маємо матрицю $X = (x_{ij})$, яка описує незалежні змінні моделі. Оскільки спостереження, що утворюють матрицю X, як правило, корельовані між собою, то можна поставити питання про кількість реально незалежних змінних, які входять до цієї матриці.

Точніше, ідея методу полягає в тому, щоб перетворити множину змінних X на нову множину попарно некорельованих змінних, серед яких перша відповідає максимально можливій дисперсії, а друга – максимально можливій дисперсії в підпросторі, який є ортогональним до першого, і т. ін.

Нехай нова змінна запишеться:

$$Z_{1i} = a_{11}x_{1i} + a_{21}x_{2i} + \ldots + a_{m1}x_{mi}, i = \overline{1,n}.$$

У матричній формі

$$Z_1 = Xa_1, \tag{3.13}$$

де Z_1 – вектор значень нової змінної;
a_1 – m-вимірний власний вектор матриці X^TX.

Суму квадратів елементів вектора подамо у вигляді

$$Z_1^T Z_1 = a_1^T X^T Xa_1. \tag{3.14}$$

Звідси необхідно вибрати такий вектор a_1, який максимізуватиме $Z_1^TZ_1$, але на вектор a_1 треба накласти обмеження, щоб він не став дуже великим. Тому ми його нормуємо, наклавши обмеження:

$$a^Ta_1 = 1. \tag{3.15}$$

Оскільки $Z_1 = Xa_1$, то максимізація a_1 буде максимізувати Z_1, а Z_1 характеризує внесок змінної Z_1 в загальну дисперсію.

Завдання тепер полягає в тому, щоб максимізувати $Z_1^TZ_1$ за умов (3.18). Побудуємо функцію Лагранжа:

$$f = a_1^T X^T X a_1 - \lambda_1 \left(a_1^T a_1 - 1 \right) \to \max, \tag{3.16}$$

де λ_1 – множник Лагранжа.

Узявши $\dfrac{\partial f}{\partial a_1} = 0$, дістанемо:

$$\left(X^T X \right) a_1 = \lambda_1 a_1. \tag{3.17}$$

Звідси бачимо, що a_1 – власний вектор матриці $X^T X$, який відповідає характеристичному числу λ_1.

Підставивши значення (3.19) у (3.16), дістанемо:

$$Z_1' Z_1 = \lambda_1 a_1^T a_1 = \lambda_1. \tag{3.18}$$

Отже, потрібно для значення λ_1 вибрати найбільший характеристичний корінь матриці $X'X$. За відсутності мультиколінеарності матриця $X'X$ буде додатно визначеною, і, відповідно, її характеристичні корені будуть додатними. Першим головним компонентом матриці X буде вектор Z_1.

Визначимо тепер $Z_2 = X \cdot a_2$. При цьому вектор a_2 має максимізувати вираз $a_2^T X^T X a^2$ за таких умов:

1) $a_2^T a^2 = 1$;
2) $a_1^T a_2 = 0$.

Друга умова забезпечить відсутність кореляції між Z_2 і Z_1, бо коваріація між Z_1 і Z_2 подається у вигляді $a^T X^T X a_2 = \lambda_1 a_1'^T a_2$, причому вона дорівнює нулю лише тоді, якщо $a_1^T a_2 = 0$.

Для розв'язування цієї задачі функцію Лагранжа запишемо у вигляді:

$$\psi = a_2^T X^T X a_2 - \lambda_2 (a_2^T a_2 - 1) - \mu (a_1^T a_2),$$

де λ_2 і μ – множники Лагранжа.

Узявши $\dfrac{\partial \psi}{\partial \lambda_2} = 0$ і $\dfrac{\partial \psi}{\partial \mu} = 0$, дістанемо $(X^T X) a_2 = \lambda_2 a_2$, де для значення λ_2 треба вибрати другий за величиною характеристичний корінь матриці $X^T X$.

Цей процес триває доти, доки всі m характеристичних значень матриці X^TX не будуть знайдені; знайдені m власних векторів матриці X^TX об'єднаємо в ортогональну матрицю:

$A = [a_1, a_2,.., a_m]$.

Отже, головні компоненти матриці X задаються матрицею

$$Z = XA \tag{3.19}$$

розміром $n \times m$.

$$Z'Z = A'X'XA = \lambda = \begin{pmatrix} \lambda_1 & 0 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \lambda_m \end{pmatrix}. \tag{3.20}$$

Вираз (3.22) означає, що головні компоненти дійсно попарно некорельовані, а їх дисперсії визначаються так:

$$Z_j^T Z_j = \lambda_j, j = \overline{1,m}; \tag{3.21}$$

$$\sum_{j=1}^{m} \lambda_j = tr(X^T X) = Z_1^T Z_1 + Z_2^T Z_2 + \ldots + Z_m^T Z_m.$$

Співвідношення $\dfrac{\lambda_1}{\sum\limits_j \lambda_j}, \dfrac{\lambda_2}{\sum\limits_j \lambda_j} \cdots \dfrac{\lambda_m}{\sum\limits_j \lambda_j}$ характеризують пропорційний внесок кожного з векторів у загальну варіацію змінних X, причому оскільки ці компоненти ортогональні, сума всіх внесків дорівнює одиниці.

Зауважимо, що вектори вихідних даних (матриця X) повинні мати однакові одиниці вимірювання, бо в протилежному разі дуже важко дати змістовне тлумачення поняттю загальної варіації змінних X і розкладанню цієї варіації на складові, виконаному відповідно до внеску кожного з векторів, якими подаються головні компоненти.

Іноді буває важко надати конкретного змісту знайденим головним компонентам. Для цього можна обчислити коефіцієнти кореляції кожного компонента з різними змінними X. Так, наприклад, візьмемо перший головний компонент Z_1 і знайдемо коефіцієнти його кореляції її з усіма змінними X. Для цього потрібно обчислити перехресні добутки між головним компонентом Z_1 і кожною з пояснювальних змінних X. Оскільки

$$X^T Z_1 = X^T X a_1 = \lambda_1 a_1,$$

маємо коефіцієнти кореляції для першого компонента:

$$r_{j1} = \frac{\lambda_1 \cdot a_{j1}}{\sqrt{\lambda_1}\sqrt{\sum\limits_{i=1}^{n} x_{ij}^2}} = \frac{a_{j1} \cdot \sqrt{\lambda_1}}{\sqrt{\sum\limits_{i=1}^{n} x_{ij}^2}}, i = \overline{1,n}; j = \overline{1,m}. \quad (3.22)$$

У загальному випадку коефіцієнт кореляції між x_j і Z_k

$$r_{kj} = \frac{a_{kj} \cdot \sqrt{\lambda_j}}{\sqrt{\sum\limits_{i=1}^{n} x_{ij}^2}} (j, k = \overline{1,m}). \quad (3.23)$$

Частка різних головних компонентів у варіації x_j визначається показником r_{kj}^2, а оскільки компоненти не корелюють один з одним, то сума їх часток дорівнює одиниці.

Визначивши всі головні компоненти і відкинувши ті з них, які відповідають невеликим значенням характеристичних коренів, знаходимо зв'язок залежної змінної Y з основними головними компонентами, а далі з допомогою оберненого перетворення повертаємося від параметрів моделі з головними компонентами до знаходження оцінок параметрів змінних X.

Алгоритм методу головних компонент

1-й крок

Нормалізація всіх пояснювальних змінних

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sigma_{x_j}}, \ i = \overline{1,n}; \ j = \overline{1,m}.$$

2-й крок

Обчислення кореляційної матриці

$$r = \frac{1}{n}\left(X^{*\prime} X^*\right).$$

3-й крок

Знаходження характеристичних чисел матриці r з рівняння

$$r = \frac{1}{n}\left(X^{*\prime} X^{*}\right).$$

4-й крок

Власні значення λ_k упорядковуються за абсолютним рівнем внеску кожного головного компонента до загальної дисперсії

5-й крок

Обчислення власних векторів a_k розв'язуванням системи рівнянь за умов

$$(r - \lambda E)a = 0$$

$$a'_j a_k = \begin{cases} 0 (j \neq k), \\ 1 (j = k). \end{cases}$$

6-й крок

Знаходження головних компонентів – векторів

$$z_k = x \cdot a_k, k = \overline{1,m}.$$

Головні компоненти мають задовольняти умови

$$\sum_{i=1}^{n} z_{k,i} = 0, i = \overline{1,n}$$

$$\frac{1}{n} z'_k z_k = \lambda_k, k = \overline{1,m}$$

$$z'_j z_k = 0, j = \overline{1,m}, j \neq k.$$

7-й крок

Визначення параметрів моделі $\hat{Y} = Z\hat{b}$

$$\hat{b} = Z^{-1} Y$$

8-й крок	Знаходження параметрів моделі $\hat{Y} = X\hat{\beta}$. Визначення параметрів моделі $\hat{Y} = Z\hat{b}$
	$\hat{\beta} = a \cdot \hat{b}$

Завдання для самоконтролю

1. Назвати, порушення яких з передумов Гаусса – Маркова виступає ознакою наявності мультиколінеарності?
2. Що означає мультиколінеарність змінних?
3. Які основні наслідки наявності явища мультиколінеарності?
4. Назвати ознаки мультиколінеарності.
5. Які статистичні критерії використовують для визначення мультиколінеарності?
6. Дайте коротку характеристику алгоритму Фаррара – Глобера.
7. Методи усунення мультиколінеарності.
8. Сутність методу головних компонент.
9. Визначити послідовність етапів застосування методу головних компонент для усунення мультиколінеарності.
10. Який статистичний критерій свідчить про наявність мультиколінеарності в масиві незалежних змінних? Описати методику визначення.
11. Який статистичний критерій свідчить про наявність мультиколінеарності кожної незалежної змінної з усіма іншими? Навести відповідні формули.
12. Який статистичний критерій свідчить про наявність мультиколінеарності кожної пари незалежних змінних? Виконання якої умови свідчить про наявність мультиколінеарності?

Практичне заняття 3

Дослідження мультиколінеарності

Постановка завдання

1 етап	Побудова інформаційного масиву дослідження шляхом збору статистичних даних та розробка моделі парної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), анаіз моделі
2 етап	Побудова множинної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), дослідження моделі
3 етап	**Виявлення, дослідження та усунення мультиколінеарності факторів у моделі множинної лінійної регресії методом Фаррара – Глобера**
4 етап	Виявлення ефекту гетероскедастичності на основі застосування параметричного тесту Гольдфельда – Квандта та усунення даного ефекту шляхом використання узагальненого методу найменших квадратів (методу Ейткена)
5 етап	Побудова адитивної та мультиплікативної моделі часового ряду на основі проведення декомпозиційного аналізу: визначення структури ряду (тенденції та сезонних (циклічних) коливань), проведення автокореляційного аналізу, визначення якіості моделі, розрахунок прогнозних значень
6 етап	Групування елементів економічної системи шляхом кластерного аналізу, побудова деревоподібної дендограми, графічне подання елементів кластеризації
7 етап	Перевірка наявності автокореляції залишків за допомогою критерію Дарбіна – Уотсона, оцінка параметрів рівняння множинної лінійної регресії методом Ейткена
8 етап	Побудова економіко-математичної моделі методом інструментальних змінних; побудова економетричної моделі на основі даних, що містять помилки вимірювання змінних, оцінка параметрів моделі за допомогою алгоритму Уоліса
9 етап	Побудова моделі розподіленого лага, оцінка параметрів моделей з лагами в незалежних змінних, оцінювання параметрів авторегресійних моделей; перетворення ARMA і ARIMA
10 етап	Структурне моделювання (моделювання на основі системи структурних рівнянь), оцінка невідомих параметрів моделі, визначення адекватності моделі щодо опису структури вхідних даних

***Мета роботи*:** набуття навичок виявлення, дослідження та усунення мультиколінеарності факторів у моделі множинної лінійної регресії.

Початкові дані

Стабільність банківської системи визначається значною кількістю показників певного досягнутого рівня та впливає на формування основних коефіцієнтів функціонування банківських установ. У межах дослідження мультиколінеарності розглянемо статистичну сукупність спостережень факторів (обсяги власного капіталу, зобов'язань, активів по банківській системі в цілому) на індикатор стабільності банківської системи («1» – факт нестабільного стану, «0» – факт стабільного стану) (табл. 3.1). Самостійно зібрана інформація – на оцінку 5 балів; за табл. 3.1, *відкориговану на кількість k* (на оцінку 3 бали).

Таблиця 3.1. Фактори, що впливають на індикатор стабільності банківської системи

Місяць	*Y*- індикатор стабільності банківської системи, бінарна характеристика	X_1 - обсяги власного капіталу, тис. грн	X_2, тис. грн	X_3, тис. грн
1	0	117081585	753551950	870633535
2	0	112597492	752097475	864694968
3	0	117968018	771990516	889958533
4	0	120207619	753241949	873449574
5	0	126646323	748243365	874964709
6	1	127162304	758093409	885255711
7	1	132802031	784695450	917497465
8	0	137725113	804358877	942083994
9	1	138434527	856598661	995033185
10	0	147816550	871994493	1019811043
11	0	151866259	877296268	1029162518
12	1	155486926	898785345	1054272287
13	0	162236166,3	920236938,9	1082473105
14	1	163775906,4	940619353,1	1104395259
15	1	165810129,5	951635752,2	1117445882
16	0	170196261,8	956983117,5	1127179379

Необхідно дослідити наявність мультиколінеарності між пояснювальними змінними, скориставшись алгоритмом Фаррара – Глобера.

Порядок виконання роботи:

1. Згідно з номером вашого варіанта оберіть умову задачі.
2. Дослідіть наявність мультиколінеарності між пояснювальними змінними, скориставшись алгоритмом Фаррара – Глобера.
3. Розв'язання задачі необхідно супроводжувати коментарями з наведенням формул, результатів обчислень та висновків за цими результатами.

Для вирішення задачі використати: інструктивні матеріали, надбудову «Аналіз даних» в MS Excel.

Методичні вказівки до розв'язання задачі

1. Обчислимо середні значення та стандартні відхилення пояснювальних змінних X_1, X_2, X_3. Для цього можна скористатися стандартними функціями MS Excel. У майстрі функцій знайдемо категорію "статистичні" і в ній функції "СРЗНАЧ" та "СТАНДОТКЛ". Дані величини можна також розрахувати за формулами:

$$\bar{X}_j = \frac{\sum X_{ij}}{n}, \tag{3.26}$$

$$\delta_j = \sqrt{\frac{\sum \left(X_{ij} - \bar{X}_j\right)^2}{n-1}}, \tag{3.27}$$

де $\bar{X}_j$ — середнє значення j-ї пояснювальної змінної;
X_{ij} — індивідуальне значення j-ї пояснювальної змінної;
j — номер пояснювальної змінної;
i — номер точки спостереження (місяця);
δ_j — стандартне відхилення j-ї пояснювальної змінної;
n — кількість спостережень.

Додаткові розрахунки наведено в табл. 3.2 та 3.3.

Таблиця 3.2. Проміжні розрахунки

Місяць	Y	$X1$	$X2$	$X3$	$(X_{1i} - X_{1ср})^2$	$(X_{2i} - X_{2ср})^2$	$(X_{3i} - X_{3ср})^2$
1	0	117081585	753551950	870633535	5,48E+14	7,05E+15	1,15E+16
2	0	112597492	752097475	864694968	7,78E+14	7,3E+15	1,28E+16
3	0	117968018	771990516	889958533	5,07E+14	4,29E+15	7,75E+15
4	0	120207619	753241949	873449574	4,11E+14	7,1E+15	1,09E+16
5	0	126646323	748243365	874964709	1,92E+14	7,97E+15	1,06E+16
6	1	127162304	758093409	885255711	1,78E+14	6,31E+15	8,61E+15
7	1	132802031	784695450	917497465	5,91E+13	2,79E+15	3,66E+15
8	0	137725113	804358877	942083994	7,64E+12	1,1E+15	1,29E+15
9	1	138434527	856598661	995033185	4,22E+12	3,64E+14	2,89E+14
10	0	147816550	871994493	1019811043	5,37E+13	1,19E+15	1,75E+15
11	0	151866259	877296268	1029162518	1,29E+14	1,58E+15	2,62E+15
12	1	155486926	898785345	1054272287	2,25E+14	3,75E+15	5,81E+15
13	0	162236166,3	920236938,9	1082473105	4,73E+14	6,84E+15	1,09E+16

Продовження табл. 3.2

Місяць	Y	$X1$	$X2$	$X3$	$(X_{1i}-X_{1ср})^2$	$(X_{2i}-X_{2ср})^2$	$(X_{3i}-X_{3ср})^2$
14	1	163775906,4	940619353,1	1104395259	5,42E+14	1,06E+16	1,6E+16
15	1	165810129,5	951635752,2	1117445882	6,41E+14	1,3E+16	1,94E+16
16	0	170196261,8	956983117,5	1127179379	8,83E+14	1,43E+16	2,22E+16
Усього	6	2247813211	13400422920	15648311148	5,63E+15	9,56E+16	1,46E+17

Таблиця 3.3. Проміжні розрахунки (продовження)

Показник	X_1	X_2	X_3
Середнє значення	140488325,7	837526432,5	978019446,7
Стандартне відхилення	19376128,51	79819390,98	98751997,19

2. Нормалізуємо пояснювальні змінні. Серед статистичних функцій MS Excel знайдемо функцію "НОРМАЛІЗАЦІЯ" та нормалізуємо X_1, X_2, X_3. .
Для цього можна також скористатися формулою

$$x_j^* = \frac{x_j - \bar{x}_j}{\delta_{x_j}}. \tag{3.28}$$

$$X^* = \begin{vmatrix} -1{,}208 & -1{,}052 & -1{,}087 \\ -1{,}439 & -1{,}070 & -1{,}148 \\ -1{,}162 & -0{,}821 & -0{,}892 \\ -1{,}047 & -1{,}056 & -1{,}059 \\ -0{,}714 & -1{,}119 & -1{,}044 \\ -0{,}688 & -0{,}995 & -0{,}939 \\ -0{,}397 & -0{,}662 & -0{,}613 \\ -0{,}143 & -0{,}416 & -0{,}364 \\ -0{,}106 & 0{,}239 & 0{,}172 \\ 0{,}378 & 0{,}432 & 0{,}423 \\ 0{,}587 & 0{,}498 & 0{,}518 \\ 0{,}774 & 0{,}767 & 0{,}772 \\ 1{,}122 & 1{,}036 & 1{,}058 \\ 1{,}202 & 1{,}292 & 1{,}280 \\ 1{,}307 & 1{,}430 & 1{,}412 \\ 1{,}533 & 1{,}497 & 1{,}510 \end{vmatrix}$$

Транспонуємо матрицю X^* (нормалізовану) в матрицю X^{*1}

$$X^{*1} = \begin{matrix} -1{,}208 & -1{,}439 & -1{,}162 & -1{,}047 & -0{,}714 & -0{,}688 & -0{,}397 \\ -1{,}052 & -1{,}070 & -0{,}821 & -1{,}056 & -1{,}119 & -0{,}995 & -0{,}662 \\ -1{,}087 & -1{,}148 & -0{,}892 & -1{,}059 & -1{,}044 & -0{,}939 & -0{,}613 \end{matrix}$$

Продовження матриці X^{*1}

$$\begin{matrix} -0{,}143 & -0{,}106 & 0{,}378 & 0{,}587 & 0{,}774 & 1{,}122 & 1{,}202 & 1{,}307 & 1{,}533 \\ -0{,}416 & 0{,}239 & 0{,}432 & 0{,}498 & 0{,}767 & 1{,}036 & 1{,}292 & 1{,}430 & 1{,}497 \\ -0{,}364 & 0{,}172 & 0{,}423 & 0{,}518 & 0{,}772 & 1{,}058 & 1{,}280 & 1{,}412 & 1{,}510 \end{matrix}$$

Перемножимо матриці X^{*1} та X^*:

$$X^{*1}X^* = \begin{vmatrix} 15 & 14{,}579 & 14{,}727 \\ 14{,}579 & 15 & 14{,}984 \\ 14{,}727 & 14{,}984 & 15 \end{vmatrix}$$

3. Знайдемо кореляційну матрицю r.

Для знаходження кореляційної матриці r необхідно кожний елемент матриці $X^{*1}X^*$ помножити на $\frac{1}{n-1}$ (у нашому випадку $n - 1 = 16 - 1 = 15$):

$$r = \begin{vmatrix} 1{,}0000 & 0{,}9720 & 0{,}9818 \\ 0{,}9720 & 1{,}0000 & 0{,}9989 \\ 0{,}9818 & 0{,}9989 & 1{,}0000 \end{vmatrix}$$

4. Знайдемо визначник матриці r ($\det r$).

Для знаходження ($\det r$) необхідно серед математичних функцій MS Excel знайти функцію “МОПРЕД”. Скориставшись нею, дістанемо: $(\det r) = 1{,}46 \cdot 10^{-9}$.

Оскільки $\det r$ наближається до нуля, то в масиві пояснювальних змінних може існувати мультиколінеарність.

Прологарифмуємо визначник матриці r: $\ln \det r = -20{,}34$.

5. Обчислимо критерій Пірсона χ^2 за формулою:

$$\chi^2 = -\left\{ n - 1 - \frac{1}{6}(2m+5) \right\} \ln(\det r). \qquad (3.29)$$

$$\chi^2 = -\left\{16-1-\frac{1}{6}(2*3+5)\right\}(-20.34) = 267.68 .$$

Знайдене значення χ^2 порівняємо з табличним значенням ($\chi^2_{кр} = 8,7$), коли маємо $\frac{1}{2}m(m-1) = 3$ ступенів свободи та при рівні значущості $\alpha = 0,05$.

Оскільки $\chi^2_{факт} > \chi^2_{кр}$, то в масиві пояснювальних змінних (обсяги власного капіталу, зобов'язань, активів по банківській системі в цілому) існує мультиколінеарність.

6. Обчислимо F-критерій.

Для визначення F-критеріїв необхідно знайти матрицю C, яка є оберненою до матриці r:

$$C = \begin{vmatrix} 1462324,251 & 6010559,564 & -7439819,228 \\ 6010559,564 & 24705540,220 & -30580194,629 \\ -7439819,228 & -30580194,629 & 37851792,738 \end{vmatrix}$$

Безпосередньо F-критерій обчислюється за формулою:

$$F = (C_{kk} - 1)\left(\frac{n-m}{m-1}\right), \tag{3.30}$$

де C_{kk} – діагональний елемент матриці C.

$$F_1 = (1462324.25 - 1)\left(\frac{16-3}{3-1}\right) = 9505101,13;$$

$$F_2 = (24705540.22 - 1)\left(\frac{16-3}{3-1}\right) = 160586004,93;$$

$$F_3 = (37851792,74 - 1)\left(\frac{16-3}{3-1}\right) = 246036646,30.$$

Обчислені критерії порівнюються з табличним значенням $F = 3,59$, коли є $n - m = 13$ ступенів свободи та при рівні значущості $\alpha = 0,05$.

У розглядуваному випадку $F_1 > F_{кр}$, $F_2 > F_{кр}$, $F_2 > F_{кр}$. Це означає, що кожна з пояснювальних змінних мультиколінеарна з іншими.

7. Визначимо частинні коефіцієнти кореляції *r*.

Частинні коефіцієнти кореляції показують тісноту зв'язку між двома пояснювальними змінними за умови, що всі інші змінні не впливають на цей зв'язок і обчислюються за формулою

$$r_{kj.s} = \frac{-c_{kj}}{\sqrt{c_{kk} c_{jj}}}, \tag{3.31}$$

$$r_{12.3} = \frac{6010559{,}564}{\sqrt{1462324{,}251 * 24705540{,}220}} = -0{,}9999905;$$

$$r_{13.2} = \frac{-7439819{,}228}{\sqrt{1462324{,}251 * 37851792{,}738}} = 0{,}9999938;$$

$$r_{23.1} = \frac{-30580194{,}629}{\sqrt{24705540{,}220 * 37851792{,}738}} = 0{,}99999963.$$

Отже, спираючись на здобуті нами значення окремих (частинних) коефіцієнтів кореляції, можна сказати, що: зв'язок між обсягами власного капіталу та зобов'язань є тісним, якщо не враховувати вплив активів; зв'язок між обсягами власного капіталу та активами є тісним, якщо не брати до уваги вплив зобов'язань; зв'язок між зобов'язаннями та активами по банківській системі в цілому також є тісним, якщо не враховувати обсяги власного капіталу.

8. Визначимо *t*-критерій.

Ці критерії застосовуються для визначення мультиколінеарності двох пояснювальних змінних і розраховуються за формулою:

$$t_{kj} = \frac{\left(r_{kj.s}\sqrt{n-m}\right)}{\left(\sqrt{1-r_{kj.s}^2}\right)}, \tag{3.32}$$

$$t_{12} = \frac{\left(-0{,}9999905 * \sqrt{16-3}\right)}{\left(\sqrt{1-(-0{,}999990529020)^2}\right)} = -828{,}43;$$

$$t_{13} = \frac{\left(0,9999938 * \sqrt{16-3}\right)}{\left(\sqrt{1-(0,9999938)^2}\right)} = 1025,42;$$

$$t_{23} = \frac{\left(0,99999963 * \sqrt{16-3}\right)}{\left(\sqrt{1-(0,99999963)^2}\right)} = 4214,84.$$

Обчислені t-критерії за абсолютним значенням порівнюються з табличним значенням ($t = 2,12$), коли маємо $n - m = 13$ ступенів свободи та при рівні значущості $\alpha = 0,05$.

Оскільки $t_{12} > t_{кр}$, то обсяги власного капіталу та зобов'язання по банківській системі в цілому є, відповідно, мультиколінеарними між собою.

$t_{13} > t_{кр}$, тому відповідно обсяги власного капіталу та активи по банківській системі в цілому є мультиколінеарними між собою.

$t_{23} > t_{кр}$, тому зобов'язання та активи по банківській системі в цілому є мультиколінеарними між собою.

Таким чином, дослідження, проведені за алгоритмом Фаррара – Глобера показали, що мультиколінеарність між пояснювальними змінними даного прикладу існує. Отже, для того, щоб можна було застосувати метод 1 МНК для оцінювання параметрів моделі за цією інформацією, необхідно в першу чергу звільнитися від мультиколінеарності.

Тести

1. Яка з умов Гаусса – Маркова не виконується в разі наявності явища мультиколінеарності:
   а) кожне значення випадкової складової рівняння *ui*, $i = 1, 2, ..., n$, є випадковою величиною, і математичне сподівання залишків *ui* дорівнює нулю;
   б) компоненти вектора залишків некорельовані (лінійно незалежні) між собою і мають сталу дисперсію;
   в) пояснювальні змінні (регресори, фактори моделі) некорельовані із залишками;
   г) пояснювальні змінні некорельовані між собою.

2. Визначте наслідки наявності мультиколінеарності:
   а) зміщення оцінок параметрів моделі;
   б) збільшення коваріації оцінок параметрів;
   в) отримання недопустимо великих значень дисперсії;
   г) зменшення довірчого інтервалу;
   д) незначущість параметрів побудованого рівняння регресії;
   е) значущість параметрів побудованого рівняння регресії.
   ж) зменшення коваріації оцінок параметрів

3. Достатня умова наявності мультиколінеарності полягає в такому:
   а) велике значення коефіцієнта детермінації;
   б) незначущість параметрів рівняння регресії;
   в) парні коефіцієнти кореляції свідчать про досить тісний зв’язок.

4. Визначте правильні твердження (об’єднуючи в речення одну ячейку з першого стовпця та одну комірку з другого) при визначенні тестів на наявність мультиколінеарності.

F-тест	побудова допоміжної регресії залежності кожного фактора x_i з усіма іншими факторами
-	відношення максимального власного числа матриці до її мінімального власного числа
характеристичні значення	побудова допоміжної регресії залежності кожного фактора x_i з результативною ознакою
умовний індекс	-
t-тест	власні числа матриці спостережень

5. Визначте правильні твердження (об'єднуючи в речення одну ячейку з першого стовпця та одну комірку з другого) при визначенні статистичних критеріїв виявлення мультиколінеарності:

весь масив незалежних змінних	–
весь масив залежних змінних	критерій x_i- квадрат
кожна незалежна змінна з усіма іншими	–
кожна пара незалежних змінних	F- критерій
кожна пара залежних змінних	t- критерій

6. Визначте послідовність етапів алгоритму Фаррара – Глобера:
   а) розрахувати елементи кореляційної матриці (матриці моментів нормалізованої системи нормальних рівнянь);
   б) визначити нормалізовані значення змінних $x_1, x_2, ..., x_m$ економетричної моделі;
   в) обчислити визначник кореляційної матриці R; обчислити критерій x_i-квадрат;
   г) розрахувати критерій Фішера;
   д) розрахувати матрицю похибок;
   е) знайти часткові коефіцієнти кореляції;
   ж) розрахувати критерій Стьюдента.

7. Визначте правильну відповідь:
   а) якщо фактичне значення критерію Стьюдента перевищує критичне, то між незалежними змінними *xk* і *xj* існує мультиколінеарність;
   б) якщо фактичне значення критерію Стьюдента перевищує критичне, то між залежними змінними *xk* і *xj* існує мультиколінеарність;
   в) якщо фактичне значення критерію Стьюдента менше від критичного, то між незалежними змінними *xk* і *xj* існує мультиколінеарність;
   г) якщо фактичне значення критерію Стьюдента менше від критичного, то між залежними змінними *xk* і *xj* існує мультиколінеарність.

8. Визначте правильну відповідь:
   а) якщо фактичне значення F-критерію перевищує критичне, то відповідна k-та незалежна змінна мультиколінеарна з іншими;
   б) якщо фактичне значення F-критерію перевищує критичне – відповідна k-та залежна змінна мультиколінеарна з іншими;
   в) у масиві незалежних змінних існує мультиколінеарність;
   г) у масиві залежних змінних існує мультиколінеарність.

9. Визначте методи усунення мультиколінеарності:
   а) використання додаткової або первинної інформації;
   б) агрегування інформації;
   в) нехтування змінною з високою кореляцією;
   г) перетворення даних (використання перших, других і т.д. різниць);
   д) зменшення кількості спостережень;
   е) збільшення незалежних змінних моделі;
   ж) застосування методу головних компонентів.

10. Алгоритм головних компонентів:
   а) нормалізація всіх пояснювальних змінних;
   б) розрахунок значень кореляційної матриці;
   в) визначення характеристичних чисел матриці r;
   г) власні значення упорядковуються за абсолютним рівнем внеску кожного головного компонента до загальної дисперсії;
   д) обчислення власних векторів;
   е) знаходження головних компонентів;
   ж) визначення параметрів проміжної моделі;
   з) визначення параметрів економетричної моделі.

Т Е М А 4

Узагальнені економетричні моделі

Моделі з порушенням передумов використання звичайного методу найменших квадратів • Узагальнений метод найменших квадратів • Суть гетероскедастичності • Гетероскедастичність і зважений метод найменших квадратів.

4.1. Моделі з порушенням передумов використання звичайного методу найменших квадратів

Розглянемо багатофакторну лінійну економетричну модель

$$y_{th} = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_m x_m$$

та її вибірковий аналог

$$y_{st} = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m + \varepsilon.$$

Під час реалізіції регресійного аналізу за допомогою звичайного МНК особливу увагу необхідно звернути на проблеми, пов'язані з виконанням необхідних умов для випадкових відхилень, оскільки властивості статистичних оцінок параметрів лінійної регресії перебувають у прямій залежності від цих відхилень ε_i.

Одним з основних припущень моделі класичної лінійної регресії є припущення про сталість дисперсії кожної випадкової величини ε_i (гомоскедастичність).

Формалізовано це припущення записується у вигляді:

$$\mathrm{var}(\varepsilon_i) = E\{\varepsilon_i - E(\varepsilon_i)\}^2 = \sigma_\varepsilon^2 = const. \qquad (4.1)$$

Якщо це припущення не задовольняється у якомусь окремому випадку, то має місце гетероскедастичність:

$$\mathrm{var}(\varepsilon_i) = \sigma_{\varepsilon_i}^2 \neq const. \qquad (4.2)$$

Звичайно, нас буде цікавити питання про доцільність цього припущення і про те, що відбувається, коли припущення про сталість дисперсії випадкової величини ε не задовольняється. Отже, спробуємо розглянути:

- природу, або суть, гетероскедастичності;
- наслідки гетероскедастичності;
- можливості тестування гетероскедастичності;
- корективні заходи, які слід вжити в разі порушення гетероскедастичності.

Суть припущення гомоскедастичності: варіація кожної ε_i навколо її математичного сподівання не залежить від значення x. Дисперсія кожної ε_i зберігається сталою, незалежно від малих чи великих значень факторів: σ_ε^2 не є функцією x_{ij}, тобто $\sigma_\varepsilon^2 \neq f(x_{1i}, x_{2i}, ..., x_{pi})$.

Для одержання якісних статистичних оцінок потрібно уважно стежити за виконанням передумов, що сформульовані в теоремі Гаусса – Маркова, бо при їх порушенні звичайний МНК дає статистичні оцінки, яким притаманні небажані властивості.

Однією із передумов теореми Гаусса – Маркова є:

$$\operatorname{cov}(\varepsilon_i \cdot \varepsilon_j) = \begin{cases} 0, & i \neq j \\ \sigma_\varepsilon^2 = const, & i = j, \end{cases} \quad \text{якщо } i, j = \overline{1,n},$$

де n – кількість спостережень.

Виконання цієї умови називають ***гомоскедастичністю*** залишків. Якщо порушується ця передумова, тобто

$$\operatorname{cov}(\varepsilon_i \cdot \varepsilon_j) = \begin{cases} 0, & i \neq j \\ \sigma_\varepsilon^2 \neq const, & i = j, \end{cases} \quad \text{якщо } i, j = \overline{1,n},$$

це є головною ознакою наявності гетероскедастичності моделі.

Моделі, для яких не виконуються передумови Гаусса – Маркова, можна поділити на три групи.

До **першої** належать такі моделі, для яких виконуються такі умови стосовно компонент випадкового вектора $\vec{\varepsilon}$:

1) вони мають нульові математичні сподівання:

$$M(\varepsilon_i) = 0, \quad i = \overline{1,n};$$

2) між собою є попарно некорельовані:

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} 0, & \text{якщо } i \neq j, \\ \sigma_{\varepsilon_i}^2 \neq \text{const}, & \text{якщо } i = j,\ i = \overline{1,n}. \end{cases}$$

У такому разі коваріаційна матриця випадкового вектора $\vec{\varepsilon}$ матиме такий вигляд:

$$\operatorname{cov}(\vec{\varepsilon}\cdot\vec{\varepsilon}') = M(\vec{\varepsilon}\cdot\vec{\varepsilon}') = M\left(\begin{pmatrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{pmatrix}\cdot\begin{pmatrix}\varepsilon_1 & \varepsilon_2 & \cdots & \varepsilon_n\end{pmatrix}\right) =$$

$$= M\begin{pmatrix}\varepsilon_1^2 & \varepsilon_1\varepsilon_2 & \cdots & \varepsilon_1\varepsilon_n\\ \varepsilon_2\varepsilon_1 & \varepsilon_2^2 & \cdots & \varepsilon_2\varepsilon_n\\ \ldots & \ldots. & .. & \ldots.\\ \varepsilon_n\varepsilon_1 & \varepsilon_n\varepsilon_2 & \cdots & \varepsilon_n^2\end{pmatrix} = \begin{pmatrix}M(\varepsilon_1^2) & M(\varepsilon_1\varepsilon_2) & \cdots & M(\varepsilon_1\varepsilon_n)\\ M(\varepsilon_2\varepsilon_1) & M(\varepsilon_2^2) & \cdots & M(\varepsilon_2\varepsilon_n)\\ \ldots & \ldots. & .. & \ldots.\\ M(\varepsilon_n\varepsilon_1) & M(\varepsilon_n\varepsilon_2) & \cdots & M(\varepsilon_n^2)\end{pmatrix} =$$

$$= \begin{pmatrix}\sigma_{\varepsilon_1}^2 & 0 & \cdots & 0\\ 0 & \sigma_{\varepsilon_2}^2 & \cdots & 0\\ \ldots & \ldots. & .. & \ldots.\\ 0 & 0 & \cdots & \sigma_{\varepsilon_n}^2\end{pmatrix}.$$

Отже,

$$\operatorname{cov}(\vec{\varepsilon}\cdot\vec{\varepsilon}') = \begin{pmatrix}\sigma_{\varepsilon_1}^2 & 0 & \cdots & 0\\ 0 & \sigma_{\varepsilon_2}^2 & \cdots & 0\\ \ldots & \ldots. & .. & \ldots.\\ 0 & 0 & \cdots & \sigma_{\varepsilon_n}^2\end{pmatrix}. \tag{4.3}$$

Такі моделі називають економетричними моделями з ознакою гетероскедастичності залишків.

До **другої** групи належать моделі, для яких виконуються такі умови:

1) збурення мають нульові математичні сподівання:

$$M(\varepsilon_i) = 0, \quad i = \overline{1,n};$$

2) вони є попарно корельованими:

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} k_{ij}, & \text{якщо } i \neq j, \\ \sigma^2_{\varepsilon_i} = \sigma^2_{\varepsilon} = \text{const}, & \text{якщо } i = j, \; i = \overline{1,n}. \end{cases} \tag{4.4}$$

де $k_{ij} = M(\varepsilon_i \cdot \varepsilon_j) - M(\varepsilon_i) \cdot M(\varepsilon_j) = M(\varepsilon_i \cdot \varepsilon_j) \neq 0.$

У цих моделях між випадковими відхиленями ε_i, ε_j існує кореляційний зв'язок, хоча дисперсії їх є сталими величинами.

Коваріаційна матриця в цьому разі матиме вигляд

$$\operatorname{cov}(\vec{\varepsilon} \cdot \vec{\varepsilon}') = \begin{pmatrix} \sigma^2_{\varepsilon} & k_{12} & k_{13} & \cdots & k_{1n} \\ k_{21} & \sigma^2_{\varepsilon} & k_{23} & \cdots & k_{2n} \\ k_{31} & k_{32} & \sigma^2_{\varepsilon} & \cdots & k_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ k_{n1} & k_{n2} & k_{n3} & \cdots & \sigma^2_{\varepsilon} \end{pmatrix}. \tag{4.5}$$

Слід пам'ятати, що $k_{ij} = k_{ji}$, тобто матриця є симетричною. Тому в цих моделях хоча умова гомоскедастичності (сталість дисперсій залишків) і виконується, але використання звичайного МНК не рекомендується внаслідок існування коваріаційних моментів між випадковими залишками.

До **третьої** групи належать моделі, для яких:

1) збурення мають нульові математичні сподівання

$$M(\varepsilon_i) = 0, \quad i = \overline{1,n};$$

2) елементи $\vec{\varepsilon}$ є попарно корельованими

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = \begin{cases} k_{ij}, & \text{якщо } i \neq j, \\ \sigma^2_{\varepsilon_i} \neq \text{const}, & \text{якщо } i = j, \; i = \overline{1,n}. \end{cases} \tag{4.6}$$

Слід наголосити, що для всіх трьох груп лінійних моделей з порушенням передумов застосування МНК точкові статистичні оцінки β_i^* для теоретичних параметрів β_i будуть незміщененими, але втрачають свою ефективність, тобто вони не матимуть мінімальну дисперсію, що призведе до зниження ймовірності одержання доброякісної оцінки.

Для моделей першої групи статистична оцінка параметрів здійснюється шляхом використання зваженого методу найменших квадратів. Для моделей другої та третьої груп — узагальненого методу найменших квадратів, які будуть розглянуті в наступних пунктах.

4.2. Узагальнений метод найменших квадратів

Розглянемо двофакторну економетричну модель:

$$Y_i = \beta_1 + \beta_2 X_i + u_i,$$

яку з метою спрощення подальших алгебраїчних перетворень пропонується записати у вигляді

$$Y_i = \beta_1 X_{0i} + \beta_2 X_i + u_i,$$

де $X_{0i} = 1$ для кожного i.

Зазначимо, що, незважаючи на проведені перетворення початкової економетричної моделі, обидві моделі є ідентичними. Для них гетероскедатстичні дисперсії σ_i^2 відомі. Поділимо останнє співвідношення на σi. У результаті даних обчислень отримаємо вираз

$$\frac{Y_i}{\sigma_i} = \beta_1\left(\frac{X_{0i}}{\sigma_i}\right) + \beta_2\left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_{0i}}{\sigma_i}\right),$$

який з метою проведення подальших перетворень пропонується записати у вигляді:

$$Y_1^* = \beta_1^* X_{0i}^* + \beta_2^* X_i^* + u_i^*,$$

де позначені зірочками змінні — трансфомовані змінні, які є вхідними початковими змінними, поділеними на відому компоненту σi.

Ми використовуємо позначення β_1^* та β_2^*, для параметрів трансформованої моделі з метою їх визначення на основі застосування звичайного методу найменших квадратів.

Розглянемо механізм застосування МНК для оцінки параметрів β_1^* та β_2^*. Запишемо таке співвідношення:

$$\frac{Y_i}{\sigma_i} = \beta_1^*\left(\frac{X_{0i}}{\sigma_i}\right) + \beta_2^*\left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right)$$

або

$$Y_i^* = \beta_1^* X_{oi}^* + \beta_2^* X_i^* + u_i^*.$$

Таким чином, для застосування узагальненого МНК необхідно мінімізувати:

$$\sum u_i^{2*} = \sum (Y_i^* - \beta_1^* X_{oi}^* - \beta_2^* X_i^*)^2,$$

тобто

$$\sum \left(\frac{u_i}{\sigma_i} \right)^2 = \sum \left[\left(\frac{Y_i}{\sigma_i} \right) - \beta_1^* \left(\frac{X_{0i}}{\sigma_i} \right) - \beta_2^* \left(\frac{X_i}{\sigma_i} \right) \right]^2.$$

Отже, після визначення частинних похідних наведеної вище функції за змінними (параметрами моделі) та розв’язання системи отриманих нормальних рівнянь співвідношення для розрахунку параметра β_2^* набуватиме вигляду:

$$\beta_2^* = \frac{(\sum w_i)(\sum w_i X_i Y_i) - (\sum w_i X_i)(\sum w_i Y_i)}{(\sum w_i)\left(\sum w_i X_i^2\right) - (\sum w_i X_i)^2},$$

де $w_i = 1/\sigma_i^*$.

Розглянемо ще один поширений у сучасній літературі з економетрики підхід до визначення сутності узагальненого методу найменших квадратів.

Умова гомоскедастичності є головною для лінійної класичної моделі і записується як

$$\operatorname{cov}(\vec{\varepsilon} \cdot \vec{\varepsilon}') = \sigma_\varepsilon^2 I_n. \tag{4.7}$$

Для лінійних моделей з властивістю випадкового вектора $\vec{\varepsilon}$, коли

$$\operatorname{cov}(\vec{\varepsilon} \cdot \vec{\varepsilon}') = \begin{pmatrix} \sigma_1^2 & k_{12} & k_{13} & \cdots & k_{1n} \\ k_{21} & \sigma_2^2 & k_{23} & \cdots & k_{2n} \\ k_{31} & k_{32} & \sigma_3^2 & \cdots & k_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ k_{n1} & k_{n2} & k_{n3} & \cdots & \sigma_n^2 \end{pmatrix} = \Omega \tag{4.8}$$

(матриця Ω є симетричною, додатньо визначеною матрицею n-го порядку) неможливим є використання звичайного МНК з метою визначення статистичних оцінок, як це було здійснено для лінійної класичної моделі. У такому разі використовують так званий узагальнений метод найменших квадратів (УМНК).

Нехай досліджується лінійна модель

$$\vec{y} = X\vec{\beta} + \vec{\varepsilon} \tag{4.9}$$

з порушенням умови гомоскедастичності, а саме:

$$\operatorname{cov}(\vec{\varepsilon} \cdot \vec{\varepsilon}') = \Omega. \tag{4.10}$$

Тоді додатньо визначена матриця Ω допускає існування такої невиродженої матриці π, що

$$\Omega = \pi \cdot \pi'. \tag{4.11}$$

Із (4.11) буде випливати

$$\pi^{-1} \cdot \Omega \cdot (\pi^{-1})' = I_n. \tag{4.12}$$

Таким чином, одержимо

$$\Omega^{-1} = (\pi^{-1})' \cdot \pi^{-1}. \tag{4.13}$$

Враховуючи (4.13) для моделі (4.9) здійснимо таке перетворення: ліву і праву частини рівняння помножимо зліва на матрицю π^{-1}:

$$\pi^{-1}\vec{y} = \pi^{-1}X\vec{\beta} + \pi^{-1}\vec{\varepsilon}.$$

Позначивши

$$\vec{y}^* = \pi^{-1}\vec{y},\ \ X^* = \pi^{-1}X,\ \ \vec{\varepsilon}^* = \pi^{-1}\vec{\varepsilon}, \tag{4.15}$$

одержимо

$$\vec{y}^* = X^*\vec{\beta} + \vec{\varepsilon}^*. \tag{4.16}$$

Здійснивши перевірку моделі на наявність гетероскедастичності, маємо:

1. $M(\vec{\varepsilon}^*) = M(\pi^{-1} \cdot \vec{\varepsilon}) = \pi^{-1}M(\vec{\varepsilon}) = 0$
2. $\operatorname{cov}(\vec{\varepsilon}^* \cdot (\vec{\varepsilon}^*)') = M(\vec{\varepsilon}^* \cdot (\vec{\varepsilon}^*)') = M\left(\pi^{-1} \cdot \vec{\varepsilon} \cdot (\pi^{-1} \cdot \vec{\varepsilon})'\right) =$
$= M\left(\pi^{-1} \cdot \vec{\varepsilon} \cdot \vec{\varepsilon}'(\pi^{-1})'\right) = \pi^{-1}M\left(\vec{\varepsilon} \cdot \vec{\varepsilon}'\right) \cdot (\pi^{-1})' = \pi^{-1} \cdot \Omega \cdot (\pi^{-1})' = I_n.$

Таким чином, виявилося, що перетворена модель (4.16) є гомоскедастичною, а тому для визначення статистичних оцінок цієї моделі можемо використати звичайний МНК, як для класичної лінійної моделі, і одержимо

$$\vec{\beta}^* = \left(\left(X^*\right)' \cdot X^*\right)^{-1} \cdot \left(X^*\right)' \cdot \vec{Y}^*. \tag{4.17}$$

Ураховуючи (4.15), маємо:

$$\vec{\beta}^* = \left(\left(\pi^{-1}X\right)' \cdot \pi^{-1}X\right)^{-1} \cdot \left(\pi^{-1}X\right)' \cdot \pi^{-1}\vec{Y} =$$

$$= \left(X'\left(\pi^{-1}\right)' \cdot \pi^{-1}X\right)^{-1} \cdot X'\left(\pi^{-1}\right)' \cdot \pi^{-1}\vec{Y} = \left(X' \cdot \Omega^{-1} \cdot X\right)^{-1} \cdot X' \cdot \Omega^{-1}\vec{Y}.$$

Коваріаційна матриця вектора $\vec{\beta}^*$ дорівнюватиме:

$$\operatorname{cov}\left(\vec{\beta}^* \cdot \left(\vec{\beta}^*\right)'\right) = \left(\left(X^*\right)' \cdot X^*\right)^{-1} = \left(\left(\pi^{-1}X\right)' \cdot \pi^{-1}X\right)^{-1} =$$

$$= \left(X'\left(\pi^{-1}\right)' \cdot \pi^{-1}X\right)^{-1} = \left(X' \cdot \Omega^{-1} \cdot X\right)^{-1}.$$

Отже, одержали:

$$\vec{\beta}^* = \left(X' \cdot \Omega^{-1} \cdot X\right)^{-1} \cdot X' \cdot \Omega^{-1}\vec{Y}, \tag{4.18}$$

$$\operatorname{cov}\left(\vec{\beta}^* \cdot \left(\vec{\beta}^*\right)'\right) = \left(X' \cdot \Omega^{-1} \cdot X\right)^{-1}. \tag{4.19}$$

Розглянутий метод перетворення початкової моделі (4.9) із подальшим використанням звичайного МНК до моделі (4.16) для визначення $\vec{\beta}^*$, $\operatorname{cov}(\vec{\beta}^* \cdot \left(\vec{\beta}^*\right)')$ дістав назву узагальненого методу найменших квадратів (УМНК).

При цьому слід наголосити, що для реалізації УМНК необхідно знати елементи матриці Ω, що на практиці є справою дуже складною. А тому цей метод певною мірою виконує суто ілюстративну функцію в економетрії.

Для практичного використання цього методу необхідно накласти певні умови на структуру матриці Ω.

4.3. Суть гетероскедастичності

Розглянемо моделі, що належать до першої групи моделей з порушенням передумов використання звичайного МНК.

При здійсненні вибірки ми маємо справу з конкретними реалізаціями залежної змінної Y і відповідними значеннями пояснювальних змінних (регресорів), при цьому завжди буде наявний фактор випадкових збурень, що породжують відхилення ε_i.

Випадкові величини ε_i апріорі можуть набувати довільних значень, що підпорядковані певним імовірним розподілам. Однією з головних вимог до цих розподілів є рівність їх дисперсій.

Цю вимогу потрібно розуміти так: незважаючи на те. що при кожному конкретному спостереженні випадкові відхилення ε_i будуть між собою відрізнятися, не повинно існувати причини, яка б спонукала значну розбіжність між цими величинами. Тобто похибки в середньому для всіх спостережень повинні мало відрізнятися. Звичайно, у певному розумінні тут припускається ідеалізація ситуації. Така ідеальна ситуація в реальних умовах не спостерігається. Часто при реалізації спостережень в одних і тих самих умовах відхилення ε_i будуть суттєво відрізнятися між собою, тобто в одних спостереженнях вони виявляються відносно великими, в інших – малими.

Так, наприклад, нехай залежність між результативною (Y) та факторною ознакою (X) описується парною лінійною регресією

$$Y = \beta_0 + \beta_1 X + \varepsilon. \tag{4.20}$$

Якщо умова гомоскедастичності не виконується (наявна гетероскедастичність), можуть виникати проблеми, пов'язані з ефектом масштабу (різних одиниць виміру). У часових рядах явище гетероскедастичності пов'язане з тим, що одні й ті самі показники розглядаються в різні моменти часу (наприклад, чистий експорт, темпи інфляції в певному регіоні за певний проміжок часу).

За наявності гетероскедастичності (моделі першої групи) статистична оцінка дисперсії σ_ε^2 обчислена за формулою

$$S_\varepsilon^2 = \frac{\vec{e}\,' \cdot \vec{e}}{n - m - 1}, \tag{4.21}$$

де n – кількість спостережень;

m – кількість регресорів в моделі, яка використовується для визначення дисперсій $S_{\beta_i^*}^2$ для всіх емпіричних коефіцієнтів β_i^* не буде незміщеною. Тоді t-статистика, F-статистика, інтервальні оцінки параметрів моделі стануть ненадійними.

Отже, використання звичайного МНК за наявності гетероскедастичності в моделі буде неефективним.

За МНК маємо суму квадратів похибок

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}\left(y_i^* - \beta_0^* - \beta_1^* x_i\right)^2. \tag{4.22}$$

Очевидно, що кожне конкретне значення e_i у наведеній сумі має однакову, так би мовити, "питому вагу", незалежно *від того, одержали його при значенні $X = x_i$ (де є мала дисперсія, чи при значенні $X = x_j$ (де наявна велика дисперсія),* що звичайно суперечить здоровому глузду, оскільки точка, одержана із розподілу $X = x_i$ точніше визначає напрямок (тенденцію) лінії регресії, ніж точка, одержана при $X = x_j$.

Тому, якщо поталанить врахувати "питому вагу" усіх точок e_i, це дозволить одержати ефективніші (доброякісні) (більш ефективні) статистичні оцінки.

4.4. Гетероскедастичність і зважений метод найменших квадратів

Ознаку гетероскедастичності в кожному конкретному випадку виявити складно, оскільки для цього необхідно знати величини σ_{ε_i} для кожного фіксованого значення $X = x_i$. На практиці, як правило, для кожного конкретного значення $X = x_i$ ми маємо в розпорядженні лише одне значення залежної змінної $Y = y_i$, а не цілий ряд розподілу. Це не дозволяє нам оцінити дисперсію випадкової величини Y за фіксованого значення $X = x_i$.

Існують апробовані тести, за допомогою яких можна виявити гетероскедастичність. І, як свідчить практика, їх використання дає позитивні наслідки. Такими є тести Глейзера і Гольдфельда – Квандта.

Тести Глейзера і Гольдфельда – Квандта можуть виявити наявність ознаки гетероскедастичності в моделі лише в разі порушення умови $\sigma_{\varepsilon_i}^2 = \text{const}$, тобто коли дисперсії залишків ε_i не є сталими величинами. Коли ж умова $\sigma_{\varepsilon_i}^2 = \text{const}$ виконується, і при цьому $\text{cov}(\varepsilon_i, \varepsilon_j) \neq 0$, то виявити це порушення умови гомоскедаститчності наведеними вище тестами неможливо.

У цьому разі маємо справу з лінійною моделлю з порушенням ознаки гомоскедастичності, яка належить до другої групи. Така ситуація виникає при дослідженні моделей з ознакою автокореляції.

Розглянемо приклад.

Приклад

Розглянемо залежність між прибутком банків України (Y, млн грн) та величиною їх статутного фонду (X, млн грн) на прикладі вибіркових даних, наведених в табл. 4.1.

Таблиця 4.1. Вхідні вибіркові дані прибутку та величини статутного фонду банків України

№ пор.	1	2	3	4	5	6	7	8	9	10
Y	27,54	14,87	11,69	11,49	11,46	11,38	11,07	10,76	10,60	9,73
X	100,0	102,67	75,00	85,00	70,09	60,00	54,00	57,35	52,30	52,00
№ пор.	11	12	13	14	15	16	17	18	19	20
Y	9,48	8,44	8,43	8,19	7,04	6,89	6,04	5,26	5,24	4,39
X	47,75	46,73	43,42	41,30	41,17	41,04	33,91	33,70	30,01	30,00
№ пор.	21	22	23	24	25	26	27	28	29	30
Y	3,47	2,49	2,48	1,41	1,39	1,00	0,53	0,20	0,18	0,15
X	26,82	24,64	23,37	23,82	22,26	20,50	15,50	14,20	13,54	13,41

Необхідно:

1. Обчислити оцінки параметрів моделі за методом найменших квадратів, перевірити суттєвість зв'язку в моделі та статистичну значущість розрахованих оцінок параметрів моделі.
2. Проаналізувати доцільність застосування методу найменших квадратів, перевіривши тестами Гольдфельда – Квандта та Глейзера наявність гетероскедастичності залишків.
3. Розрахувати оцінки параметрів моделі з урахуванням результатів тестів на наявність гетероскедастичності залишків.

Розв'язання

1. *Побудова моделі за МНК, перевірка моделі та її параметрів на статистичну значущість.*

Лінійна економетрична модель, оцінки параметрів якої знайдено за методом найменших квадратів для вибіркових даних (табл. 4.1), має вигляд:

$$y_i^* = -2,366 + 0,219 x_i. \qquad (4.23)$$

Коефіцієнт детермінації цієї моделі $R^2 = 0,854$, між результативною змінною Y і регресором X коефіцієнт парної кореляції $r_{xy} = 0,924$. Перевіримо статистичну значущість коефіцієнта парної кореляції за критерієм Стьюдента:

$$t^*_{cn} = \sqrt{\frac{0,854 \cdot (30-1-1)}{1-0,854}} = 12,779.$$

При рівні значущості $a = 0,05$ та ступенях свободи $k = n-m-1 = 28$ табличне значення критерію Стьюдента $t''_{кр}(\frac{\alpha}{2}, k) = 2,048$, отже, $t^*_{cn} > t''_{кр}(\frac{\alpha}{2}, k)$ зв'язок у моделі між Y та X суттєвий. Стандартні помилки оцінок параметрів і відповідні їм значення t-критерію дорівнюють:

$$s_{\beta_0^*} = 0,848; \quad s_{\beta_1^*} = 0,017;$$
$$t_{\beta_0^*} = -2,789; \quad t_{\beta_1^*} = 12,779.$$

$t_{\beta_0^*}, t_{\beta_1^*} \notin [-2,048; 2,048]$, а це дає можливість зробити висновок про суттєву відмінність від нуля розрахованих оцінок параметрів моделі.

2. *Перевірка на наявність гетероскедастичності залишків.*

Застосуємо тест Гольдфельда – Квандта для перевірки наявності гетероскедастичності залишків. Тестом перевіряється за критерієм Фішера основна гіпотеза H_0: $\sigma^2_{\varepsilon_1} = \sigma^2_{\varepsilon 2} = \ldots = \sigma^2_{\varepsilon_m}$ при альтернативній гіпотезі H_α: не H_0. Для цього упорядкуємо вхідні дані в порядку спадання значень пояснювальної змінної X (табл. 4.2).

Таблиця 4.2. Впорядковані вхідні вибіркові дані за показником прибутку банків України

№ пор.	1	2	3	4	5	6	7	8	9	10
Y	14,87	27,54	11,49	11,69	11,46	11,38	10,76	11,07	10,60	9,73
X	102,67	100,0	85,00	75,00	70,09	60,00	57,35	54,00	52,30	52,00
№ пор.	11	12	13	14	15	16	17	18	19	20
Y	9,48	8,44	8,43	8,19	7,04	6,89	6,04	5,26	5,24	4,39
X	47,75	46,73	43,42	41,30	41,17	41,04	33,91	33,70	30,01	30,00
№ пор.	21	22	23	24	25	26	27	28	29	30
Y	3,47	2,49	2,48	1,41	1,39	1,00	0,53	0,20	0,18	0,15
X	26,82	24,64	23,37	23,82	22,26	20,50	15,50	14,20	13,54	13,41

Відкинемо $c = \frac{4 \cdot n}{15} = 8$ $\left(i = \overline{12,19}\right)$ середніх спостережень (друга підвибірка), вважаючи, що дисперсія залишків для них постійна. За методом найменших квадратів знайдемо оцінки параметрів спочатку для першої моделі (підвибірка 1) з $n_1 = \frac{n-c}{2} = \frac{30-8}{2} = 11$ найбільшими значеннями регресора X, потім для другої – з $n_3 = 11$ найменшими значеннями X (третя підвибірка).

Знайдемо значення y_i^* та e_i (табл. 4.3) для першої та другої моделей та перевіримо наявність гетероскедастичності залишків на основі тесту Гольдфельда – Квандта.

Таблиця 4.3. Моделі перевірки наявності гетероскедастичності залишків на основі тесту Гольдфельда – Квандта

№ пор.	Модель 1: $y_i^* = -0{,}116 + 0{,}187x_i$					Модель 2: $y_i^* = -3{,}246 + 0{,}234x_i$				
	y_i	x_i	y_i^*	e_i	e_i^2	y_i	x_i	y_i^*	e_i	e_i^2
1	14,87	102,67	19,08	–4,21	17,688	4,39	30,00	3,78	0,61	0,375
2	27,54	100,00	18,58	8,96	80,343	3,47	26,82	3,03	0,44	0,191
3	11,49	85,00	15,77	–4,28	18,342	2,49	24,64	2,52	–0,03	0,001
4	11,69	75,00	13,90	–2,21	4,899	1,41	23,82	2,33	–0,92	0,848
5	11,46	70,09	12,99	–1,53	2,328	2,48	23,37	2,23	0,25	0,065
6	11,38	60,00	11,10	0,28	0,079	1,39	22,26	1,97	–0,58	0,331
7	10,76	57,35	10,60	0,16	0,024	1,00	20,50	1,55	–0,55	0,307
8	11,07	54,00	9,98	1,09	1,192	0,53	15,50	0,38	0,15	0,022
9	10,60	52,30	9,66	0,94	0,883	0,20	14,20	0,08	0,12	0,015
10	9,73	52,00	9,60	0,13	0,016	0,18	13,54	–0,08	0,26	0,065
11	9,48	47,75	8,81	0,67	0,449	0,15	13,41	–0,11	0,26	0,066
Σ					126,243					2,285

Для цього обчислимо суми квадратів залишків $S_1 = \sum_{i=1}^{n_1} e_i^2$ і $S_3 = \sum_{i=1}^{n_2} e_i^2$ та розрахуємо значення F_{cn}^*:

$$F_{cn}^{*} = \begin{cases} \dfrac{S_3 / n_3}{S_1 / n_1}, & \text{якщо } S_3 > S_1; \\ \dfrac{S_1 / n_1}{S_3 / n_3}, & \text{якщо } S_1 > S_3. \end{cases}$$

Розраховане значення $F_{cn}^{*} = \dfrac{126{,}243}{2{,}285} = 55{,}253$ порівняємо з табличним значенням $F_{кр}$ (α, k_1, k_2) = 3,18 =3,18 із ступенями свободи $k_1 = 11–1–1 = 9$ і $k_2 = 11–1–1 = 9$ при рівні значущості $\alpha = 0{,}05$. Оскільки $F_{cn}^{*} > F_{кр}$ (α, k_1, k_2), гіпотезу H_0 про відсутність гетероскедастичності залишків відхиляємо.

Проаналізуємо наявність гетероскедастичності за тестом Глейзера. Для моделі (4.20) обчислимо значення залишків та запишемо їх за абсолютною величиною (табл. 4.4).

Таблиця 4.4. Значення залишків та їх абсолютних величин у розрізі перевірки наявності гетероскедастичності за тестом Глейзера

№ пор.	1	2	3	4	5	6	7	8	9	10
e_i	7,96	–5,29	–2,40	–4,80	–1,55	–0,58	–1,59	–0,54	–1,49	–0,69
$\lvert e_i \rvert$	7,96	5,29	2,40	4,80	1,55	0,58	1,59	0,54	1,49	0,69
№ пор.	11	12	13	14	15	16	17	18	19	20
e_i	1,37	0,55	1,27	1,49	0,37	0,25	0,97	0,23	1,02	0,17
$\lvert e_i \rvert$	1,37	0,55	1,27	1,49	0,37	0,25	0,97	0,23	1,02	0,17
№ пор.	21	22	23	24	25	26	27	28	29	30
e_i	–0,05	–0,55	–0,28	–1,45	–1,13	–1,13	–0,51	–0,55	–0,43	–0,43
$\lvert e_i \rvert$	0,05	0,55	0,28	1,45	1,13	1,13	0,51	0,55	0,43	0,43

Коефіцієнт детермінації моделі $R^2 = 0{,}628$, між $\lvert e_i \rvert$ та X коефіцієнт парної кореляції $r = 0{,}792$. Стандартні помилки оцінок параметрів і відповідні їм значення t-критерію дорівнюють:

$$s_{\alpha_0^*} = 0{,}403; \qquad s_{\alpha_1^*} = 0{,}008;$$

$$t_{\alpha_0^*} = -2{,}604; \qquad t_{\alpha_1^*} = 6{,}874.$$

За рівня значущості $\alpha = 0{,}05$ і ступенів свободи $k = n-m-1 = 28$ табличне значення критерію Стьюдента $t''_{кр}(\frac{\alpha}{2},k)$ $t_{\alpha/2,\,k} = 2{,}048$. Порівняємо розраховані значення $t_{\alpha_j^*}$ з табличним:

$$t_{\alpha_0^*}, t_{\alpha_1^*} \notin [-2{,}048; 2{,}048].$$

Отже, приймаємо за тестом Глейзера гіпотезу про наявність гетероскедастичності залишків.

3. *Застосуємо зважений метод найменших квадратів для знаходження оцінок параметрів моделі з гетероскедастичними регресійними залишками.*

На основі тесту Гольдфельда – Квандта та Глейзера припускаємо, що $\sigma_{\varepsilon_i}^2 = \sigma^2 x_i^2$. Використовуючи (4.15), маємо

$$\vec{\beta}^* = \left(\left(X^*\right)' \cdot X^*\right)^{-1} \cdot \left(X^*\right)' \cdot \vec{y}^*,$$

де $$\vec{y}^* = \begin{pmatrix} \frac{y_1}{x_1} \\ \frac{y_2}{x_2} \\ \ldots \\ \frac{y_{30}}{x_{30}} \end{pmatrix}; \quad X^* = \begin{pmatrix} 1 & \frac{1}{x_1}1 \\ 1 & \frac{1}{x_2} \\ \ldots & \ldots \\ 1 & \frac{1}{x_{30}} \end{pmatrix}; \quad \vec{\beta}^* = \begin{pmatrix} \beta_0^* \\ \beta_1^* \end{pmatrix}.$$

На основі даних табл. 4.5 знаходимо, що

$$\vec{\beta}^* = \begin{pmatrix} -3{,}194 \\ 0{,}241 \end{pmatrix}.$$

Для визначення s_{β_0} та s_{β_1} використовується формула

$$\operatorname{cov}(\vec{\beta}^* \cdot (\vec{\beta}^*)') = s_\varepsilon^2((X^*)' \cdot X^*)^{-1},$$

Таблиця 4.5. Проміжні розрахунки для застосування зваженого методу найменших квадратів

№ пор.	1	2	3	4	5	6	7	8	9	10
$\frac{y_i}{x_i}$	0,2754	0,1448	0,1558	0,1352	0,1634	0,1897	0,2050	0,1877	0,2026	0,1870
$\frac{1}{x_i}$	0,0100	0,0097	0,0133	0,0118	0,0143	0,0167	0,0185	0,0174	0,0191	0,0192
№ пор.	11	12	13	14	15	16	17	18	19	20
$\frac{y_i}{x_i}$	0,1986	0,1806	0,1941	0,1941	0,1710	0,1678	0,1782	0,1561	0,1747	0,1465
$\frac{1}{x_i}$	0,0209	0,0214	0,0230	0,0242	0,0243	0,0244	0,0295	0,0297	0,0333	0,0333
№ пор.	21	22	23	24	25	26	27	28	29	30
$\frac{y_i}{x_i}$	0,1295	0,1011	0,1060	0,0592	0,0623	0,0487	0,0343	0,0143	0,0132	0,0110
$\frac{1}{x_i}$	0,0373	0,0406	0,0428	0,0420	0,0449	0,0488	0,0645	0,0704	0,0738	0,0746

де $s_\varepsilon^2 = \dfrac{\vec{e}'\cdot\vec{e}}{n-m-1} = \dfrac{\sum_{i=1}^{30}(y_i-(-3{,}194+0{,}241x_i))^2}{28} = \dfrac{0{,}02785}{28} = 0{,}00995$ та

$$((X^*)'\cdot X^*)^{-1} = \begin{pmatrix} 0{,}131 & -3{,}063 \\ -3{,}063 & 96{,}328 \end{pmatrix}.$$ Тут $n-m-1 = 30-1-1 = 28$.

Таким чином, одержано

$$s_{\beta_0^*} = 0{,}310;\ s_{\beta_1^*} = 0{,}011.$$

Перевіримо статистичну значущість оцінок параметрів на основі t-критерію:

$$t_{\beta_0^*} = 10{,}311; \qquad t_{\beta_1^*} = 21{,}146.$$

За рівня значущості $\alpha = 0{,}05$ та ступенів свободи $k = n - m - 1 = 28$ табличне значення критерію Стьюдента $t''_{кр}(\frac{\alpha}{2},k) =$ 2,048.

$t_{\beta_0^*}, t_{\beta_1^*} \notin [-2{,}048; 2{,}048]$, розраховані оцінки параметрів моделі суттєво відрізняються від нуля.

Порівняно з моделлю (4.23) середньоквадратичні помилки оцінок параметрів зменшилися при відповідному збільшенні значення $t_{\beta_j^*}$.

Отже, економетрична модель з урахуванням гетероскедастичності залишків має такий вигляд:

$$y_i^* = -3{,}194 + 0{,}241 x_i,$$

тобто майже чверть величини статутного фонду визначає щорічний прибуток банку, що може бути характеристикою ефективності використання початкового (акціонерного) капіталу українських банків.

Завдання для самоконтролю

1. Дайте визначення гомоскедастичності і гетероскедастичності.
2. Як впливає явище гетероскедастичності на оцінку параметрів моделі?
3. Назвіть методи визначення гетероскедастичності.
4. Як перевіряється гетероскедастичність згідно з критерієм μ?
5. Як застосовується параметричний тест для визначення гетероскедастичності?
6. У чому сутність непараметричного тесту?
7. Як визначається гетероскедастичність з допомогою регресії залишків?
8. Опишіть методи формування матриці S в умові $M(uu') = \sigma^2 S$.
9. Як використовується матриця S у методі Ейткена?
10. Які властивості повинна мати матриця S?
11. Запишіть формулу обчислення матриці коваріацій параметрів моделі. Чим вона відрізняється від формули при застосуванні 1 МНК?
12. Як отримати незміщену оцінку дисперсії залишків за наявності гетероскедастичності?
13. Суть та наслідки мультиколінеарності. Методи виявлення та усунення з моделі ознаки мультиколінеарності.
14. Суть та наслідки гетероскедастичності. Методи виявлення та усунення з моделі ознаки гетероскедастичності.
15. Суть тесту Гольдфельда – Квандта. Послідовність його виконання.
16. Узагальнений метод найменших квадратів (метод Ейткена): сутність і використання.
17. Явище автокореляції: причини, наслідки. Алгоритм Дарбіна – Уотсона.

Практичне заняття 4

Виявлення та усунення ефекту гетероскедастичності з використанням MS Excel і пакету STATISTICA

Постановка завдання

1 етап	Побудова інформаційного масиву дослідження шляхом збору статистичних даних та розробка моделі парної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), анаіз моделі
2 етап	Побудова множинної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), дослідження моделі
3 етап	Виявлення, дослідження та усунення мультиколінеарності факторів у моделі множинної лінійної регресії методом Фаррара – Глобера
4 етап	**Виявлення ефекту гетероскедастичності на основі застосування параметричного тесту Гольдфельда – Квандта та усунення даного ефекту шляхом використання узагальненого методу найменших квадратів (методу Ейткена)**
5 етап	Побудова адитивної та мультиплікативної моделі часового ряду на основі проведення декомпозиційного аналізу: визначення структури ряду (тенденції та сезонних (циклічних) коливань), проведення автокореляційного аналізу, визначення якіості моделі, розрахунок прогнозних значень
6 етап	Групування елементів економічної системи шляхом кластерного аналізу, побудова деревоподібної дендограми, графічне подання елементів кластеризації
7 етап	Перевірка наявності автокореляції залишків за допомогою критерію Дарбіна – Уотсона, оцінка параметрів рівняння множинної лінійної регресії методом Ейткена
8 етап	Побудова економіко-математичної моделі методом інструментальних змінних; побудова економетричної моделі на основі даних, що містять помилки вимірювання змінних, оцінка параметрів моделі за допомогою алгоритму Уоліса
9 етап	Побудова моделі розподіленого лага, оцінка параметрів моделей з лагами в незалежних змінних, оцінювання параметрів авторегресійних моделей; перетворення ARMA і ARIMA
10 етап	Структурне моделювання (моделювання на основі системи структурних рівнянь), оцінка невідомих параметрів моделі, визначення адекватності моделі щодо опису структури вхідних даних

Мета роботи: набуття навичок виявлення та усунення ефекту гетероскедастичності з використанням MS Excel.

Зміст роботи: за допомогою табличного процесора MS Excel перевірити гіпотезу про наявність гетероскедастичності для вихідних даних.

Початкові дані

На основі даних, які наведені в табл. 4.6—4.25 (номер варіанта відповідає номеру таблиці) необхідно:

1. Перевірити гіпотезу про наявність гетероскедастичності для вихідних даних.
2. Оцінити параметри моделі за методом 1 МНК і методом Ейткена.
3. Визначити матриці коваріацій оцінок та їх стандартні помилки.
4. Перевірити істотність зв'язку, що описується моделями, оцінки параметрів яких розраховані за методом 1 МНК і методом Ейткена.
5. Дати порівняльний аналіз оцінок параметрів, отриманих різними методами, зробити висновки.
6. Для вирішення задачі використати: інструктивні матеріали, вбудовані функції, надбудову «Аналіз даних» в MS Excel.

Вимоги до оформлення звіту

Звіт про проведення даної лабораторної роботи оформлюється, як й інші звіти цього курсу, в окремому зошиті згідно зі встановленими вимогами до оформлення звітів і містить:

- назву, тему, завдання, опис мети лабораторної роботи;
- вихідні дані варіанта;
- результати рішення у вигляді таблиці MS Excel;
- висновок про істотність зв'язку, що описується моделями;
- оцінки параметрів, які розраховані за методом 1 МНК і методом Ейткена;
- короткий опис технології вирішення задачі в MS Excel та Statistica.

1. Для побудови економетричної моделі, що характеризує залежність між витратами на реалізацію продукції, обсягом товарообігу та середнім рівнем товарних запасів, необхідно перевірити гіпотезу про наявність гетероскедастичності для вихідних даних, які наведені в табл. 4.6—4.12.

Таблиця 4.6. Вхідні статистичні дані для побудови економетричної моделі (варіант 1)

№ складу	Витрати на реалізацію продукції, млн грн	Обсяг товарообігу, млн т	Середній рівень товарних запасів, млн т
1	300	25	5
2	280	20	4
3	350	30	6
4	340	30	7
5	330	28	7
6	320	28	5
7	310	25	6
8	300	24	4
9	320	27	5
10	280	22	4
11	340	35	6
12	360	30	7
13	320	29	7
14	300	28	5
15	310	25	6
16	350	26	4

Таблиця 4.7. Вхідні статистичні дані для побудови економетричної моделі (варіант 2)

№ складу	Витрати на реалізацію продукції, млн грн	Обсяг товарообігу, млн т	Середній рівень товарних запасів, млн т
1	350	26	5
2	280	22	4
3	350	30	6
4	340	30	7
5	300	29	7
6	320	28	5
7	320	25	6
8	280	24	4
9	300	23	6
10	380	21	5
11	340	30	6
12	360	32	7
13	330	28	8
14	320	29	5
15	340	25	6
16	300	24	9

Таблиця 4.8. Вхідні статистичні дані для побудови економетричної моделі (варіант 3)

№ складу	Витрати на реалізацію продукції, млн грн	Обсяг товарообігу, млн т	Середній рівень товарних запасів, млн т
1	400	25	5
2	380	20	4
3	350	30	6
4	360	30	7
5	430	28	7
6	420	28	5
7	310	25	6

Таблиця 4.9. Вхідні статистичні дані для побудови економетричної моделі (варіант 4)

№ складу	Витрати на реалізацію продукції, млн грн	Обсяг товарообігу, млн т	Середній рівень товарних запасів, млн т
1	320	15	5
2	280	10	4
3	350	20	6
4	340	20	7
5	330	18	7
6	320	18	5
7	310	15	6

Продовження табл. 4.8

№ складу	Витрати на реалізацію продукції, млн грн	Обсяг товарообігу, млн т	Середній рівень товарних запасів, млн т
8	400	24	4
9	350	25	5
10	380	20	4
11	350	30	6
12	340	30	7
13	320	28	7
14	320	28	5
15	310	25	6
16	380	24	4

Продовження табл. 4.9

№ складу	Витрати на реалізацію продукції, млн грн	Обсяг товарообігу, млн т	Середній рівень товарних запасів, млн т
8	300	14	4
9	310	15	5
10	290	10	4
11	350	20	6
12	340	20	7
	330	18	7
14	320	18	5
15	310	15	6
16	300	14	4

Таблиця 4.10. Вхідні статистичні дані для побудови економетричної моделі (варіант 5)

№ складу	Витрати на реалізацію продукції, млн грн	Обсяг товарообігу, млн т	Середній рівень товарних запасів, млн т
1	300	25	5
2	280	20	4
3	350	20	6
4	340	10	7
5	380	28	7
6	320	28	5
7	310	15	6
8	400	24	4
9	350	25	5
10	280	20	4
11	350	20	6
12	340	20	7
13	380	28	7
14	320	28	5
15	310	15	6
16	320	14	4

Таблиця 4.11. Вхідні статистичні дані для побудови економетричної моделі (варіант 6)

№ складу	Витрати на реалізацію продукції, млн грн	Обсяг товарообігу, млн т	Середній рівень товарних запасів, млн т
1	200	15	5
2	180	20	4
3	250	30	6
4	240	30	7
5	230	38	7
6	220	38	5
7	210	25	6
8	250	34	4
9	200	15	5
10	190	25	4
11	250	30	6
12	240	30	7
13	230	18	7
14	280	28	5
15	210	25	6
16	200	14	4

Таблиця 4.12. Вхідні статистичні дані для побудови економетричної моделі (варіант 7)

№ складу	Витрати на реалізацію продукції, млн грн	Обсяг товарообігу, млн т	Середній рівень товарних запасів, млн т
1	300	8	5
2	280	10	4
3	350	20	6
4	340	15	7
5	330	18	7
6	320	18	5
7	310	15	6
8	300	14	4
9	300	9	5
10	280	10	4
11	350	15	6
12	340	20	7
13	330	21	7
14	320	15	5
15	310	17	6
16	300	20	4

2. Для побудови економетричної моделі, що характеризує залежність між обсягом споживання на душу населення, ціною за одиницю продукції та доходом на душу населення, перевірити гіпотезу про відсутність гетероскедастичності для вихідних даних, які наведено в табл. 4.13—4.17.

Таблиця 4.13. Вхідні статистичні дані для побудови економетричної моделі (варіант 8)

№ спостереження	Обсяг споживання продукції на душу населення, т	Ціна за одиницю продукції, тис. грн	Дохід на душу населення, тис. грн
1	100	30	90
2	120	35	75
3	130	40	85
4	125	30	90
5	140	45	105

Таблиця 4.14. Вхідні статистичні дані для побудови економетричної моделі (варіант 9)

№ спостереження	Обсяг споживання продукції на душу населення, т	Ціна за одиницю продукції, тис. грн	Дохід на душу населення, тис. грн
1	200	30	80
2	220	35	75
3	230	40	85
4	225	35	90
5	240	45	95

Продовження табл. 4.13

№ спостереження	Обсяг споживання продукції на душу населення, т	Ціна за одиницю продукції, тис. грн	Дохід на душу населення, тис. грн
6	150	30	93
7	155	55	97
8	160	60	100
9	100	30	80
10	120	35	75
11	130	40	85
12	125	30	100
13	140	45	95
14	150	50	93
15	155	55	97
16	160	60	100

Продовження табл. 4.14

№ спостереження	Обсяг споживання продукції на душу населення, т	Ціна за одиницю продукції, тис. грн	Дохід на душу населення, тис. грн
6	250	50	93
7	255	55	97
8	260	62	100
9	200	30	80
10	220	38	75
11	230	40	85
12	225	36	90
13	240	45	95
14	250	50	93
15	255	55	97
16	260	60	98

Таблиця 4.15. Вхідні статистичні дані для побудови економетричної моделі (варіант 10)

№ спостереження	Обсяг споживання продукції на душу населення, т	Ціна за одиницю продукції, тис. грн	Дохід на душу населення, тис. грн
1	110	20	20
2	130	25	95
3	140	30	85
4	125	20	100
5	140	35	95
6	160	20	103
7	155	45	97
8	180	50	100
9	120	20	90
10	120	25	85
11	130	30	85
12	135	20	90
13	145	35	105
14	150	20	193
15	155	45	197
16	160	50	200

Таблиця 4.16. Вхідні статистичні дані для побудови економетричної моделі (варіант 11)

№ спостереження	Обсяг споживання продукції на душу населення, т	Ціна за одиницю продукції, тис. грн	Дохід на душу населення, тис. грн
1	100	30	8
2	120	35	17
3	130	40	18
4	125	30	19
5	140	45	19
6	150	50	19
7	155	55	19
8	160	60	20
9	100	30	9
10	120	35	7
11	130	40	28
12	125	30	29
13	140	45	9
14	150	50	9
15	155	55	19
16	160	60	30

Таблиця 4.17. Вхідні статистичні дані для побудови економетричної моделі (варіант 12)

№ спостереження	Обсяг споживання продукції на душу населення, т	Ціна за одиницю продукції, тис. грн	Дохід на душу населення, тис. грн
1	100	30	28
2	120	35	27
3	130	40	25
4	125	30	20
5	140	45	35
6	150	50	40
7	155	55	45
8	160	60	50
9	150	65	120
10	140	70	125
11	130	35	30
12	180	45	145
13	120	40	50
14	135	55	55
15	155	50	88
16	160	40	68

3. Для побудови економетричної моделі, що характеризує залежність між споживанням продукту, ціною його та витратами на виробництво одиниці продукції, необхідно перевірити гіпотезу про відсутність гетероскедастичності для вихідних даних, які наведено в табл. 4.18–4.23.

Таблиця 4.18. Вхідні статистичні дані для побудови економетричної моделі (варіант 13)

№ спостереження	Споживання продукту на душу населення, т	Ціна за одиницю продукції, тис. грн	Витрати на виробництво 1т продукту млн грн
1	50	10	60
2	45	12	62
3	55	9	65
4	50	10	60

Таблиця 4.19. Вхідні статистичні дані для побудови економетричної моделі (варіант 14)

№ спостереження	Споживання продукту на душу населення,т	Ціна за одиницю продукції, тис. грн	Витрати на виробництво 1т продукту, млн грн
1	50	10	60
2	45	12	62
3	55	9	62
4	50	10	60

Продовження табл. 4.18

№ спостереження	Споживання продукту на душу населення, т	Ціна за одиницю продукції, тис. грн	Витрати на виробництво 1т продукту млн грн
5	60	8	55
6	70	16	50
7	52	10	70
8	47	12	62
9	55	9	62
10	50	10	60
11	65	18	85
12	70	9	50

Продовження табл. 4.19

№ спостереження	Споживання продукту на душу населення,т	Ціна за одиницю продукції, тис. грн	Витрати на виробництво 1т продукту, млн грн
5	60	8	55
6	70	6	50
7	55	5	20
8	57	6	22
9	56	7	20
10	59	8	25
11	51	9	20
12	58	10	15

Таблиця 4.20. Вхідні статистичні дані для побудови економетричної моделі (варіант 15)

№ спостереження	Споживання продукту на душу населення, т	Ціна за одиницю продукції, тис. грн	Витрати на виробництво 1т продукту млн грн
1	50	10	70
2	45	12	62
3	55	9	72
4	50	10	60
5	60	18	55
6	70	6	60
7	50	15	60
8	45	12	82
9	55	19	62
10	50	10	60
11	60	9	85
12	70	14	60

Таблиця 4.21. Вхідні статистичні дані для побудови економетричної моделі (варіант 16)

№ спостереження	Споживання продукту на душу населення,т	Ціна за одиницю продукції, тис. грн	Витрати на виробництво 1т продукту, млн грн
1	50	10	90
2	45	12	62
3	55	19	82
4	50	10	60
5	60	8	85
6	70	16	50
7	50	10	70
8	45	12	62
9	55	9	62
10	50	10	80
11	60	12	65
12	70	14	60

Таблиця 4.22. Вхідні статистичні дані для побудови економетричної моделі (варіант 17)

	Спожи-вання продукту на душу населення, т	Ціна за одиницю продукції, тис. грн	Витрати на виробництво 1т продукту млн грн
1	50	10	70
2	45	12	62
3	55	9	62
4	50	10	60
5	60	8	55
6	70	6	50
7	80	8	82
8	85	10	65
9	75	12	74
10	90	7	87
11	45	15	62
12	50	14	60

Таблиця 4.23. Вхідні статистичні дані для побудови економетричної моделі (варіант 18)

№ спостереження	Спожи-вання продукту на душу населення,т	Ціна за одиницю продукції, тис. грн	Витрати на виробництво 1т продукту, млн грн
1	55	5	100
2	57	6	110
3	56	7	112
4	59	8	115
5	51	9	120
6	58	10	135
7	60	8	140
8	50	9	120
9	45	15	110
10	65	8	130
11	70	11	150
12	52	12	170

4. Для побудови економетричної моделі, що характеризує залежність між споживанням продукту, доходом на душу населення та ціною, необхідно перевірити гіпотезу про відсутність гетероскедастичності для вихідних даних, які наведено в табл. 4.24–4.25.

Таблиця 4.24. Вхідні статистичні дані для побудови економетричної моделі (варіант 19)

№ спостереження	Спожи-вання продукту на душу населення, т	Ціна за одиницю продукції, тис. грн	Дохід на душу населенння
1	75	5	100
2	67	6	110
3	56	7	112
4	59	8	115
5	51	9	120
6	58	10	135
7	60	7	120

Таблиця 4.25. Вхідні статистичні дані для побудови економетричної моделі (варіант 20)

№ спостереження	Спожи-вання продукту на душу населення, т	Ціна за одиницю продукції, тис. грн	Дохід на душу населення
1	55	5	160
2	87	6	170
3	56	7	112
4	59	8	115
5	51	9	120
6	58	10	135
7	60	11	145

Продовження табл. 4.24

№ спостереження	Споживання продукту на душу населення, т	Ціна за одиницю продукції, тис. грн	Дохід на душу населеннння
8	65	8	130
9	45	9	110
10	40	10	100
11	50	15	130
12	65	14	140
13	80	7	150
14	82	5	170
15	58	10	100

Продовження табл. 4.25

№ спостереження	Споживання продукту на душу населення, т	Ціна за одиницю продукції, тис. грн	Дохід на душу населення
8	65	5	100
9	70	6	110
10	50	4	120
11	55	5	130
12	62	7	150
13	53	8	140
14	50	9	120
15	55	10	130

Довідкові матеріали

1. Основні положення теми «Гетероскедастичність».
2. Застосування параметричного тесту Гольдфельда – Квандта для визначення гетероскедастичності.
3. Оцінка параметрів моделі на основі узагальненого методу найменших квадратів (методу Ейткена).

Основні положення теми «Гетероскедастичність»

Передумови, які висуваються при оцінці параметрів моделі за методом 1 МНК, на практиці часто можуть порушуватися. Однією з таких передумов є незмінність дисперсії залишків для всіх спостережень вихідної сукупності. Це явище називається ***гомоскедастичністю***. У практичних дослідженнях воно часто порушується. Наприклад, в економетричній моделі, що характеризує залежність витрат на споживання від доходу, дисперсія залишків може змінюватись для спостережень, які відносяться до різних груп населення за величиною доходів.

Якщо дисперсія залишків в економетричному моделюванні змінюється для кожного спостереження або для груп спостережень, то це явище називається ***гетероскедастичністю***.

Наявність гетероскедастичності спричиняє порушення властивостей оцінок параметрів моделі при розрахунку їх за методом 1 МНК. Тому завжди виникає необхідність вивчати це явище, і, якщо воно існує, для оцінки параметрів моделі використовувати узагальнений метод найменших квадратів (метод Ейткена).

Для визначення гетероскедастичності застосовуються чотири критерії:

1) критерій μ;
2) параметричний тест Гольдфельда – Квандта;
3) непараметричний тест Гольдфельда – Квандта;
4) тест Глейзера.

1. *Критерій* μ

Цей метод застосовується в разі, якщо вихідна сукупність спостережень досить велика. Розглянемо цей алгоритм.

Крок 1. Вихідні дані залежної змінної Y поділяються на k груп $(r = \overline{1,k})$ згідно із зміною рівня величини Y.

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

$$S_r = \sum_{i=1}^{n_r} (y_{ir} - \overline{y_r})^2.$$

Крок 3. Розраховується сума квадратів відхилень у цілому по всій сукупності спостережень:

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

Крок 4. Обчислюється параметр λ:

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

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

Крок 5. Розраховується критерій μ:

$\mu = -2\ln \lambda,$

який наближено буде відповідати розподілу X_2 за ступенів свободи $k - 1$, коли дисперсія всіх спостережень однорідна. Тобто, якщо значення менше від табличного значення X_2 при вибраному рівні довіри і ступені свободи $k - 1$, то явище гетероскедастичності відсутнє.

2. *Параметричний тест Гольдфельда – Квандта*

Коли сукупність спостережень невелика, то розглянутий метод 1 застосовувати неможливо.

Тоді Гольдфельд і Квандт розглянули випадок, коли $M(uu') = \sigma^2 x_{ij}^2$, тобто дисперсія залишків зростає пропорційно квадрату однієї із незалежних змінних моделі:

$Y = XA + u.$

Вони запропонували для виявлення наявності гетероскедастичності параметричний тест, в якому треба виконати такі кроки.

Крок 1. Упорядкувати спостереження згідно з величиною елементів вектора x_j.

Крок 2. Відкинути c спостережень, які будуть знаходитись у центрі вектора. На основі експериментальних розрахунків автори розрахували оптимальні співвідношення між параметрами c і n, де n — кількість елементів вектора x_j.

$$\frac{c}{n} = \frac{4}{15}.$$

Крок 3. Побудувати дві економетричні моделі на основі 1 МНК за двома створеними сукупностями спостережень ${(n_1 - c)}/{2}$ за умови, що ${(n_2 - c)}/{2}$ перевищує кількість змінних m.

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

$$S_1 = \hat{u}_1' \, \hat{u}_1,$$

де $\hat{u}_1$ – залишки по моделі (1);

$$S_2 = \hat{u}_2' \, \hat{u}_2,$$

де $\hat{u}_2$ – залишки по моделі (2);

Крок 5. Розрахувати критерій R:

$$R = \frac{S_2}{S_1},$$

який при виконанні гіпотези про гомоскедастичність буде відповідати F-розподілу з $\frac{n_1 - c - 2m)}{2}$, $\frac{n_2 - c - 2m)}{2}$ ступенями свободи. Це означає, що розраховане значення R^* порівнюється з табличним значенням F-критерію за ступенів свободи $\frac{n_1 - c - 2m)}{2}$, $\frac{n_2 - c - 2m)}{2}$ і вибраного рівня довіри. Якщо $R \le F_{табл}$, то гетероскедастичність відсутня.

3. *Непараметричний тест Гольдфельда — Квандта*.

Гольфельд і Квандт запропонували також для оцінки наявності гетероскедастичності непараметричний тест. Цей тест базується на числі піків у величині залишків після упорядкування спостережень по x_{ij}.

4. *Тест Глейзера.*

Ще один тест для перевірки гетероскедастичності запропонував Глейзер: розглядати регресію абсолютних значень залишків $|u_i|$, які відповідають регресії найменших квадратів як деяку функцію від x_j, де x_j є тією незалежною змінною, яка відповідає зміні дисперсії σ_u^2. Для цього використовуються такі види функцій:

1) $|u| = a_0 + a_1 x_j$;

2) $|u| = a_0 + a_1 x_j^{-1}$;

3) $|u| = a_0 + a_1 x_j^{\frac{1}{2}}$ і т.п.

Рішення про відсутність гетероскедастичності залишків приймається на основі статистичної значущості коефіцієнтів a_0 й a_1. Переваги цього тесту визначаються можливістю розрізняти випадок чистої і змішаної гетероскедастичності. Чистій гетероскедастичності відповідають значення параметрів $a_0 = 0$, $a_0 \neq 0$; а змішаній – $a_0 \neq 0$, $a_1 \neq 0$. Залежно від цього необхідно користуватись різними матрицями S. Нагадаємо, що:

$$M(uu') = \sigma_u^2 S.$$

Якщо при економетричному моделюванні для певних вихідних даних буде виявлено явище гетероскедастичності, то оцінку параметрів моделі треба виконувати на основі узагальненого методу найменших квадратів. Оператор оцінювання цим методом запишеться:

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

де $$S^{-1} = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

У даній матриці залежно від висунутої гіпотези

або $\lambda_i = \dfrac{1}{x_{ij}}$;

або $\lambda_i = \dfrac{1}{x_{ij}^2}$;

або $\lambda_i = \left\{ \left|\hat{u}_i\right| \right\}$.

Прогноз на основі економетричної моделі, у якій оцінка параметрів виконана узагальненим методом найменших квадратів, можна отримати на основі такого співвідношення:

$$\hat{Y}_{np} = X_0 A + W' V^{-1} U,$$

де u – вектор залишків, який відповідає оцінці параметрів моделі на основі 1 МНК;
W' – транспонований вектор коваріацій поточних і прогнозних значень залишків;
$V^{-1} = S^{-1}$, а $V = \sigma_U^2 S$.

Застосування параметричного тесту Гольдфельда – Квандта для визначення гетероскедастичності

Приклад 1. Нехай треба побудувати економетричну модель, яка характеризує залежність заощаджень від доходів населення. Для побудови цієї моделі використовується вихідна сукупність даних, яка охоплює 18 спостережень. Ці дані та розрахунки на їх основі наведені в табл. 4.26. Виходячи із сутності взаємозв'язку величини заощаджень та доходу населення, можна припустити, що дисперсія залишків не є постійною для кожного спостереження, тобто тут може існувати явище гетероскедастичності. Тому, щоб правильно вибрати метод для оцінки параметрів моделі, необхідно перевірити, чи властива гетероскедастичність для наведених вихідних даних.

Розв'язання

1. Ідентифікація змінних:

$$Y = f(X, u),$$

де Y – залежна змінна (заощадження);
X – незалежна змінна (дохід);
u – стохастична складова.

Таблиця 4.26. Вхідні та розрахункові дані взаємозв'язку величини заощаджень та доходів населення

Рік	Заощадження Y	Дохід* X		X^2	XY	$\hat{Y}$	$Y-\hat{Y}$	$(Y-\hat{Y})^2$
1-й	1,36	13,8		190,44	18,768	1,1974	0,1626	0,0264
2-й	1,20	14,4		207,36	17,280	1,2016	–0,0016	0,0000025
3-й	1,08	15,0	*n*2	225,00	16,200	1,2057	–0,1257	0,0158
4-й	1,20	15,6		243,36	18,720	1,2098	–0,0098	0,000096
5-й	1,10	16,0		256,00	17,600	1,2126	–0,1126	0,0127
6-й	1,12	16,9		285,61	18,928	1,2188	–0,0988	0,0098
7-й	1,41	17,7		313,29	24,957	1,2243	0,1857	0,0345
8-й	1,50	18,5						
9-й	1,43	19,3						
10-й	1,59	20,5						
11-й	1,90	21,7						
12-й	1,95	22,7		515,29	44,265	1,8401	0,1099	0,0121
13-й	1,82	23,6		556,96	42,925	1,9885	–0,1685	0,0284
14-й	2,04	24,7		610,09	50,388	2,1699	–0,1299	0,0169
15-й	2,53	26,1	*n*1	681,21	66,033	2,4008	0,1292	0,0167
16-й	2,94	27,8		772,84	81,732	2,6811	0,2589	0,0670
17-й	2,75	28,9		835,21	79,475	2,8625	–0,1125	0,0127
18-й	2,99	30,2		912,04	90,298	3,0769	–0,0869	0,0076
Σ	31,91	373,4						

Продовження табл. 4.26

Рік	$\lambda_i = 1/X_i$	$\hat{Y}$	$Y-\hat{Y}$	$(Y-\hat{Y})^2$	$Y-\bar{Y}$	$(Y-\bar{Y})^2$
1-й	0,0725	0,9865	0,3735	0,1395	–0,4128	0,1704
2-й	0,0694	1,0542	0,1458	0,2126	–0,5728	0,3281
3-й	0,0667	1,1219	–0,0419	0,0176	–0,6928	0,4799
4-й	0,0641	1,1896	0,0104	0,0001	–0,5728	0,3281
5-й	0,0625	1,2347	–0,1347	0,0181	–0,6728	0,4527
6-й	0,0592	1,3362	–0,2162	0,0467	–0,6528	0,4261
7-й	0,0565	1,4265	–0,0165	0,0003	–0,3628	0,1316
8-й	0,0541	1,5167	–0,0167	0,0003	–0,2728	0,0744
9-й	0,0518	1,6069	–0,1769	0,0313	–0,3428	0,1175

* У таблиці дані впорядковані за величиною доходу, починаючи від меншого до більшого значення.

Продовження табл. 4.26

Рік	$\lambda_i = 1/X_i$	$\hat{Y}$	$Y-\hat{Y}$	$(Y-\hat{Y})^2$	$Y-\overline{Y}$	$(Y-\overline{Y})^2$
10-й	0,0488	1,7423	–0,1523	0,0232	–0,1828	0,0334
11-й	0,0461	1,8777	0,0223	0,0005	0,1272	0,0162
12-й	0,0441	1,9905	–0,0405	0,0016	0,1772	0,0314
13-й	0,0424	2,0919	0,2719	0,0739	0,0472	0,0022
14-й	0,0405	2,2161	–0,1761	0,0310	0,2672	0,0714
15-й	0,0383	2,3739	0,1561	0,0244	0,7572	0,5734
16-й	0,0359	2,5657	0,3743	0,1401	1,1672	1,3624
17-й	0,0346	2,6898	0,0602	0,0036	0,9772	0,9549
18-й	0,0331	2,8365	0,1535	0,0236	1,2172	1,4816
Σ				0,7884		7,0357

2. Специфікація моделі:

$Y = a_0 + a_1X + u,$

$\hat{Y} = \hat{a}_0 + \hat{a}_1X,$

$u = Y - \hat{Y}.$

3. Визначимо наявність гетероскедастичності. Для цього застосуємо алгоритм Гольдфельда – Квандта. Дану сукупність спостережень впорядкуємо по X від меншого до більшого значення. Відшукуємо C спостережень, які знаходяться в середині сукупності:

$$\frac{C}{n} = \frac{5}{15}, \quad n = 18, \quad \frac{C}{18} = \frac{4}{15}, \quad C = \frac{4 \cdot 18}{15}.$$

Тоді $n_1 = n_2 = 7.$

3.1. Розрахуємо економетричну модель для сукупності $n_1 = 7$. Оцінимо кількісно параметри моделі на основі 1 МНК.

$$\begin{cases} n\hat{a}_0 + \hat{a}_1 \sum x = \sum y, \\ \hat{a}_0 \sum x + a_1 \sum x^2 = \sum xy. \end{cases}$$

$\Sigma x = 109{,}4.$

$\Sigma y = 8{,}47.$

$\Sigma x^2 = 1721{,}06.$

$\Sigma xy = 132{,}453.$

$$\begin{cases} 7\hat{a}_0 + 109{,}4\hat{a}_1 = 8{,}47, \\ 109{,}4\hat{a}_0 + 1721{,}06\hat{a}_1 = 132{,}453, \end{cases} \left| \begin{array}{l} *109{,}4 \\ *7 \end{array} \right.$$

$$\begin{cases} 765{,}8\hat{a}_0 + 11968{,}36\hat{a}_1 = 926{,}618, \\ 765{,}8\hat{a}_0 + 12047{,}42\hat{a}_1 = 927{,}171, \end{cases}$$

$79{,}06\,\hat{a}_1 = 0{,}553.$
$\hat{a}_1 = 0{,}0069.$

$7\hat{a}_0 + 109{,}4 * 0{,}0069 = 8{,}47.$
$7\hat{a}_0 = 8{,}47 - 0{,}7549.$
$7\hat{a}_0 = 7{,}7151.$
$\hat{a}_0 = 1{,}1022.$

$\hat{Y}_1 = 1{,}1022 + 0{,}0069X$ – перша економетрична модель.

На основі моделі можна дійти висновку: якщо дохід виросте на 1, то заощадження збільшаться на 0,0069 грн.

3.2. Розрахуємо економетричну модель для сукупності $n^2 = 7$.
Оцінимо кількісно параметри моделі на основі 1 МНК.

$$\begin{cases} n\hat{a}_0 + \hat{a}_1 \sum x = \sum y, \\ \hat{a}_0 \sum x + \hat{a}_1 \sum x^2 = \sum xy. \end{cases}$$

$\Sigma x = 184.$
$\Sigma y = 17{,}02.$
$\Sigma x^2 = 4883{,}64.$
$\Sigma xy = 455{,}142.$

$$\begin{cases} 7\hat{a}_0 + 184\hat{a}_1 = 17{,}02, \\ 184\hat{a}_0 + 4883{,}64\hat{a}_1 = 455{,}143, \end{cases} \left| \begin{array}{l} *184 \\ *7 \end{array} \right.$$

$$\begin{cases} 1288\hat{a}_0 + 33856\hat{a}_1 = 3131{,}68, \\ 1288\hat{a}_0 + 34185{,}48\hat{a}_1 = 3186{,}001, \end{cases}$$

$329{,}48\,\hat{a}_1 = 54{,}321.$
$\hat{a}_1 = 0{,}1649.$

$7\hat{a}_0 + 184{*}0{,}1649 = 17{,}02.$
$7\hat{a}_0 = 17{,}02 - 30{,}3416.$
$7\hat{a}_0 = -13{,}3216.$
$\hat{a}_0 = -1{,}9231.$

$\hat{Y}_1 = -1{,}9031 + 0{,}1649X$ – друга економетрична модель.

На основі моделі можна дійти висновку: якщо дохід виросте на 1, то заощадження збільшаться на 0,1649 грн для даної сукупності спостережень.

3.3. Для кожної моделі знайдемо суму квадратів залишків:

$$S_1 = u_1{}'u_1 = \sum(Y_1 - \hat{Y}_1)^2,$$

$$S_2 = u_2{}'u_2 = \sum(Y_2 - \hat{Y}_2)^2,$$

$S_1 = 0{,}0993,$
$S_2 = 0{,}1614.$

3.4. Знаходимо критерій R:

$$R = \frac{S_2}{S_1}; \quad R = \frac{0{,}1614}{0{,}0993} = 1{,}625.$$

Порівняємо цей критерій з табличним значенням критерію Фішера за ступенів свободи $\frac{n-c-2m}{2} = 5$ і рівня довіри $\alpha = 0{,}05$ $F_{табл} = 5{,}05$. Гетероскедастичність відсутня, тому що $R < F_{табл}$.

Оцінка параметрів моделі на основі узагальненого методу найменших квадратів (методу Ейткена)

Приклад 2. Необхідно оцінити параметри економетричної моделі, яка характеризує залежність витрат на харчування від загальних затрат на основі даних, що наведені в табл. 4.27.

Виходячи з особливостей вихідної інформації, можна припустити, що порушується гіпотеза про незмінність дисперсії.

Таблиця 4.27. Вхідні статистичні дані характеристик залежності витрат на харчування від загальних затрат

Номер спостереження	Витрати на харчування	Загальні витрати	$\hat{Y}$	u	u^2
1	2,30	15	2,16	0,14	0,020
2	2,20	15	2,16	0,04	0,002
3	2,08	16	2,20	–0,12	0,015
4	2,20	17	2,25	–0,05	0,002
5	2,10	17	2,25	–0,15	0,022
6	2,32	18	2,29	0,26	0,0007
7	2,45	19	2,34	0,11	0,012
8	2,50	20			
9	2,20	20			
10	2,50	22			
11	3,10	64			
12	2,40	68	2,37	0,13	0,016
13	2,82	72	2,52	0,29	0,085
14	3,04	80	2,68	0,36	0,128
15	2,70	85	2,99	–0,29	0,084
16	3,91	90	3,18	0,76	0,573
17	3,10	95	3,38	–0,28	0,076
18	3,99	100	3,57	0,42	0,178

Розв'язання

1. Ідентифікуємо змінні моделі:

Y – витрати на харчування, залежна змінна;

X – загальні витрати, незалежна змінна;

$$Y = f\,(X, u).$$

2. Перевіримо наявність гетероскедастичності для наведених вихідних даних на основі параметричного тесту Гольдфельда – Квандта.

2.1. Упорядкуємо значення незалежної змінної X від меншого до більшого і відкинемо C значень, які знаходяться всередині впорядкованого ряду:

$$\frac{C}{18} = \frac{4}{15};\ C \approx 4.$$

2.2. На основі отриманих двох сукупностей спостережень (від першого до сьомого включно і від одинадцятого до вісімнадцятого значення) побудуємо дві економетричні моделі за методом 1 МНК.

1-ша модель: $\hat{Y}_1 = 1,475 + 0,046X$;

2-га модель: $\hat{Y}_1 = -0,093 + 0,039X$.

2.3. Визначимо залишки за цими двома моделями:

$u_1 = Y_1 + \hat{Y}_1$;

$u_2 = Y_2 + \hat{Y}_2$;

Залишки та квадрати залишків наведені в табл. 4.27.

2.4. Розрахуємо залишкові дисперсії та знайдемо їх співвідношення R:

$$R = \frac{\sigma^2_{u_2}}{\sigma^2_{u_1}} = \frac{1,14/7}{0,074/7} = 15,41.$$

2.5. Порівняємо критерій R з критичним значенням F-критерію за $\gamma_1 = 5$ і $\gamma_2 = 5$ ступенів свободи і рівня довіри $\alpha = 0,01$ $F = 11$. Оскільки $R > F_{крит}$, вихідні дані мають гетероскедастичність.

3. За наявності гетероскедастичності оцінку параметрів моделі виконаємо методом Ейткена:

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

3.1. Запишемо матриці змінних, які входять в оператор Ейткена:

$$X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 15 & 15 & 16 & 17 & 17 & 18 & 19 & 68 & 72 & 80 & 85 & 90 & 95 & 100 \end{pmatrix}.$$

Визначимо матрицю S^{-1}, користуючись гіпотезою: $\lambda_i = \dfrac{1}{x_i}$, тобто

$$S^{-1} = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \lambda_4 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \lambda_5 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & \lambda_{14} \end{pmatrix}$$

$$S^{-1}=\begin{pmatrix}
0{,}0667 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0{,}0667 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0{,}0625 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0{,}0588 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0{,}0588 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0{,}0555 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0{,}0526 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}05 & 0 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}05 & 0 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}0154 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}0156 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}0148 & 0 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}0139 & 0\\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}0125
\end{pmatrix}$$

Визначимо добутки матриць:

$$X'S^{-1}=\begin{pmatrix}
0{,}0667 & 0{,}0667 & 0{,}0625 & 0{,}0589 & 0{,}0589 & 0{,}0555 & 0{,}0526 & 0{,}05 & 0{,}05 & 0{,}0154 & 0{,}0156 & 0{,}0148 & 0{,}0139 & 0{,}0125\\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{pmatrix}$$

$$X'S^{-1}X = \begin{pmatrix} 0{,}6672 & 18\\ 18 & 833\end{pmatrix}.$$

3.2. Знайдемо обернену матрицю:

$$(X'S^{-1}X)^{-1}= \begin{pmatrix} 3{,}5934 & -0{,}0776\\ -0{,}0776 & 0{,}0029\end{pmatrix}$$

і вектор:

$$X'S^{-1}Y = \begin{pmatrix} 1{,}5998\\ 48{,}04\end{pmatrix}.$$

3.3. Обчислимо вектор оцінок параметрів моделі:

$$\hat{A} = \begin{pmatrix} 3{,}5934 & -0{,}0776\\ -0{,}0776 & 0{,}0029\end{pmatrix} * \begin{pmatrix} 1{,}5998\\ 48{,}04\end{pmatrix} = \begin{pmatrix} 2{,}0187\\ 0{,}0141\end{pmatrix}.$$

Звідси $\hat{a}_0 = 2{,}0187$;

$\hat{a}_1 = 0{,}0141$.

Економетрична модель витрат на харчування запишеться так: $\hat{Y} = 2{,}0187 + 0{,}0141X$.

4. Економічний аналіз характеристик економетричної моделі.

4.1. Коефіцієнт детермінації $R^2 = 0{,}722$. Це означає, що на 72,2% варіація витрат на харчування залежить від варіації загальних витрат.

4.2. Коефіцієнт кореляції $R = \sqrt{R^2} = 0{,}85$ свідчить про досить тісний зв'язок витрат на харчування і загальних затрат.

4.3. Залишкова дисперсія $\sigma_U^2 = 0{,}083$ показує, що розрахункові значення витрат на харчування дуже близькі до фактичних.

4.4. Параметр моделі $\hat{a}_1 = 0{,}0141$ свідчить про те, що збільшення загальних витрат на одиницю сприятиме граничному зростанню витрат на харчування на 0,014 одиниць.

5. Розрахуємо матрицю коваріацій оцінок параметрів моделі:

$$\mathrm{var}(\hat{A}) = \sigma_U^2 (X'S^{-1}X)^{-1} = 0{,}083 \begin{pmatrix} 3{,}5934 & -0{,}0776 \\ -0{,}0776 & 0{,}0029 \end{pmatrix} = \begin{pmatrix} 0{,}298 & -0{,}0064 \\ -0{,}0064 & 0{,}0002 \end{pmatrix}.$$

Діагональні елементи цієї матриці є дисперсіями оцінок параметрів моделі, інші елементи характеризують коваріацію між оцінками.

6. Визначимо стандартні помилки оцінок параметрів і знайдемо їх довірчі інтервали:

$$S_{\hat{a}_j} = \sqrt{\mathrm{var}(\hat{A})} = \sigma_U \sqrt{c_{jj}};$$
$$S_{\hat{a}_0} = \sqrt{0{,}298} = 0{,}546;$$
$$S_{\hat{a}_1} = \sqrt{0{,}0002} = 0{,}014.$$

Для побудови довірчих інтервалів оцінок параметрів моделі знайдемо t-критерій за ступенів свободи $n - m = 12$ і рівня довіри $a = 0{,}05$; $t_{крит} = 2{,}179$.

Довірчі інтервали оцінок:

$$\hat{a}_0 - t_{0,05} S_{\hat{a}_0} \le \hat{a}_0 \le \hat{a}_0 + t_{0,05} S_{\hat{a}_0}.$$

$$2{,}0187 - 1{,}102 \le a_0 \le 2{,}0187 + 1{,}102.$$

$$0{,}917 \le a_0 \le 3{,}121.$$

$$\hat{a}_1 - t_{0,05} S_{\hat{a}_1} \le \hat{a}_1 \le \hat{a}_1 + t_{0,05} S_{\hat{a}_1}.$$

$$0{,}016 \le a_1 \le 0{,}045.$$

Рівень стандартних помилок та довірчі інтервали оцінок параметрів моделі свідчать про те, що отримані оцінки є неефективними та зміщеними.

Тести

1. Визначити умови Гаусса – Маркова:
   а) математичне сподівання випадкових відхилень має дорівнювати нулю;
   б) математичне сподівання випадкових відхилень має бути постійною величиною;
   в) дисперсія випадкових відхилень має бути сталою величиною;
   г) дисперсія випадкових відхилень має дорівнювати нулю;
   д) випадкові відхилення мають бути незалежними одне від одного;
   е) випадкові відхилення повинні бути залежними одне від одного;
   ж) випадковий вектор відхилень має бути незалежним від регресорів;
   з) випадковий вектор відхилень має бути залежним від регресорів;
   и) компоненти випадкового вектора мають мати нормальний закон розподілу.

2. Моделі, для яких виконується умова сталості дисперсії випадкових відхилень, називають:
   а) гомоскедастичними;
   б) гетероскедастичними;
   в) класичними.

3. Для моделей першої групи статистична оцінка параметрів здійснюється шляхом використання:
   а) зваженого методу найменших квадратів;
   б) узагальненого методу найменших квадратів;
   в) звичайного методу найменших квадратів.

4. Тести Глейзера і Гольдфельда – Квандта можуть виявити наявність ознаки гетероскедастичності в моделі лише в разі порушення умови:
   а) коли дисперсії залишків не є сталими величинами;
   б) коли дисперсії залишків є сталими величинами;
   в) компоненти випадкового вектора мають нульові математичні сподівання;
   г) компоненти випадкового вектора мають ненульові математичні сподівання.

5. Визначити алгоритм застосування тесту Гольдфельда – Квандта:
   а) упорядковуємо вхідні дані в порядку спадання значень пояснювальної змінної;
   б) відкинемо *c* середніх спостережень (друга підвибірка), вважаючи, що дисперсія залишків для них постійна;
   в) знайдемо оцінки параметрів для першої моделі (підвибірка 1) з найбільшими значеннями регресора;

г) знайдемо оцінки параметрів для другої моделі (підвибірка 2) з найменшими значеннями регресора;
д) обчислимо суми квадратів залишків для першої та другої моделей;
е) розрахуємо значення критерію Фішера;
ж) розраховане значення критерію Фішера порівняємо з табличним значенням, зробимо відповідні висновки.

6. Визначити алгоритм застосування тесту Глейзера:
   а) знайдемо значення критерію Стьюдента для оцінок параметрів;
   б) розрахуємо значення статистичних оцінок параметрів рівняння регресії за умови, що залежною змінною є значення залишків за абсолютною величиною;
   в) обчислимо значення залишків та запишемо їх за абсолютною величиною;
   г) знайдемо стандартні помилки оцінок параметрів;
   д) розраховані значення критерію Стьюдента порівняємо з табличним значенням, зробимо відповідні висновки.

7. Сутність зваженого МНК полягає в тому, що:
   а) конкретне значення випадкового відхилення має однакову “питому вагу”, незалежно від того одержали його при значенні, де є мала дисперсія, чи при значенні, де наявна велика дисперсія;
   б) конкретне значення випадкового відхилення має відповідну “питому вагу”, залежно від того одержали його при значенні, де є мала дисперсія, чи при значенні, де наявна велика дисперсія.

8. Визначте правильну відповідь.

тест Гольдфельда – Квандта	базується на побудові регресійної функції, що характеризує залежність величини залишків за модулем від пояснювальної змінної *xj*, яка може зумовити зміну дисперсії залишків
тест Глейзера	базується на встановленні кількості піків значень залишків після впорядкування (ранжування) спостережень за *xij*. Якщо для всіх значень змінної *xij* залишки розподіляються приблизно однаково, то дисперсія їх однорідна і гетероскедастичність відсутня
-	базується на встановленні наявності систематичного зв'язку квадратів залишків регресійної моделі, побудованої на основі припущення про відсутність гетероскедастичності

ТЕМА 5

Економетричні моделі динаміки

Основні поняття і попередній аналіз рядів динаміки. Поняття часового ряду. Основні характеристики динаміки часового ряду. Систематичні та випадкові компоненти часового ряду • Перевірка гіпотези про існування тренда • Методи фільтрації сезонної компоненти. Проблеми аналізу сезонності (та/або циклічності). Фільтрація сезонної компоненти за допомогою індексу сезонності. Метод декомпозиції часового ряду • Методи прогнозування часових рядів. Методи соціально-економічного прогнозування. Прогнозування тенденцій часового ряду за середніми характеристиками. Прогнозування тенденцій часового ряду за механічними методами. Прогнозування тенденцій часового ряду за аналітичними методами.

5.1. Основні поняття і попередній аналіз рядів динаміки

5.1.1. Поняття часового ряду

При побудові эконометрічної моделі використовують два типи даних: дані, що характеризують сукупність різних об'єктів у певний момент часу; дані, що характеризують один об'єкт за ряд послідовних моментів часу.

Моделі, побудовані на підставі даних першого типу дістали назву *просторових моделей*. Моделі, побудовані на підставі даних другого типу, називаються *моделями часових рядів*.

Часовий ряд (ряд динаміки) – сукупність значень будь-якого показника за кілька послідовних моментів або періодів часу.

Характерним для часового ряду $y = y_1, y_1, \dots, y_n$ є те, що порядок в послідовності $t_1, t_1, \dots, t_n$ є суттєвим для аналізу, тобто час становить собою один із визначальних чинників. Це відрізняє часовий ряд від випадкової вибірки, де індекси служать лише для зручності ідентифікації.

Складовими ряду спостережень є числові значення показника, які називають ***рівнями ряду***, та моменти або інтервали часу, до яких належать рівні. Часовий ряд (ЧР) можна записати в стислому вигляді:

$$y_t, t = 1, 2, \ldots, n, \tag{5.1}$$

де t – рівновіддалені моменти спостережень (година, доба, місяць, рік тощо).

Довжина ряду – час, що минув від першого до останнього моменту спостереження. Часто довжиною ряду називають кількість рівнів n, які утворюють часовий ряд.

На формування кожного рівня часового ряду справляє вплив велика кількість факторів, які умовно можна поділити на три групи:

а) фактори формування тенденції ряду;
б) фактори формування циклічних коливань ряду;
в) випадкові фактори.

Ряди, у яких рівні коливаються навколо постійної середньої, називаються ***стаціонарними*** (рис. 5.1). Економічні ряди, як правило, нестаціонарні. Більшості з них характерна систематична зміна рівнів з нерегулярними коливаннями, коли піки й спади чергуються з різною інтенсивністю.

Економічні цикли повторюються з різною тривалістю і різною амплітудою коливань.

Рис. 5.1. Стаціонарний часовий ряд процентної маржі банків України

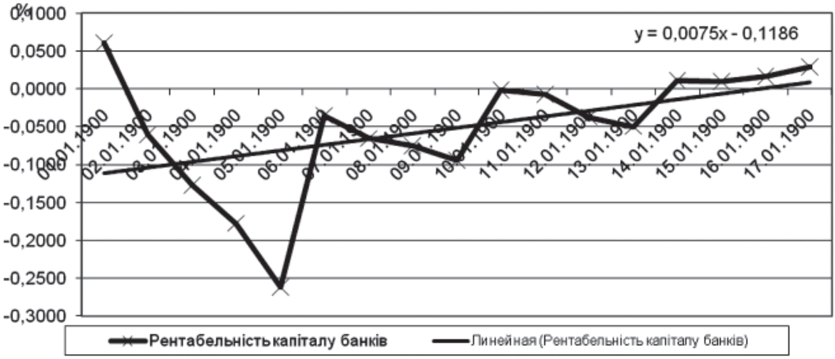


Рис. 5.2. Часовий ряд, що містить тенденцію до зростання

Більшість часових рядів економічних показників має тенденцію, що характеризує сукупну тривалу дію множини чинників на динаміку показника, що вивчається. Усі ці фактори можуть справляти різнонапрямлений вплив на досліджуваний показник. Проте в сукупності вони формують його зростаючу або спадаючу тенденцію. На рис. 5.2. наведено гіпотетичний часовий ряд, що містить тенденцію до зростання.

Досліджуваний показник також може зазнавати циклічних коливань. Ці коливання можуть мати сезонний характер, оскільки економічна діяльність низки галузей економіки залежить від пору року (наприклад, ціни на сільськогосподарську продукцію в літній час вищі, ніж у зимовий, рівень безробіття в курортних містах у зимовий період вищий порівняно з літнім). За наявності більших масивів даних за тривалі проміжки часу можна визначити циклічні коливання, пов'язані із загальною динамікою кон'юнктури ринку. На рис. 5.3. наведений гіпотетичний часовий ряд, що містить тільки сезонну компоненту.

Деякі часові ряди не містять тенденції і циклічної компоненти, а кожний наступний рівень утворюється як сума середнього рівня ряду і деякої (додатної чи від'ємної) випадкової компоненти. Приклад ряду, що містить лише випадкову компоненту, наведений на рис. 5.4.

Зрозуміло, що реальні дані не випливають стовідсотково з якихось моделей, що описані вище. Частіш за все вони містять усі три компоненти. Кожний їх рівень формується в результаті тенденції, сезонних коливань і випадкової компоненти.

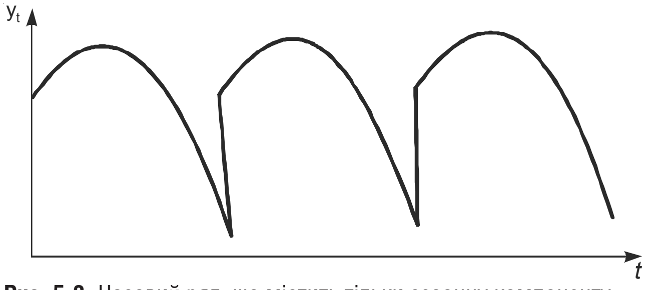


Рис. 5.3. Часовий ряд, що містить тільки сезонну компоненту

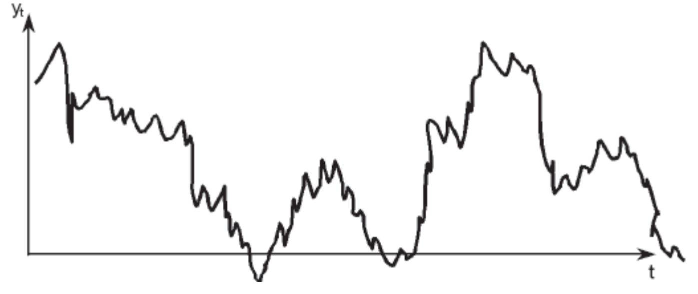


Рис. 5.4. Часовий ряд, що містить випадкову компоненту

Фактичний рівень часового ряду здебільшого можна подати як суму або добуток трендової, циклічної та випадкової компоненти.

Адитивна модель часового ряду – модель, у якій часовий ряд поданий як сума складових компонент.

Мультиплікативна модель часового ряду – модель, у якій часовий ряд поданий як добуток складових компонент.

Основним завданням економетричного дослідження окремого часового ряду є виявлення і кількісне подання кожної з наведених вище компонент з метою використання отриманої інформації для прогнозування майбутніх значень ряду або при побудові моделей взаємозв'язку двох або більше часових рядів.

Залежно від характеру досліджуваних соціально-економічних показників часові ряди поділяють на моментні, інтервальні та похідні.

Моментні часові ряди – часові ряди, утворені показниками, що характеризують економічне явище на певні моменти часу; прикладом такого ряду є дані щодо виданих позичок відділенням Держбанку (табл. 5.1).

Приклад

Розглянемо один із показників функціонування відділення комерційного банку України – величина наданої позички, значення якої впорядковані залежно від дати укладення відповідного договору, зафіксованої в жовтні та листопаді року, який розглядається.

Таблиця 5.1. Моментний часовий ряд

Дата надання позички	01.10	05.10	12.10	23.10	03.11	07.11
Величина наданої позички, тис. грош. од.	3747	3710	3839	3783	3747	3710

Інтервальні часові ряди – якщо рівні часового ряду утворюються шляхом агрегування за певний проміжок (інтервал) часу; приклад наведено в табл. 5.2.

Приклад

Розглянемо один із показників функціонування відділення комерційного банку України – обсяг залучених депозитів, значення якого отримані шляхом визначення суми всіх договорів, укладених протягом кожного місяця поточного року. Упорядкуємо отриманий набір показників залежно від місяця укладення відповідних договорів та подамо у вигляді часового ряду (табл. 5.2).

Таблиця 5.2. Інтервальний часовий ряд

Місяць	Січень	Лютий	Березень	Квітень	Травень
Обсяг залучених депозитів, млн грн	6578	7016	7353	7353	7941

Часові ряди можуть бути створені як абсолютних значень економічних показників, так і з середніх або відносних величин – це *похідні* ряди; приклад такого ряду наведено в табл. 5.3.

Приклад

Розглянемо статистичині дані середньої заробітної плати працівників розглянутої банківської установи за січень-травень поточного року. Зазначені дані отримані на основі використання величин даного показника за місяць за попередні роки функціонування банківської установи, на основі яких розраховані похідні від вихідного часового ряду – середні значення, наведені в табл. 5.3.

Таблиця 5.3. Часовий ряд, утворений із середніх значень показника

Місяць	Січень	Лютий	Березень	Квітень	Травень
Середня заробітня плата працівників банківської установи, грн/міс.	152,2	153,7	165,8	161,6	163,71

5.1.2. Основні характеристики динаміки часового ряду

Для аналізу соціально-економічних показників абсолютні рівні моментних або інтервальних часових рядів, а також рівні середніх величин часто доводиться перетворювати на відносні величини. До найпоширеніших характеристик динаміки розвитку соціально-економічних процесів та їхніх розрахунків належать: абсолютний приріст, коефіцієнти зростання та приросту, темпи зростання та приросту, середні арифметична, хронологічна тощо (табл. 5.4).

Для визначення змін, що відбуваються з досліджуваним явищем, передусім обчислюють швидкість розвитку цього явища за часом. Показником швидкості слугує абсолютний приріст, який характеризує величину зміни показника за інтервал часу між порівнюваними періодами й обчислюється за формулою:

$$\Delta y_i = y_i - y_{i-k}, \qquad (5.2)$$

де y_t – i-й рівень часового ряду (i = 2, 3, ..., n);
k – індекс початкового рівня; k = 1, 2, ..., n – 1 і може бути обраний будь-яким залежно від мети дослідження: за k = 1 отримують ланцюгові показники, за $k = i - 1$ отримують базові показники із базовим початковим рівнем ряду тощо.

Точніше, швидкість зміни показника характеризує приріст за одиницю часу; ця величина має назву середнього абсолютного приросту:

Таблиця 5.4. Характеристики динаміки часового ряду

Характеристики	Розрахункові формули
Абсолютний приріст	$\Delta y_i = y_i - y_{i-k}$
Коефіцієнт зростання	$K_{i(зр)} = \frac{y_i}{y_{i-k}}$
Коефіцієнт приросту	$K_{i(пр)} = \frac{y_i - y_{i-k}}{y_{i-k}}$
Темп зростання	$T_{i(зр)} = \frac{y_i}{y_{i-k}} \cdot 100\% = K_{i(зр)} \cdot 100\%$
Темп приросту	$T_{i(пр)} = T_{i(зр)} - 100\ \%$, або $T_{i(пр)} = \frac{y_i - y_{i-k}}{y_{i-k}} \cdot 100\%$
Середня арифметична	$\bar{y} = \frac{\sum y_i}{n}$
Середня хронологічна	$\bar{y} = \frac{\sum y_i t}{\sum t}$
Середній абсолютний приріст	$\overline{\Delta y}_k = \frac{y_i - y_{i-k}}{k}$
Середній темп зростання	$\bar{T}(зр) = \sqrt[n-1]{\frac{y_n}{y_1}}$
Середній темп приросту	$\bar{T}(пр) = \bar{T}(зр) - 100\%$

$$\overline{\Delta y}_k = \frac{y_i - y_{i-k}}{k}. \tag{5.3}$$

Зокрема, середній абсолютний приріст за весь період спостереження для заданого часового ряду дорівнює:

$$\overline{\Delta y} = \frac{y_n - y_1}{n-1} \tag{5.4}$$

і характеризує середню швидкість зміни часового ряду, де n — індекс останнього спостереження.

Для визначення відносної швидкості зміни економічного явища як одиницю часу використовують відносні показники: коефіцієнти зростання й приросту (якщо ці показники виражені у відсотках, їх називають, відповідно, темпами зростання й приросту). Зазначимо, що в усіх наступних формулах індекс початкового рівня, стосовно якого здійснюють порівняння, також визначають за допомогою індексу k, як і раніше, для показника абсолютного приросту.

Коефіцієнт зростання для i-го періоду обчислюють за формулою:

$$K_{i(\text{зр})} = \frac{y_i}{y_{i-k}}. \tag{5.5}$$

$K_{i(зр)} > 1$, якщо рівень підвищується; $K_{i(зр)} < 1$, якщо рівень зменшується; за $K_{i(зр)} = 1$ рівень не змінюється.

Коефіцієнт приросту дорівнює:

$$K_{i(\text{пр})} = K_{i(\text{зр})} - 1 \quad \text{або} \quad K_{i(\text{ппр}} = \frac{y_i - y_{i-k}}{y_{i-k}}. \tag{5.6}$$

На практиці часто застосовують показники темпу зростання й темпу приросту:

$$T_{i(\text{зр})} = \frac{y_i}{y_{i-k}} \cdot 100\,\%, \tag{5.7}$$

де $T_{i(зр)}$ – темп зростання для i-го періоду;

$$T_{i(\text{пр})} = T_{i(\text{зр})} - 100\,\% \ \text{ або } \ T_{t(\text{пр})} = \frac{y_i - y_{i-k}}{y_{i-k}} \cdot 100\,\%, \tag{5.8}$$

де $T_{i(пр)}$ – темп приросту для і-го періоду. Темп приросту показує, на скільки відсотків рівень одного періоду збільшився стосовно рівня іншого періоду, тобто цей показник характеризує відносну величину приросту у відсотках.

Порівняння абсолютного приросту та темпу приросту за той самий інтервал часу показує, що в реальних економічних процесах уповільнення темпу приросту часто не супроводжується зменшенням абсолютних приростів.

Абсолютне значення одного відсотка приросту визначають як відношення абсолютного приросту Δy_i до темпу приросту у відсотках $T_{i(пр)}$.

Середню швидкість зміни показника, що вивчається, за певний період характеризує також середній темп зростання. Його розраховують за формулою середньої геометричної:

$$\overline{T}_{(\text{зр})} = \sqrt[n-1]{T_{1(\text{зр})} \cdot T_{2(\text{зр})} \cdot \ldots \cdot T_{n(\text{зр})}} = \sqrt[n-1]{\frac{y_n}{y_1}} \cdot 100\,\%, \tag{5.9}$$

де $T_{1(зр)}, T_{2(зр)}, .., T_{n(зр)}$ – середні темпи зростання за окремі інтервали часу.

Відповідно середній темп приросту визначають як

$$\overline{T}_{(\text{пр})} = \overline{T}_{(\text{зр})} - 100\,\% . \tag{5.10}$$

Показник середнього темпу зростання, обчислюваний за формулою середньої геометричної (5.9), має суттєві недоліки, оскільки ґрунтується на зіставленні останнього та початкового рівнів часового ряду, проміжні рівні до уваги не беруться. У разі суттєвого коливання рівнів використання середнього геометричного темпу зростання для статистичного аналізу може призвести до серйозних помилок, унаслідок чого реальна тенденція часового ряду буде викривлена.

Сучасні способи розрахунків середнього темпу зростання певною мірою позбавлені недоліків середньої геометричної. Наприклад, для розрахунків середнього темпу зростання пропонується використовувати формулу:

$$\overline{T}_{(\text{зр})} = \sqrt[n-1]{\frac{\hat{y}_n}{\hat{y}_1}} , \tag{5.11}$$

де $\hat{y}_1$, $\hat{y}_n$ – згладжені за рівнянням тренду (рівнянням кривої зростання) перший та останній рівні часового ряду.

Якщо тенденція часового ряду не змінюється, використовують характеристику середнього рівня ряду. В інтервальному ряду динаміки з однаково розташованими в часі рівнями середній рівень ряду обчислюють за формулою простої середньої арифметичної (тут і далі додавання ведеться за всіма періодами спостережень):

$$\overline{y} = \frac{\sum y_i}{n} , \tag{5.13}$$

де t – кількість періодів часу, для яких значення рівня y_i не змінюється.

Для моментального ряду з однаково розташованими в часі рівнями середню хронологічну розраховують за формулою:

$$\overline{y} = \frac{1/2\, y_1 + y_2 + y_3 + \ldots + y_{n-1} + 1/2\, y_n}{n-1} , \tag{5.14}$$

де n – кількість рівнів ряду.

Таблиця 5.5. Основні статистичні характеристики випадкової вибірки

Характеристики	Оцінки вибіркових значень
1. Середні значення:	
– арифметичне	$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$
– геометричне	$y_G = \sqrt[n]{y_1 \cdot y_2 \cdot ... y_n} = (y_1 \cdot y_2 \cdot ... y_n)^{1/n}$
– гармонійне	$\frac{1}{y_H} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{y_i}$
2. Дисперсія	$\sigma^2 = \frac{\sum_{i-1}^{n}(y_1 - \bar{y})^2}{n-1}$ (незміщена оцінка)
Середньоквадратичне відхилення (СКВ)	$\sigma = \sqrt{\sigma^2}$
3. Середнє абсолютне лінійне відхилення (*MAD*)	$MAD = \frac{\sum_{i=1}^{n}\lvert y_i - \bar{y}\rvert}{n}$
4. Початкові моменти: другого, третього, четвертого порядку	$v_2 = \frac{\sum_{i=1}^{n} y_i^2}{n}$; $v_3 = \frac{\sum_{i=1}^{n} y_i^3}{n}$; $v_4 = \frac{\sum_{i=1}^{n} y_i^4}{n}$
5. Моменти центральні:	
– другого	$m_2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n}$;
– третього	$m_3 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^3}{n}$
– четвертого порядку	$m_4 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^4}{n}$
9. Коефіцієнт асиметрії:	$A = \frac{m_1}{m_2^{3/2}}$
– його незміщена оцінка	$A_H = \left[((n-1)n)^{1/2} / (n-2)\right] A$
– СКВ	$S_H = \left[6n(n-1)/(n-2)(n+1)(n+3)\right]^{1/2}$

Продовження табл. 5.5

Характеристики	Оцінки вибіркових значень
10. Показник ексцесу:	$E = \frac{m_4}{(m_2)^2} - 3$
– його незміщена оцінка	$E_H = [(n-1)/(n-2)(n-3)] \cdot [(n+1)E + 6]$
– СКВ	$S_E = [24n(n-1)^2 / (n-3)(n-2)(n+3)(n+5)]^{1/2}$
11. Коефіцієнти варіації:	
– за розмахом	$\frac{R}{\overline{y}}$
– за середнім абсолютним лінійним відхиленням	$\frac{MAD}{\overline{y}}$
– за СКВ	$\frac{\sigma}{y}$
– медіана	$m_e = \frac{y_n}{2}$
– мода	m_0 – характеризує величину, яка найчастіше спостерігається
– мінімальне значення ряду	$y_{\min}$
– максимальне значення ряду	$y_{\max}$
розмах	$R = y_{\max} - y_{\min}$

Середню хронологічну для моментального часового ряду з неоднаково розташованими в часі рівнями розраховують за формулою

$$\overline{y} = \frac{(y_1 + y_2)t_1 + (y_2 + y_3)t_2 + \ldots + (y_{n-1} + y_n)t_{n-1}}{2\sum t}, \qquad (5.15)$$

де n – кількість рівнів ряду, а t – період часу, що відокремлює 1-й рівень ряду від $(t + 1)$-го рівня.

Засоби описової статистики та обчислення їх за даними вибіркових спостережень наведено в табл. 5.5.

5.1.3. Систематичні та випадкові компоненти часового ряду

При дослідженні часового ряду виникає питання існування систематичної та випадкової компонент, які дозволяють ідентифікувати природу даного ряду, побудувати адекватну математичну модель та розрахувати прогнозні значення досліджуваного показника. Отже, існує дві основні *мети аналізу* часових рядів (рис. 5.5).

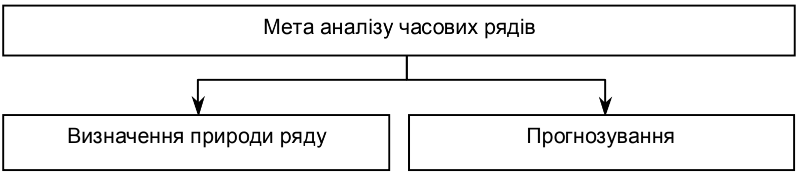


Рис. 5.5. Складові мети аналізу часових рядів

Обидві ці цілі вимагають, щоб модель ряду була ідентифікованою і більш-менш формально описаною. Як тільки модель визначена, з її допомогою можна інтерпретувати дані, що розглядаються. Не звертаючи уваги на глибину розуміння та справедливість теорії, можна екстраполювати потім ряд на основі знайденої моделі, тобто передбачати його майбутні значення.

Розгляд реальних статистичних даних дозволяє дійти до висновку, що типові часові ряди можуть бути подані як декомпозиція з чотирьох структуротворних елементів (рис. 5.6).

Часовий ряд y_t, $t = 1, 2, ..., n$ можна подати у вигляді суми складових:

$$Y_t = U_t + S_t + V_t + E_t\,,\ t = 1, 2, ..., n \qquad (5.18)$$

або різноманітних комбінацій окремих функцій. Однак завжди припускають обов'язкову наявність випадкової складової.

Моделі тренду й сезонності (тренд-сезонні) можуть відображати як відносно постійну сезонну хвилю (цикл), так і динамічно змінювану залежно від тренду (рис. 5.7). Перша форма (5.19) – (5.21) належить до ***адитивних***, друга ($Y_t = U_t \cdot S_t \cdot V_t \cdot E_t$, $t = 1, 2, .., n$) – до ***мультиплікативних*** моделей.

Систематичні компоненти часового ряду – тренд, сезонна і циклічна компоненти.

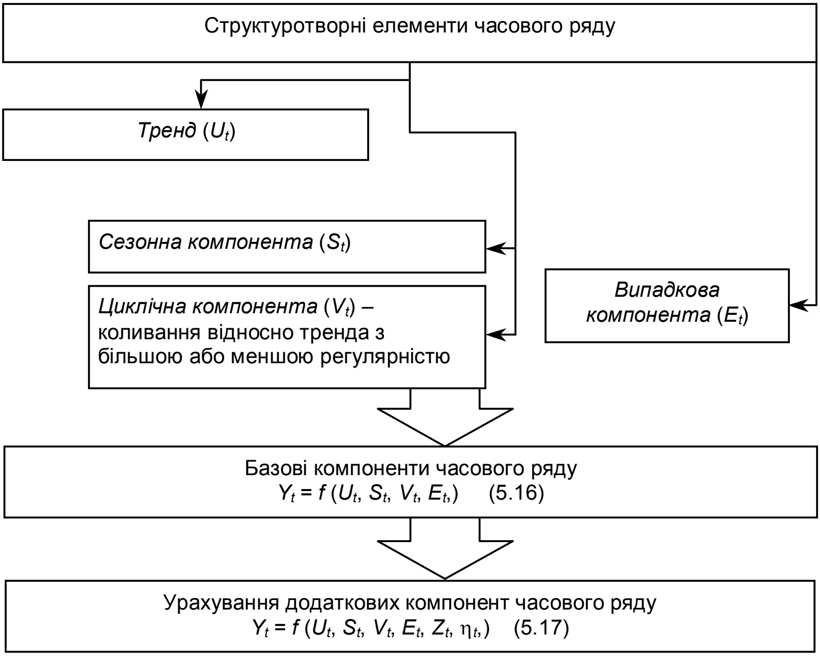


Рис. 5.6. Послідовність визначення елементів часового ряду[1]

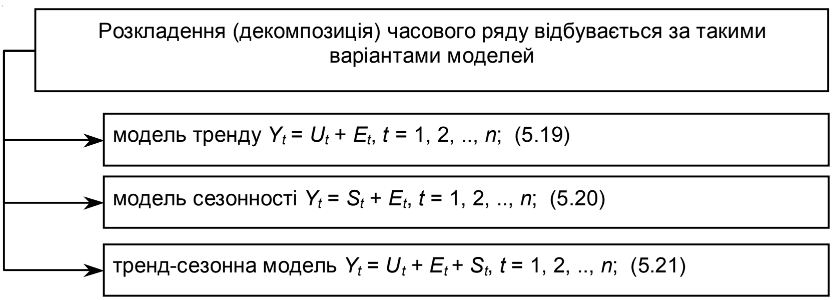


Рис. 5.7. Види моделей часового ряду

[1] Z_t – компонента, що забезпечує зіставність елементів динамічного ряду, η_t – керівна компонента, за допомогою якої впливають на значення членів динамічного ряду для формування в майбутньому бажаної траєкторії.

Випадкова компонента (залишки, помилки) E_t – складова частина часового ряду, що залишається після вилучення з нього систематичних компонент.

Оскільки випадкові відхилення неминуче супроводжують будь-яке макроекономічне явище, випадкова компонента є обов'язковою складовою часового ряду і визначає стохастичний (несистематичний) характер його елементів Y_t. Якщо побудована «якісна» модель прогнозування, то E_t є близькою до нуля, випадковою, незалежною, нормально розподіленою компонентою, інакше модель вважається поганою.

Будь-який ряд можна описати у вигляді однієї з таких складових або комбінації з них.

Завдання розкладення часового ряду полягає в аналізі чинників, що впливають на значення його рівнів, у вирізненні з-поміж них головних і другорядних (випадкових), а потім серед головних – еволюційних та періодичних (сезонних тощо).

Еволюційні чинники визначають загальний напрям розвитку економічного показника, провідну його тенденцію.

Тенденція, тренд – невипадкова складова часового ряду, яка змінюється повільно і описується за допомогою певної функції U_t, яку називають функцією тренду.

Тренд відображає вплив на економічний показник деяких постійних чинників, дія яких акумулюється в часі.

Тренд у широкому розумінні – будь-який упорядкований процес, що відрізняється від випадкового, тобто функцію $f(t)$; зміщення у часі математичного сподівання.

Відносно $U(t)$ припускається, що це певна гладка функція, ступінь гладкості якої заздалегідь не відомий.

Під ступенем гладкості розуміють мінімальний ступінь поліному, що найкраще згладжує компоненту $U(t)$.

З-поміж чинників, що визначають *регулярні коливання* ряду, розрізняють такі:

1. *Сезонні*, що відповідають коливанням, які мають періодичний або близький до нього характер упродовж одного року. Сезонні чин-

ники можуть охоплювати причини, пов'язані з діяльністю людини (свята, відпустки, релігійні традиції тощо). Так, у ряду щомісячних даних слід очікувати наявності сезонних коливань із періодом 12, у квартальних рядах – із періодом 4. Результат дії сезонних чинників моделюють за допомогою функції S_t.

2. *Циклічні* (кон'юнктурні) коливання схожі на сезонні, але виявляються на більш тривалих інтервалах часу. Циклічні коливання пояснюються дією довготермінових циклів економічної, демографічної або астрофізичної природи. Наприклад, за багаторічними спостереженнями активність сонця має циклічність у 10,5–11 років, причому сплески сонячної радіації впливають на врожайність зернових культур, репродуктивну властивість тварин тощо. Отже, динаміка показника міситиме характерні зміни, що повторюються з однаковою циклічністю.

Тренд, сезонна й циклічна компоненти не є випадковими, тому їх називають ***систематичними компонентами часового ряду***.

Випадкові чинники не підлягають вимірюванню, але неминуче супроводжують будь-який економічний процес і визначають стохастичний характер його елементів. До випадкових чинників можна віднести помилки вимірювання, випадкові збурення тощо. Деякі часові ряди, наприклад *стаціонарні*, не мають тенденції та сезонної складової, кожен наступний рівень їх утворюється як сума середнього рівня ряду і випадкової (додатної або від'ємної) компоненти. Результат впливу випадкових чинників позначається випадковою компонентою Et, яку обчислюють як залишок або похибку, що залишається після вилучення з часового ряду систематичних компонент. Це не означає, що така складова не підлягає подальшому аналізу, оскільки містить лише хаос.

Розрахунок складових часвого ряду необхідно здійснювати в таких випадках:

1) якщо здійснюється операція фільтрації, тобто окремий розрахунок компонент U_t, S_t, V_t, E_t, Z_t;
2) якщо здійснюється необхідно процедура згладжування, тобто розрахунок значень тренду разом із сезонною складовою, тобто $U_t + S_t$, отриманий при цьому ряд називають тренд-сезонним часовим рядом.

Побудова адитивної та мультипликативної моделей полягає в розрахунку значень T, S і E для кожного рівня ряду.

Послідовність етапів побудови економіко-математичної моделі часового ряду можна представити у вигляді алгоритму (рис. 5.8).

Процес побудови моделі включає наступні кроки

Вирівнювання вихідного ряду за методом змінної середньої.

↓

Розрахунок значень сезонної компоненти *S*.

↓

Видалення сезонної компоненти із вихідних рівнів ряду й отримання вирівняних даних ($T + E$) у адитивній або ($T \cdot E$) у мультипликативній моделі.

↓

Аналітичне вирівнювання рівнів ($T + E$) або ($T \cdot E$) та розрахунок значень T з використанням отриманого рівняння тренду.

↓

Розрахунок отриманих за моделлю значень ($T + E$) або ($T \cdot E$)

↓

Розрахунок абсолютних і/або відносних помилок. Якщо отримані значення помилок не містять автокореляції, ними можна замінити вихідні рівні ряду і у подальшому використовувати часовий ряд помилок E для аналізу взаємозв'язку вихідного ряду й інших часових рядів.

Рис. 5.8. Алгоритм побудови адитивної та мультиплоікативної моделі часового ряду

Приклад

Розглянемо часовий ряд даних рентабельності капіталу банків України, розрахованих за кожен квартал з 2009 по 2013 рік на основі даних, взятих на офіційному сайті Світого банку. Значення рентабельності капіталу банків України впорядковані залежно від часового періоду, до якого належать, і розташовані у відповідному порядку.

Розглянемо приклад фільтрації компонент деякого зазначеного часового ряду, зображеного на рис. 5.9.

Нехай нам відомі відфільтровані компоненти ряду, що графічно зображені на рис. 5.10 а—в.

Урахування трендової і циклічної компонент часового ряду рентабельності капіталу банків дозволяє побудувати таке теоретичне рівняння:

$$ROE_t = \alpha + \beta \cdot ROE_{t-m} + F(t) \cdot S =$$
$$= 0{,}002 + 0{,}31 \cdot ROE_{t-4} + 0{,}60 \cdot (0{,}01t - 0{,}16)(-0{,}24)^{d_1} 0{,}98^{d_2} 1{,}26^{d_3} 1{,}99^{d_4}, \quad (5.22)$$

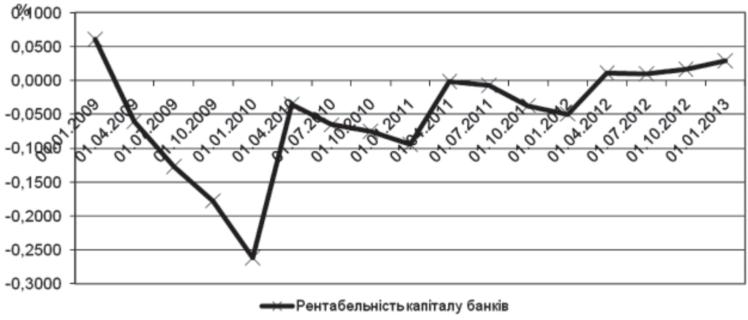


Рис. 5.9. Часовий ряд рентабельності капіталу банків

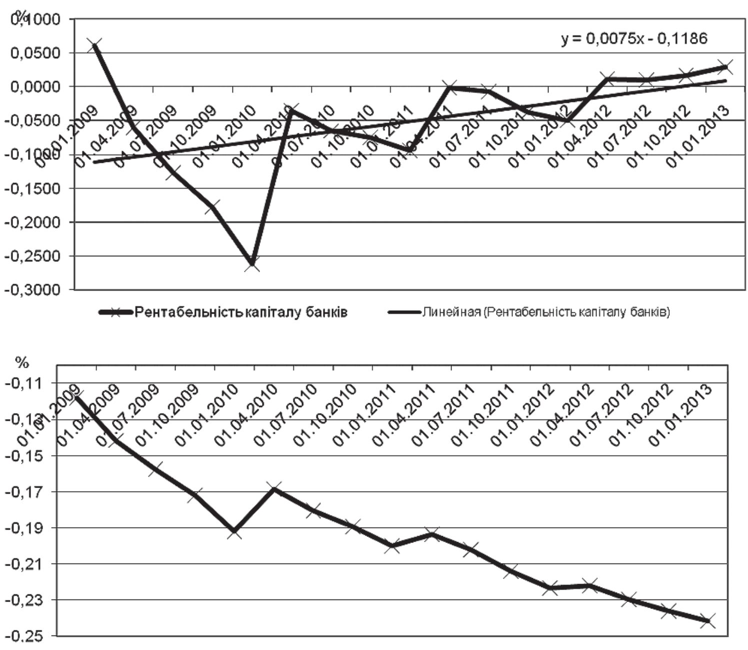


Рис. 5.10. Трендова та циклічна складові часового ряду рентабельності капіталу банків: а – трендова компонента: визначена шляхом побудови теоретичного рівняння парної лінійної регресії (методом найменших квадратів), яке описує основну зростаючу тенденцію розглянутого часового ряду;
б – циклічна компонента: визначена шляхом різниці від кожного рівня часового ряду рентабельності капіталу банків України трендової складової, тобто значень, розрахованих шляхом підстановки в рівняння $y = 0{,}0075 \cdot x - 0{,}1186$ замість x послідовно значень 1, 2, 3 і т.д., які відповідають часовим періодам аналізу.

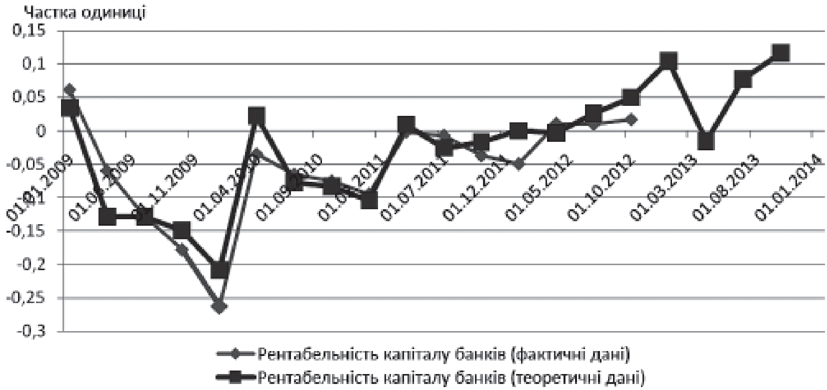


Рис. 5.11. Зіставлення фактичних і прогнозних даних часового ряду рентабельності капіталу банків

де ROE_t – рентабельність капіталу банків в момент часу t;

d_1 (d_2, d_3, d_4) – індикатор першого (відповідно, другого, третього і четвертого) кварталу, який набуває значення: «1», якщо часовий період розрахунку належить до першого (відповідно, другого, третього і четвертого) кварталу, «0» – в іншому випадку.

Спроба визначити тренд за вихідним часовим рядом дуже часто дає не зовсім адекватний результат (рис. 5.11), якщо порівняти рівняння тренду реального та отриманого (розрахованого на основі використання рівняння 5.22 шляхом підстановки змінної t та відповідного кварталу.

Тому проблема адекватної декомпозиції часового ряду є досить серйозною.

5.2. Перевірка гіпотези про існування тренду

За наявності у часовому ряду тенденції й циклічних коливань значення кожного наступного рівня ряду залежать від попередніх. Кореляційну залежність між послідовними рівнями часового ряду називають автокореляцією рівнів ряду.

Кількісно її можна виміряти за допомогою лінійного коефіцієнта корреляції між рівнями вихідного часового ряду і рівнями цього ряду, зсунутих на декілька кроків у часі.

Формула розрахунку коефіцієнта автокореляції має вигляд:

$$r_1 = \frac{\sum_{t=2}^{n}\left(y_t - \overline{y}_1\right)\left(y_{t-1} - \overline{y}_2\right)}{\sqrt{\sum_{t=2}^{n}\left(y_t - \overline{y}_1\right)^2 \sum_{t=2}^{n}\left(y_{t-1} - \overline{y}_2\right)^2}}, \qquad (5.23)$$

де $\overline{y}_1 = \frac{1}{n-1}\sum_{t=2}^{n} y_t, \qquad \overline{y}_2 = \frac{1}{n-1}\sum_{t=2}^{n} y_{t-1}.$

Цю величину називають коефіцієнтом автокореляції рівнів ряду першого порядку, так як він вимірює залежність між сусідніми рівнями ряда t і y_{t-1}.

Аналогічно можна визначити коефіцієнти автокореляції другого і більш високих порядків. Так, коефіцієнт автокореляції другого порядку характеризує тісноту зв'язку між рівнями y_t і y_{t-2} і визначається за формулою:

$$r_1 = \frac{\sum_{t=3}^{n}(y_t - \overline{y}_3)(y_{t-2} - \overline{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \overline{y}_3)^2 \sum_{t=3}^{n}(y_{t-2} - \overline{y}_4)^2}}, \qquad (5.24)$$

де $\overline{y}_3 = \frac{1}{n-2}\sum_{t=3}^{n} y_t\,, \ \overline{y}_4 = \frac{1}{n-2}\sum_{t=3}^{n} y_{t-2}$ (5.25)

Під час досліджень економічних показників результат впливу факторів на їх значення простежується не відразу, а через певний проміжок часу, що викликає необхідність введення лага.

Лаг – це запізнення впливу факторів на досліджуваний показник. Математично виражається як кількість періодів, за якими розраховується коефіцієнт автокореляції.

Із збільшенням лага кількість спостережень, за якими розраховується коефіцієнт автокореляції, зменшується. Для забезпечення статистичної достовірності коефіцієнтів автокореляції необхідно використовувати наступне правило – максимальний лаг має бути не більше $\frac{n}{4}$.

Властивості коефіцієнта автокореляції.

1. Він базується за аналогією з лінійним коефіцієнтом автокореляції і таким чином характеризує тісноту лише лінійного зв'язку поточного і попереднього рівнів ряду. Тому за коефіцієнтом автокореляції можна судити про наявність лінійної (або близької до лінейної) тенденції. Для деяких часових рядів, що мають сильну нелінійну тенденцію (наприклад, параболу другого порядку або експоненту), коефіцієнт автокореляції рівней вихідного ряду може наближуватись до нуля.
2. За знаком коефіцієнта автокореляції не можна робити висновок про зростаючу чи спадаючу тенденцію у рівнях ряду. Більшість часових рядів економічних даних містять додатню автокореляцію рівнів, однак при цьому можуть мати спадаючу тенденцію.

Автокореляційна функція часового ряду – послідовність коефіцієнтів автокореляції рівнів першого, другого тощо порядків.

Корелограма – графік залежності значень автокореляційної функції від величини лага (порядка коефіцієнта автокореляції).

Аналіз автокореляційної функції і корелограми дозволяє визначити лаг, при якому автокореляція найбільш висока, а отже, і лаг, при якому зв'язок між поточним і попередніми рівнями ряду найбільш тісна, тобто за допомогою аналізу автокореляційної функції і корелограми можна виявити структуру ряду.

Якщо найбільш високим виявився коефіцієнт автокореляції першого порядку, досліджуваний ряд містить тільки тенденцію. Якщо найбільш високим виявився коефіцієнт автокореляції порядку τ, то ряд містить циклічні коливання з періодичністю в τ моментів часу. Якщо жоден із коефіцієнтів автокореляції не є значимим, можна зробити одне з двох припущень відносно структури цього ряду: або ряд не містить тенденції і циклічних коливань, або ряд містить сильну нелінійну тенденцію, для виявлення якої треба провести додатковий аналіз. Тому коефіцієнт автокореляції рівнів і автокореляційну функцію доцільно використовувати для виявлення у часовому ряді наявності чи відсутності трендової компоненти і циклічної (сезонної) компоненти.

5.3. Методи фільтрації сезонної компоненти

5.3.1. Проблема аналізу сезонності (та/або циклічності)

Проблема аналізу сезонності (та/або циклічності) полягає в дослідженні сезонних коливань й у вивченні зовнішнього циклічного механізму, який породжує. Для дослідження суто сезонних коливань необхідно виокремити з часового ряду y_t сезонну компоненту s_t і потім аналізувати її динаміку. Більшість методів фільтрації побудована так, що попередньо виокремлюється тренд, а потім сезонна компонента. Тренд у чистому вигляді необхідний і для аналізу динаміки сезонної хвилі. Оскільки індекси сезонності сезонної хвилі величини безрозмірні і не змінюються з року в рік, то їх можна використовувати для визначення рівня сезонності в часовому ряду. У разі використання квартальних даних їх буде чотири, а місячних спостережень – 12.

Для повного дослідження тренд-сезонного часового ряду потрібно розв'язати сукупність завдань у такій послідовності (рис. 5.12).

Під час дослідження сезонної хвилі s_t найчастіше припускають, що вона не змінюється з року в рік, тобто $s_{i+m \cdot j} = s_{ij}$, $i + m \leq k$. Насправді таке припущення далеке від дійсності, принаймні для більшості еко-

Рис. 5.12. Схема алгоритму дослідження тренд-сезонного часового ряду

номічних процесів. Для сезонної хвилі характерна зміна з часом як її розмаху, так і форми. Зрештою, виникає потреба в аналізі та передбаченні змін сезонної хвилі.

Найбільш поширеними методами фільтрації є ітераційні та гармонічного аналізу.

5.3.2. Фільтрація сезонної компоненти за допомогою індексу сезонності

Найпростішим способом, який характеризує коливання рівнів досліджуваного показника, є розрахунок питомої ваги кожного рівня в загальному річному обсязі, або індексу сезонності.

Часові ряди з інтервалом менше року (місяць, квартал), як правило, містять сезонність. Сезонна компонента s_t має період m: $s_{t+m} = s_t$ (m = 12 для ряду місячних даних; $m = 4$ – для ряду квартальних даних). Крім того, відомо, що m кратне n, тобто $n = k \cdot m$, k – ціле число. Очевидно, якщо m – кількість місяців або кварталів у році, то k – кількість років, представлених у часовому ряду $\{y_t\}$. Тому вхідні дані тренд-сезонного часового ряду часто подають у вигляді матриці $\{y_{ij}\}$ розміру [k х m].

Розглянемо таку модель:

$$y_{ij} = u_{ij} \cdot I_j + \varepsilon_{ij}, \tag{5.26}$$

де u_{ij} – «річна» складова (тренд);
I_j – індекс сезонності, або стала пропорційності для j-го кварталу (місяця), яка є безрозмірною величиною та не змінюється з року в рік.

Індекс сезонності I_j характеризує ступінь відхилення рівня сезонного часового ряду від ряду середніх *ui* (тренду) або, інакше кажучи, ступінь коливань відносно 100%.

Наближені оцінки індексів обчислюють як:

$$I_j = \frac{\sum_{i=1}^{k} I_{ij}}{k} \text{ або } I_j = \frac{\sum_{i=1}^{k} I_{ij}}{k} \cdot 100\%, \tag{5.27}$$

де $$I_{ij} = \frac{y_{ij}}{\bar{y}_i} \text{ та } \bar{y}_i = \frac{\sum_{j=1}^{m} y_{ij}}{m}. \tag{5.28}$$

Якщо відомі оцінки тренду $\hat{u}_{ij}$ і сезонної компоненти $\hat{s}_{ij}$ в адитивній моделі, то I_{ij} можна оцінити точніше:

$$I_{ij} = \frac{\hat{u}_{ij} + \hat{s}_{ij}}{\hat{u}_{ij}} = \frac{\hat{y}_{ij}}{\hat{u}_{ij}}. \tag{5.29}$$

Останнє свідчить про можливості оцінювання рівня сезонності незалежно від того, яку модель розглядають – адитивну чи мультиплікативну. Недоліком цього підходу є те, що він не враховує наявності випадкових коливань і тенденцію зміни середньорічного рівня й сезонної хвилі.

5.3.3. Метод декомпозиції часового ряду

Загальна процедура методу для адитивної або мультиплікативної моделей майже однакова. Спочатку виявляють та прогнозують кожну компоненту окремо (етап декомпозиції), а потім отримують загальний прогноз шляхом певного об'єднання отриманих результатів.

Побудову прогнозової адитивної або мультиплікативної тренд-сезонної моделі здійснюють за таким алгоритмом (рис. 5.13).

Зазначимо, що для реалізації ***третього етапу*** знаходять середні значення $\overline{s_j}$ для кожного періоду j:

$$\overline{s_j} = \frac{\sum_{j=1}^{k} s_{ij}}{k}, \quad j = 1,2,\ldots,m \tag{5.30}$$

і середнє сезонне значення $\bar{s}$

$$\bar{s} = \sum_{j=1}^{m} \bar{s}_j. \tag{5.31}$$

При цьому припускають, що сезонні впливи за весь річний цикл гасять одне одного, тобто $\sum_{j=1}^{m} \bar{s}_j = 0$ для адитивної моделі та $\sum_{j=1}^{m} \bar{s}_j = m$ для мультиплікативної моделі. Якщо ці умови не виконуються, то середні оцінки сезонної компоненти ($\bar{s}_j$) коригують.

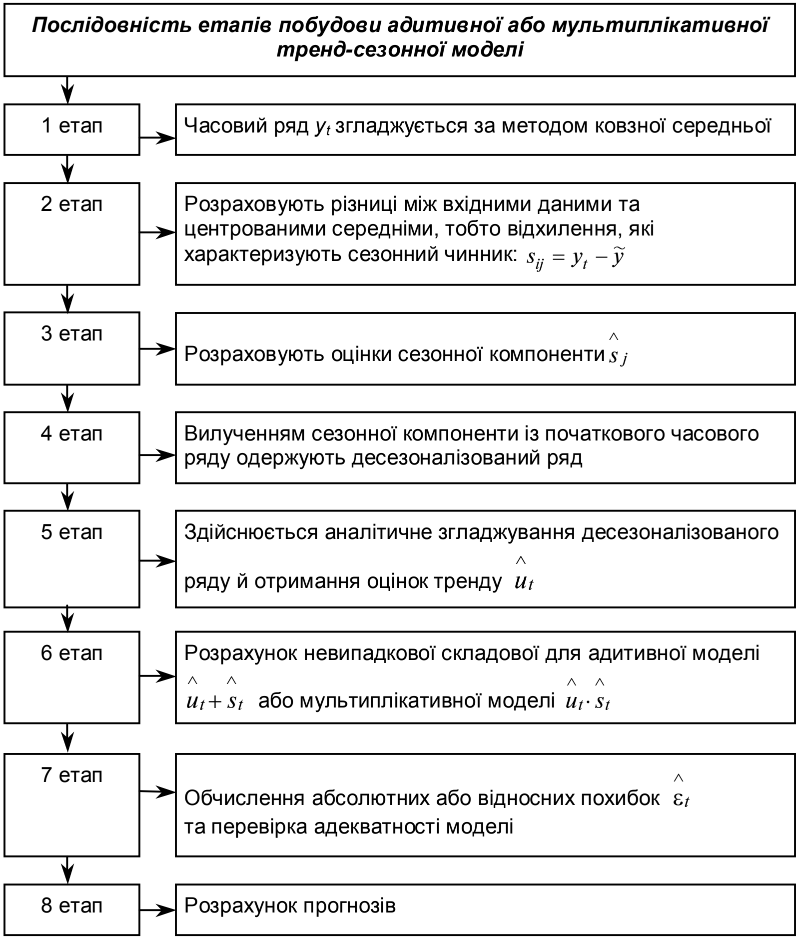


Рис. 5.13. Етапи побудови адитивної або мультиплікативної тренд-сезонної моделі

Для адитивної моделі відкоригована оцінка сезонної компоненти вимірюється в абсолютних величинах і дорівнює:

$$\hat{s}_j = \bar{s}_j - a, \quad a = \frac{\bar{s}}{m}.$$

Для мультиплікативної моделі це значення таке:

$$\hat{s}_j = \bar{s}_j \cdot a, \quad a = \frac{m}{\bar{s}}.$$

Детальніше побудову моделі розглянемо на прикладі.

Приклад

Побудувати прогноз рентабельності капіталу банків, користуючись декомпозиційним аналізом часових рядів. Оцінити ймовірність прогнозу.

У табл. 5.6 у стовпчику 3 відображено поведінку ряду впродовж п'яти років на підставі квартальних даних. Графічний аналіз динаміки наданих кредитів (рис. 5.14) вказує на наявність лінійного сезонно-адитивного або мультиплікативного тренду. Розрахуємо його компоненти.

Спочатку розглянемо *мультиплікативну модель* $Y_t = T_t \cdot S_t \cdot \varepsilon_t$.

Крок 1. Вирівнюємо початкові рівні ряду методом ковзної середньої з вікном згладжування 4 (згладжування ефекту квартального впливу на тенденцію ряду). Узгодимо згладжені значення з фактичними моментами часу, для чого знайдемо середні значення із двох послідовних ковзних середніх – центровані ковзні середні (стовп. 4 табл. 5.6).

Крок 2. Знайдемо оцінки сезонної хвилі для кожного кварталу. Для цього шляхом віднімання від рівнів фактичного ряду (Y_t) центрованих ковзних середніх (T_t) розрахуємо сезонну й випадкову компоненти (стовп. 5 табл. 5.6) і за їхніми значеннями проведемо в допоміжній табл. 5.7 виокремлення сезонної складової. У моделях із сезонною компонентою припускається, що сезонні впливи за кожен річний період взаємно погашаються. У мультиплікативній моделі це виражається в тому, що сума значень сезонної компоненти за всіма кварталами має дорівнювати чотирьом. Для цієї моделі маємо:

–0,249 + 1,040 + 1,336 + 2,101 = 4,228.

Визначимо коефіцієнт коригування: $\alpha = \hat{s}_j = \bar{s}_j \cdot a = 0{,}946$.

Розрахуємо скориговані оцінки сезонної компоненти як різницю між її середньою оцінкою та коефіцієнтом коригування: $\hat{s}_j = \bar{s}_j \cdot a$. Оцінки $\hat{s}_t$ зведено в табл. 5.6 (стовп. 6).

Крок 3. Вилучимо вплив сезонної компоненти із ряду y_t за формулою $y_t / \hat{s}_t = v_t \cdot \varepsilon_t$ (стовп. 7 табл. 5.6).

Крок 4. Визначимо оцінку тренду $\hat{v}_t$. Для цього проведемо аналітичне згладжування ряду $v_t \cdot \varepsilon_t$ за допомогою лінійного тренду. Остаточно отримаємо:

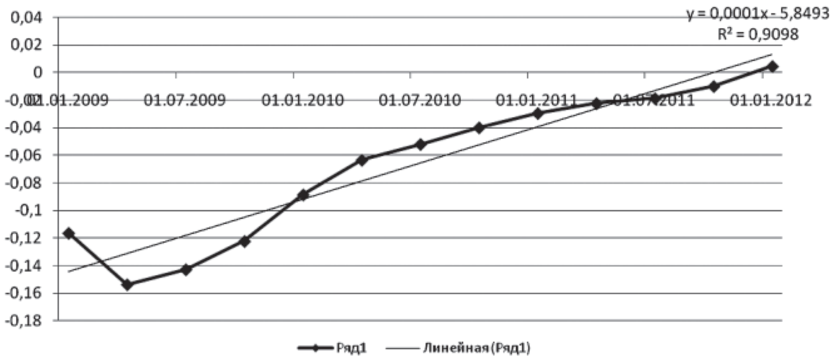


Рис. 5.14. Згладжування ряду за допомогою лінійного тренду

Підставляючи в отримане рівняння значення $t = 1, 2, \ldots 20$ знайдемо теоретичні рівні тренду як для кожного заданого моменту часу, так і для прогнозованого періоду випередження (стовп. 8 табл. 5.6).

Крок 5. Знайдемо теоретичні рівні ряду за формулою $\hat{v}_t \cdot \hat{s}_t$ (стовп. 9 табл. 5.6).

Таблиця 5.6. Розрахунок оцінок компонент мультиплікативної моделі декомпозиції часового ряду рентабельності капіталу банку

Рік	Квартал	Обсяги наданих кредитів	Ковзні середні	Сезонна та випадкова компоненти	Оцінка сезонної компоненти	$y_t / \hat{s}_t = v_t$	Лінійний тренд	$\hat{v}_t \cdot \hat{s}_t$
01.01.2009	1	0,061				–0,144	–0,236	0,034
01.04.2009	2	–0,060	–0,076			–0,131	0,984	–0,129
01.07.2009	3	–0,127	–0,157	–0,116	1,094	–0,118	1,264	–0,149
01.10.2009	4	–0,178	–0,150	–0,154	1,156	–0,105	1,988	–0,208
01.01.2010	1	–0,262	–0,135	–0,143	1,836	–0,092	–0,236	0,022
01.04.2010	2	–0,035	–0,109	–0,122	0,286	–0,079	0,984	–0,077

Продовження табл. 5.6

Рік	Квартал	Обсяги наданих кредитів	Ковзні середні	Сезонна та випадкова компоненти	Оцінка сезонної компоненти	$y_t / \hat{s}_t = v_t$	Лінійний тренд	$\hat{v}_t \cdot \hat{s}_t$
01.07.2010	3	–0,065	–0,068	–0,088	0,738	–0,066	1,264	–0,083
01.10.2010	4	–0,075	–0,059	–0,063	1,188	–0,052	1,988	–0,104
01.01.2011	1	–0,095	–0,045	–0,052	1,822	–0,039	–0,236	0,009
01.04.2011	2	–0,002	–0,035	–0,040	0,038	–0,026	0,984	–0,026
01.07.2011	3	–0,007	–0,024	–0,029	0,244	–0,013	1,264	–0,017
01.10.2011	4	–0,037	–0,021	–0,022	1,663	0,000	1,988	0,000
01.01.2012	1	–0,050	–0,017	–0,019	2,646	0,013	–0,236	–0,003
01.04.2012	2	0,011	–0,003	–0,010	–1,072	0,026	0,984	0,026
01.07.2012	3	0,009	0,012	0,005	2,083	0,039	1,264	0,050
01.10.2012	4	0,017				0,052	1,988	0,104
01.01.2013	1					0,066	–0,236	–0,015

Таблиця 5.7. Розрахунок скоригованих оцінок сезонної компоненти

j / i	1	2	3	4	Середньосезонне значення $\bar{s}$
2009		1,094	1,156	1,836	
2010	0,286	0,738	1,188	1,822	
2011	0,038	0,244	1,663	2,646	
2012	–1,072	2,083			
Середня оцінка сезонної компоненти для *j*-го кварталу, $\bar{s}_j$	–0,249	1,040	1,336	2,101	1,057
Скоригована оцінка сезонної компоненти, $\hat{s}_j$	–0,236	0,984	1,264	1,988	1,000

Крок 7. Прогнозування за мультиплакативною моделлю. Прогнозові значення за адитивною моделлю розраховують за формулою

$$\hat{y}_t = \hat{v}_t \cdot \hat{s}_t \text{ (за квартал)},$$

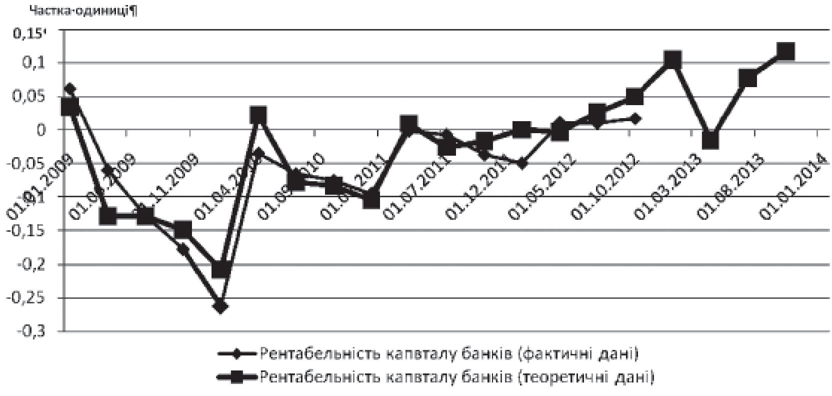


Рис. 5.15. Графічний аналіз динаміки рентабельності капіталу банків за мультиплакативною моделлю

де тренд обчислюють за рівнянням $\hat{v}_t = 0.0001 \cdot t - 5.4893$, коли t дорівнює прогнозованим кварталам, а оцінки сезонної компоненти не змінюються. Оцінки прогнозових значень обсягів наданих кредитів на чотири квартали 2009 року наведено в табл. 5.6 (ст. 4).

Не слід забувати: чим віддаленішим є період випередження, тим прогноз менш переконливий (рис. 5.15). У цьому випадку ми припускаємо, що тенденція, виявлена за ретроспективними даними, поширюється і на майбутній період. Для порівняно невеликих періодів випередження таке припущення справді може бути реальним, однак воно стає менш імовірним, якщо прогноз складають на віддалену перспективу.

5.4. Методи прогнозування часових рядів

5.4.1. Методи соціально-економічного прогнозування

Одним з основних завдань дослідження та аналізу часових рядів є визначення головних тенденцій їх зміни та перспектив, основних напрямів розвитку в майбутньому, що дозволяють отримати результати прогнозів.

Прогноз – це науково обґрунтоване судження стосовно можливих станів об'єкта в майбутньому, альтернативні шляхи й терміни їх здійснення.

Прогноз має випадковий характер, та оскільки він будується на підставі аргументованих наукових уявлень про стан і розвиток об'єкта, здійснення його є доволі ймовірним. Самі розробники прогнозу оцінюють його як очікуваний, імовірний стан об'єкта в майбутньому.

Прогнозування – це процес розроблення прогнозів.

Об'єктом соціально-економічного прогнозування є сукупність економічних і соціальних процесів (СЕП) формування та функціонування соціально-економічної системи, які характеризують динаміку зміни її параметрів на певному рівні господарювання.

Завданням соціально-економічного прогнозування, з одного боку, є з'ясування перспективи найближчого або віддаленішого майбутнього, зважаючи на реальні процеси сьогодення, а з іншого – сприяння розробленню оптимальних програм і планів економічного та соціального розвитку об'єкта, що має ґрунтуватися на пропонованому прогнозі йураховувати оцінку прийнятого рішення з позицій його наслідків у прогнозованому періоді.

Необхідність прогнозування розвитку тієї чи іншої ситуації, майбутніх змін тих чи інших обставин, ставить нас перед непростою проблемою вибору конкретного методу прогнозування (рис. 5.16).

Цей вибір залежить від багатьох факторів, а саме: наявність даних, планований момент виконання та точність прогнозу, а також вартісні та часові витрати на його складання.

Аналіз часових рядів заснований на припущенні, у відповідності з яким те, що відбулося в минулому, дає гарне наближення в оцінці майбутнього. І є способом виявлення тенденцій минулого та їх продовження в майбутнє.

Каузальні методи застосовуються в тому разі, коли шуканий стан залежить не тільки від часу, а від декількох, навіть багатьох змінних.

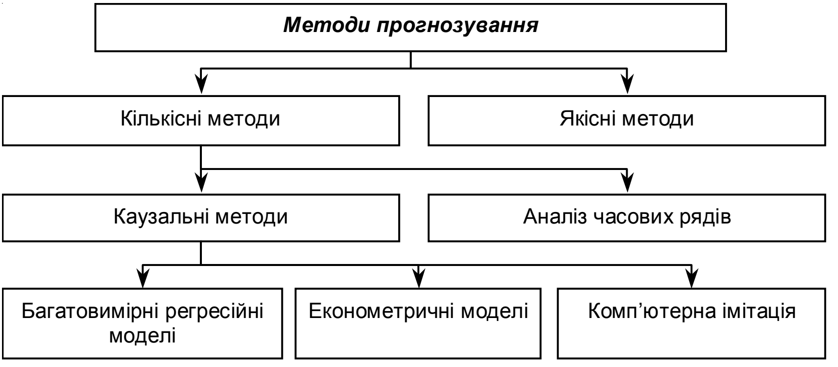


Рис. 5.16. Методи складання прогнозів

Методи згладжування часових рядів об'єднуються у дві основні групи:

1) механічне згладжування окремих рівнів часового ряду, яке не потребує знань про аналітичний вид згладженої функції;
2) аналітичне згладжування з використанням кривої, проведеної між певними рівнями ряду так, щоб вона відбивала тенденцію, притаманну ряду, і одночасно позбавляла його незначних коливань.

5.4.2. Прогнозування тенденції часового ряду за середніми характеристиками

Найпростішим способом прогнозування вважається підхід, який визначає прогнозову оцінку від фактично досягнутого рівня за допомогою середнього рівня, середнього приросту, середнього темпу зростання.

Екстраполяція на основі середнього рівня ряду. Під час екстраполяції соціально-економічних процесів на основі середнього рівня ряду прогнозоване значення беруть як середнє арифметичне значення попередніх рівнів ряду, тобто точковий прогноз $\hat{y}_n(\tau)$, зроблений у момент часу $t = n$ на період упередження τ, розраховують за формулою

$$\hat{y}_n(\tau) = \bar{y}. \qquad (5.32)$$

Інтервал надійності для прогнозованої оцінки ряду дорівнює:

$$\hat{y}_n(\tau) \pm t_\alpha \hat{\sigma}\sqrt{1+\frac{1}{n}}. \qquad (5.33)$$

Екстраполяцію за середнім абсолютним приростом можна виконати в тому разі, коли загальна тенденція розвитку вважається лінійною. Прогнозову оцінку $\hat{y}_n(\tau)$ одержують за формулою

$$\hat{y}_n(\tau) = y_n + \tau \cdot \overline{\Delta y}, \qquad (5.34)$$

де $\overline{\Delta y}$ – середній абсолютний приріст.

Екстраполяцію за середнім темпом зростання можна виконувати в разі, коли є підстави вважати, що загальна тенденція динамічного ряду характеризується експоненціальною кривою. Прогноз $\hat{y}_n(\tau)$, зроблений у момент часу $t = n$ на період випередження τ, у цьому разі розраховують за формулою

$$\hat{y}_n(\tau) = y_n \cdot \bar{T}_{\text{зр}}^{\ \tau}, \qquad (5.35)$$

де $\bar{T}_{зр}$ – середній темп зростання, розрахований за середньою геометричною.

Інтервал надійності прогнозу за середнім абсолютним приростом і середнім темпом зростання можна одержати лише в разі, якщо ці середні визначаються за допомогою статистичного оцінювання параметрів, відповідно, лінійної та експоненціальної кривої.

5.4.3. Прогнозування тенденції часового ряду за механічними методами

Механічні методи згладжування часових рядів використовують фактичні значення сусідніх рівнів ряду і не досліджують аналітичний вид згладженої функції. Вони мають механізм автоматичного налагодження на зміну досліджуваного показника. Завдяки цьому модель постійно пристосовується до зміни інформації й наприкінці інтервалу прогнозової бази відображає тенденцію, що склалася на поточний момент. До механічних методів належать: згладжування за двома точками, метод простої ковзної середньої, метод зваженої ковзної середньої, метод експоненційного згладжування.

Метод ковзної середньої (*MA*(*m*) – *moving average*).

Згладжування за допомогою ковзної середньої ґрунтується на тому, що в середніх величинах взаємно гасяться випадкові відхилення. Саме зменшення випадкового розкиду (дисперсії) й означає згладжування відповідної траєкторії.

При згладжуванні за допомогою ковзної середньої початкові рівні часового ряду y_t замінюють його середніми (згладженими) величинами $\hat{y}_t$, розрахованими для певної кількості рівнів ряду. Кількість даних, які входять до інтервалу, називають порядком ковзної середньої.

$$\hat{y}_t = \frac{\sum_{i=-k}^{k} y_{t+i}}{m} \tag{5.36}$$

Для визначення згладжених значень $\hat{y}_t$ у k перших і k останніх крайніх точках усього часового ряду можна використати відповідні значення локально апроксимаційних поліномів, побудованих, відповідно, за $2k + 1$ першими та $2k + 1$ останніми точками часового ряду y_1, y_2, y_3, ..., y_n.

Точніші результати згладжування дає застосування *зваженої ковзної середньої*. Її оцінку $\hat{y}_t$ в середині кожного інтервалу згладжування описує поліном p-го ступеня:

$$\hat{y}_t = a_0 + \sum_{i=1}^{p} a_i t^i \tag{5.37}$$

Параметри цього рівняння знаходять за методом найменших квадратів. Ковзну середню в обраному інтервалі визначають як зважене середнє усіх попередніх рівнів, причому ваги спостережень мають неоднакові значення.

Наприклад, якщо в інтервал згладжування входять п'ять спостережень, а тенденцію можна подати як поліном другого ступеня, то згладжений середній рівень у взятому інтервалі виражатиме значення тенденції на початку відліку. За $t = 0$ початок відліку, як випливає із формули (5.37), дорівнює α_0.

Для цього випадку

$$a_0 = \frac{1}{35}(-3y_{t-2} + 12y_{t-1} + 17y_t + 12y_{t+14} - 3y_{t+2}), \tag{5.38}$$

де коефіцієнти за y_{t+i} (позначимо їх ω_i) характеризують «вагу», що надається рівню ряду, розташованому на відстані i від моменту t. Наприклад, $w_1 = \frac{12}{35}$.

Обчислені значення вагових коефіцієнтів ω_i, $i = -k, -k+1, \ldots, -1, 0$ для різної довжини відрізків усереднення k (або порядку згладжування $m = 2k + 1$) і порядку апроксимаційного полінома p (5.37) наведено в табл. 5.8 [24]. Зазначимо, що, по-перше, значення ω_i для додатних i не наводяться, оскільки коефіцієнти симетричні стосовно середини відрізка згладжування, тобто $\omega_i = \omega_{-i}$; по-друге, за однакової довжини інтервалів згладжування $m = 2p + 1$, ваги ω_i у формулі (5.37) для поліномів парного ступеня будуть такими самими, що й для поліномів ступеня, більшого на одиницю (непарного).

Метод ковзної середньої набув поширення для короткотермінового прогнозування. За невеликої кількості спостережень метод часто призводить до викривлення тенденцій, а вибір величини інтервалу згладжування важко обґрунтувати, хоча від цього залежить форма кривої. Одночасно зі зменшенням дисперсії у згладженому ряду можуть з'явитися систематичні коливання, зумовлені автокореляцією його послідовних значень аж до порядку $2k$ (ефект Слуцького – Юла), тобто

Таблиця 5.8. Значення вагових коефіцієнтів w залежно від довжини відрізків усереднення k та порядку апроксимаційних поліномів p

k	p	w_{-k}	w_{-k+1}	…	w_0
3	0 або 1	$\frac{1}{2k+1}$	$\frac{1}{2k+1}$	…	$\frac{1}{2k+1}$
5	2 або 3	$-\frac{3}{35}$	$\frac{12}{35}$	–	$\frac{17}{35}$
7	2 або 3	$-\frac{2}{21}$	$\frac{3}{21}$	$\frac{6}{21}$	$\frac{7}{21}$
9	2 або 3	$-\frac{21}{231}$	$\frac{14}{231}$	$\frac{39}{231}; \frac{54}{231}$	$\frac{59}{231}$
7	4 або 5	$\frac{5}{231}$	$-\frac{30}{231}$	$\frac{75}{231}$	$\frac{131}{231}$
9	4 або 5	$\frac{15}{429}$	$-\frac{55}{429}$	$\frac{30}{429}; \frac{135}{429}$	$\frac{179}{429}$

методи ковзної середньої можуть спричинити автокореляцію залишків, навіть якщо вона була відсутня в початкових даних.

Метод експоненціального згладжування. Метод експоненціального згладжування дає можливість описати такий перебіг процесу, коли найбільшої ваги надають останньому спостереженню, а вага решти спостережень спадає геометрично. Так, для спостережень y_t, $t \le n$ прогноз наступного значення y_{n+1} має вигляд:

$$\hat{y}_n(1) = \alpha(y_n + (1-\alpha)y_{n-1} + (1-\alpha)^2 y_{n-2} + \dots), 0 < \alpha < 1, \qquad (5.39)$$

де підсумок усіх ваг дорівнює 1, а α — параметр згладжування.

Практичний розрахунок експоненціальної середньої здійснюють за рекурентною формулою

$$\hat{y}_t(1) = \alpha y_t + (1-\alpha)\hat{y}_t$$

або (5.40)

$$\hat{y}_t(1) = \hat{y}_t + \alpha(y_t - \hat{y}_t),$$

тобто в розрахунку нової експоненціальної середньої беруть попередню експоненціальну середню та частку (α) від різниці між попереднім спостереженням і його згладженим значенням, тобто похибки $e_t = y_t - \hat{y}_t$.

Використання методу експоненціального згладжування передбачає розв'язання трьох питань: вибір постійної згладжування α, вибір початкового рівня згладжування ряду $\hat{y}_0$, вибір початкового моменту згладжування (довжини бази згладжування). Аналітичного розв'язку поставлених завдань наразі не існує, і він навряд чи можливий. Вибір характеристик згладжування має ґрунтуватися на експериментальних розрахунках і здійснюватися в кожному конкретному випадку по-різному.

Вибір параметра згладжування α є основною та досить складною проблемою. Для різних значень α результати прогнозування відрізнятимуться. Якщо значення α близьке до одиниці, то під час прогнозування зважають здебільшого на основному вплив останніх спостережень; якщо близьке до нуля, то вплив рівнів ряду спадає повільно, що уможливлює врахування попередніх значень.

Від вибору початкового рівня згладжування $\hat{y}_0$ залежить поведінка наступної згладженої послідовності. Найчастіше він або дорівнює значенню першого рівня ряду y_1, або береться на рівні середньої арифметичної ряду. Зазначимо, що чим довший ряд, тим менший вплив на результат згладжування справляє вибір $\hat{y}_0$.

Вибір початкового моменту згладжування (довжини бази згладжування). Проблема вибору початкової точки згладжування зумовлена проблемою вибору сталої згладжування α. Чим ближче початкова точка до поточної, тим менше інформації знадобиться для побудови прогнозу і тим ближче α до 1; чим далі початкова точка до поточної, тим менш чутливим буде прогноз до нових даних, і тим ближче α до 0.

Метод експоненціального згладжування застосовують під час короткотермінового прогнозування. Для побудови прогнозу необхідно задати лише початкову оцінку прогнозу, подальші розрахунки здійснюються автоматично мірою надходження нових даних спостережень, і прогноз не потрібно обчислювати спочатку. За цим методом згладжування не втрачаються ані початкові, ані останні рівні заданого часового ряду, тут немає точки, на якій ряд обривається. Чутливість експоненціально зваженого середнього з метою підвищення адекватності прогнозової моделі можна в будь-який момент змінити, якщо зробити іншою величину α.

Медіанне згладжування. Основною перевагою медіанного згладжування є стійкість до викидів. В основі методу лежить розрахунок ковзної медіани[1]. Для того щоб знайти значення ковзної медіани в точці t, розраховується медіана значень ряду в часовому інтервалі $(t - q, t + q)$. Медіана ряду в часовому інтервалі визначається як центральний член варіаційного ряду — послідовності значень ряду, що входять в цей часовий проміжок і впорядковані за зростанням. Відповідне значення називається $(2q + 1)$ точковою ковзною медіаною. На відміну від вибіркового середнього, вибіркова медіана значно більш стійка відносно наявності викидів та інших випадкових викривлень даних.

Однак, якщо момент часу t знаходиться від початку чи кінця ряду менш ніж на q точок, розрахунок стає неможливим. Для того щоб не звужувати зону визначення згладженого ряду відносно вихідного, ці граничні ефекти усувають різними методами. Наприклад, для таких точок, за винятком кінцевих, розраховують значення ковзної медіани меншого, максимально можливого порядку.

---

[1] ***Медіана*** (англ. *median*): 1. У статистиці величина ознаки, що знаходиться посередині ранжованого ряду вибірки, тобто це величина, що знаходиться всередині ряду величин, розташованих у зростаючому або спадному порядку. 2. У теорії ймовірності характеристика розподілення випадкової величини. Медіана ділить ряд значень ознаки на дві рівні частини, по обидві частини від неї розміщується однакова кількість одиниць сукупності.

5.4.4. Прогнозування тенденції часового ряду за аналітичними методами

Аналітичні методи згладжування часових рядів ґрунтуються на припущенні, що відомий загальний вигляд невипадкової складової часового ряду. Вони реалізуються за допомогою регресійних та адаптивних методів.

Регресійний аналіз. Оцінювання параметрів кривих зростання здійснюють на підставі побудови моделі регресії, у якій пояснювальною змінною є час:

$$y_t = v_t + \varepsilon_t \quad t = 1, 2, ..., n, \tag{5.41}$$

де v_t – функція тренду (крива зростання);
ε_t – невідомі випадкові похибки.

Виходячи з теоретичних міркувань крива зростання може описуватися будь-якою математичною функцією v_t. Оцінювання цієї функціональної залежності здійснюють за вибірковими спостереженнями $\{t, y_t\}$, $t = 1,2..., n$, а вибір методу оцінювання залежить від виду кривої й стохастичного походження випадкових похибок ε_t.

Методи прогнозування, що спираються на методи регресії, використовують для короткотермінового та середньострокового прогнозування. Вони не допускають адаптації: з отриманням нових даних процедура побудови прогнозу має повторюватися спочатку. Оптимальна довжина періоду випередження визначається окремо для кожного економічного процесу з урахуванням його статистичної нестабільності. Ця довжина, як правило, не перевищує для рядів річних спостережень третини обсягу даних, а для квартальних і місячних рядів – двох років.

Регресійні методи є основою побудови кривих зростання. Для відображення економічних процесів існує велика кількість видів кривих зростання. Щоб правильно підібрати найдоцільнішу криву для моделювання й прогнозування економічного явища, необхідно знати особливості кожного виду кривих. Криві зростання описують різні тенденції економічних процесів, наприклад, життєвий цикл товару, процес нагромадження капіталу, маркетингові зусилля фірм тощо. В економічній практиці вже набутоий певний досвід і розроблено певні типи кривих, які найчастіше використовують у соціально-економічних дослідженнях. До таких кривих належать: поліноміальні, експоненціальні та S-подібні криві зростання.

Алгоритм методу послідовних різниць

1 етап

Розраховують різниці (прирости) до d-го порядку включно. Для апроксимації економічних процесів зазвичай розраховують різниці до четвертого порядку.

$$\Delta_t^1 = y_t - y_{t-1}\,;$$
$$\Delta_t^2 = \Delta_t^1 - \Delta_{t-1}\,; \qquad (5.42)$$
$$\Delta_t^d = \Delta_t^{d-1} - \Delta_{t-1}^{d-1}\,.$$

2 етап

Для вхідного ряду та для кожного різницевого ряду обчислюють дисперсії:

– для вхідного ряду – $s_0^2 = \dfrac{\sum\limits_{t=1}^{n} y_t^2 - \dfrac{1}{n}\left(\sum\limits_{t=1}^{n} y_t\right)^2}{n-1}$; (5.43)

– для різницевого ряду d-го порядку (d = 1, 2, ...) –

$$s_d^2 = \frac{\sum\limits_{t=d+1}^{n} (\Delta_t^d)^2}{(n-d)c_{2d}^d}\,, \qquad (5.44)$$

де c_{2d}^d – біноміальний коефіцієнт.

3 етап

Порівнюють значення кожної наступної дисперсії з попередньою, тобто розраховують різниці $\left|s_d^2 - s_{d-1}^2\right|$, і якщо для будь-якого k ця величина не перевищує певної наперед заданої додатної величини, тобто порядок величин дисперсій однаковий, то ступінь апроксимаційного полінома має дорівнювати $d - 1$.

Рис. 5.17. Алгоритм методу послідовних різниць

Вибір форми кривої для згладжування певною мірою залежить від мети згладжування – інтерполяції або екстраполяції. У першому випадку метою є досягнення найбільшої близькості до фактичних рівнів часового ряду. У другому – виявлення основної закономірності розвитку явища, стосовно якої можна припустити, що в майбутньому вона збережеться.

На практиці під час попереднього аналізу часового ряду обирають, як правило, дві-три криві зростання для подальшого дослідження і побудови трендової моделі часового ряду.

Метод послідовних різниць (Тінтнера). Цей метод може бути використаний для визначення порядку (ступеня) апроксимаційного полінома, якщо, по-перше, рівні часового ряду складаються лише з двох компонент – тренду та випадкової, і, по-друге, тренд є досить гладеньким, щоб його можна було згладити поліномом певного ступеня. Алгоритм застосування методу ідентичний алгоритму визначення порядку інтеграції нестаціонарного процесу і передбачає такі кроки (рис. 5.17).

Необхідно зазначити, що для визначення тренду в економічних часових рядах не слід використовувати поліноми дуже великого порядку, оскільки отримані в такий спосіб функції згладжування відображатимуть випадкові відхилення, а не детерміновану складову, що суперечить поняттю тенденції.

Метод характеристик приросту є універсальним методом попереднього вибору кривих зростання. Він ґрунтується на використанні окремих характерних властивостей кривих, розглянутих вище. За цього методу вхідний часовий ряд попередньо згладжують методом простої змінної середньої. Наприклад, для інтервалу згладжування $t = 3$ згладжені рівні розраховують за формулою

$$\bar{y}_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}, \tag{5.45}$$

причому щоб не втратити перший та останній рівні, їх згладжують за формулами

$$\bar{y}_1 = \frac{5y_1 + 2y_2 - y_3}{6} \quad \text{та} \tag{5.46}$$

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

Далі обчислюють перші середні прирости

$$\overline{\Delta}_t = \frac{\overline{y}_{t+1} - \overline{y}_{t-1}}{2} \quad t = 2, 3,...1, n-1, \tag{5.47}$$

другі середні прирости

$$\overline{\Delta}_t^{(2)} = \frac{\overline{\Delta}_{t+1} - \overline{\Delta}_{t-1}}{2}. \tag{5.48}$$

Відповідно до характеру зміни середніх приростів і похідних показників обирають вид кривої зростання для вхідного часового ряду, при цьому використовують відомості з табл. 5.9.

Адаптивні методи прогнозування. Адаптивні методи прогнозування застосовуються в ситуації зміни зовнішніх умов, коли найбільш важливими стають останні реалізації досліджуваного процесу. Загальна схема побудови адаптивних методів може бути подана так:

1) за кількома першими рівнями ряду будується модель і оцінюються її параметри;

Таблиця 5.9. Вибір кривої зростання за характером зміни показника

Показник	Характер зміни показника з часом	Вид кривої зростання
Перший середній приріст $\overline{\Delta}_t$	Майже однаковий	Поліном першого порядку (пряма)
Перший середній приріст $\overline{\Delta}_t$	Змінюється лінійно	Поліном другого порядку (парабола)
Другий середній приріст $\overline{\Delta}_t^{(2)}$	Змінюється лінійно	Поліном третього порядку (кубічна парабола)
Відношення першого середнього приросту до середнього рівня ряду $\frac{\overline{\Delta}_t}{\overline{y}_t}$	Майже однаковий	Проста експонента
τ	Змінюється лінійно	Модифікована експонента
$\log \frac{\overline{\Delta}_t}{\overline{y}_t}$	Змінюється лінійно	Крива Гомперця
$\log \frac{\overline{\Delta}_t}{\overline{y}_t^2}$	Змінюється лінійно	Логістична крива

2) на основі побудованої моделі розраховується прогноз на один крок уперед, причому його відхилення від фактичного рівня ряду розцінюється як помилка прогнозування, яка враховується відповідно до прийнятої схеми коригування моделі;
3) за моделлю з відкоригованими параметрами розраховується прогнозна оцінка на наступний момент часу тощо.

Схему такого процесу зображено на рис. 5.18.

Після надходження фактичного значення обчислюється помилка, розбіжність між фактичним і прогнозованим рівнем $e_{t+1} = y_{t+1} - \hat{y}_t$.

У моделі передбачається, що зміна фактичного рівня є деякою часткою ($0 \le |\gamma| \le 1$) від очікуваної зміни $y_t - y_{t-1} = \gamma(\hat{y}_t - y_{t-1})$. Параметр γ називається коригувальним коефіцієнтом або параметром адаптації. За критерій оптимальності під час вибору параметра адаптації можна взяти мінімум середнього квадрата помилок прогнозування. Чим

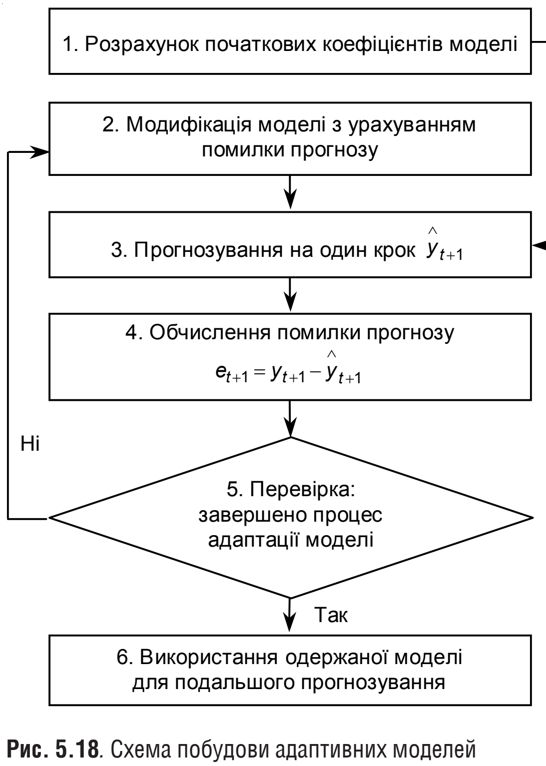


Рис. 5.18. Схема побудови адаптивних моделей

ближчий γ до одиниці, тим більше сподівання економічних суб'єктів відповідають реальній динаміці часового ряду, і, навпаки, чим ближче до нуля – тим менше володіємо ситуацією, тому треба вносити корективи.

Таким чином, модель постійно вбирає в себе нову інформацію і до кінця періоду навчання відбиває тенденцію розвитку процесу, що існує на даний момент. Прогноз отримується як екстраполяція останньої тенденції. Численні адаптивні методи відрізняються один від одного лише способами числової оцінки параметрів моделі і визначення параметрів адаптації. Базовими адаптивними методами вважаються методи Хольта, Брауна і Хольта – Уїнтерса.

Модель Брауна. Застосовуєтьтся при проведенні короткострокового прогнозування на основі розрахунку прогнозного рівня часового ряду лише на один крок уперед. За даним підходом новий прогноз будується з урахуванням принципу адаптації, тобто визначається шляхом коригування попереднього прогнозного значення з урахуванням його помилки (різниці фактичного і теоретичного рівня ряду). Особливість моделі Брауна полягає в тому, що враховує експоненційне згладжування при виявленні тренду і дозволяє досить швидко відобразити нові зміни шляхом надання більшої ваги даним, які характеризують останні проведені спостереження. У той самий час згладження випадкових відхилень вимагає зменшення зазначених вагів, що є суперечністю цієї моделі і складає завдання оптимізації..

Якщо є часовий ряд спостережень y_t, $t = 1, ..., n,$ то прогноз у момент часу t на τ кроків уперед можна здійснити за формулою:

$$\hat{y}_t(\tau) = a_{1,t} + a_{2,t} \cdot \tau, \tag{5.49}$$

де $a_{1,t}$, $a_{2,t}$ – поточні оцінки коефіцієнтів адаптивного поліному.

У моделі Брауна модифікація (адаптація) коефіцієнтів лінійної моделі здійснюється в такий спосіб:

$$\begin{aligned} a_{1,t} &= a_{1,t-1} + a_{2,t-1} + \left(1 - \beta^2\right) e_t \\ a_{2,t} &= a_{2,t-1} + \left(1 - \beta^2\right) e_t, \end{aligned} \tag{5.50}$$

де β – коефіцієнт дисконтування даних;
e_t – похибка прогнозу ($e_t = y_t - \hat{y}_{t-1}$).

Початкові значення параметрів моделі можна визначити за методом МНК на основі кількох перших спостережень. Оптимальне значення параметра дисконтування знаходиться в межах [0; 1], визначається методом чисельної оптимізації і є постійним для всього періоду спостережень. За рахунок оператора *B* можна зрушувати всю послідовність на один крок назад: $By(t) = y(t-1)$. Застосування оператора B до спостережень і до коефіцієнтів адаптивного поліному дозволяє виразити модель Брауна у вигляді

$$(1-B)^2 y_t = \left(1 - 2B\beta + B^2\beta^2\right)e_t. \tag{5.51}$$

Тоді модель Брауна можна трактувати як модель авторегресії *ARIMA* (0, 2, 2)-моделлю: $(1-B)^2 y_t = (1-2B\beta + B^2\beta^2)e_t$ з коефіцієнтами ковзної середньої – 2β та β^2.

Точковий прогноз розраховують після підстановки значення τ в оцінювану модель. Межі інтервалу надійності прогнозу можна визначити за формулою:

$$\hat{y}_t(\tau) \mp \sigma_e \sqrt{1 + C(\tau)}, \tag{5.52}$$

де величини $C(\tau) = (1-\beta)(1{,}25 + (1-\beta)\cdot\tau)$.

Модель Хольта. У разі опису статистичних даних лінійним трендом модель Хольта не дає можливості отримати адекватні прогнозні значення, що призводить до необхідності розгляду моделі Хольта. Особливість даної моделі полягає в тому, що вона враховує лінійний тренд без сезонних коливань і застосовується для прогнозування часових рядів, які мають тенденцію до зростання чи спадання. Коефіцієнти прогнозного рівняння розраховуються на основі коефіцієнтів рівня, побудованого на даних без урахування інформації останніх спостережень.

У моделі коефіцієнти лінійної моделі

$$\hat{y}_t(\tau) = a_{1,t} + a_{2,t}\cdot\tau \tag{5.53}$$

модифікуються наступним чином:

$$a_{1,t} = a_{1,t-1} + a_{2,t-1} + \alpha_1 e_t, \tag{5.54}$$

$$a_{2,t} = a_{2,t-1} + \alpha_2 e_t.$$

Початкові значення параметрів моделі можна визначити за методом МНК на основі кількох перших спостережень. Оптимальні значення параметрів згладжування знаходяться в межах [0; 1], визначаються методом багатовимірної чисельної оптимізації і є постійним для всього періоду спостережень.

Аналогічно моделі Брауна модель Хольта в термінах *ARIMA* (0, 1, 1) можна подати у вигляді

$$(1-B)^2 y_t = \left[1-(2-(\alpha_1+\alpha_1\alpha_2)B+(1-\alpha_1))B^2\right]e_t. \tag{5.55}$$

Завдання для самоконтролю

1. Поняття часового ряду.
2. Поняття структури ряду.
3. Що називається декомпозицією часового ряду?
4. Адитивна модель часового ряду.
5. Мультиплікативна модель часового ряду.
6. Які висновки можна зробити з аналізу корелограми?
7. Параметри оцінки якості моделі.
8. Побудова прогнозного значення часового ряду.

Практичне заняття 5

Часові ряди

Постановка завдання

1 етап	Побудова інформаційного масиву дослідження шляхом збору статистичних даних та розробка моделі парної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), анаіз моделі
2 етап	Побудова множинної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), дослідження моделі
3 етап	Виявлення, дослідження та усунення мультиколінеарності факторів у моделі множинної лінійної регресії методом Фаррара – Глобера
4 етап	Виявлення ефекту гетероскедастичності на основі застосування параметричного тесту Гольдфельда – Квандта та усунення даного ефекту шляхом використання узагальненого методу найменших квадратів (методу Ейткена)
5 етап	**Побудова адитивної та мультиплікативної моделі часового ряду на основі проведення декомпозиційного аналізу: визначення структури ряду (тенденції та сезонних (циклічних) коливань), проведення автокореляційного аналізу, визначення якості моделі, розрахунок прогнозних значень**
6 етап	Групування елементів економічної системи шляхом кластерного аналізу, побудова деревоподібної дендограми, графічне подання елементів кластеризації
7 етап	Перевірка наявності автокореляції залишків за допомогою критерію Дарбіна – Уотсона, оцінка параметрів рівняння множинної лінійної регресії методом Ейткена
8 етап	Побудова економіко-математичної моделі методом інструментальних змінних; побудова економетричної моделі на основі даних, що містять помилки вимірювання змінних, оцінка параметрів моделі за допомогою алгоритму Уоліса
9 етап	Побудова моделі розподіленого лага, оцінка параметрів моделей з лагами в незалежних змінних, оцінювання параметрів авторегресійних моделей; перетворення ARMA і ARIMA
10 етап	Структурне моделювання (моделювання на основі системи структурних рівнянь), оцінка невідомих параметрів моделі, визначення адекватності моделі щодо опису структури вхідних даних

Якщо ваш номер варіанта парний, ви виконуєте завдання 1–6.

Якщо ваш номер варіанта непарний, ви виконуєте завдання 1, 2, 7–10.

1. За даними вашого варіанта побудувати кореляційне поле та зробити попередній висновок про наявність чи відсутність сезонних коливань.

2. Розрахувати коефіцієнти автокореляції різних порядків і побудувати корелограму. Зробити висновок про структуру ряду.
3. Побудувати адитивну модель часового ряду.
4. Побудувати графіки фактичних і теоретичних значень рівнів часового ряду, отриманих за адитивною моделлю.
5. Оцінити якість адитивної моделі за коефіцієнтом детермінації.
6. Розрахувати за адитивною моделлю прогноз на два періоди вперед.
7. Побудувати мультиплікативну модель часового ряду.
8. Побудувати графіки фактичних і теоретичних значень рівнів часового ряду, отриманих за мультиплікативною моделлю.
9. Оцінити якість мультиплікативної моделі за коефіцієнтом детермінації.
10. Розрахувати за мультиплікативною моделлю прогноз на два періоди вперед.

Варіанти 1, 2

t	y_t	t	y_t
1	5,8	9	7,9
2	4,5	10	5,5
3	5,1	11	6,3
4	9,1	12	10,8
5	7,0	13	9,0
6	5,0	14	6,5
7	6,0	15	7,0
8	10,1	16	11,1

Варіанти 3, 4

t	y_t	t	y_t
1	5,5	9	8,0
2	4,6	10	5,6
3	5,0	11	6,4
4	9,2	12	10,9
5	7,1	13	9,1
6	5,1	14	6,4
7	5,9	15	7,2
8	10,0	16	11,0

Варіанти 5, 6

t	y_t	t	y_t
1	5,3	9	8,2
2	4,7	10	5,5
3	5,2	11	6,5
4	9,1	12	11,0
5	7,0	13	8,9
6	5,0	14	6,5
7	6,0	15	7,3
8	10,1	16	11,2

Варіанти 7, 8

t	y_t	t	y_t
1	5,5	9	8,3
2	4,8	10	5,4
3	5,1	11	6,4
4	9,0	12	10,9
5	7,1	13	9,0
6	4,9	14	6,6
7	6,1	15	7,5
8	10,0	16	11,2

Варіанти 9, 10

t	y_t	t	y_t
1	5,6	9	8,2
2	4,7	10	5,6
3	5,2	11	6,4
4	9,1	12	10,8
5	7,0	13	9,1
6	5,1	14	6,7
7	6,0	15	7,5
8	10,2	16	11,3

Приклад виконання

Нехай маємо дані про обсяг квартальних витрат на рекламу деякого банку за чотири роки.

1. Побудувати кореляційне поле та зробити попередній висновок про наявність чи відсутність сезонних коливань.
   Вхідні дані внесемо на лист «Вхідні дані».

Таблиця 5.10. Вхідні статистичні дані обсягу квартальних витрат на рекламу розглянутого банку за чотири роки

Рік	Квартал	t	Витрати на рекламу, y_t
2007	I	1	375
	II	2	371
	III	3	869
	IV	4	1015
2008	I	5	357
	II	6	471
	III	7	992
	IV	8	1020
2009	I	9	390
	II	10	355
	III	11	992
	IV	12	905
2010	I	13	461
	II	14	454
	III	15	920
	IV	16	927

Побудуємо кореляційне поле на окремому листі Excel з відповідною назвою (рис. 5.19).

З аналізу графіку 1, зображеного на рис 5.19, можна дійти висновку, що, оскільки значення y утворюють пилоподібну фігуру, то, можливо, присутня сезонна компонента.

2. Розрахувати коефіцієнти автокореляції різних порядків і побудувати корелограму. Зробити висновок про структуру ряду.

Для перевірки цього розрахуємо кілька послідовних коефіцієнтів автокореляції. Складемо на листі Excel «Коефіцієнти автокореляції» першу допоміжну таблицю для розрахунку коефіцієнта автокореляції з лагом 1 (табл. 5.11), де $y_{1\,\text{сер}} = \frac{1}{n-1}\sum_{t=2}^{n} y_t$, $y_{2\,\text{сер}} = \frac{1}{n-1}\sum_{t=2}^{n} y_{t-1}$.

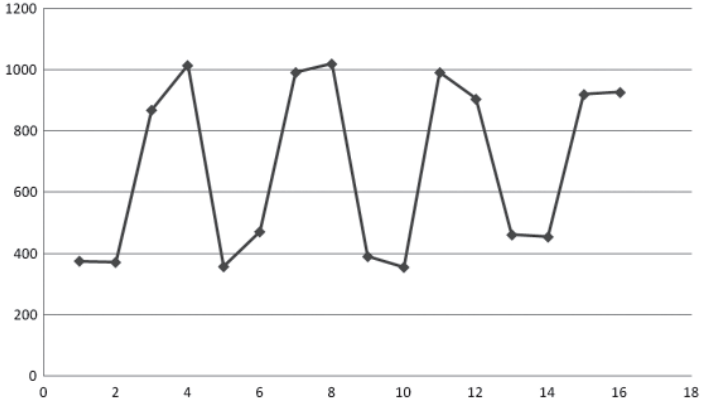


Рис. 5.19. Кореляційне поле

Таблиця 5.11. Розрахунок коефіцієнта автокореляції першого порядку

t	y_t	$y_t - 1$	$y_t - y_{1сер}$	$y_t - 1 - y_{2сер}$	$(y_t - y_{1сер})$* $(y_t - 1 - y_{2сер})$	$(y_t - y_{1сер})^2$	$(y_t - 1 - y_{2сер})^2$
1	375						
2	371	375,00	–328,93	–288,13	94776,66	108197,14	83020,82
3	869	371,00	169,07	–292,13	–49390,01	28583,54	85341,88
4	1015	869,00	315,07	205,87	64861,72	99267,00	42381,08
5	357	1015,00	–342,93	351,87	–120666,81	117603,27	123810,15
6	471	357,00	–228,93	–306,13	70084,12	52410,47	93717,62
7	992	471,00	292,07	–192,13	–56115,74	85302,94	36915,22
8	1020	992,00	320,07	328,87	105259,26	102442,67	108153,28
9	390	1020,00	–309,93	356,87	–110604,88	96058,67	127353,82
10	355	390,00	–344,93	–273,13	94212,79	118979,00	74601,82
11	992	355,00	292,07	–308,13	–89995,48	85302,94	94946,15
12	905	992,00	205,07	328,87	67439,59	42052,34	108153,28
13	461	905,00	–238,93	241,87	–57790,01	57089,14	58499,48
14	454	461,00	–245,93	–202,13	49711,32	60483,20	40857,88
15	920	454,00	220,07	–209,13	–46023,28	48429,34	43736,75
16	927	920,00	227,07	256,87	58325,86	51559,27	65980,48
Сума	10499	9947,00			74085,13	1153760,93	1187469,73
Середнє значення	699,93	663,13					

Коефіцієнт автокореляції першого порядку	0,063294

Слід зазначити, що середнє значення розраховують для $n - 1 = 15$ значень.

3. Коефіцієнт автокореляції першого порядку розраховуємо за формулою (1):

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \overline{y}_1)(y_{t-1} - \overline{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \overline{y}_1)^2 \sum_{t=2}^{n}(y_{t-1} - \overline{y}_2)^2}}, \qquad (1)$$

$$r_1 = \frac{74085{,}16}{\sqrt{1153756{,}39 \cdot 1187469{,}73}} = 0{,}063294\,.$$

Складаємо на цьому самому листі Excel допоміжну таблицю для розрахунку коефіцієнта автокореляції другого порядку (табл. 5.12).

Таблиця 5.12. Розрахунок коефіцієнта автокореляції другого порядку

	y_t	$y_t - 2$	$y_t - y_{3сер}$	$y_t - 2 - y_{4сер}$	$(y_t - y_{3сер}) \cdot (y_t - 2 - y_{4сер})$	$(y_t - y_{3сер})^2$	$(y_t - 2 - y_{4сер})^2$
	375						
	371						
	869	375	145,57	–269,79	–39273,09	21191,04	72784,33
	1015	371	291,57	–273,79	–79828,09	85013,90	74958,62
	357	869	–366,43	224,21	–82158,52	134269,90	50272,05
	471	1015	–252,43	370,21	–93452,66	63720,18	137058,62
	992	357	268,57	–287,79	–77291,02	72130,61	82820,62
	1020	471	296,57	–173,79	–51539,88	87954,61	30201,47
	390	992	–333,43	347,21	–115771,16	111174,61	120557,76
	355	1020	–368,43	375,21	–138239,66	135739,61	140785,76
	992	390	268,57	–254,79	–68428,16	72130,61	64915,76
	905	355	181,57	–289,79	–52616,81	32968,18	83975,76
	461	992	–262,43	347,21	–91118,95	68868,76	120557,76
	454	905	–269,43	260,21	–70109,16	72591,76	67711,47
	920	461	196,57	–183,79	–36127,02	38640,33	33777,19
	927	454	203,57	–190,79	–38838,52	41441,33	36399,19
Сума	10128	9027			–1034792,71	1037835,43	1116776,36
Середнє значення	723,43	644,79					

Коефіцієнт автокореляції другого порядку	–0,9612

Аналогічно знаходимо коефіцієнти кореляції більш високих порядків і заносимо їх значення у зведену таблицю (табл. 5.13) на лист «Зведена таблиця коефіцієнтів автокореляції».

Таблиця 5.13. Зведена таблиця коефіцієнтів автокореляції

Лаг	Коефіцієнт автокореляції рівнів
1	0,0633
2	– 0,9612
3	– 0,0363
4	0,9647
5	0,0506
6	– 0,9765
7	– 0,0694
8	0,9646
9	0,1683
10	– 0,9729
11	– 0,0653
12	0,9858

4. Розраховуємо границі значущості коефіцієнтів кореляції за формулою $\pm\frac{1}{\sqrt{n}} = \pm\frac{1}{\sqrt{16}} = \pm 0{,}25$.

Результат розташовуємо на листі «Зведена таблиця коефіцієнтів кореляції».

На рис. 5.20 наведено корелограму – послідовність коефіцієнтів кореляції першого, другого та інших рівнів. На цей графік також винесено границі значущості коефіцієнтів кореляції. Дану корелограму будуємо на окремому листі Excel з відповідною назвою.

З аналізу корелограми можна дійти висновку, що даний часовий ряд містить сезонні коливання з періодичністю в чотири квартали, оскільки найбільші значення коефіцієнтів кореляції спостерігаються для лагів 4, 8 та 12.

Побудувати адитивну модель часового ряду.

Крок 1. Проведемо вирівнювання початкових даних методом ковзної середньої. Для цього:

1.1. Скопіюємо вхідні дані на лист «Оцінка сезонної комп (адитивна)» у вигляді, аналогічному стовпчикам 1 і 2 табл. 5.14.

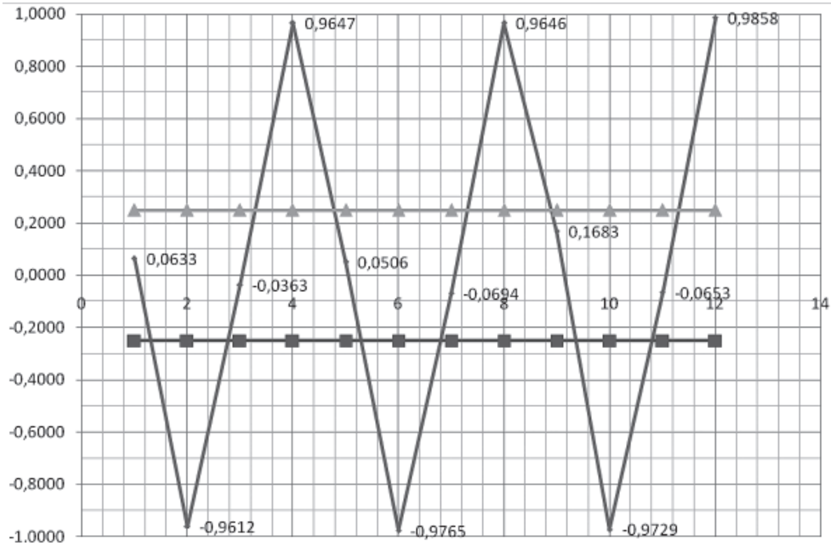


Рис. 5.20. Корелограма

1.2. Знайдемо суму рівнів ряду за кожний квартал послідовно із зсувом на один момент часу і визначимо умовні річні витрати на рекламу (стовп. 3 табл. 5.14).

1.3. Поділимо отримані суми на 4, знайдемо ковзні середні (стовп. 4 табл. 5.14). Отримані таким чином вирівняні значення вже не містять сезонної компоненти.

1.4. Приведемо ці значення у відповідність з фактичними моментами часу, для чого знайдемо середні значення з двох послідовних ковзних середніх – центровані ковзні середні (стовп. 5 табл. 5.14).

Таблиця 5.14. Проміжні розрахунки побудови адитивної моделі часового ряду

№ квартала	Кількість витрат на рекламу	Разом за 4 квартали	Ковзна середня	Центрована ковзна середня	Оцінка сезонної компоненти	Квадрат відхилення від середнього
1	2	3	4	5	6	7
1	375					92796,39063
2	371	2630	657,50			95249,39
3	869	2612	653,00	655,25	213,75	35862,89
’4	1015	2712	678,00	665,50	349,50	112476,39

Продовження табл. 5.14

№ квар-тала	Кількість витрат на рекламу	Разом за 4 квартали	Ковзна середня	Центрована ковзна середня	Оцінка сезонної компоненти	Квадрат відхилення від середнього
1	2	3	4	5	6	7
5	357	2835	708,75	693,38	– 336,38	104086,89
6	471	2840	710,00	709,38	– 238,38	43524,39
7	992	2873	718,25	714,13	277,88	97578,14
8	1020	2757	689,25	703,75	316,25	115855,14
9	390	2757	689,25	689,25	– 299,25	83882,64
10	355	2642	660,50	674,88	– 319,88	105381,39
11	992	2713	678,25	669,38	322,63	97578,14
12	905	2812	703,00	690,63	214,38	50793,89
13	461	2740	685,00	694,00	– 233,00	47796,89
14	454	2762	690,50	687,75	– 233,75	50906,64
15	920					57780,14
16	927					61194,39
Середнє	679,625				Сума	1252743,75

Крок 2. Знайдемо оцінки сезонної компоненти як різницю між фактичними рівнями ряду і центрованими ковзними середніми (стовп. 6 табл. 5.14).

На окремому листі Excel «Скоригована сезонна комп(ад)» сформуємо таблицю вхідних даних з оцінками сезонної компоненти по роках і кварталах, як це показано у табл. 5.15.

Таблиця 5.15. Оцінка сезонної компоненти адитивної моделі часового ряду

Рік	№ квартала, *i*			
	I	II	III	IV
2007	–	–	213,75	349,5
2008	– 336,75	– 238,375	277,875	316,25
2009	– 299,25	– 319,875	322,625	214,375
2010	– 233	– 233,75	–	–

Знайдемо середні за кожен квартал (за всіма роками) оцінки сезонної компоненти S_i. У моделях із сезонною компонентою сезонні впливи за період взаємогасяться. Для адитивної моделі сума значень сезонної компоненти за всіма кварталами має дорівнювати нулю (табл. 5.16).

Таблиця 5.16. Проміжні розрахунки визначення скоригованої сезонної компоненти адитивної моделі розглянутого часового ряду

Показник	Рік	№ квартала					
		I	II	III	IV		
	2007			213,75	349,50		
	2008	–336,38	–238,38	277,88	316,25		
	2009	–299,25	–319,88	322,63	214,38		
	2010	–233,00	–233,75				
Всього за *i*-й квартал		–868,63	–792,00	814,25	880,13		
Середня оцінка сезонної компоненти для і-го кварталу		–289,54	–264,00	271,42	293,38	Коригувальний коефіцієнт	2,8125
Скоригована сезонна компонента		–292,35	–266,81	268,60	290,56	Сума значень сезонної компоненти	0

Для даної моделі маємо коригувальний коефіцієнт $k = 2{,}8125$.

6. Розраховуємо скориговані значення сезонної компоненти $S_i = \overline{S}_i - k$ і заносимо отримані дані в табл. 5.16.

   Перевіряємо рівність нулю суми значень сезонної компоненти.

Крок 3. На окремому листі Excel «Адитивна модель + прогноз» сформуємо таблицю вхідних даних, аналогічну стовпчикам 1–3 табл. 5.17.

Виключимо вплив сезонної компоненти, віднімаючи її з кожного рівня вхідного часового ряду. Отримаємо величини $T + E = Y - S$ (стовп. 4 табл. 5.17). Ці значення містять тільки тенденцію і випадкову компоненту.

Крок 4. Визначимо компоненту T (тренд) даної моделі. Для цього проведемо аналітичне вирівнювання ряду $(T + E)$ за допомогою лінійного тренду. Проміжні обчислення розташуємо на листі «Аналітичне вирівнювання ряду (адитивна модель)» (див. табл. 5.18). Підставляючи в отримане рівняння лінійної регресії значення $t = 1, 2, \dots 16$, знайдемо рівні T для кожного моменту часу (стовп. 5 табл. 5.17).

Таблиця 5.17. Проміжні розрахунки побудови адитивної моделі часового ряду та прогнозування

t	y_t	S_i	$y_t – S_i$	T	$T+S$	$E=y_t–(T+S)$	E_2
1	2	3	4	5	6	7	8
1	375	– 292,35	667,35	672,68	380,33	– 5,33	28,41
2	371	– 266,81	637,81	673,61	406,80	– 35,80	1281,41
3	869	268,60	600,40	674,53	943,14	– 74,14	5496,59
4	1015	290,56	724,44	675,46	966,02	48,98	2398,77
5	357	– 292,35	649,35	676,39	384,03	– 27,03	730,71
6	471	– 266,81	737,81	677,31	410,50	60,50	3660,40
7	992	268,60	723,40	678,24	946,84	45,16	2039,34
8	1020	290,56	729,44	679,16	969,72	50,28	2527,60
9	390	– 292,35	682,35	680,09	387,73	2,27	5,14
10	355	– 266,81	621,81	681,01	414,20	– 59,20	3504,73
11	992	268,60	723,40	681,94	950,54	41,46	1718,69
12	905	290,56	614,44	682,86	973,43	– 68,43	4682,22
13	461	– 292,35	753,35	683,79	391,44	69,56	4839,21
14	454	– 266,81	720,81	684,72	417,90	36,10	1303,02
15	920	268,60	651,40	685,64	954,24	– 34,24	1172,71
16	927	290,56	636,44	686,57	977,13	– 50,13	2512,88
Сума							37901,81

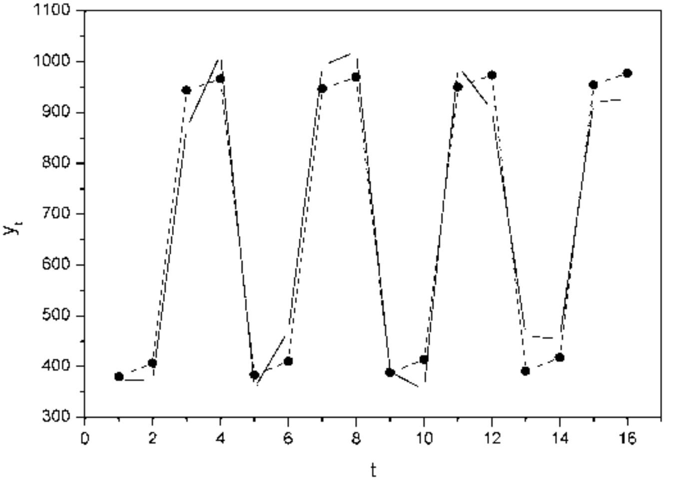


Рис. 5.21. Графіки фактичних і теоретичних значень часового ряду, отриманих за адитивною моделлю

Таблиця 5.18. Проміжні розрахунки кількісної оцінки трендової компоненти адитивної моделі розглянутого часового ряду

№ пор.	x_i	y_i	x_i-X	(x_i-X)2	y_i-Y	(y_i-Y)2	(x_i-X)(y_i-Y)	Y_i=b_0+b_1x_i	Y_i-y_i	$(Y_i-y_i)^2$	Y_i-Y	$(Y_i-Y)^2$
1	1	667,3542	–7,50	56,2500	–12,271	150,5734	92,03125	672,683824	5,329657	28,40524	–6,941	48,17993
2	2	637,8125	–6,50	42,2500	–41,813	1748,285	271,7813	673,609314	35,79681	1281,412	–6,016	36,18848
3	3	600,3958	–5,50	30,2500	–79,229	6277,261	435,7604	674,534804	74,13897	5496,587	–5,090	25,9101
4	4	724,4375	–4,50	20,2500	44,813	2008,16	–201,656	675,460294	–48,9772	2398,767	–4,165	17,34478
5	5	649,3542	–3,50	12,2500	–30,271	916,3234	105,9479	676,385784	27,03162	730,7084	–3,239	10,49252
6	6	737,8125	–2,50	6,2500	58,188	3385,785	–145,469	677,311275	–60,5012	3660,398	–2,314	5,353326
7	7	723,3958	–1,50	2,2500	43,771	1915,886	–65,6562	678,236765	–45,1591	2039,341	–1,388	1,927197
8	8	729,4375	–0,50	0,2500	49,813	2481,285	–24,9063	679,162255	–50,2752	2527,6	–0,463	0,214133
9	9	682,3542	0,50	0,2500	2,729	7,448351	1,364583	680,087745	–2,26642	5,136667	0,463	0,214133
10	10	621,8125	1,50	2,2500	–57,813	3342,285	–86,7188	681,013235	59,20074	3504,727	1,388	1,927197
11	11	723,3958	2,50	6,2500	43,771	1915,886	109,4271	681,938725	–41,4571	1718,692	2,314	5,353326
12	12	614,4375	3,50	12,2500	–65,188	4249,41	–228,156	682,864216	68,42672	4682,215	3,239	10,49252
13	13	753,3542	4,50	20,2500	73,729	5435,99	331,7813	683,789706	–69,5645	4839,214	4,165	17,34478
14	14	720,8125	5,50	30,2500	41,188	1696,41	226,5313	684,715196	–36,0973	1303,015	5,090	25,9101
15	15	651,3958	6,50	42,2500	–28,229	796,8859	–183,49	685,640686	34,24485	1172,71	6,016	36,18848
16	16	636,4375	7,50	56,2500	–43,188	1865,16	–323,906	686,566176	50,12868	2512,884	6,941	48,17993
?	136	10874	0	340	SST =	38193,03	314,6667		SSE=	37901,81	SSR=	291,2209
	X	Y		var(x)		var(y)	cov(x,y)					
Σ/n	8,500	679,625	0,000	21,250		2387,065	19,667					

Коефіцієнти рівняння лінійної регресії: b_1 = cov(x,y)/var(x) = 0,9254902 b_0 = Y – –b_1X = 671,7583

Рівняння лінійної регресії матиме вигляд: y = b_0+b_1x = 671,7583+ 0,9254902x

Крок 5. Знайдемо значення рівнів ряду, отримані за адитивною моделлю. Для цього додамо до рівнів T значення сезонної компоненти для відповідних кварталів (стовп. 6 табл. 5.17).

7. Побудувати графіки фактичних і теоретичних значень рівнів часового ряду.

На окремому листі Excel «Графік адитивної моделі і факт» побудуємо графіки фактичних і теоретичних значень часового ряду, отриманих за адитивною моделлю (див. рис. 5.21).

Оцінка якості адитивної моделі за коефіцієнтом детермінації

Для оцінки якості побудованої моделі на листі «Адитивна модель+прогноз» розраховуємо величину $(1-\dfrac{\sum E^2}{\sum(y_t-y_{cp})^2})\cdot 100\%$. У даному випадку ця величина дорівнює 96,97%, тобто адитивна модель пояснює 96,97% загальної варіації рівнів часового ряду витрат на рекламу за чотири роки.

8. Розрахувати за адитивною моделлю прогноз на два періоди вперед.

Прогнозне значення F_t рівня часового ряду в адитивній моделі є сума трендової і сезонної компонент. Обчислення здійснюватимемо на листі «Адитивна модель+прогноз».

Для визначення трендової компоненти скористаємося рівнянням тренду, підставивши в нього значення $x = 17$, $x = 18$. Отримаємо:

$T_{17} = 687{,}4917;$

$T_{18} = 688{,}4172.$

Значення сезонних компонент за відповідні квартали дорівнюють $S_1 = -292{,}3542$ і $S_2 = -266{,}8125$.

Отже,

$F_{17} = 687{,}4917 - 292{,}3542 = 395{,}1375;$

$F_{18} = 688{,}4172 - 266{,}8125 = 421{,}6046.$

Тобто у перші два квартали наступного року необхідно очікувати витрати на рекламу в сумі 395 і 421 тис. грн відповідно.

9. Побудувати мультиплікативну модель часового ряду.

Крок 1. Методика, що застосовується на цьому кроці, повністю збігається з методикою адитивної моделі. Тому скопіюємо на окремий лист «Оцінка сезонної компоненти (мультиплікативна модель)» стовп. 1–5 з листа «Оцінка сезонної компоненти (адитивна модель)» (див. табл. 5.19).

Таблиця 5.19. Проміжні розрахунки побудови мультиплікативної моделі часового ряду

№ квартала	Кількість витрат на рекламу	Разом за 4 квартали	Ковзна середня	Центрована ковзна середня	Оцінка сезонної компоненти
1	2	3	4	5	6
1	375				
2	371	2630	657,5		
3	869	2612	653	655,25	1,32621137
4	1015	2712	678	665,5	1,525169046
5	357	2835	708,75	693,375	0,514872904
6	471	2840	710	709,375	0,663964758
7	992	2873	718,25	714,125	1,38911255
8	1020	2757	689,25	703,75	1,44937833
9	390	2757	689,25	689,25	0,565832427
10	355	2642	660,5	674,875	0,526023338
11	992	2713	678,25	669,375	1,481979458
12	905	2812	703	690,625	1,31040724
13	461	2740	685	694	0,66426513
14	454	2762	690,5	687,75	0,660123591
15	920				
16	927				
Середнє	679,625				

Крок 2. Знайдемо оцінки сезонної компоненти як частку від ділення фактичних рівнів ряду на центровані ковзні середні (стовп. 6 табл. 5.19). Ці оцінки використаємо для розрахунку сезонної компоненти S, для чого на окремому листі Excel «Скоригована сезонна комп(мульт)» сформуємо таблицю вхідних даних у вигляді табл. 5.20.

Знайдемо середні за кожен квартал оцінки сезонної компоненти S_i (табл. 5.21).

Як і в адитивній моделі, сезонні впливи в мультиплікативній моделі взаємогасяться. У мультиплікативній моделі це виражається в

Таблиця 5.20. Оцінка сезонної компоненти мультиплікативної моделі часового ряду

Рік	№ квартала			
	I	II	III	IV
2007			1,3262114	1,52516905
2008	0,5148729	0,663965	1,3891126	1,44937833
2009	0,56583243	0,526023	1,4819795	1,31040724
2010	0,66426513	0,660124		

Таблиця 5.21. Проміжні розрахунки визначення скоригованої сезонної компоненти мультиплікативної моделі розглянутого часового ряду

Показники	Рік	№ квартала					
		I	II	III	IV		
	2007			1,33	1,53		
	2008	0,51	0,66	1,39	1,45		
	2009	0,57	0,53	1,48	1,31		
	2010	0,66	0,66				
Усього за *i*-й квартал		1,74	1,85	4,20	4,28		
Середня оцінка сезонної компоненти для *i*-го квартала		0,58	0,62	1,40	1,43	Коригувальний коефіцієнт	0,99
Скоригована сезонна компонента		0,58	0,61	1,39	1,42	Сума значень сезонної компоненти	4

тому, що сума значень сезонної компоненти за всіма кварталами має дорівнювати кількості періодів у циклі (в даному прикладі – чотирьом).

Маємо: 0,5816 + 0,6167 + 1,3991 + 1,4283 = 4,0257.

Визначаємо коригувальний коефіцієнт:

$$k = \frac{4}{4{,}0257} = 0{,}9936 .$$

Скориговані значення сезонної компоненти отримують при множенні її середньої оцінки $\bar{S}_i$ на коригувальний коефіцієнт k.

Умова рівності чотирьом суми значень сезонної компоненти виконується:

0,5779 + 0,6128 + 1,3901 + 1,4192 = 4.

Крок 3. На окремому листі Excel «Мультиплікативна модель+прогноз» сформуємо таблицю, аналогічну до табл. 5.22. Розділимо кожен рівень початкового ряду на відповідні значення сезонної компоненти. Отримаємо величини $TE = Y/S$ (стовп. 4 табл. 5.22), які містять тільки тенденцію і випадкову компоненту.

Таблиця 5.22. Проміжні розрахунки побудови мультиплікативної моделі часового ряду та прогнозування

t	y_t	S_i	y_t/S_i	T	$T \cdot S$	$E=y_t-(T \cdot S)$	E^2
1	2	3	4	5	6	7	8
1	375	0,57793204	648,8652	654,9157	378,4968	–3,496776862	12,22745
2	371	0,61275468	605,4625	658,197	403,3133	–32,31330333	1044,15
3	869	1,39014165	625,1162	661,4783	919,5485	–50,54854826	2555,156
4	1015	1,41917163	715,206	664,7596	943,408	71,59202533	5125,418
5	357	0,57793204	617,7197	668,0409	386,0822	–29,08224401	845,7769
6	471	0,61275468	768,66	671,3222	411,3558	59,64417576	3557,428
7	992	1,39014165	713,5963	674,6035	937,7944	54,2055802	2938,245
8	1020	1,41917163	718,7291	677,8848	962,0349	57,9651298	3359,956
9	390	0,57793204	674,8198	681,1661	393,6677	–3,667711154	13,45211
10	355	0,61275468	579,3509	684,4474	419,3983	–64,39834515	4147,147
11	992	1,39014165	713,5963	687,7287	956,0403	35,95970865	1293,101
12	905	1,41917163	637,696	691,01	980,6618	–75,66176572	5724,703
13	461	0,57793204	797,6716	694,2913	401,2532	59,7468217	3569,683
14	454	0,61275468	740,9164	697,5726	427,4409	26,55913394	705,3876
15	920	1,39014165	661,8031	700,8539	974,2862	–54,2861629	2946,987
16	927	1,41917163	653,198	704,1352	999,2887	–72,28866125	5225,651
Сума							**43064,47**

Крок 4. На окремому листі «Аналітичне вирівнювання ряду (мультиплікативна модель)» розрахуємо компоненту T мультиплікативної моделі, використовуючи рівні TE (табл. 5.22). Підставляючи в отримане рівняння значення $t = 1, 2, .., 16$, знайдемо рівні T для кожного моменту часу (стовп. 9 табл. 5.23). Скопіюємо отримані значення на лист «Мультиплікативна модель + прогноз» у стовпчику для T.

Крок 5. Знайдемо рівні ряду, помноживши значення T на відповідні значення сезонної компоненти (стовп. 6 табл. 5.22).

Таблиця 5.23. Проміжні розрахунки кількісної оцінки трендової компоненти адитивної моделі розглянутого часового ряду

№ пор.	x_i	y_i	$x_i–X$	$(x_i–X)2$	y_i-Y	$(y_i–Y)2$	$(x_i–X)(y_i–Y)$	$Y_i=b_0+b_1x_i$	$Y_i–y_i$	$(Y_i–y_i)^2$	$Y_i–Y$	$(Y_i–Y)^2$
1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	648,87	– 7,50	56,25	– 30,66	940,05	229,95	654,92	6,05	36,61	– 24,61	605,63
2	2	605,46	– 6,50	42,25	– 74,06	5485,32	481,41	658,20	52,73	2780,92	– 21,33	454,90
3	3	625,12	– 5,50	30,25	– 54,41	2960,37	299,25	661,48	36,36	1322,20	– 18,05	325,69
4	4	715,21	– 4,50	20,25	35,68	1273,10	– 160,52	664,76	– 50,45	2544,83	– 14,77	218,03
5	5	617,72	– 3,50	12,25	– 61,81	3819,95	216,32	668,04	50,32	2532,22	– 11,49	131,89
6	6	768,66	– 2,50	6,25	89,14	7944,96	– 222,83	671,32	– 97,34	9474,64	– 8,20	67,29
7	7	713,60	– 1,50	2,25	34,07	1160,83	– 51,11	674,60	– 38,99	1520,44	– 4,92	24,23
8	8	718,73	– 0,50	0,25	39,20	1536,93	– 19,60	677,88	– 40,84	1668,26	– 1,64	2,69
9	9	674,82	0,50	0,25	– 4,71	22,14	– 2,35	681,17	6,35	40,28	1,64	2,69
10	10	579,35	1,50	2,25	– 100,17	10034,93	– 150,26	684,45	105,10	11045,20	4,92	24,23
11	11	713,60	2,50	6,25	34,07	1160,83	85,18	687,73	– 25,87	669,14	8,20	67,29
12	12	637,70	3,50	12,25	– 41,83	1749,71	– 146,40	691,01	53,31	2842,38	11,49	131,89
13	13	797,67	4,50	20,25	118,14	13958,52	531,66	694,29	– 103,38	10687,50	14,77	218,03
14	14	740,92	5,50	30,25	61,39	3768,85	337,65	697,57	– 43,34	1878,68	18,05	325,69
15	15	661,80	6,50	42,25	– 17,72	314,08	– 115,19	700,85	39,05	1524,96	21,33	454,90
16	16	653,20	7,50	56,25	– 26,33	693,14	– 197,45	704,14	50,94	2594,60	24,61	605,63
Σ	136	10872,4	0	340	SST =	56823,71	1115,64		SSE =	53162,9	SSR =	3660,7
	X	Y		var(x)		var(y)	cov(x,y)					
Σ/**n**	8,500	679,525	0,000	21,250		3551,482	69,728					

Коефіцієнти рівняння лінійної регресії: b_1 = c ov(x,y)/var(x) = 3,281297 b_0 = Y – –b_1X = 651,6344

Рівняння лінійної регресії матиме вигляд: y = b_0+b_1x = 651,634+3,2812*x

10. Побудувати графіки фактичних і теоретичних значень рівнів часового ряду, отриманих за мультиплікативною моделлю.

На окремому листі Excel «Графік мультмоделі і факт» побудуємо графіки фактичних і теоретичних значень часового ряду, отриманих за мультиплікативною моделлю (рис. 5.22).

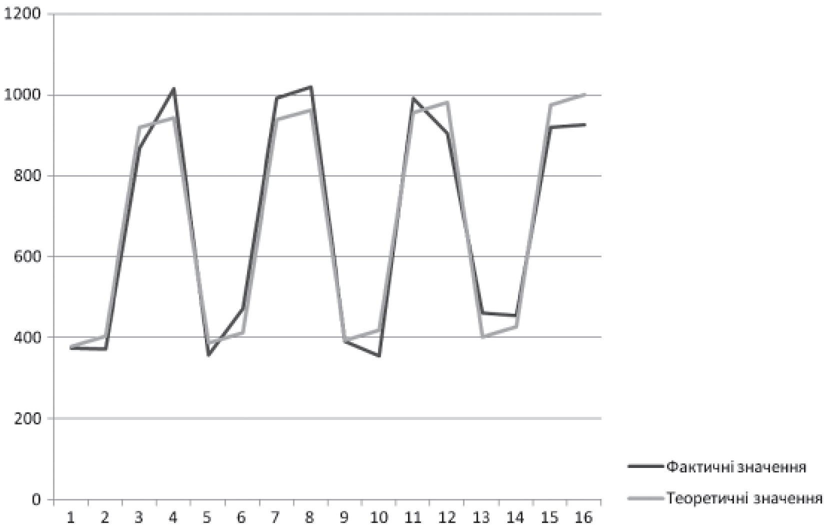


Рис. 5.22. Графіки фактичних і теоретичних значень часового ряду, отриманих за мультиплікативною моделлю

11. Оцінити якість мультиплікативної моделі.

Розрахунок помилки в мультиплікативній моделі здійснюється за формулою $E = \dfrac{Y}{(T \cdot S)}$ (стовп. 8 табл. 5.21).

Для порівняння мультиплікативної моделі та інших моделей часового ряду можна, аналогічно до адитивної моделі, використати суму квадратів абсолютних похибок $(y_t - TS)^2$ (стовп. 9 табл. 5.21).

За формулою $R^2 = 1 - \dfrac{(y_t - T \cdot S)^2}{(y_t - \bar{y})^2}$ на листі «Мультиплікативна модель + прогноз» розрахуємо значення R^2. Отримаємо $R^2 = 0{,}9656$.

Порівнюючи коефіцієнти детермінації адитивної та мультиплікативної моделей, можна дійти висновку, що вони приблизно однаково апроксимують початкові дані.

12. Розрахувати за мультиплікативною моделлю прогноз на два періоди вперед.

Прогнозне значення F_t рівня часового ряду в мультиплікативній моделі розраховується як добуток трендової і сезонної компонент.

Отже, скористаємося рівнянням тренду для $x = 17,18$.

$T_{17} = 707,4165$,

$T_{18} = 710,6978$.

Значення сезонних компонент за відповідні квартали дорівнюють $S_1 = 0,57793204$ і $S_2 = 0,61275468$. Таким чином,

$F_{17} = 408,8386$,

$F_{18} = 435,4834$.

Тобто, у перші два квартали наступного року необхідно очікувати витрати на рекламу в сумі 409 і 435 тис. грн відповідно.

Отже, адитивна і мультиплікативна моделі дають приблизно однаковий прогнозний результат.

Тести

1. Визначити правильну відповідь

Часовий ряд	одержують із вхідного ряду завдяки впорядкуванню за величиною рівня ряду
Динамічний ряд	спостереження над деяким явищем, характер якого змінюється в часі, породжує впорядковану послідовність
Варіаційний ряд	час, що минув від першого до останнього моменту спостереження
Рівні ряду	сукупність спостережень одного показника, впорядкованих залежно від значень іншого показника, що послідовно зростають або спадають
Довжина часового ряду	рівновіддалені моменти спостережень (година, доба, місяць, рік тощо)

2. Визначте, яка характеристика динаміки часового ряду показує, на скільки відсотків рівень одного періоду збільшився стосовно рівня іншого періоду, тобто цей показник характеризує відносну величину приросту у відсотках:

а) абсолютний приріст;
б) коефіцієнт зростання;
в) коефіцієнт приросту;
г) темп зростання;
д) темп приросту;
е) середня арифметична;
ж) середня хронологічна;
з) середній абсолютний приріст;
і) середній темп зростання.

3. Визначте характеристики моделі виду $Y_t = U_t \cdot S_t \cdot E_t$:

а) адитивна;
б) мультиплікативна;
в) модель тренду;
г) модель сезонності;
д) тренд-сезонна модель.

4. Систематичними компонентами часового ряду є:

а) тренд;
б) сезонна компонента;

в) циклічна компонента;
г) випадкова компонента.

5. Визначте правильну відповідь

Еволюційні чинники	визначають загальний напрям розвитку економічного показника
Тенденція	невипадкова складова часового ряду, яка змінюється повільно і описується за допомогою певної функції, яку називають функцією тренду або просто трендом
Сезонні коливання	відповідають коливанням, які мають періодичний або близький до нього характер упродовж одного року
Циклічні (кон'юнктурні) коливання	-
Випадкові чинники	чинники не підлягають вимірюванню, але неминуче супроводжують будь-який економічний процес і визначають стохастичний характер його елементів
-	часові ряди, які не мають тенденції та сезонної складової, кожен наступний рівень їх утворюється як сума середнього рівня ряду і випадкової компоненти

6. Проводячи ідентифікацію детермінованого тренду та сезонності, визначте правильні твердження:

а) ряд не має тренду, якщо коефіцієнти автокореляції між рівнями ряду не залежать від часового лага (статистично незначущі) і не мають певної закономірності зміни;
б) ряд не має тренду, якщо коефіцієнти автокореляції між рівнями ряду залежать від часового лага (статистично незначущі) і мають певні закономірності зміни;
в) ряд має лінійний адитивний тренд у разі, коли автокореляційний аналіз указує на лінійну залежність зміни коефіцієнтів автокореляції від часового лага, а перехід до перших різниць виключає цю залежність;
г) ряд має лінійний адитивний тренд у разі, коли автокореляційний аналіз указує на нелінійну залежність зміни коефіцієнтів автокореляції від часового лага, а перехід до других різниць виключає цю залежність;

д) ряд містить сезонну складову, якщо не існує лінійної залежності зміни коефіцієнтів автокореляції від часового лага, але корелограма містить велику кількість значущих максимальних і мінімальних значень коефіцієнтів автокореляцій, що свідчить про значну залежність між спостереженнями, зрушеними на однаковий часовий інтервал;

е) ряд містить тренд і сезонну складову, якщо не існує лінійної залежності зміни коефіцієнтів автокореляції від часового лагу, але корелограма містить велику кількість значущих максимальних і мінімальних значень коефіцієнтів автокореляцій, що свідчить про значну залежність між спостереженнями, зміщеними на однаковий часовий інтервал.

3. Укажіть пропущене слово: число періодів, за якими розраховується коефіцієнт автокореляції, називають...

4. Послідовність коефіцієнтів автокореляції рівней першого, другого і т.д. порядків називають:

а) автокореляційна функція часового ряду;
б) порядок коефіцієнтів автокореляції;
в) корелограмма.

5. До механічних методів прогнозування тенденцій часового ряду відносяться:

а) метод ковзної середньої;
б) метод зваженої ковзної середньої;
в) метод експоненціального згладжування;
г) регресійний аналіз;
д) екстраполяція за середнім абсолютним приростом;
е) екстраполяція за середнім темпом зростання;
ж) екстраполяція на основі середнього рівня ряду;
з) згладжування по двох точках.

6. Визначте, для якого методу прогнозування характерне: дає можливість описати такий перебіг процесу, коли найбільшої ваги надають останньому спостереженню, а вага решти спостережень спадає геометрично:

а) метод ковзної середньої;
б) метод зваженої ковзної середньої;
в) метод експоненціального згладжування;
г) регресійний аналіз;
д) екстраполяція за середнім абсолютним приростом;
е) екстраполяція за середнім темпом зростання;
ж) екстраполяція на основі середнього рівня ряду;
з) згладжування по двох точках.

ТЕМА 6

Економетричні методи кількісного аналізу на основі статистичних рівнянь

Вибір форми моделі. Ознаки “гарної” моделі • Види помилок специфікації • Виявлення та коригування помилок специфікації • Дослідження залишкового члена моделі • Проблеми специфікації. Сутність, типологізація та прикладна спрямованість завдань класифікації об’єктів • Класифікація з навчанням

6.1. Вибір форми моделі. Ознаки “гарної” моделі

Різноманіття і складність економічних процесів зумовлює різноманіття моделей, які використовуються для економетричного аналізу. Це значною мірою ускладнює процес знаходження максимально адекватної формули залежності. Для випадку парної регресії підбір моделі зазвичай здійснюється за виглядом розташування точок спостережень на кореляційному полі. Проте нерідкі ситуації, коли розташування точок приблизно відповідає кільком функціям і необхідно з них виявити найкращу. Наприклад, криволінійні залежності можуть апроксимуватися поліноміальною, показовою, степеневою, логарифмічною функціями. Ще більш неоднозначна ситуація для множинної регресії, оскільки наочне подання статистичних даних в цьому разі неможливе.

Правильний вибір виду економічної моделі є відправною точкою для якісного її аналізу. Безперечно, на практиці невідомо, яка модель є правильною, і часто підбирають таку модель, яка найточніше відповідає реальним даним. При цьому необхідно враховувати, що ідеальної моделі не існує. Тому, щоб вибрати якісну модель, необхідно відповісти на низку запитань, що виникають при її аналізі.

У ряді випадків досить очевидно, яка модель краща. В інших випадках для прийняття обгрунтованого рішення доводиться проводити достатньо глибокий та ґрунтовний порівняльний аналіз. Для цього необхідно вибрати критерії, які дозволять зробити обгрунтований

Вибір якісної моделі вимагає відповідей на такі запитання:

- Які ознаки «гарної» (якісної) моделі?
- Які помилки специфікації мають місце і які наслідки цих помилок?
- Як виявити помилку специфікації?
- Яким чином можна виправити помилку специфікації і перейти до кращої (якісною) моделі?

Рис. 6.1. Особливості вибору якісної моделі

висновок. Зазвичай для побудови «гарної» працездатної моделі і порівняння її з іншими можливими моделями необхідно враховувати відповіді на запитання, наведені на рис. 6.1.

Якісна модель – це модель, яка з достатнім рівнем відповідності описує зв'язки між системотвірними складовими досліджуваної системи і має такі властивості.

Простота моделі: модель повинна бути максимально простою. Ця властивість визначається тим фактом, що модель не відображає дійсність ідеально, а є її спрощенням. Тому з двох моделей, що приблизно однаково відображають реальність, перевага надається моделі, що містить меншу кількість пояснювальних змінних. Моделі мають бути «настільки простими, наскільки це можливо, але не простіше» (*А. Эйнштейн*).

Однозначність моделі: для будь-якого набору статистичних даних визначувані коефіцієнти повинні обчислюватися однозначно.

Максимальна відповідність моделі: рівняння тим краще, чим більшу частину розкиду залежної змінної воно може пояснити. Тому прагнуть побудувати рівняння з максимально можливим скоректованим коефіцієнтом детерміації R_2.

Узгодженість теорії з моделлю: модель обов'язково повинна спиратися на теоретичний фундамент, оскільки інакше результат використання регресійного рівняння може бути досить поганим. Наприклад, якщо у функції попиту коефіцієнт при ціні позитивний, то навіть значна величина коефіцієнта детерміації (наприклад 0,7) не дозволить визнати рівняння задовільним.

Прогнозні якості моделі: модель може бути визнана якісною, якщо отримані на її основі прогнози підтверджуються реальністю.

Іншим критерієм прогнозних якостей оціненої моделі регресії може служити таке відношення:

$$V = \frac{S}{\bar{y}}, \tag{6.1}$$

де $S = \sqrt{\frac{\sum \varepsilon_i^2}{n - m - 1}}$ – стандартна похибка регресії;

$\bar{y}$ – середнє значення залежної змінної рівняння регресії.

Якщо величина V прямує до нуля і економетрична модель відповідає умовам Гаусса – Маркова (див. підрозділ 2.6), то прогнозна якість моделі є високою.

Якщо рівняння регресії використовується для прогнозування, то величина V розраховується не для того періоду, на якому оцінювалося рівняння, а для деякого наступного за ним часового інтервалу, для якого відомі значення залежної і пояснювальних змінних. Тим самим на практиці перевіряються прогнозні якості моделі. У разі позитивного рішення, якщо можна спрогнозувати значення пояснювальних змінних на деякий подальший період, побудована модель обґрунтовано може бути використана для прогнозу значень залежної змінної Y. При цьому слід пам'ятати, що період прогнозування має бути в 3 рази коротший за період, за яким оцінювалося рівняння регресії.

Оскільки не існує якогось єдиного правила побудови регресійних моделей, аналіз перерахованих властивостей дозволяє будувати якісні економетричні моделі.

6.2. Види помилок специфікації

Одним із базових припущень побудови якісної моделі є правильна специфікація рівняння регресії.

Специфікація – визначення й перелік характерних особливостей об'єкта дослідження; математичний опис форми залежності між показниками моделі.

Специфікація рівняння регресії передбачає два основні підходи – відкидання незначущої змінної та додавання значущої змінної (рис. 6.2)

Правильна специфікація рівняння регресії означає, що воно в цілому правильно відображає співвідношення між економічними показниками, що беруть участь у моделі. Це є необхідною передумовою подальшого якісного оцінювання.

Помилки специфікації – це неправильний вибір функціональної форми або набору пояснювальних змінних.

Розглянемо основні типи помилок специфікації.

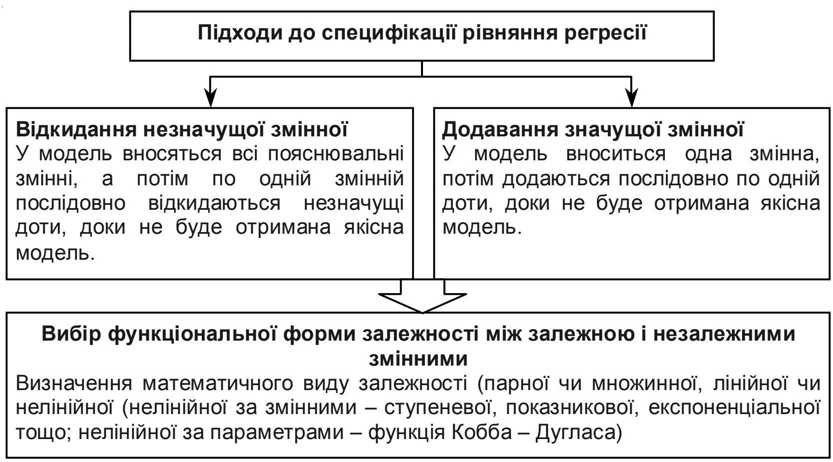


Рис. 6.2. Послідовність специфікації економетричної моделі

1. *Відкидання значущої змінної.*

Суть даної помилки і її наслідку наочно ілюструються таким прикладом. Нехай теоретична модель, що відображає дану економічну залежність, має вигляд:

$$y = a_0 + a_1 x_1 + a_2 x_2. \tag{6.2}$$

Даній моделі відповідає таке емпіричне рівняння регресії:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \varepsilon. \tag{6.3}$$

Дослідник за певних причин (недостатність інформації, поверхове знання про предмет дослідження тощо) вважає, що на змінну y впливає лише змінна x_1. У цьому разі обмежуємося розглядом моделі

$$y = \gamma_0 + \gamma_1 x_1 + \varepsilon. \tag{6.4}$$

При цьому не розглядається як пояснювальна зміна x_2, що свідчить про помилку відкидання істотної змінної.

2. *Додавання незначущої змінної.*

У деяких випадках у рівняння регресії вносять дуже багато пояснювальних змінних, причому не завжди обґрунтовано. Наприклад, нехай теоретична модель має такий вигляд:

$$y = a_0 + a_1 x_1. \tag{6.5}$$

Нехай дослідник замінює її більш складною моделлю:

$$y = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \varepsilon, \tag{6.6}$$

додаючи змінну, яка не здійснює реальної дії на y, пояснюючу змінну x_2. У такому разі спостерігається помилка врахування неістотної змінної.

3. *Вибір неправильної функціональної форми.*

Суть помилки проілюструємо таким прикладом. Нехай правильна регресійна модель має вигляд

$$Y = a_0 + a_1 x_1 + a_2 x_2. \tag{6.7}$$

Будь-яке емпіричне рівняння регресії з тими самими змінними, але іншого функціонального виду, призводить до викривлення дійсної залежності. Наприклад, у таких рівняннях:

$$\ln y = a_0 + a_1 x_1 + a_2 x_2, \tag{6.8}$$

$$y = c_0 + c_1 \ln x_1 + c_2 \ln x_2 \tag{6.9}$$

здійснена помилка вибору неправильної функціональної форми рівняння регресії. Здебільшого така помилка призводить або до отримання зміщених оцінок, або до погіршення статистичних властивостей оцінок коефіцієнтів регресії та інших показників якості рівняння. Прогнозні якості моделі в такому разі є дуже низькими.

6.3. Виявлення та коригування помилок специфікації

При побудові рівнянь регресії, особливо на початкових етапах, нерідко мають місце помилки специфікації (рис. 6.3).

Важливо виявити й виправити ці помилки. Складність процедури визначається типом помилки і знаннями про досліджуваний об'єкт.

Якщо в рівнянні регресії є одна неістотна змінна, то вона буде виявлена за низькою *t-статистикою.* Надалі цю змінну виключають з розгляду.

Якщо в рівнянні декілька статистично незначущих пояснювальних змінних, то слід побудувати інше рівняння регресії без цих незначущих змінних. Потім за допомогою *F-статистики* порівнюються коефіцієнти детерміації для первинного і додаткового рівнянь регресій:

$$F = \frac{R_1^2 - R_2^2}{1 - R_1^2} \cdot \frac{n - m - 1}{k}, \tag{6.10}$$

де n – кількість спостережень;
m – кількість пояснювальних змінних у початковому рівнянні;
k – кількість відкинутих з початкового рівняння пояснювальних змінних.

За наявності кількох неістотних змінних, можливо, має місце мультиколінеарність.

Рис. 6.3. Помилки початкових етапів побудови економетричної моделі

Проте здійснення вказаних перевірок має сенс лише в разі правильного підбору виду (функціональної форми) рівняння регресії, що можна здійснити, якщо погоджувати його з теорією. Наприклад, при побудові кривої Філіпса, яка вказує, що залежність між заробітною платою y_i і безробіттям x є зворотною, можливі такі моделі:

$$
\begin{aligned}
&y = a + \alpha x + \varepsilon, \alpha < 0,\\
&\ln y = a + \alpha \ln x + \varepsilon, \alpha < 0,\\
&y = a + \alpha \frac{1}{x+\gamma} + \varepsilon, \alpha > 0,\\
&y = a + a^{\alpha x} + \varepsilon, \alpha < 0.
\end{aligned}
\tag{6.11}
$$

Зазначимо, що вибір моделі далеко не завжди здійснюється однозначно, і надалі потрібно порівнювати модель як з теоретичними, так і з емпіричними даними, удосконалювати її. При визначенні якості моделі зазвичай аналізуються параметри, наведені на рис. 6.4.

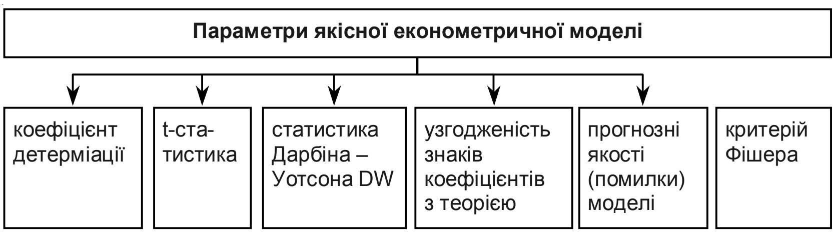


Рис.6.4. Аналіз якості економетричної моделі[1]

---

[1] ***Коефіцієнт детермінації*** (*R^2 – R-квадрат)* – це частка дисперсії залежної змінної, яка пояснюється пояснювальними змінними. Більш точно – це одиниця мінус частка непоясненої дисперсії (дисперсії випадкової помилки моделі, або умовної за факторами дисперсії залежної змінної) в дисперсії залежної змінної. Його розглядають як універсальну міру зв'язку однієї випадкової величини й безлічі інших. В окремому випадку лінійної залежності коефіцієнт детермінації є квадратом так званого множинного коефіцієнта кореляції між залежною змінною і пояснювальними змінними. Зокрема, для моделі парної лінійної регресії коефіцієнт детермінації дорівнює квадрату звичайного коефіцієнта кореляції між y і x.

Для того, щоб була можливість порівнювати моделі з різною кількістю факторів так, щоб кількість регрессрів (факторів) не впливала на статистику R^2 зазвичай використовується скоригований коефіцієнт детермінації, у якому використовуються незміщені оцінки дисперсій: $R^2_{adj} = 1 - (1 - R^2)\frac{n-1}{n-k} \le R^2$, де n – кількість спостережень, а k – кількість параметрів.

Якщо всі ці показники задовільні, то ця модель може бути запропонована для опису досліджуваного реального процесу. Якщо ж якась із описаних вище характеристик не є задовільною, тобто існують підстави сумніватися щодо якості цієї моделі (неправильно вибрана функціональна форма рівняння; не врахована важлива пояснювальна змінна; є пояснювальна змінна, що не справляє значного впливу на залежну змінну).

Для більш докладного аналізу адекватності моделі може бути запропоноване дослідження залишкового члена моделі.

6.4. Дослідження залишкового члена моделі

Графічне подання поведінки залишкового члена e (тобто графічне подання випадкових відхилень e_i, $i = 1,2,...,n$) дозволяє перш за все проаналізувати наявність автокореляції й гетероскедастичності (непостійність дисперсій відхилень).

Крім того, за допомогою графічного подання відхилень ε може бути також виявлена неправильна специфікація рівняння. Для цього будується графік залежності величин відхилень ε від номера спостереження i (рис. 6.5). Якщо залежність, зображена на цьому графіку,

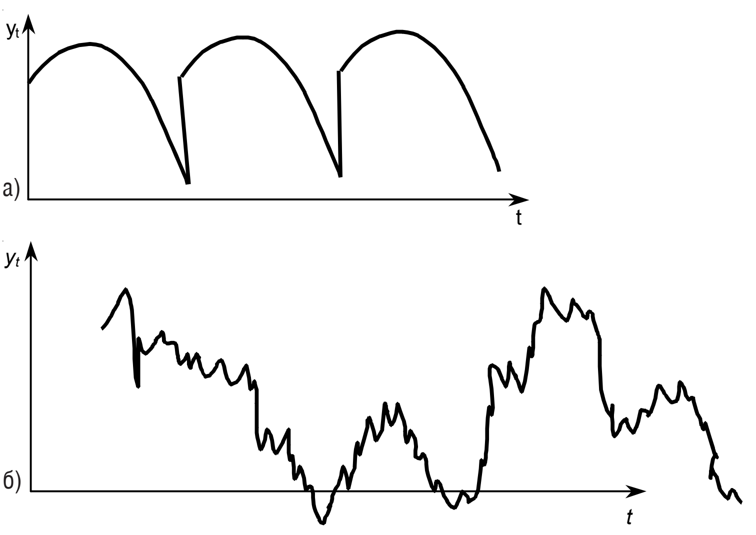


Рис 6.5. Графічне зображення залежності величин відхилень ε від номера спостереження *i:* а – функціональний зв'язок; б – випадковий зв'язок

має регулярний (невипадковий) характер (рис. 6.5а), то це означає, що досліджуване рівняння регресії неправильно специфіковане.

Існуть й інші тести виявлення помилок специфікації, серед яких можна виділити:

1. Тест Рамсея RESET (*Regression specification error test)*[1].
2. Тест (критерій) максимальної правдоподібності (*The Likelihood Ratio test)*[2].
3. Тест Вальда (*The Wald test*)[3].

---

[1] RESET – тест Рамсея (скор. від англ. Regression Equation Specification Error Test) – процедура, яка використовується в економетриці для тестування функціональної форми (специфікації) моделі. Тест, запропонований Рамсеєм в 1969 році, заснований на допоміжній регресії залежної змінної на фактори вихідної моделі плюс різні ступені оцінених за вихідною моделлю значень залежної змінної

$$y_t = x_t^T b + a_2 \hat{y}_t^2 + a_3 \hat{y}_t^3 + \ldots + a_m \hat{y}_t^m + u_t .$$

Далі необхідно перевірити гіпотезу: $H_0 : a_2 = a_3 = \ldots = a_m = 0$. Якщо значення статистики більше критичного, то нульова гіпотеза відкидається і специфікація моделі визнається неправильною. В іншому випадку функціональна форма моделі є прийнятною. Метод максимальної правдоподібності або метод найбільшої правдоподібності (ММП, ML, MLE – Maximum Likelihood Estimation) у математичній статистці – це метод оцінювання невідомого параметра шляхом максимізації функції правдоподібності. Заснований на припущенні про те, що вся інформація про статистичну вибірку міститься у функції правдоподібності.

[2] Метод максимальної правдоподібності був проаналізований, рекомендований і значно популяризований Р. Фішером у 1912–1922 роках (хоча раніше його використовували Гаусс, Лаплас та ін.). Наприклад, ви цікавитеся таким антропометричним параметром, як зростання населення України. Припустимо, у вас є дані про зростання деякої кількості людей, а не всього населення. Крім того передбачається, що зростання є нормально розподіленою величиною з невідомою дисперсією і середнім значенням. Середнє значення і дисперсія зростання у вибірці є максимально правдоподібними до середнього значення і дисперсії всього населення.

[3] Тест Вальда (англ. Wald test) – статистичний тест, який використовується для перевірки обмежень на параметри статистичних моделей, оцінених на основі вибіркових даних. Є одним з трьох базових тестів перевірки обмежень поряд з тестом відношення правдоподібності і тестом множників Лагранжа. Тест є асимптотичним, тобто для достовірності висновків потрібен досить великий обсяг вибірки.

Нехай є економетрична модель з вектором параметрів b. Необхідно перевірити за вибірковими даними гіпотезу H_0: $g(b) = 0$, де g – сукупність (вектор) деяких параметрів функцій. Ідея тесту полягає в тому, що якщо нульова гіпотеза правильна, то і вибірковий вектор $g(\hat{b})$ повинен бути в певному сенсі близький до нуля.

4. Тест множника Лагранжа (*The Lagrange multiplier test*)[1].
5. Тест Хаусмана (*The Hausman test*)[2].
6. Box-cox перетворення (*Box-Cox transformation*)[3].

За наявності хоча б однієї незадовільної відповіді за яким-небудь тестом модель удосконалюється з метою усунення виявленого недоліку.

У разі позитивних відповідей за всіма проведеними тестами модель вважається якісною. Вона використовується для аналізу і прогнозу з'ясовної змінної.

---

[1] Метод множників Лагранжа – метод знаходження умовного екстремуму функції $f(x)$, де $x \in R^n$ щодо m обмежень $\varphi_i(x) = 0$, де i змінюється від одиниці до m.

Складемо функцію Лагранжа у вигляді лінійної комбінації функцій f і функцій φ_i, узятих з коефіцієнтами, які називаються множниками Лагранжа – λ_i:

$$L(x,\lambda) = f(x) + \sum_{i=1}^{m} \lambda_i \varphi_i(x) \text{, де } \lambda = (\lambda_1, ..., \lambda_m) .$$

Складемо систему $n + m$ рівнянь, прирівнявши до нуля частинні похідні функції Лагранжа $L(x, \lambda)$ по x_j та λ_i.

Якщо отримана система має рішення щодо параметрів x_j' і λ_i', тоді точка x' може бути умовним екстремумів, то є рішенням вихідної задачі. Зауважимо, що це умова носить необхідний, але не достатній характер.

[2] Тест Хаусмана, званий також тестом Ву – Хаусмана або Дарбіна – Ву – Хаусмана – тест, який застосовується в економетриці для порівняння моделей, оцінених різними методами, один з яких дозволяє отримати спроможні оцінки і при нульовій і при альтернативній гіпотезі, а інший – тільки при нульовій гіпотезі.

Нульова гіпотеза полягає в тому, що фактори моделі є екзогенними, альтернатива – ендогенними. Якщо статистика перевищує критичне значення, регресори моделі не можна вважати екзогенними, тому краще використовувати метод інструментальних змінних. В іншому випадку можна вважати, що регресори є не гіршими, ніж інструменти, і застосовувати звичайний МНК. На першому етапі оцінюється МНК – регресія факторів на інструменти. На другому – оцінюється (також за допомогою МНК) регресія, що пояснюється змінною на вихідні фактори й оцінки цих факторів, отриманих на першому етапі (або залишки регресій, отриманих на першому етапі). Якщо коефіцієнти при додаткових змінних спільно значимі (перевіряється стандартними тестами – тест Вальда, F-тест , t-статистики), то регресори є ендогенними.

[3] Для вихідної послідовності X довжиною N $X = x_0, x_1, ..., x_{N-1}$

Однопараметричне Бокс – Кокс перетворення визначається наступним чином:

$$x_i(\lambda) = \begin{cases} \dfrac{x_i^{\lambda} - 1}{\lambda} \\ \ln(x_i) \end{cases} \text{, } \lambda \neq 0, \lambda = 0, \text{ де } i = 0, 1, 2, ..., N-1.$$

Як бачимо, це перетворення має єдиний параметр – лямбда. При значенні лямбда рівному нулю здійснюється логарифмічне перетворення вхідної послідовності, при значенні лямбда відмінному від нуля – ступеневий. Якщо параметр лямбда дорівнює одиниці, то закон розподілу вихідної послідовності не змінюється, хоча при цьому послідовність отримає зрушення за рахунок віднімання одиниці з кожного її значення.

Проте необхідно застерегти від абсолютизації отриманого результату, оскільки навіть якісна модель є підгонкою специфікації моделі під наявний набір даних. Тому цілком реальна картина, коли дослідники, які володіють різними наборами даних, будують різні моделі для пояснення однієї і тієї самої змінної. Проблематичним є і використання моделі для прогнозування значень залежної змінною. Іноді гарні з погляду діагностичних тестів моделі мають досить низькі прогнозні якості.

Один із головних напрямів економічного аналізу – постійне вдосконалення моделей. Тут слід зазначити, що якогось глобального підходу, який визначає заздалегідь можливі шляхи вдосконалення, немає і, швидше за все, бути не може. Дослідник повинен пам'ятати, що досконалої моделі не існує. Через мінливі умови перебігу економічних процесів, не може бути й постійно якісних моделей. Нові умови вимагають перегляду навіть досить стійких моделей.

До цього часу достатньо спірним є питання, як будувати моделі:

а) починати з найпростішою і постійно ускладнювати її;
б) починати з максимально складної моделі і спрощувати її на основі досліджень, що здійснюються.

І той, і інший підхід мають як переваги, так і недоліки. Наприклад, якщо слідувати схемі а), то відбувається звичайна підгонка моделі під емпіричні дані. При теоретично більш виправданому підході б) пошук можливих напрямів вдосконалення моделі часто зводиться до повного перебору, що робить аналіз неефективним. На етапах спрощення моделі можливе також відкидання пояснювальних змінних, які були б корисні в спрощеній моделі. Отже, побудова моделі є індивідуальною в кожній конкретній ситуації і спирається на серйозні знання економічної теорії і статистичного аналізу.

Проте зазначимо, що попри всі недоліки моделей прийняття на їх основі рішень забезпечує в цілому набагато точніші результати, ніж при прийнятт+і рішень лише на основі інтуїції й економічної теорії.

Комбіноване групування – формування груп за двома або більше ознаками, які були взяті в певному поєднанні, називається комбінованим групуванням. При цьому групувальні ознаки прийнято розміщувати, починаючи з пріоритетного, у певній послідовності, виходячи із логіки взаємозв'язків показників.

Прикладом комбінованого групування може слугувати розділення сформованих груп за формами господарювання на підгрупи за рівнем рентабельності (дохідності) чи за іншими ознаками (продуктивність праці, фондовіддача та ін.).

Проте в міру зростання обсягів інформації, що переробляється, і, зокрема, числа об'єктів, що класифікуються, і ознак, які їх характеризують, можливість ефективної реалізації подібної логіки дослідження ставала все менш реальною. Саме електронно-обчислювальна техніка стала тим головним інструментом, який дозволив по-новому підійти до вирішення цієї важливої проблеми і, зокрема, конструктивно скористатися розробленим до цього часу могутнім апаратом багатовимірного статистичного аналізу: методами розпізнавання образів «з навчанням» (аналіз дискримінанта) і «без навчання» (автоматична класифікація, або кластер-аналіз).

Розвиток електронно-обчислювальної техніки як засоби обробки великих масивів даних стимулювало здійснення останніми роками широких комплексних досліджень складних соціально-економічних, технічних, медичних й інших процесів і систем, таких як спосіб і рівень життя населення, удосконалення організаційних систем, регіональна диференціація соціально-економічного розвитку, планування й прогнозування галузевих систем, закономірності виникнення перебоїв (у техніці) або захворювань (у медицині) і т. п. У зв'язку з багатоплановістю та складністю цих об'єктів і процесів дані про них мають багатовимірний і різний характер, оскільки до їх аналізу зазвичай буває неясно, наскільки важлива та чи інша властивість для конкретної мети. За цих умов виходять на перший план проблеми побудови угруповань і класифікацій за багатовимірними даними (тобто проблеми класифікації *багатовимірних спостережень)*, причому з'являється можливість оптимізації цієї побудови з погляду найбільшої відповідності отриманого результату поставленій кінцевій меті класифікації (рис. 6.7).

Цілі класифікації значною мірою розширюються, і одночасний зміст самого процесу класифікації стає набагато складнішим. Він, зокрема, доповнюється *проблемою побудови самої процедури класифікації*, що раніше мала суто технічний характер.

Класифікація – це поділ певної сукупності об'єктів або явищ на однорідні[1] групи або віднесення кожного із заданої безлічі об'єктів до одного із визначених класів.

[1] Коефіцієнт варіації має дорівнювати не більше ніж 33%, значення досліджуваної ознаки для кожного об'єкта виділеної групи має незначною мірою відрізнятися від середньої величини.

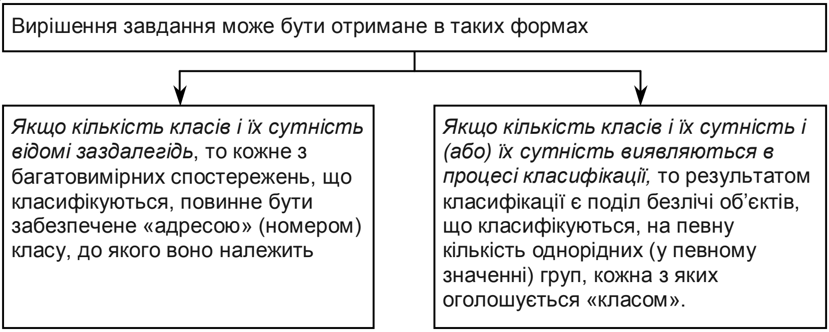


Рис. 6.7. Форми класифікації об'єктів дослідження

Одновимірна класифікація – це поділ певної сукупності об'єктів або явищ на однорідні групи за одним критерієм.

Багатовимірна класифікація – поділ об'єктів сукупності, у якому кожна група може виділятися більш ніж за однією ознакою.

6.6. Класифікація з навчанням

Якщо дослідник використовує для вирішення завдання не тільки дані, що класифікуються, а й навчальні вибірки, то вирішується *завдання класифікації за наявності навчальних вибірок (“класифікація з навчанням”)*; в іншому випадку мова йде про завдання “*класифікації без навчання*”.

Приклад

Виділення типів споживчої поведінки населення, аналіз стратегій кожного з типів поведінки, прогноз структури споживання.

Як початкову інформаційної базу використовують дані бюджетних обстежень сімей. Кожна сім'я характеризується, з одного боку, певним набором *X чинників-детермінантів* (соціально-демографічні й інші ознаки, що описують умови життєдіяльності сім'ї), а з іншого – набором *Y* параметрів («змінних поведінки»), у яких відбиваються її фактичні потреби.

Як соціально-демографічні чинники, що мають істотне значення для вивчення споживчих аспектів соціального життя, доцільно використовувати, наприклад, суспільну і національну належність, рівень освіти і кваліфікацію, характер праці, демографічний тип і вік сім'ї, тип населеного пункту і характер житла, величину і структуру майна, рівень доходів.

Відмінності в потребах, які складаються під впливом соціально-демографічних і природно-кліматичних умов, є такими, що об'єктивно існують; вони формують поведінку споживача в конкретно-історичних умовах і, у свою чергу, породжують однорідні типи споживачів, орієнтовані на істотно різне споживання.

Весь комплекс соціально-демографічних та інших чинників, що істотно впливають на структуру споживання, називається *типотвірним*. Вони мають визначальне значення, тоді як усі інші дають лише випадкову варіацію в межах однієї групи (типу) споживчої поведінки.

Як ознаку поведінки *Y* можна розглядати три групи параметрів:

а) рівень і структуру споживання;
б) характер (обсяг та зміст) використання вільного часу;
в) інтенсивність зміни соціального, трудового, демографічного статусу.

Отже, розглядаються числові характеристики і градації типотвірних й одночасно поведінкових ознак кожної сім'ї з аналізованої сукупності.

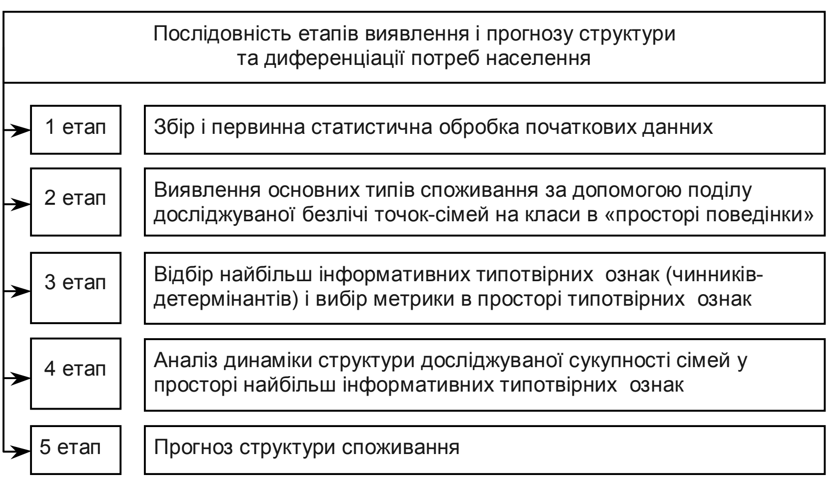


Рис. 6.8. Послідовність етапів типології споживчої поведінки населення, аналіз, прогнозу структури споживання

Вирішення загальної проблеми, пов'язаної з виявленням і прогнозом структури і диференціації потреб населення, розпадається на такі етапи.

1. *Збір і первинна статистична обробка початкових даних.* Досліджувані об'єкти (сім'ї) поступають як багатовимірні спостереження або точки у двох багатовимірних просторах ознак. Фіксуючи координати цих точок значеннями (або градаціями) типотвірних змінних X (тобто чинників-детермінантів), розглядаємо їх у «просторі стану» $П(X)$ – просторі, координатами якого служать основні показники життєдіяльності сімей. Фіксуючи координати тих самих об'єктів, значення показників Y їх споживчої поведінки, розглядаємо їх у «просторі поведінки» $П(Y)$. Очевидно, *за належного вибору метрики*[1] *в просторах $П(X)$ і $П(Y)$ геометрична близькість двох точок у $П(X)$ означатиме схожість умов життєдіяльності відповідних двох сімей, так само як і геометрична близькість точок в $П(Y)$ означатиме схожість їх споживчої поведінки*. Серед методів первинної статистичної обробки аналізованих даних досить поширеними й корисними є методи вивчення різних одно-, дво- і тривимірних емпіричних розподілів, які зводяться до побудови і різних подань (графічних, табличних) згаданих вище комбінованих групувань. Приклад табличного подання одного з таких двовимірних комбінованих групувань наведений в табл. 6.1.

Це комбіноване групування побудоване на основі статистичного обстеження 400 сімей за двома ознаками з простору $П(X)$: за $x^{(1)}$ (грн) – величиною середньодушового сімейного доходу (з трьома градаціями: «низький», «середній» і «високий») і за $x^{(2)}$ – якістю житлових умов (з чотирма градаціями: «низька», «задовільна», «добра» і «дуже доб-

Таблиця 6.1. Двовимірне комбіноване групування статистичного обстеження сімей

Градація ознаки $x^{(1)}$ (дохід)	Градація ознаки $x^{(2)}$ (якість житлових умов)				Сума
	низька	задовільна	добра	дуже добра	
Низький	24	12	4	0	40
Середній	20	100	140	20	280
Високий	4	8	28	40	80
Сума	48	120	172	60	400

[1] Метрика – функція, яка визначає відстань у метричному просторі.
Метричним простором називається множина, у якій визначена відстань між будь-якою парою елементів.

ра»). Кожна клітинка таблиці відповідає класу, отриманому в результаті проведеного комбінованого групування; усередині клітинки позначено кількість сімей, що мають дане поєднання градацій аналізованих ознак (такі таблиці називають також «таблицями зв'язаності»).

2. *Виявлення основних типів споживання за допомогою поділу досліджуваної множини точок-сімей на класи в «просторі поведінки» П(Y)*. Гіпотеза існування «природних», об'єктивно обумовлених типів поведінки, тобто якоїсь невеликої кількості класів сімей, таких, що сім'ї одного класу характеризуються порівняно схожою, однотипною споживчою поведінкою, геометрично означає розпад досліджуваної в «просторі поведінки» сукупності точок-сімей на відповідну кількість «згустків» або «скупчень» точок. Виявивши за допомогою відповідних методів багатовимірного статистичного аналізу (кластер-аналізу, таксономії) ці класи-згустки, тим самим визначимо основні типи споживчої поведінки.

3. *Відбір найбільш інформативних типотвірних ознак (чинників-детермінантів) і вибір метрики в просторі типотвірних ознак*. Очевидно, неправомірно розраховувати на те, що діапазони можливих значень кожного з кандидатів в типотвірні ознаки виявляться непересічними для сімей з різним типом споживчої поведінки. Інакше кажучи, значення кожної з ознак $x^{(i)}$ окремо і їх набору в сукупності схильні до деякого неконтрольованого розкиду при аналізі сімей усередині кожного з типів споживання. Природно вважати найбільш інформативними ті чинники-детермінанти або ті їх набори, різниця в законах розподілу яких є найбільшою при переході від одного класу споживчої поведінки до іншого. Ця ідея і покладена в основу методу відбору найбільш інформативних (типотвірних) ознак-детермінантів.

4. *Аналіз динаміки структури досліджуваної сукупності сімей у просторі найбільш інформативних типотвірних ознак*. Кінцевою метою цього етапу є прогноз тих поступових перетворень класифікаційної структури сукупності споживачів (сімей, що розглядаються в просторі типотвірних ознак), які повинні відбутися з часом.

5. *Прогноз структури споживання*. На цьому етапі дослідження спираємося на результати, отримані під час проведення попереднього етапу, тобто виходимо із заданої класифікаційної структури споживачів у період часу, що цікавить нас, у майбутньому. Відновлюючи класифікаційну структуру споживання (класифікаційну структуру сукупності сімей у просторі ознак *П(Y)*, що характеризують споживчу поведінку сім'ї) за класифікаційною структурою споживачів (за класифікаційною структурою тієї самої сукупності, але в просторі типотвір-

них ознак), будемо відносити кожну конкретну сім'ю до того типу споживання, для якого значення тих, що характеризують її типотвірні ознаки є найбільш типовими.

Завдання для самоконтролю

1. Які аспекти дослідження відображає якісно побудована економіко-математична модель?
2. Назвати та схарактеризувати основні властивості «гарної» моделі, яку з даних властивостей вважають набільш пріоритетною.
3. Які виділяють основні помилки специфікації?
4. Схарактеризувати методики коригування помилок специфікації.
5. Назвати основні тести виявлення помилок специфікації, навести методики їх розрахунків.
6. Перерахувати основні етапи дослідження проблеми специфікації, зазначити математичне забезпечення кожного етапу.
7. Навести приклад класифікації з навчанням.

Практичне заняття 6

Кластерний аналіз

Постановка завдання

1 етап	Побудова інформаційного масиву дослідження шляхом збору статистичних даних та розробка моделі парної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), анаіз моделі
2 етап	Побудова множинної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), дослідження моделі
3 етап	Виявлення, дослідження та усунення мультиколінеарності факторів у моделі множинної лінійної регресії методом Фаррара – Глобера
4 етап	Виявлення ефекту гетероскедастичності на основі застосування параметричного тесту Гольдфельда – Квандта та усунення даного ефекту шляхом використання узагальненого методу найменших квадратів (методу Ейткена)
5 етап	Побудова адитивної та мультиплікативної моделі часового ряду на основі проведення декомпозиційного аналізу: визначення структури ряду (тенденції та сезонних (циклічних) коливань), проведення автокореляційного аналізу, визначення якості моделі, розрахунок прогнозних значень
6 етап	**Групування елементів економічної системи шляхом кластерного аналізу, побудова деревоподібної дендограми, графічне подання елементів кластеризації**
7 етап	Перевірка наявності автокореляції залишків за допомогою критерію Дарбіна – Уотсона, оцінка параметрів рівняння множинної лінійної регресії методом Ейткена
8 етап	Побудова економіко-математичної моделі методом інструментальних змінних; побудова економетричної моделі на основі даних, що містять помилки вимірювання змінних, оцінка параметрів моделі за допомогою алгоритму Уоліса
9 етап	Побудова моделі розподіленого лага, оцінка параметрів моделей з лагами в незалежних змінних, оцінювання параметрів авторегресійних моделей; перетворення ARMA і ARIMA
10 етап	Структурне моделювання (моделювання на основі системи структурних рівнянь), оцінка невідомих параметрів моделі, визначення адекватності моделі щодо опису структури вхідних даних

Мета роботи: набуття навичок групування елементів економічної системи з використанням пакету STATISTICA.

Зміст роботи: за допомогою пакету STATISTICA провести кластерний аналіз обраних для аналізу банків на основі вихідних даних.

Початкові дані

На основі самостійно зібраної інформації з обраної сукупності банків (на оцінку 5 балів) необхідно:

- провести кластеризацію країн за відповідними показниками;
- кластеризацію виконати для різної кількості кластерів (3–5) різними методами;
- порівняти склад кластерів, що були отримані при застосуванні даних методів;
- побудувати деревоподібні дендограми;
- надати графічне представлення елементів проведеної кластеризації банків.

Для вирішення задачі використати: інструктивні матеріали, можливості пакету STATISTICA.

Вимоги до оформлення звіту

Звіт про проведення даної лабораторної роботи оформлюється разом з іншими звітами цього курсу в окремому зошиті згідно з встановленими вимогами до оформлення звітів і містить:

- назву, тему, завдання, опис мети лабораторної роботи;
- вихідні дані варіанта;
- результати рішення;
- висновок про результати групування банків на основі застосування кластерного аналізу;
- короткий опис технології вирішення задачі в Statistica.

Хід виконання

Кластерний аналіз – один із методів багатовимірного аналізу, призначений для групування (кластеризації) сукупності елементів, які характеризуються багатьма чинниками, і отримання однорідних груп (кластерів). Поділ на кластери відбувається за допомогою деякої метрики, наприклад, евклідової відстані. Завдання кластерного аналізу полягає в поданні початкової інформації про елементи в стислому вигляді без її істотної втрати.

Приклад

Розглянемо двадцять банків, акції яких котуються на ринку, що надали таку інформацію (табл. 6.2), де X – витрати за минулий період, Y – прибуток за минулий період. Необхідно з'ясувати, акції яких банків має сенс придбати (*Buy*), яких – притримати (*Hold*), а від яких – позбавитися (*Sell*).

Графічне подання початкових даних наводиться на рис. 6.9. Порівнюючи першу (1, 20, 7,16, 8, 17, 11) і другу групу банків (2, 15, 6, 3, 12, 19), можна сказати, що друга група має переваги, оскільки при одних і тих самих витратах друга група одержує більше прибутку. Порівнюючи першу групу банків з третьою (4, 13, 5, 9, 10, 14, 18), бачимо, що переважає перша група, оскільки при одному і тому самому прибутку витрати в неї менші.

Таблиця 6.2. Вхідна інформація характеристики досліджуваної сукупності банків

Номер банку	Витрати *X*	Прибуток *Y*	Рекомендація
1	4	2	Hold
2	6	10	Buy
3	5	7	Buy
4	12	3	Sell
5	17	4	Sell
6	3	10	Buy
7	6	1	Hold
8	6	3	Hold
9	15	1	Sell
10	15	4	Sell
11	5	4	Hold
12	3	8	Buy
13	13	5	Sell
14	15	3	Sell
15	5	9	Sell
16	8	3	Hold
17	6	4	Hold
18	14	3	Sell
19	4	8	Buy
20	4	3	Hold

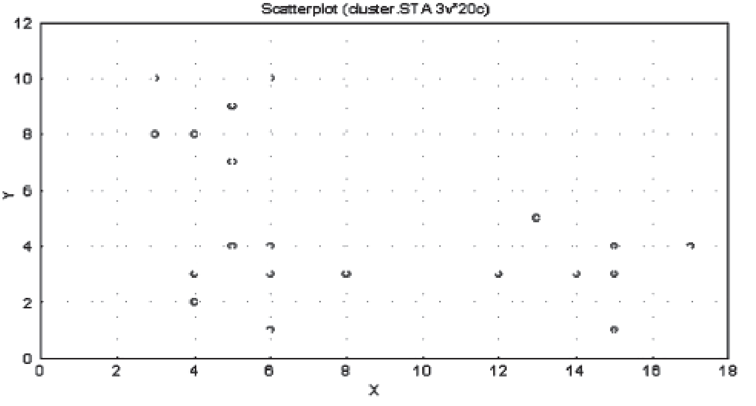


Рис. 6.9. Графічне подання вхідного масиву даних за показниками характеристики банків

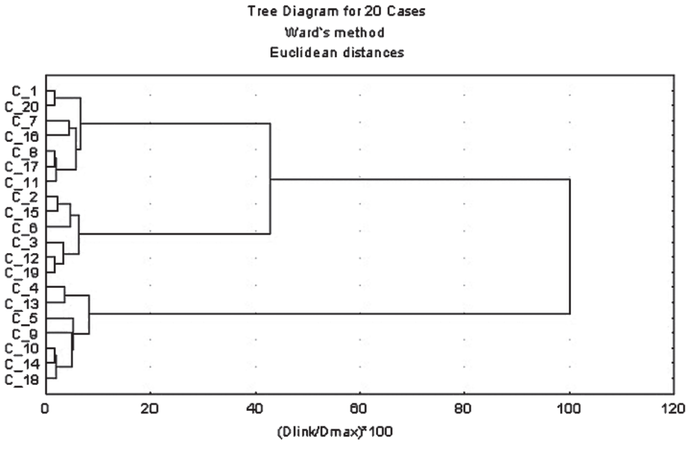


Рис. 6.10. Деревоподібна дендограма кластеризації банків

Аналогічні результати були отримані з використанням пакету STATISTICA і його модуля "Cluster modules" (рис. 6.10).

Проте на практиці не все так просто, як здається. При розв'язанні задач кластерного аналізу доводиться мати справу з низкою проблем:

- елементи (у нашому випадку банки) характеризуються великою кількістю чинників, які мають різні одиниці вимірювання і різні абсолютні величини, що буквально не зіставні один з одним і мають різний обсяг інформації;
- спочатку невідома кількість кластерів, на яку необхідно поділити початкову сукупність елементів, і візуальні спостереження в багатовимірному випадку просто не приводять до успіху;
- які метрики використовувати як міру відстані (заходи близькості) між елементами;
- яку цільову функцію або метод використовувати для об'єднання елементів у кластери.

Нормування (стандартизація) даних

У кластерному аналізі поділ на кластери значною мірою залежить від абсолютних значень початкових даних. Цю проблему вирішують за допомогою нормування (стандартизації). Для цього з усіх значень за кожним чинником віднімають вибіркове середнє цього чинника і одержані різниці ділять на середнє квадратичне відхилення.

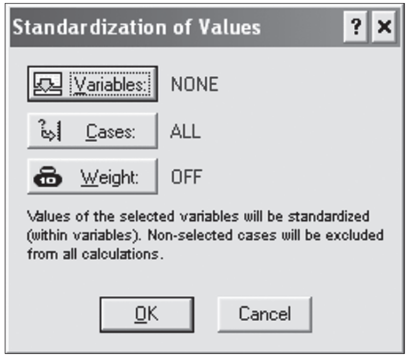


Рис. 6.11. Форма вікна в пакеті STATISTICA стандартизації даних

При цьому стандартизованные значення матимуть вибіркові середні, що дрівнюють нулю, а вибіркові дисперсії дорівнюють одиниці. Інакше кажучи, ми всі чинники звели в одну вагову категорію, як боксерів перед змаганнями. Для здійснення цієї операції в пакеті STATISTICA потрібно викликати модуль Data Management (Данные/ Стандартизация) (рис. 6.11).

Методи кластерного аналізу

Передбачається, що матриця початкових даних має вигляд *X[n,k]*, де *n* – виміри, *k* – кількість чинників. Тому при кластеризації елементів у пакеті STATISTICA слід вибирати режим: *cases* (*rows*) – рядки, а при кластеризації чинників: *variables* (*columns*) – стовпці.

Як основні методи аналізу пакет STATISTICA пропонує *Joining* (*tree clustering*) – групу ієрархічних методів, які використовуються в тому разі, якщо кількість кластерів наперед невідома, і *K-Means Clustering* (метод К-середніх), у якому користувач наперед визначає кількість кластерів.

До групи ієрархічних методів входить Ward's method – метод Уорда, який добре працює з невеликою кількістю елементів і націлений на вибір кластерів з приблизно однаковою кількістю членів.

Як метрику відстані пакет пропонує різні заходи, але найбільш споживаними є *Euclidean distance* (евклідова відстань), або *Squared Euclidean distance* (квадратична евклідова відстань).

Стосовно цільової функції, то однією з найбільш поширених є внутрішньогрупова сума квадратів. При використанні такої цільової функції алгоритм кластерного аналізу може зводитися до такого: якщо є *n* елементів і матриця відстаней між ними, то спочатку вважається, що кожен елемент є окремий кластер. Потім на кожному кроці об'єднуються такі два кластери, які приводять до мінімального збільшення цільової функції.

Різноманіття алгоритмів кластерного аналізу часто дезорієнтує користувача. Тому він може вдатися до застосування кількох алгоритмів і надати перевагу якомусь висновку на підставі комплексної оцінки сукупності результатів роботи.

Перевірка результатів контрольного прикладу в Cluster Analysis
Приклад 2

У даному прикладі буде досліджено 16 відомих інвестиційних фондів з метою оцінки їх стану. Як змінні використовуються такі характеристики: прибутковість за п'ятирічний період, ризик, щорічний відсоток доходу, витратна частина і податкові рейтинги. Початкові дані про інвестиційні фонди наведені в табл. 6.3.

Після введення даних та їх стандартизації (рис. 6.12) викликається модуль кластерного аналізу.

Таблиця 6.3. Вхідна інформація характеристики досліджуваної сукупності інвестиційних фондів

Fund	Five	Risk	Per90	Per91	Per92	Per93	Per94	Expens	Tax	Recom
F Chip	16476	2	10	25	6	55	4	1,22	89	Buy
FContra	15476	2	-1	21	16	55	4	1,03	90	Buy
F Destiny	14757	3	4	26	15	39	-3	0,7	69	Buy
Vista A	15145	4	-1	20	13	71	-6	1,49	96	Hold
Berger 100	15596	5	-7	21	9	89	-6	1,7	95	Hold
Gab Assett	13640	1	0	22	15	18	-6	1,33	85	Buy
Neub Focus	14081	3	1	16	21	25	-6	0,85	75	Buy
F Magellan	13827	3	-2	25	7	41	-5	0,96	73	Buy
Janus	13187	2	-1	11	7	43	-1	0,91	85	Sell
L Mason	13029	4	1	12	11	35	-17	1,82	92	Hold
Gabelli Gr.	12301	3	-3	11	4	34	-2	1,41	80	Sell
Franklin	11793	2	3	7	3	27	2	0,77	90	Sell
Janus 20	12441	4	-7	3	2	69	1	1,02	95	Sell
AARP	11728	4	-10	16	5	41	-16	0,97	68	Sell
Kemper	11386	4	-6	2	-2	67	4	1,09	86	Sell
20th Cent Gr	11258	4	-8	15	-4	32	0	1	60	Sell

1 Five	2 Risk	3 Per90	4 Per91	5 Per92	6 Per93	7 Per94	8 Expens	9 Tax	10 Recom
1,81029	-1,0342	2,27629	1,20194	-0,2924	0,44335	1,151	0,23987	0,54591	-1,1441
1,20045	-1,0342	0,1339	0,67865	1,16941	0,44335	1,151	-0,3435	0,63689	-1,1441
0,76197	-0,1149	1,10772	1,33276	1,02323	-0,3732	0,04919	-1,3567	-1,2738	-1,1441
0,99859	0,80437	0,1339	0,54782	0,73088	1,25989	-0,423	1,06884	1,1828	-0,0673
1,27363	1,72364	-1,0347	0,67865	0,14618	2,17849	-0,423	1,7136	1,09181	-0,0673
0,08077	-1,9535	0,32866	0,80947	1,02323	-1,4449	-0,423	0,5776	0,18197	-1,1441
0,34971	-0,1149	0,52343	0,02453	1,90029	-1,0877	-0,423	-0,8961	-0,7279	-1,1441
0,19481	-0,1149	-0,0609	1,20194	-0,1462	-0,2711	-0,2656	-0,5584	-0,9098	-1,1441
-0,1955	-1,0342	0,1339	-0,6296	-0,1462	-0,169	0,36399	-0,7119	0,18197	1,00947
-0,2919	0,80437	0,52343	-0,4988	0,43853	-0,5773	-2,1544	2,08203	0,81886	-0,0673
-0,7358	-0,1149	-0,2556	-0,6296	-0,5847	-0,6284	0,20659	0,82322	-0,273	1,00947
-1,0456	-1,0342	0,91295	-1,1529	-0,7309	-0,9856	0,8362	-1,1418	0,63689	1,00947
-0,6504	0,80437	-1,0347	-1,6762	-0,8771	1,15782	0,6788	-0,3742	1,09181	1,00947
-1,0853	0,80437	-1,619	0,02453	-0,4385	-0,2711	-1,997	-0,5277	-1,3648	1,00947
-1,2938	0,80437	-0,8399	-1,807	-1,4618	1,05576	1,151	-0,1593	0,27295	1,00947
-1,3719	0,80437	-1,2294	-0,1063	-1,7541	-0,7304	0,52139	-0,4356	-2,0926	1,00947

Рис. 6.12. Подання інформаційного забезпечення кластерного аналізу інвестиційних фондів у табличному вигляді в пакеті STATISTICA

Після його запуску в діалоговому вікні виберемо команду метод Analysis/Joining (tree clustering). Визначимо всі змінні, метод і міру відстані. Натиснемо OK (рис. 6.13).

З'являється відповідне діалогове вікно, у якому необхідно визначити розташування графіка (вертикальне або горизонтальне). Натиснемо OK (рис. 6.14).

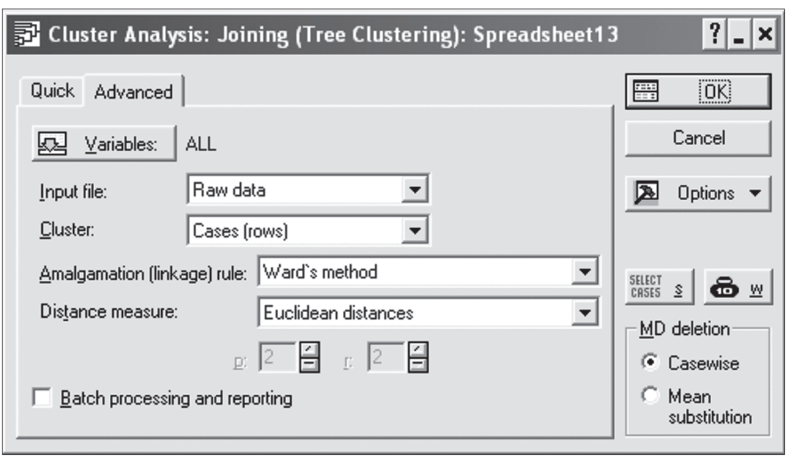


Рис. 6.13. Кластерний аналіз методу Analysis/Joining (tree clustering)

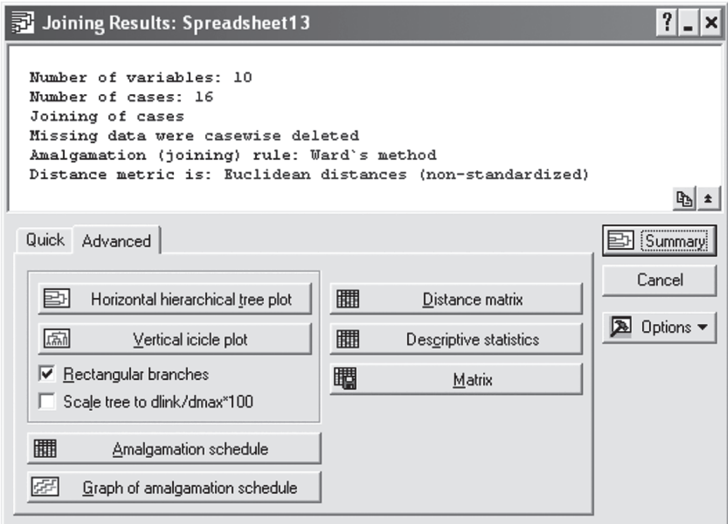


Рис. 6.14. Вибір опцій кластерного аналізу методом Analysis/Joining (tree clustering)

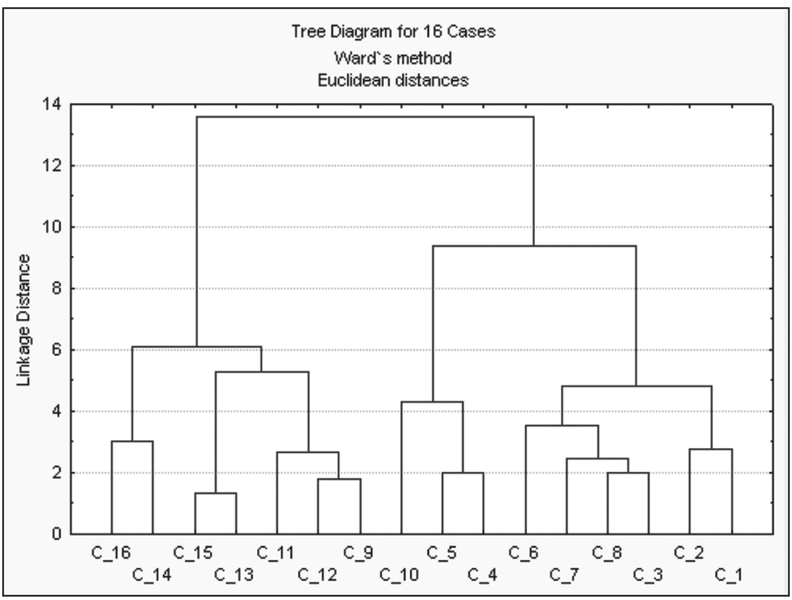


Рис. 6.15. Вертикальна деревоподібна дендограма кластеризації інвестиційних фондів

На графіку (рис. 6.15) чітко виявляються три групи кластерів. Аналіз цих кластерів (які елементи (інвестиційні фонди) віднести до перспективних, які – до безперспективних) вимагає спеціальних знань у цій сфері. Можна керуватися такими правилами:

- у першому кластері (правом) видно, що витрати були розумними: за низьких доходів у 1990 році в наступні роки стан фондів цього кластера помітно поліпшувався. При невисокому рейтингу ризику податкові збори були також досить низькими, акції цих фондів доцільно купувати;
- у другому кластері (середньому) були найбільші витрати, хоча за п'ятирічний період доходи були високими. Оцінка ризику і податкові збори виявилися максимальними серед всіх кластерів, це означає – акції цих фондів слід притримати;
- про третій кластер можна сказати, що він посідає друге місце за витратами щодо доходів за п'ятирічний період. Оцінка ризику – найвища, проте податкові збори значно нижчі, ніж у першого кластера, тому акції цих фондів доцільно продати.

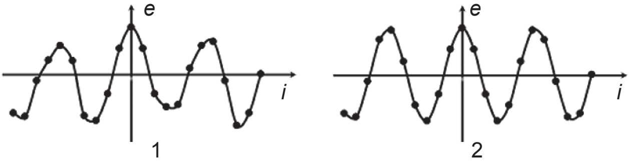
Тести

1. Які з наведених нижче параметрів зазвичай аналізують при визначенні якості побудованої моделі?
   а) скоректований коефіцієнт детерміації;
   б) t-статистики;
   в) статистика Дарбіна – Уотсона DW;
   г) узгодженість знаків коефіцієнтів з теорією;
   д) прогнозні помилки моделі;
   е) значення критерію ксі-квадрат;
   ж) значення критерію Колмогорова – Смирнова.

2. До основних властивостей якісної моделі належать:
   а) простота;
   б) єдиність;
   в) максимальна відповідність;
   г) узгодженість з теорією;
   д) прогнозна якість;
   е) емерджентність;
   ж) полярність;
   з) множинність.

3. Як відомо, для перевірки прогнозної якості моделі використовують так званий критерій V, значення якого розраховують за формулою $V = \frac{S}{y}$. Серед наведеного переліку тверджень відносно його застосування оберіть правильні:
   а) якщо величина V мала і відсутня автокореляція, то прогнозна якість моделі є високою;
   б) величина стандартної похибки регресії є обернено пропорційною до якості моделі;
   в) якщо величина V мала, то прогнозна якість моделі є високою;
   г) якщо величина V є значною і відсутня автокореляція, то прогнозна якість моделі є високою;
   д) якщо величина V мала і наявна автокореляція, то прогнозна якість моделі є високою.

4. Серед наведеного переліку можливих помилок оберіть ті, що належать до помилок специфікації моделі:
   а) відкидання значущої змінної;
   б) додавання незначущої змінної;
   в) вибір неправильної функціональної форми;
   г) відкидання незначущої змінної;
   д) додавання значущої змінної.



5. На наведенному вище рисунку подана залежність відхилень *e* від номера спостережень і для двох рівнянь. Серед наведеного переліку оберіть правильні твердження:

а) неправильно специфіковане лише рівняння якому відповідає залежність, подана на графіку 2;
б) неправильно специфіковане лише рівняння якому відповідає залежність, подана на графіку 1;
в) обидва рівняння є неправильно специфікованими.

6. Серед наведеного переліку тестів оберіть ті, які використовуються для виявлення помилок специфікації:

а) тест Рамсея;
б) тест максимальної правдоподібності;
в) тест Вальда;
г) тест множника Лагранжа;
д) тест Хаусмана;
е) тест Гольдфельда – Квандта;
ж) тест Чоу;
з) тест Парка;
и) тест Уайта.

7. Чи правильним є твердження, що відповідно до методу комбінаційного групування два об'єкти належать до однієї групи при точному збігу зареєстрованих на них градацій хоча б за однією ознакою, що їх характеризує?

а) так;
б) ні.

8. Розташуйте в правильному порядку етапи прогнозування структури споживання на основі класифікації споживачів:

а) збір і первинна статистична обробка початкових даних;
б) виявлення основних типів споживання за допомогою розбиття досліджуваної безлічі споживачів на класи в "просторі поведінки";
в) відбір найбільш інформативних типотвірних ознак (чинників-детермінантів) і вибір метрики в просторі типотвірних ознак;

г) аналіз динаміки структури досліджуваної сукупності споживачів у просторі найбільш інформативних типотвірних ознак;
д) прогноз структури споживання.

9. Які з моделей можливі при побудові кривої Філіпса, яка вказує, що залежність між заробітною платою *Y* і безробіттям *X* є зворотною?

а) $Y = a + \beta X + \varepsilon,\ \beta < 0;$
б) $\ln Y = a + \beta \ln X + \varepsilon,\ \beta < 0;$
в) $Y = a + a^{\beta X} + \varepsilon, \beta < 0,\ \beta > 0;$
г) $Y = a + \beta X + \varepsilon,\ \beta > 0;$
д) $\ln Y = a + \beta \ln X + \varepsilon,\ \beta > 0;$
е) $Y = a + a^{\beta X} + \varepsilon,\ \beta > 0.$

10. При проведенні прогнозування майбутніх значень результуючої ознаки за побудованою регресійною моделлю, яким повинен бути максимально допустимий період прогнозування?

а) $Y = a + a^{\beta X} + \varepsilon,\ \beta > 0$ у 3 рази коротше за період, за яким оцінювалося рівняння регресії;
б) $Y = a + a^{\beta X} + \varepsilon,\ \beta > 0$ не повинен перевищувати трьох майбутніх періодів;
в) $Y = a + a^{\beta X} + \varepsilon,\ \beta > 0$ не повинен перевищувати п’яти майбутніх періодів;
г) $Y = a + a^{\beta X} + \varepsilon,\ \beta > 0$ у 2 рази коротше за період, за яким оцінювалося рівняння регресії.

ТЕМА 7

Побудова економетричної моделі з автокореольованими залишками

Природа і наслідки автокореляції • Методи визначення автокореляції. Критерій Дарбіна – Уотсона. Критерій фон Неймана • Коефіцієнти автокореляції та їх застосування • Моделі з автокорельованими залишками • Метод оцінювання параметрів Ейткена • Метод Кочрена – Оркатта • Метод перетворення вихідної інформації. Метод Дарбіна.

7.1. Природа і наслідки автокореляції

Розглянемо класичну лінійну багатофакторну модель

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \ldots + \alpha_m x_m + \varepsilon \qquad (7.1)$$

або у матричному вигляді

$$Y = X\alpha + \varepsilon, \qquad (7.2)$$

де Y – вектор-стовпець залежної змінної розмірності $(n \times 1)$;
X – матриця незалежних змінних розмірності $(n \times (m + 1))$;
α – вектор-стовпець невідомих параметрів розмірності $((m \times 1) \times 1)$;
ε – вектор-стовпець випадкових помилок розмірності .

$$\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0, i \neq j. \qquad (7.3)$$

Після оцінки багатофакторної моделі (7.1), тобто виявлення алгебраїчної залежності

$$y = a_0 + a_1 x_1 + a_2 x_2 + \ldots + a_m x_m\,, \qquad (7.4)$$

здійснюється перевірка припущень Гаусса – Маркова стосовно вільного члена. На даному етапі виникає необхідність перевірки третього припущення відносно того, що випадкові відхилення ε_i та ε_j, $i \neq j$ повинні бути незалежними одне від одного. У разі порушення даної передумови застосування методу найменших квадратів для визначення параметрів регресійного рівняння виникає автокореляція залишків.

Автокореляція залишків – це явище, яке виникає в разі порушення припущення класичного регресійного аналізу про незалежність випадкових величини ε_i, $i = 1, ..., n$ (незважаючи на те, що дисперсія залишків є сталою, наявна гомоскедастичність).

Важливо зрозуміти, що спричинює автокореляцію, які її практичні та теоретичні наслідки, чи є ефективними методи тестування наявності автокореляції, чи змінюються методи знаходження невідомих параметрів моделі в умовах автокореляції.

Автокореляція залишків виникає найчастіше в разі, якщо економетрична модель будується на основі часових рядів. Якщо існує кореляція між послідовними значеннями деякої незалежної змінної, то відбуватиметься й кореляція послідовних значень залишків, так звані лагові затримки (запізнювання) в економічних процесах.

Автокореляція може виникати через інерційність і циклічність багатьох економічних процесів. Провокувати автокореляцію також може неправильно специфікована функціональна залежність у регресійних моделях.

Припустимо, модель (7.1) має автокорельовані залишки, тобто випадкові величини u_i залежні між собою (математичне сподівання

Таблиця 7.1. Порівняльний аналіз порушення припущень Гаусса – Маркова при гетероскедастичності й автокореляції залишків

Умова Гаусса – Маркова	Гетероскедастичність	Автокореляція залишків
Дисперсія залишків[1]	змінюється	постійна
Коваріація залишків[2]	відсутня	наявна

[1] Якість моделі лінійної регресії пов'язана з адекватністю (відповідністю) моделі з даними, за якими ведеться спостереження. Перевірка адекватності моделі регресії проводиться на основі аналізу регресійних залишків, а саме: на основі аналізу дисперсії залишків, тобто є відношення залишкової суми квадратів до різниці між кількістю спостережень і факторних ознак: $S^2 = \dfrac{RSS}{n-k}$, де $RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ – залишкова сума квадратів (сума квадратів відхилення i-го значення фактичного і теоретичного).

[2] Коваріація – статистична характеристика, яка показує міру схожості (або різниці) двох величин, які розглядаються, у динаміці. Оцінка коваріації проводиться на основі коефіцієнта кореляції.

Наслідки при ігноруванні матриці Ω при визначенні дисперсії залишків шляхом оцінювання параметрів моделі методом МНК

- Оцінки параметрів моделі можуть бути незміщеними, але неефективними, тобто вибіркові дисперсії вектора оцінок *а* можуть бути невиправдано великими
- Статистичні критерії *t*- і *F*-статистик, які отримані для класичної лінійної моделі, не можуть бути використані для дисперсійного аналізу, бо їх розрахунок не враховує наявності коваріації залишків
- Неефективність оцінок параметрів економетричної моделі, як правило, призводить до неефективних прогнозів, тобто прогнозні значення матимуть велику вибіркову дисперсію

Примітка:

$$\Omega = \begin{pmatrix} 1 & p & p^2 & p^3 & \dots & p^{n-1} \\ p & 1 & p & p^2 & \dots & p^{n-2} \\ p^2 & p & 1 & p & \dots & p^{n-3} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ p^{n-1} & p^{n-2} & p^{n-3} & p^{n-4} & \dots & 1 \end{pmatrix}, \qquad (7.6)$$

де параметр p характеризує коваріацію кожного наступного значення залишків із попереднім.

Рис. 7.1. Наслідки наявності автокореляції залишків

добутку випадкових відхилень в i-му та j-му спостереженнях): $M(\varepsilon_i \varepsilon_j) \neq 0, i \neq j$. Для кращого розуміння сутності автокореляції залишків пропонується провести порівняльний аналіз порушення припущень Гаусса – Маркова при гетероскедастичності і автокореляції залишків у вигляді табл. 7.1.

Так, якщо для залишків записати авторегресійну модель першого порядку

$$\varepsilon_t = p\varepsilon_{t-1} + v_t, \qquad (7.5)$$

то p характеризує силу зв’язку величин залишків у період t з величинами залишків у період $t-1$.

За наявності автокореляції поширеним методом оцінювання невідомих параметрів є узагальнений метод найменших квадратів, який було розглянуто в попередньому розділі. Отримані за допомогою УМНК оцінки є незміщеними та ефективними (рис. 7.1).

7.2. Методи визначення автокореляції. Критерій Дарбіна – Уотсона. Критерій фон Неймана

Тестування наявності автокореляції, як правило, здійснюється за *d*-тестом Дарбіна – Уотсона (рис. 7.2), хоча існують й інші не менш відомі тести: критерій фон Неймана, нециклічний коефіцієнт автокореляції, циклічний коефіцієнт автокореляції.

У критерії фон Неймана розраховується

$$Q = Q_{факт} = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t^2} \frac{n}{n-1}. \tag{7.8}$$

Звідси $Q = DW \dfrac{n}{n-1}$. (7.9)

Отже, якщо $n \to \infty$, $Q = DW$.

Фактичне значення критерію фон Неймана порівнюється з табличним при вибраному рівні значущості a і заданій кількості спостережень: $Q_{табл} = Q_{(a,\,n)}$.

Якщо $Q_{факт} < Q_{табл}$, то існує додатна автокореляція.

7.3. Коефіцієнти автокореляції та їх застосування

Нехай побудована регресійна теоретична модель (7.4), яка відповідає емпіричній моделі (7.1). Розрахуємо випадкові залишки, тобто відхилення фактичних і теоретичних значень для кожного *i*-го спостереження, які розташуємо в такій послідовності:

$$\varepsilon_1, \varepsilon_2, \varepsilon_3, ..., \varepsilon_m. \tag{7.10}$$

Проведемо зсунення значень випадкових відхилень на один елемент, отримаємо наступну послідовність:

$$\varepsilon_{1+1}, \varepsilon_{2+1}, \varepsilon_{3+1}, ..., \varepsilon_{m-1}. \tag{7.11}$$

Етапи здійснення d-теста Дарбіна – Уотсона на тестування наявності автокореляції

Крок 1

Розраховується значення *d*-статистики за формулою (7.7)
Зауваження.
Зауваження: доведено, що значення *d*-статистики Дарбіна – Уотсона перебуває в межах $0 < DW < 4$.

$$DW = d = \frac{\sum_{t=2}^{n}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n}\varepsilon_t^2} \qquad (7.7)$$

Крок 2

Задаємо рівень значущості *a*. За таблицею Дарбіна – Уотсона при заданому рівні значущості *a*, кількості факторів *m* і кількості спостережень *n* знаходимо два значення DW_1 і DW_2

- Якщо $0 < DW < DW_1$, то наявна додатна автокореляція.
- Якщо $DW_1 < DW < DW_2$, або $4 - DW_2 < DW < 4 - DW_1$, ми не можемо зробити висновки ані про наявність, ані про відсутність автокореляції (*DW* потрапляє в зону невизначеності).
- Якщо $4 - DW_1 < DW < 4$, маємо від'ємну автокореляцію.
- Якщо $DW_2 < DW < 4 - DW_2$, то автокореляція відсутня.

Графічне зображення розподілу ілюструє рис.7.2а

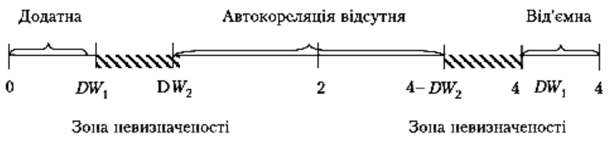


Рис. 7.2а. Зони автокореляційного зв'язку за критерієм Дарбіна – Уотсона

Рис. 7.2. Послідовність етапів тестування економетричної моделі на наявність автокореляції

На основі застосування послідовностей випадкових залишків (7.10) і (7.11) розрахуємо між їх значеннями коефіцієнт кореляції $r_{\varepsilon_i,\, \varepsilon_{i+1}}$, який називається ***коефіцієнтом автокореляції першого порядку***, оскільки визначає залежність між значеннями випадкових відхилень і значеннями тих самих відхилень, але зсунутих на один елемент:

$$\varepsilon_{1+k}, \varepsilon_{2+k}, \varepsilon_{3+k}, \ldots, \varepsilon_{m-k} \tag{7.12}$$

Проведемо зсунення значень випадкових відхилень на два елементи, отримаємо наступну послідовність:

$$\varepsilon_{1+2}, \varepsilon_{2+2}, \varepsilon_{3+2}, \ldots, \varepsilon_{m-2}. \tag{7.13}$$

На основі застосування послідовностей випадкових залишків (7.10) і (7.13) розрахуємо між їх значеннями коефіцієнт кореляції $r_{\varepsilon_i,\, \varepsilon_{i+2}}$, який називається ***коефіцієнтом автокореляції другого порядку***, оскільки визначає залежність між значеннями випадкових відхилень і значеннями тих самих відхилень, але зсунутих на два елементи:

$$r_{\varepsilon_i,\varepsilon_{i+2}} \; \frac{\overline{\varepsilon_i \cdot \varepsilon_{i+2}} - \overline{\varepsilon_i} \cdot \overline{\varepsilon_{i+2}}}{\overline{\varepsilon_{i+2}^{\,2}} - \overline{\varepsilon_{i+2}}^{\,2}} \cdot \frac{\sigma_{\varepsilon_{i+2}}}{\sigma_{\varepsilon_i}}. \tag{7.14}$$

Аналогічно проведемо зсунення значень випадкових відхилень на k елементів, отримаємо таку послідовність:

$$\varepsilon_{1+k}, \varepsilon_{2+k}, \varepsilon_{3+k}, \ldots, \varepsilon_{m-k}. \tag{7.15}$$

На основі застосування послідовностей випадкових залишків (7.10) і (7.15) розрахуємо між їх значеннями коефіцієнт кореляції $r_{\varepsilon_i,\, \varepsilon_{i+k}}$, який називається ***коефіцієнтом автокореляції k-го порядку***, оскільки визначає залежність між значеннями випадкових відхилень і значеннями тих самих відхилень, але зсунутих на k елементів:

$$r_{\varepsilon_i,\varepsilon_{i+k}} \; \frac{\overline{\varepsilon_i \cdot \varepsilon_{i+k}} - \overline{\varepsilon_i} \cdot \overline{\varepsilon_{i+k}}}{\overline{\varepsilon_{i+k}^{\,2}} - \overline{\varepsilon_{i+k}}^{\,2}} \cdot \frac{\sigma_{\varepsilon_{i+k}}}{\sigma_{\varepsilon_i}}. \tag{7.16}$$

Окрім статистик Дарбіна – Уотсона та фон Неймана, для перевірки автокореляції застосовують також нециклічний коефіцієнт автокореляції r^*, який відображає ступінь взаємозв'язку рядів ε_1, ε_2, ..., ε_{n-1} і обчислюється за формулою:

$$r^* = \frac{\sum_{t=2}^{n} \varepsilon_t \varepsilon_{t-1} - \frac{1}{n-1}\left(\sum_{t=2}^{n} \varepsilon_t\right)\left(\sum_{t=2}^{n} \varepsilon_{t-1}\right)}{\left[\sum_{t=-2}^{n} \varepsilon_t^2 - \frac{1}{n-1}\left(\sum_{t=2}^{n} \varepsilon_t\right)\right]^{1/2}\left[\sum_{t=2}^{n} \varepsilon_{t-1}^2 - \frac{1}{n-1}\left(\sum_{t=2}^{n} \varepsilon_{t-1}\right)^2\right]^{1/2}}. \quad 7.17)$$

Коефіцієнт r^* може набувати значень в інтервалі (1; –1). Його від'ємні значення свідчать про від'ємну автокореляцію залишків, а додатні – про додатну автокореляцію. Значення, що лежать у деякій критичній області поблизу нуля, підтверджують нульову гіпотезу про відсутність автокореляції в залишках. Оскільки ймовірнісний розподіл r^* встановити важко, то на практиці замість r^* обчислюють циклічний коефіцієнт автокореляції r^0. Загалом, якщо часовий ряд має циклічний характер, тобто припускається, що після значення ε_t загальний характер зміни членів ряду повторюється, то автокореляцію визначають за допомогою коефіцієнта r^0, запровадженого Андерсоном.

У цьому разі автокореляція визначається між послідовностями, зсунутими на період τ:

$$\varepsilon_1, \varepsilon_2, ..., \varepsilon_\tau \text{ і } \varepsilon_{t+1}, \varepsilon_{t+2}, ..., \varepsilon_{2\tau}, \quad (7.18)$$

$$\varepsilon_1, \varepsilon_2, ..., \varepsilon_n \text{ і } \varepsilon_{t+1}, ..., \varepsilon_n, \varepsilon_2, ..., \varepsilon_\tau.$$

Якщо період $\tau = 1$, то маємо коефіцієнт циклічної автокореляції першого порядку, який відбиває інтенсивність взаємозв'язку між послідовностями

$$\varepsilon_1, \varepsilon_2, ..., \varepsilon_{n-1}, \varepsilon_n \text{ та } \varepsilon_2, \varepsilon_3, ..., \varepsilon_n, \varepsilon_1. \quad (7.19)$$

Для досить довгих рядів вплив циклічних членів стає незначним, тому ймовірнісний розподіл коефіцієнта r^* наближається до ймовірнісного розподілу коефіцієнта циклічної автокореляції r^0, який обчислюється за формулою:

$$r^0 = \frac{\sum_{t=2}^{n} \varepsilon_t \varepsilon_{t-1} + \varepsilon_n \varepsilon_1 - \frac{1}{n}\left(\sum_{t=1}^{n} \varepsilon_t\right)^2}{\sum_{t=1}^{n} \varepsilon_t^2 - \frac{1}{n}\left(\sum_{t=1}^{n} \varepsilon_t\right)^2}. \quad (7.20)$$

Якщо останній член ряду дорівнює першому, тобто $\varepsilon_1 = \varepsilon_n$, то нециклічний коефіцієнт автокореляції дорівнює циклічному. Очевидно, якщо залишки не містять тренду, то припущення про рівність $\varepsilon_1 = \varepsilon_n$ недалеке від дійсності й циклічний коефіцієнт автокореляції близький до нециклічного. Крім того, припускаючи, що середня залишків дорівнює нулю, тобто $\bar{\varepsilon} = 0$, а отже,

$$\sum_{t=2}^{n} \varepsilon_t \approx \sum_{t=2}^{n} \varepsilon_{t-1} \approx 0, \tag{7.21}$$

отримуємо приблизну формулу для обчислення циклічного коефіцієнта автокореляції:

$$r = \frac{\sum_{t=2}^{n} \varepsilon_t \varepsilon_{t-1}}{\sum_{t=1}^{n} \varepsilon_t^2}. \tag{7.22}$$

Причому $r \in (-1;1)$. Значення r використовується при оцінюванні параметрів моделі.

7.4. Моделі з автокорельованими залишками

Розглянемо методи побудови економетричних моделей у випадку наявності автокореляції залишків.

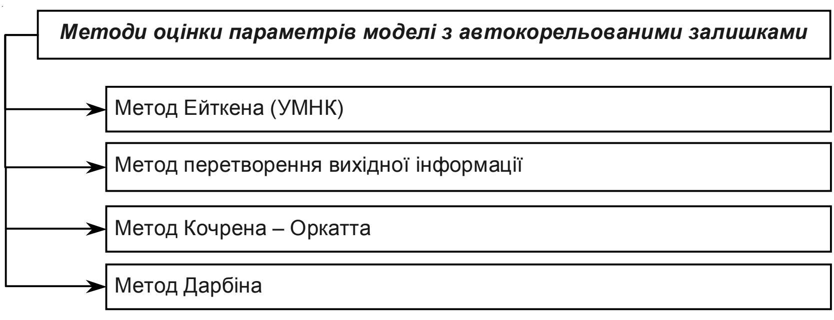


Рис. 7.3. Методи оцінки параметрів моделі з автокорельованими залишками

Перші два методи доцільно застосовувати в разі, якщо залишки описуються авторегресійною моделлю першого порядку

$$\varepsilon_t = \rho\varepsilon_{t-1} + v_t. \tag{7.23}$$

Ітераційні методи Кочрена – Оркатта і Дарбіна можна застосовувати для оцінки параметрів економетричної моделі й тоді, коли залишки описуються авторегресійною моделлю вищого порядку:

$$\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + v_t, \qquad (7.24)$$

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

7.5. Метод оцінювання параметрів Ейткена

Як зазначалося, оператор оцінювання УМНК можна записати так:

$$\hat{a} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y, \qquad (7.25)$$

або

$$\hat{a} = (X'S^{-1}X)^{-1}X'S^{-1}Y, \qquad (7.26)$$

де Ω^{-1} – матриця, обернена до диснерсійно-коваріаційної матриці залишків Ω;

S^{-1} – матриця, обернена до матриці $S = \sigma_\varepsilon^2\Omega$.

Оскільки в Ω коваріація залишків $\rho^s \to 0$ при $S > 2$, то матриця Ω^{-1} матиме вигляд

$$\Omega^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho & 0 & 0 & 0 & \ldots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 & \ldots & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 & \ldots & 0\,0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & & \ldots & 1 \end{pmatrix}. \qquad (7.27)$$

На практиці для розрахунку ρ використовується співвідношення

$$\rho \approx r \approx \frac{\sum_{t=2}^{n}\varepsilon_t\varepsilon_{t-2}}{\sum_{t=1}^{n}\varepsilon_t^2} \qquad (7.28)$$

або

$$\rho \approx r' \approx \frac{\sum_{t=2}^{n}\varepsilon_t\varepsilon_{t-1}}{\sum_{t=1}^{n}\varepsilon_t^2}\frac{n}{n-1}. \qquad (7.29)$$

7.6. Метод Кочрена – Оркатта

Зауваження. Метод Кочрена – Оркатта є ітераційним методом наближеного пошуку оцінок параметрів моделі з автокорельованими залишками, який базується на МНК.

Етапи реалізації методу Корчена – Оркатта

Крок 1

Довільно вибрати значення параметра ρ, наприклад $\rho = r_1$. Підставивши його у

$$\sum_{t=2}^{n} \varepsilon_t^2 = \sum_{t=2}^{n} \left[\left(y_t - \rho y_{t-1} \right) - a_0 \left(1 - \rho \right) - a_1 \left(x_t - \rho x_{t-1} \right) \right]^2, \qquad (7.23)$$

обчислити $\hat{a}_0^{(1)}$ і $\hat{a}_1^{(1)}$.

Крок 2

Покласти $\hat{a}_0 = \hat{a}_0^{(1)}$ і $\hat{a}_1 = \hat{a}_1^{(1)}$; підставивши їх у рівняння (7.23), обчислити $\rho = r_1$.

Крок 3

Підставити в рівняння (7.23) значення $\rho = r_2$, знайти $\hat{a}_0^{(2)}$ і $\hat{a}_1^{(2)}$.

Крок 4

Використати $\hat{a}_0 = \hat{a}_0^{(2)}$ і $\hat{a}_1 = \hat{a}_1^{(2)}$ для мінімізації суми квадратів залишків (7.23) за невідомим параметром $\rho = r_3$. Процедуру повторювати доти, доки наступні значення параметрів $\hat{a}_0, \hat{a}_1$ і ρ не відрізнятимуться менш як на задану величину.

Рис. 7.4. Послідовність етапів упровадження методу Корчена – Оркатта

Зазначимо, що наведений метод завжди забезпечує:

- знаходження глобального оптимуму;
- порівняно добру збіжність.

7.7. Метод перетворення вихідної інформації. Метод Дарбіна. Метод перетворення вихідної інформації

Випадок, коли залишки задовольняють авторегресійну модель першого порядку, допускає альтернативний підхід до пошуку оцінок параметрів моделі за допомогою двокрокової процедури (рис. 7.5).

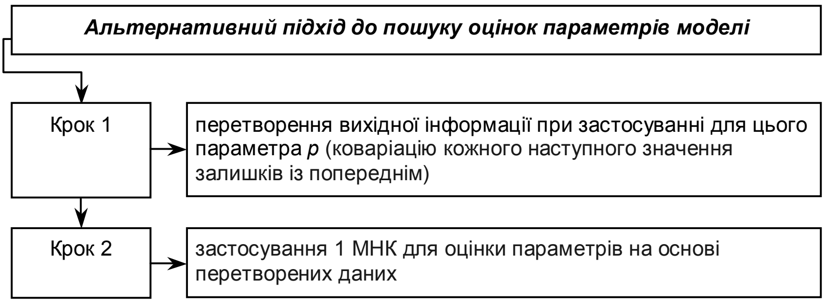


Рис. 7.5. Послідовність етапів упровадження альтернативного підходу до пошуку оцінок параметрів моделі

Для цього треба знайти матрицю перетворення T, щоб модель

$$TY = TXA + T\varepsilon \tag{7.30}$$

мала скалярну дисперсійну матрицю $M(T\varepsilon\varepsilon' T') = \sigma_\varepsilon^2 E$.

Розглянемо матрицю T_1 розміром $n \times n$:

$$T_1 = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ 0 & 0 & -\rho & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}. \tag{7.31}$$

Безпосереднім множенням легко переконатися, що $M(T_1\varepsilon\varepsilon' T'_1) = \sigma_\varepsilon^2 E$.

А це означає, що можна застосувати 1 МНК до перетворених даних T_1Y і T_1X, які мають вигляд

$$T_1Y = \begin{pmatrix} \sqrt{1-\rho^2}\, y_1 \\ y_2 - \rho y_1 \\ y_3 - \rho y_2 \\ y_4 - \rho y_3 \\ \dots \\ y_n - \rho y_{n-1} \end{pmatrix} \tag{7.32}$$

$$T_1X = \begin{pmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}\, x_1^1 & \dots & \sqrt{1-\rho^2}\, x_1^m \\ 1-\rho & x_2^1 - \rho x_1^1 & \dots & x_2^m - \rho x_1^m \\ \dots & \dots & \dots & \\ 1-\rho & x_n^1 - \rho x_n^1 & \dots & x_n^m - \rho x_n^m \end{pmatrix}. \tag{7.33}$$

Іноді для перетворення вихідної інформації використовується матриця T_2 розміром $(n - 1) \times n$, яка отримується з матриці T_2 внаслідок викреслювання першого рядка:

$$T_2 = \begin{pmatrix} -\rho & 1 & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1 & 0 & \dots & 0 \\ 0 & 0 & -\rho & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} \tag{7.34}$$

Неважко показати, що застосування 1 МНК до даних T_1Y і T_1X дає таку саму оцінку параметрів моделі, як і метод Ейткена, а для даних T_2Y і T_2X забезпечує порівняно 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 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 (рис. 7.6).

Етапи реалізації методу Дарбіна

Крок 1

Підставимо значення залишків, яке підпорядковане авторегресійній моделі першого порядку $u_t = pu_{t-1} + \varepsilon_t$ до економетричної моделі $y_t = a_0 + a_1x_t + u_t$. Тоді дістанемо $y_t = a_0 + a_1x_t + pu_{t-1} + \varepsilon_t$, де $u_{t-1} = y_{t-1} - \hat{a}_0 - \hat{a}_1x_{t-1}$.

Звідси $y_t = a_0(1 - p) + py_{t-1} + a_1x_t - a_1px_{t-1} + \varepsilon_t$, де ε_t має скалярну матрицю дисперсій.

Згідно з 1МНК визначаються параметри цієї моделі, куди входить і коефіцієнт p. У результаті обчислень маємо $p = r$.

Крок 2

Значення $p = r$ використовується для перетворення змінних $(y_t - ry_{t-1})$ і $(x_t - rx_{t-1})$, а 1 МНК застосовується до перетворених даних. Коефіцієнт при $(x_t - rx_{t-1})$ є оцінкою параметра a_1, а вільний член, поділений на $-r$, оцінює параметр a_0.

Рис. 7.6. Послідовність етапів впровадження метода Дарбіна

Метод Дарбіна дуже просто поширюється на випадок кількох незалежних змінних і для автокореляції вищих порядків.

Нехай задано модель

$$y_t = a_0 + a_1x_{1t} + a_2x_{2t} + \ldots + a_mx_{mt} + u_t, \tag{7.35}$$

де $\varepsilon_t = \rho_1\varepsilon_{t-1} + \rho_2\varepsilon_{t-2} + \varepsilon_t$.

Підставивши значення ε_t в (7.29), маємо:

$$\begin{aligned} y_t &= \rho_1 y_{t-1} + \rho_2 y_{t-2} + \hat{a}_1x_{1t} + \ldots + \hat{a}_m x_{mt} - \\ &- \rho_1\hat{a}_1x_{1,t-1} - \ldots - \rho_1\hat{a}_m x_{m,t-1} - \rho_2\hat{a}_1x_{1,t-2} - \ldots - \rho_2\hat{a}_m x_{m,t-1} + \varepsilon_t. \end{aligned} \tag{7.36}$$

Застосувавши 1 МНК, обчислимо параметри цієї моделі. Коефіцієнти $\rho_1 = r$ і $\rho_2 = r_2$ використаємо для перетворення даних:

$$(y_t - r_1y_{t-1} - r_2y_{t-2}), (x_{jt} - r_1x_{j,t-1} - r_2x_{j,t-2}),\ j = \overline{1,m}, t = \overline{1,n}.$$

Знову застосуємо 1 МНК для цих перетворених даних і знайдемо оцінки параметрів моделі $\hat{a}_0$, $\hat{a}_j (j = \overline{1, m})$.

Описаний щойно ітеративний метод Кочрена – Оркатта і розглянута двокрокова процедура Дарбіна за наявності автокореляції залишків асимптотично більш ефективні, ніж 1 МНК.

Висновок 1. 1 МНК дає менш ефективні оцінки порівняно з іншими методами, якщо сукупність спостережень $n = 20$ одиниць, а $|p| > 0{,}3$.

Висновок 2. Якщо $|p| > 0{,}3$, то зниження ефективності оцінок 1 МНК порівняно зі складнішими процедурами невелике.

Висновок 3. Метод Дарбіна забезпечує найкращу оцінку для ширшого кола параметрів порівняно з іншими методами.

Висновок 4. Нелінійний метод оцінювання параметрів не дає відчутних переваг порівняно з двокроковою процедурою Дарбіна.

Завдання для самоконтролю

1. Природа і наслідки автокореляції.
2. Методи визначення автокореляції.
3. Критерій Дарбіна – Уотсона.
4. Критерій фон Неймана.
5. Коефіцієнти автокореляції та їх застосування.
6. Моделі з автокорельованими залишками.
7. Метод оцінювання параметрів Ейткена.
8. Метод Кочрена – Оркатта
9. Метод перетворення вихідної інформації.
10. Метод Дарбіна.

Практичне заняття 7

Дослідження автокореляції

Постановка завдання

1 етап	Побудова інформаційного масиву дослідження шляхом збору статистичних даних та розробка моделі парної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), анаіз моделі
2 етап	Побудова множинної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), дослідження моделі
3 етап	Виявлення, дослідження та усунення мультиколінеарності факторів у моделі множинної лінійної регресії методом Фаррара – Глобера
4 етап	Виявлення ефекту гетероскедастичності на основі застосування параметричного тесту Гольдфельда – Квандта та усунення даного ефекту шляхом використання узагальненого методу найменших квадратів (методу Ейткена)
5 етап	Побудова адитивної та мультиплікативної моделі часового ряду на основі проведення декомпозиційного аналізу: визначення структури ряду (тенденції та сезонних (циклічних) коливань), проведення автокореляційного аналізу, визначення якості моделі, розрахунок прогнозних значень
6 етап	Групування елементів економічної системи шляхом кластерного аналізу, побудова деревоподібної дендограми, графічне подання елементів кластеризації
7 етап	**Перевірка наявності автокореляції залишків за допомогою критерію Дарбіна – Уотсона, оцінка параметрів рівняння множинної лінійної регресії методом Ейткена**
8 етап	Побудова економіко-математичної моделі методом інструментальних змінних; побудова економетричної моделі на основі даних, що містять помилки вимірювання змінних, оцінка параметрів моделі за допомогою алгоритму Уоліса
9 етап	Побудова моделі розподіленого лага, оцінка параметрів моделей з лагами в незалежних змінних, оцінювання параметрів авторегресійних моделей; перетворення ARMA і ARIMA
10 етап	Структурне моделювання (моделювання на основі системи структурних рівнянь), оцінка невідомих параметрів моделі, визначення адекватності моделі щодо опису структури вхідних даних

Статистичні дані про залежність витрат на рекламу від прибутку на деякому підприємстві протягом 15 років наведені в табл. 7.2.

Таблиця 7.2. Статистичні дані про залежність витрат на рекламу від прибутку на деякому підприємстві

Рік	1	2	3	4	5	6	7	8
X	$11*N_2-4$	$11*N_4-6$	$11*N_2-9$	$11*N_4-6$	$11*N_2-7$	$11*N_1-6$	$11*N_2-8$	$11*N_3-5$
Y	$88+2*N_3$	$63+2*N_3$	$39+2*N_3$	$64+2*N_1$	$57+2*N_1$	$68+2*N_4$	$52+2*N_2$	$78+2*N_4$

Рік	9	10	11	12	13	14	15
X	$11*N_2-8$	$11*N_4-9$	$11*N_4-4$	$11*N_3-6$	$11*N_4-8$	$11*N_1-4$	$11*N_4-5$
Y	$58+2*N_3$	$43+2*N_4$	$86+2*N_2$	$68+2*N_3$	$52+2*N_3$	$86+2*N_4$	$77+2*N_1$

Таблиця 7.3. Розподіл варіантів виконання лабораторної роботи

Варіант	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
N_1	7	7	2	7	4	3	3	4	7	8	7	4	3	4	3
N_2	6	7	6	6	8	2	3	5	4	8	3	3	6	8	3
N_3	1	2	8	4	3	6	3	6	2	1	6	7	6	5	4
N_4	7	2	6	7	4	8	7	5	9	8	3	7	4	3	2

Вважаючи, що величина витрат на рекламу залежить від величини отриманого прибутку необхідно:

- оцінити параметри рівняння взаємозв'язку між обсягом витрат на рекламу і обсягом отриманого прибутку;
- перевірити наявність автокореляції залишків;
- до оцінки параметрів регресії застосувати метод Ейткена.

Методичні вказівки до розв'язання задачі

Таблиця 7.4. Статистичні дані про залежність витрат на рекламу від прибутку на деякому підприємстві

Рік	1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Прибуток підприємства, млн грн, X	2	6	4	1	4	3	4	2	5	2	1	6	4	2	6	5
Витрати на рекламу, тис. грн, Y	*3*	99	74	50	75	68	79	63	89	69	54	97	79	63	97	88

1. Для знаходження оцінок параметрів лінійної регресії скористаємося формулою

   $$\alpha = \left(X^T X\right)^{-1} X^T y.$$

   Розрахуємо матрицю моментів $X^T X$:

   $$\left(X^T X\right) = \begin{vmatrix} 15 & 55 \\ 55 & 245 \end{vmatrix}.$$

   Розрахуємо вектор:

   $$\left(X^T y\right) = \begin{vmatrix} 1144 \\ 4569 \end{vmatrix}.$$

   Оцінки параметрів будуть дорівнювати

   $$\alpha = \left(X^T X\right)^{-1} X^T y = \begin{vmatrix} 44,29 \\ 8,72 \end{vmatrix}.$$

   Економетрична модель має вигляд:

   $$\hat{y}_t = a + bx_t + \varepsilon_t;$$

   $$\hat{y}_t = 44,29 + 8,72x_t + \varepsilon_t.$$

2. На основі економетричної моделі визначимо вектор збурення ε_t, який є різницею між розрахованим $\hat{y}_t$ та фактичним y_t значенням витрат на рекламу:

   $$\varepsilon_t = y_t - \hat{y}_t.$$

   Розрахунки наведено в табл. 7.5.

   Розрахуємо критерій Дарбіна – Уотсона:

   $$DW = \frac{\sum_{t}^{15}\left(\varepsilon_t - \varepsilon_{t-1}\right)^2}{\sum_{t}^{15}\varepsilon_t^2};$$

   $$DW = \frac{152,14}{123,27} = 1,23.$$

Таблиця 7.5. Проміжні розрахунки

Рік	X	Y	X^2	Y^2	XY	$\hat{Y}_x$	ε_t	ε_t^2	$\varepsilon_t - \varepsilon_{t-1}$	$(\varepsilon_t - \varepsilon_{t-1})^2$	$\varepsilon_t\ \varepsilon_{t-1}$
1	2	3	4	5	6	7	8	9	10	11	12
1	6	99	36	9801	594	96,42	2,58	6,64	-	-	-
2	4	74	16	5476	296	79,15	-5,15	26,48	-7,72	59,65	-13,26
3	1	50	1	2500	50	53,23	-3,23	10,44	1,92	3,67	16,63
4	4	75	16	5625	300	79,15	-4,15	17,19	-0,92	0,84	13,40
5	3	68	9	4624	204	70,51	-2,51	6,29	1,64	2,68	10,40
6	4	79	16	6241	316	79,15	-0,15	0,02	2,36	5,58	0,37
7	2	63	4	3969	126	61,87	1,13	1,28	1,28	1,63	-0,17
8	5	89	25	7921	445	87,78	1,22	1,48	0,08	0,01	1,37
9	2	69	4	4761	138	61,87	7,13	50,85	5,92	34,99	8,67
10	1	54	1	2916	54	53,23	0,77	0,59	-6,36	40,47	5,49
11	6	97	36	9409	582	96,42	0,58	0,33	-0,19	0,04	0,44
12	4	79	16	6241	316	79,15	-0,15	0,02	-0,72	0,52	-0,08
13	2	63	4	3969	126	61,87	1,13	1,28	1,28	1,63	-0,17
14	6	97	36	9409	582	96,42	0,58	0,33	-0,55	0,31	0,65
15	5	88	25	7744	440	87,78	0,22	0,05	-0,36	0,13	0,12
Разом	55	1144	245	90606	4569	1144	0,00	123,27	-2,36	152,14	43,86

Для прийняття рішення про наявність автокореляції залишків використаємо двосторонній d-тест для таких значень:

$\alpha = 0{,}05;\ m = 1;\ t = 15;$

$d_i = 1{,}08;\ d_{\hat{a}} = 1{,}36.$

У нашому випадку $d_i \le d \le d_{\hat{a}}$,

$1{,}08 \le 1{,}23 \le 1{,}36.$

Цей результат указує на те, що гіпотеза $H_0 : \rho = 0$ не може бути ні прийнятою, ні відхиленою (зона невизначеності). Тому зробити висновок щодо наявності чи відсутності автокореляції залишків за d-тестом неможливо. Для цього необхідно скористатися критерієм фон Неймана:

$$Q_{розр} = DW\frac{n}{n-1};$$

$$Q_{розр} = 1{,}23\frac{15}{14} = 1{,}32.$$

Це значення порівняємо з табличним $Q_{табл} = 1{,}18$, якщо $n = 15$; $\alpha = 0{,}10$. Оскільки $Q_{розр} > Q_{табл}$, то можна вважати, що додатна автокореляція залишків відсутня.

Визначимо авторегресійний параметр ρ:

$$\hat{\rho} = \frac{\sum\limits_{t=2}^{15} U_t U_{t-1}}{\sum\limits_{t=2}^{15} U_t^2},$$

де $\hat{\rho} = \dfrac{43{,}49}{123{,}27} = 0{,}36$ – циклічний коефіцієнт автокореляції.

Коригування ρ на величину зміщення:

$$\hat{\rho}' = \frac{n}{n-1}\hat{\rho},$$

де $\hat{\rho}' = 0{,}36 * 1{,}11 = 0{,}38$.

3. Визначимо оцінки параметрів регресії за допомогою методу Ейткена:

$$\hat{\hat{\alpha}} = \left(X^{*T} X^{*}\right)^{-1} X^{*T} y^{*},$$

де $X^{*T}X^{*}$ – матриця моментів у допоміжній матриці T^A – матриці перетворень:

$$\left[y^{*} : X^{*}\right] = \left[T^A y : T^A X\right],$$

де матриця T^A при авторегресійному процесі першого порядку має такий вигляд:

$$T^A = \begin{vmatrix} \sqrt{(1-\hat{\rho})^2} & 0 & 0 & 0 & \ldots & 0 & 0 \\ -\rho & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & -\rho & 1 & 0 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & \ldots & -\rho & 1 \end{vmatrix},$$

Де перші елементи розраховуються за формулами:

$$y_1^* = y_1 + \sqrt{1-\hat{\rho}'^2};$$

$$x_1^* = x_1 + \sqrt{1-\hat{\rho}'^2}.$$

А всі інші:

$$y_t^* = y_t - \hat{\rho}' y^*_{t-1};$$

$$x_t^* = x_t - \hat{\rho}' x^*_{t-1}.$$

Матриця перетворень буде мати вигляд (табл. 7.6):

Таблиця 7.6. Скориговані розрахункові дані

Рік	X	Y	X^2	Y^2	XY	$\hat{Y}_x$	U_t	U^2_t	$U_t - U_{t-1}$	$(U_t - U_{t-1})^2$
1	6,93	99,93	48,09	9986,92	693,00	94,69	5,24	27,48	-	-
2	1,53	38,45	2,35	1478,11	58,93	45,85	-7,40	54,76	-12,64	159,83
3	0,45	36,32	0,21	1319,29	16,51	36,10	0,23	0,05	7,63	58,16
4	3,84	62,08	14,73	3853,64	238,27	66,69	-4,62	21,30	-4,84	23,44
5	1,63	45,91	2,67	2108,15	75,05	46,76	-0,85	0,72	3,77	14,18
6	3,42	62,66	11,69	3926,90	214,22	62,90	-0,23	0,05	0,62	0,38
7	0,78	40,71	0,61	1656,95	31,90	39,07	1,63	2,67	1,87	3,48
8	4,72	74,52	22,29	5552,95	351,81	74,68	-0,16	0,03	-1,79	3,21
9	0,32	42,49	0,10	1805,28	13,61	34,88	7,61	57,86	7,77	60,31
10	0,89	38,88	0,79	1511,95	34,45	40,00	-1,11	1,24	-8,72	76,04
11	5,68	83,17	32,32	6916,63	472,78	83,39	-0,22	0,05	0,89	0,79
12	1,98	49,41	3,91	2441,53	97,71	49,87	-0,46	0,21	-0,23	0,05
13	1,30	45,42	1,68	2063,04	58,89	43,71	1,71	2,93	2,17	4,70
14	5,54	80,84	30,68	6535,21	447,76	82,07	-1,23	1,51	-2,94	8,66
15	3,03	59,24	9,18	3509,29	179,46	59,38	-0,14	0,02	1,09	1,19
Разом	42,05	860,04	181,29	54665,85	2984,37	860,04	0,00	170,89	-5,38	414,40

Розрахуємо матрицю моментів $X^{*T}X^*$:

$$\left(X^{*T}X^*\right) = \begin{vmatrix} 15 & 42,052 \\ 42,052 & 181,289 \end{vmatrix}.$$

Розрахуємо вектор:

$$\left(X^{*T} y^{*}\right)=\begin{vmatrix}860{,}035\\2984{,}368\end{vmatrix}.$$

Оцінки параметрів будуть дорівнювати:

$$\alpha^{*}=\left(X^{*T}X^{*}\right)^{-1}X^{*T}y^{*}=\begin{vmatrix}31{,}98\\8{,}04\end{vmatrix}.$$

Економетрична модель має вигляд:

$$\hat{\hat{y}}_t=\hat{\hat{a}}+\hat{\hat{b}}x^{*}{}_t+\hat{\hat{\varepsilon}}_t;$$

$$\hat{\hat{y}}_t=31{,}98+8{,}04x^{*}{}_t+\hat{\hat{\varepsilon}}_t.$$

Розрахуємо критерій Дарбіна – Уотсона:

$$DW=\frac{414{,}40}{170{,}89}=2{,}42\,.$$

Висновок: авторегресійний процес першого порядку відсутній, тобто після застосування методу Ейткена прогноз моделі покращився.

Тести

1. Визначити послідовність етапів здійснення d-теста Дарбіна — Уотсона на тестування наявності автокореляції:
   а) задаємо рівень значущості a;
   б) за таблицею Дарбіна — Уотсона при заданому рівні значущості a, кількості факторів m і кількості спостережень n знаходимо два значення DW_1 і DW_2;
   в) розраховується значення d-статистики.

2. Коефіцієнт автокореляції може набувати значень в інтервалі:
   а) (–1; 1)
   б) (0; 1)
   в) (–1; 0)
   г) (0; ∞)

3. Для методу Кочрена — Оркатта характерним є:
   а) ітераційний метод наближеного пошуку оцінок параметрів моделі;
   б) застосовується для моделей з автокорельованими залишками;
   в) базується на МНК.

4. Визначити, чи правильним є наведене твердження:
   Автокореляція залишків — це явище, яке виникає в разі порушення припущення класичного регресійного аналізу про незалежність випадкових величин ε_i, $i = 1, \ldots, n$ (незважаючи на те, що дисперсія залишків є сталою, наявна гомоскедастичність).
   а) так;
   б) ні.

5. Назвати основні наслідки автокореляції:
   а) оцінки параметрів моделі можуть бути незміщеними, але неефективними, тобто вибіркові дисперсії вектора оцінок a можуть бути невиправдано великими;
   б) статистичні критерії t- і F-статистик, які отримані для класичної лінійної моделі, не можуть бути використані для дисперсійного аналізу, бо їх розрахунок не враховує наявності коваріації залишків;
   в) неефективність оцінок параметрів економетричної моделі, як правило, призводить до неефективних прогнозів, тобто прогнозні значення матимуть велику вибіркову дисперсію.

6. Якщо за критерієм Дарбіна — Уотсона виконуються нерівності $DW_1 < DW < DW_2$ або $4 - DW_2 < DW < 4 - DW_1$:
   а) наявна додатна автокореляція;

б) ми не можемо зробити висновки ані про наявність, ані про відсутність автокореляції;
в) маємо від'ємну автокореляцію;
г) автокореляція відсутня.

7. Визначити методи оцінки параметрів моделі з автокорельованими залишками:
а) метод Ейткена (УМНК);
б) метод перетворення вихідної інформації;
в) метод Кочрена – Оркатта;
г) метод Дарбіна.

8. Назвати, який метод забезпечує знаходження глобального оптимуму та порівняно гарну збіжність:
а) метод Ейткена (УМНК);
б) метод перетворення вихідної інформації;
в) метод Кочрена – Оркатта;
г) метод Дарбіна.

ТЕМА 8

Методи інструментальних змінних

Сутність методу інструментальних змінних • Оператор оцінювання Вальда • Особливості оцінювання методом Бартлета • Оператор оцінювання Дарбіна • Побудова економетричної моделі на основі даних, що містять помилки вимірювання змінних • Оцінка параметрів моделі за допомогою алгоритму Уоліса

8.1. Сутність методу інструментальних змінних

Іще одним способом усунення корелювання пояснювальної змінної з випадковим відхиленням є метод інструментальних змінних (ІЗ).

Метод інструментальних змінних: сутність цього методу полягає в заміні змінної, що корелює із залишками, інструментальною змінною (ІЗ), яка повинна мати такі властивості: 1) корелювати (бажано значною мірою) із заміненою пояснюючою змінною; 2) не корелювати з випадковим відхиленням.

Опишемо схему використання ІЗ на прикладі парної регресії, у якій

$$Y = \alpha_0 + \alpha_1 X + \varepsilon. \tag{8.1}$$

Змінну X замінюють змінною Z такою, що cov $(Z; X) \neq 0$ і cov $(Z; \varepsilon) = 0$. Принципи використання ІЗ передбачають виконання таких умов:

$$M(\varepsilon_t) = 0, \operatorname{cov}(z_t, \varepsilon_t). \tag{8.2}$$

Відповідні вибіркові оцінки даних умов такі:

$$\begin{cases} \frac{1}{T}\sum e_t = 0, \\ \frac{1}{T}\sum z_t e_t = 0. \end{cases} \tag{8.3}$$

У розгорненому вигляді остання система має вигляд:

$$\begin{cases} \sum \left(y_t - b_0^{із} - b_1^{із} x_t \right) = 0, \\ \sum z_t \left(y_t - b_0^{із} - b_1^{із} x_t \right) = 0. \end{cases} \tag{8.4}$$

Звідки

$$\begin{cases} b_1^{із} = \dfrac{\sum (z_t - \bar{z})(y_t - \bar{y})}{\sum (z_t - \bar{z})(x_t - \bar{x})}. \\ b_0^{із} = \bar{y} - b_1^{із} \bar{x}. \end{cases} \tag{8.5}$$

Нехай зі збільшенням обсягу вибірки *D(X)* прямує до деякої скінченної межі σ_x^2, а коваріація cov*(Z; X)* – до скінченної межі $\sigma_{zx} \neq 0$.

Покажемо, що в цьому разі b_1 прямує до істинного значення β_1. З останньої системи маємо

$$\begin{aligned} b_1^{із} &= \frac{\operatorname{cov}(Z,Y)}{\operatorname{cov}(Z,X)} = \frac{\operatorname{cov}(Z, \alpha_0 + \alpha_1 X + \varepsilon)}{\operatorname{cov}(Z,X)} = \\ &= \frac{\operatorname{cov}(Z,\alpha_0)}{\operatorname{cov}(Z,X)} + \frac{\operatorname{cov}(Z,\alpha_1 X)}{\operatorname{cov}(Z,X)} + \frac{\operatorname{cov}(Z,\varepsilon)}{\operatorname{cov}(Z,X)} = \\ &= \alpha_1 + \frac{\operatorname{cov}(Z,\varepsilon)}{\operatorname{cov}(Z,X)} \xrightarrow[n \to \infty]{} \alpha_1 + \frac{0}{\sigma_{zx}} = \alpha_1. \end{aligned} \tag{8.6}$$

Тут ми скористалися такими співвідношеннями:

$$\begin{aligned} &\operatorname{cov}(Z, \alpha_0) = 0; \\ &\alpha_0 = \text{const}; \operatorname{cov}(Z, \alpha_1 X) = \alpha_1 \operatorname{cov}(Z, X). \end{aligned} \tag{8.7}$$

За великих обсягів вибірки розподіл $b_1^{із}$ прямує до нормального:

$$b_1^{із} \approx N\left(\alpha_1, S^2\left(b_1^{із}\right)\right), \tag{8.8}$$

де $S^2\left(b_1^{із}\right) = \dfrac{S^2 \sum (z_t - \bar{z})}{\left[\sum (z_t - \bar{z})(x_t - \bar{x})\right]^2}$; $S^2 = \dfrac{1}{T} \sum e_t^2$;

$$e_t = y_t - b_0^{із} - b_1^{із} x_t.$$

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

8.2. Оператор оцінювання Вальда

Нехай економетрична модель має вигляд

$$Y_t = a_0 + a_1 x_t + \varepsilon_t \qquad (8.9)$$

Послідовність дій для визначення другого рядку матриці (8.10)

1 етап

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

$$x' = \begin{pmatrix} 1 & 1 & 1 & 1 & \ldots & 1 \\ x_1 & x_2 & x_3 & x_4 & \ldots & x_n \end{pmatrix}. \qquad (8.11)$$

2 етап

Величини відхилень, що мають знак «плюс», замінюються на одиниці, величини відхилень, що мають знак «мінус», – на одиниці з цим знаком.
Використовуючи оператор оцінювання:

$$\hat{a} = (Z'x)^{-1} Z'Y, \qquad (8.12)$$

маємо

$$\hat{a} = \begin{pmatrix} n & 0 \\ 0 & \frac{n}{2}(\overline{x}_2 - \overline{x}_1) \end{pmatrix} \begin{pmatrix} n\overline{Y} \\ \frac{n}{2}(\overline{Y}_2 - \overline{Y}_1) \end{pmatrix}, \qquad (8.13)$$

де $\bar{x}_2$ і $\bar{x}_1$ характеризують середні відхилення значень X відповідно вгору і вниз від медіани, а $\bar{Y}_2$ і $\bar{Y}_1$ – середні значення залежної змінної, які відповідають середнім $\bar{x}_2$ і $\bar{x}_1$. Звідси

$$\hat{a} = \begin{pmatrix} \overline{Y} \\ \dfrac{\overline{Y}_2 - \overline{Y}_1}{\overline{x}_2 - \overline{x}_1} \end{pmatrix}, \qquad (8.14)$$

Це означає, що параметр a, у моделі (8.9) подається у вигляді

$$\hat{a} = \frac{\overline{Y}_2 - \overline{Y}_1}{\overline{x}_2 - \overline{x}_1}, \qquad (8.15)$$

причому

$$\hat{a}_0 = \overline{Y} - a_1 \overline{x} \qquad (8.16)$$

Рис. 8.1. Послідовність дій для визначення другого рядку матриці (8.10)

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

$$Z' = \begin{pmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ -1 & -1 & -1 & 1 & \dots & 1 \end{pmatrix}. \qquad (8.10)$$

Щоб визначити другий рядок цієї матриці, необхідно виконати такі дії.

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

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

8.3. Особливості оцінювання методом Бартлета

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

$$\hat{a}_1 = \frac{\bar{Y}_2 - \bar{Y}_1}{\bar{x}_2 - \bar{x}_1}, \qquad (8.17)$$

де $\bar{x}_i$, $\bar{Y}_i$ – середні величини для спостережень, які потрапили в дві крайні групи. Вільний член оцінюється так само, як і в (8.16).

Поділ вибіркової сукупності спостережень на три рівні групи має задовольняти вимоги прикладних досліджень, оскільки немає можливості дістати точну інформацію про закон розподілу значень X.

8.4. Оператор оцінювання Дарбіна

Дарбін запропонував упорядковувати значення вектора X у порядку зростання і ввів як інструментальну змінну порядковий номер (ранг), тобто числа 1, 2, 3, 4, ..., n. Учений показав, що для великих вибіркових сукупностей ефективність застосування такого методу оцінювання досягає майже 96% ефективності оцінок 1 МНК, 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 (візьмемо Z' із 8.10):

$$x' = \begin{pmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ x_1 & x_2 & x_3 & x_4 & \dots & x_n \end{pmatrix}, \tag{8.18}$$

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

Оператор оцінювання

$$\hat{a}_1 = \frac{\sum_{i=1}^{n} iY_i}{\sum_{i=1}^{n} ix_i}. \tag{8.19}$$

Причому $\hat{a}_0 = \overline{Y} - \hat{a}_1\bar{x}$, де i – порядковий номер, Z – матриця 8.10.

Дисперсія оцінок параметрів

$$\operatorname{var}(\hat{a}) = \sigma_u^2 (Z'x)^{-1} (Z'Z)(x'Z)^{-1}, \tag{8.20}$$

де Z', x' – транспоновані[1] матриці Z і x відповідно;
$(Z'x)^{-1}$, $(x'Z)^{-1}$ – обернена[2] матриця до матриці $Z'x$, $x'Z^1$ відповідно.

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

---

[1] ***Транспонована матриця*** – матриця A^T, що виникає з матриці A в результаті унарної операції транспонування: заміни її рядків на стовпчики.

[2] ***Обернена матриця*** – матриця (позначається A^{-1}), яка існує для кожної невиродженої квадратної матриці A, розмірності n x n, причому $AA^{-1} = A^{-1}A = I_n$, де I_n – одинична n x n матриця.

8.5. Побудова економетричної моделі на основі даних, що містять помилки вимірювання змінних

Раніше ми припускали, що змінні вимірюються без помилок, і лише відхилення ε – це єдина припустима форма помилок. Останнє було пов'язане з наміром урахувати вплив різних пояснювальних змінних, які не входять до економетричної моделі в явному вигляді.

Проте досить часто при вимірюванні змінних, які належать до економетричної моделі, припускаються помилок. Тоді постає запитання, як наявність помилок змінних може вплинути на оцінку параметрів моделі.

Щоб відповісти на це запитання, розглянемо матрицю незалежних змінних X, елементи якої містять помилки.

Нехай

$$x = \tilde{x} + V, \tag{8.21}$$

де $\tilde{x}$ – матриця розміром $n \times m$ справжніх (фактичних) значень;
V – матриця помилок вимірювання.

Тоді модель має вигляд:

$$Y = \tilde{x}a + \varepsilon \tag{8.22}$$

або

$$Y = xa + (\varepsilon - Va). \tag{8.23}$$

Оцінка параметрів для цієї моделі 1 МНК матиме вигляд

$$\hat{a} = a + \left(x'x\right)^{-1} x'\left(\varepsilon - Va\right),$$

де $\left(x'x\right)^{-1} x'\left(\varepsilon - Va\right)$ – величина зміщення оцінки.

Обґрунтованість цієї оцінки залежить від того, чи дорівнює нулю

$$P\lim_{n\to\infty}\left[\frac{1}{n}x'\left(\varepsilon - Va\right)\right]. \tag{8.25}$$

Запишемо:

$$P\lim_{n\to\infty}\left[\frac{1}{n}x'\left(\varepsilon - Va\right)\right] = P\lim_{n\to\infty}\left(\frac{1}{n}x'\varepsilon\right) - P\lim_{n\to\infty}\left(\frac{1}{n}x'V\right)a. \tag{8.26}$$

За припущення, що залишки *u* не корелюють гранично зі змінними X (як зі справжніми значеннями, так і з їх помилками), можна стверджувати таке:

$$P\lim_{n\to\infty}\left(\frac{1}{n}x'\varepsilon\right)=0\,. \tag{8.27}$$

Проте

$$P\lim_{n\to\infty}\left(\frac{1}{n}x'V\right)=P\lim_{n\to\infty}\left(\frac{1}{n}x'V\right)+P\lim_{n\to\infty}\left(\frac{1}{n}V'V\right). \tag{8.28}$$

Отже, навіть тоді, коли помилки вимірювання змінних X не корелюють зі справжніми значеннями цих змінних і перший доданок у правій частині дорівнює нулю, другий доданок, який характеризує матрицю коваріацій помилок, здебільшого не дорівнює нулю. А це означає, що за наявності помилок вимірювання змінних оцінка параметрів моделей 1 МНК є необґрунтованою і асимптотичне зміщення визначається формулою:

$$P\lim_{n\to\infty}(\hat{a}-a)=-P\lim_{n\to\infty}(x'x)^{-1}\cdot P\lim_{n\to\infty}\left(\frac{1}{n}V'V\right)a. \tag{8.29}$$

Наприклад, якщо ми оцінюємо параметри моделі з двома змінними 1 МНК, то зміщення

$$P\lim_{n\to\infty}(\hat{a}-a)=-\frac{\sigma_v^2 a}{\sigma_x^2+\sigma_v^2} \tag{8.30}$$

або

$$P\lim_{n\to\infty}\hat{a}=-\frac{a}{1+\dfrac{\sigma_v^2}{\sigma_x^2}}\,, \tag{8.31}$$

де $\sigma_v{}^2$ – дисперсія помилки вимірювання X;
$\sigma_x{}^2$ – дисперсія справжніх значень X, причому припускаємо, що помилки вимірювання не корелюють із цими значеннями X.

Рівняння (8.31) показує, що оцінка справжнього значення параметра моделі занижена. Наприклад, як би ми не збільшували сукупність спостережень, якщо $\sigma_v{}^2 = 10\%$ від $\sigma_x{}^{\mathbf{2}}$, то оцінка параметра $\hat{a}$ відрізнятиметься від справжнього значення також майже на 10%, тобто за наявності помилок вимірювання змінних збільшення сукупності спостережень не компенсує зміщення.

Тому при оцінюванні параметрів економетричної моделі, коли трапляються помилки вимірювання змінних, доцільно застосувати метод інструментальних змінних, який ми розглянули раніше.

8.6. Алгоритм Уоліса

Одна зі складнощів моделі – це існування кореляції y_{t-1} з v_t. Однак, ураховуючи зроблене припущення, коли пояснювальні змінні ймовірніше за все не корелюють з y_{t-1}, оцінку параметрів моделі

$$y_{t-1} = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + \ldots + \varepsilon_t \tag{8.32}$$

можна знайти за допомогою 1 МНК. Кількість лагових значень X, які включаються в цю модель, можна вибрати залежно від обсягу вибірки і від їх здатності пояснити поводження залежної змінної y_t. Якщо значення змінної X має високу автокореляцію, то навряд чи потрібно брати більше, ніж два її лагові значення. Записане вище співвідношення змістимо на один період назад, щоб дістати y_{t-1}, і підставимо вираз у праву частину моделі (8.33) замість y_{t-1}. Після цього застосовується 1 МНК для оцінки параметрів a. Ці оцінки будуть обґрунтованими, бо всі пояснювальні змінні гранично не корельовані із залишками, але вони будуть неефективними, оскільки при оцінюванні параметрів не була врахована автокореляція залишків.

Проведені Уолісом експерименти показали, що його метод оцінювання приводить до значно менших величин зміщення і до меншої суми квадратів залишків, ніж застосування методу Ейткена (рис. 8.2).

Алгоритм Уоліса

1 етап

Оцінюються параметри моделі

$$y_t = a_0 + a_1 y_{t-1} + a_2 x_t + v_t \ , \qquad (8.33)$$

де x_{t-1} використовується як інструментальна змінна для y_{t-1}. Таким чином, визначають:

$$Z = \begin{pmatrix} 1 & x_0 & x_1 \\ 1 & x_1 & x_2 \\ 1 & x_2 & x_3 \\ ... & ... & ... \\ 1 & x_{n-1} & x_n \end{pmatrix} \text{ і } X = \begin{pmatrix} 1 & y_0 & y_1 \\ 1 & y_1 & y_2 \\ 1 & y_2 & y_3 \\ ... & ... & ... \\ 1 & y_{n-1} & y_n \end{pmatrix}, \ Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ ... \\ y_{n-1} \end{pmatrix}. \qquad (8.34)$$

2 етап

Для залишків цієї моделі $v = Y - XA$ розраховують коефіцієнт автокореляції першого порядку з урахуванням поправки на зміщення:

$$r = \frac{\dfrac{\sum_{t=2}^{n} v_t v_{t-1}}{n-1}}{\dfrac{\sum_{t=1}^{n} v_t^2}{n}} + \frac{m+1}{n}, \qquad (8.35)$$

де $r \approx p$.

3 етап

За допомогою оцінки, здобутої для ρ, формують матрицю:

$$S = \begin{pmatrix} 1 & r & r^2 & ... & r^{n-1} \\ r & 1 & r & ... & r^{n-2} \\ r^2 & r & 1 & ... & r^{n-3} \\ ... & ... & ... & ... & ... \\ r^{n-1} & r^{n-2} & r^{n-3} & ... & 1 \end{pmatrix} \qquad (8.36)$$

і обчислюють оцінку вектора A узагальненим методом найменших квадратів:

$$A = (X'S^{-1}X)^{-1} X'S^{-1}Y. \qquad (8.37)$$

де S^{-1} – обернена матриця, X' – транспонована

Рис. 8.2. Послідовність етапів алгоритму Уоліса

Завдання для самоконтролю

1. Назвати основні властивості, які повинна мати інструментальна змінна.
2. Описати схему використання інструментальних змінних на прикладі парної регресії.
3. Описати особливості та методику розрахунків у розрізі оператора оцінювання Вальда.
4. Наведіть основні переваги та недоліки оцінювання параметрів економетричної моделі методом Бартлета.
5. Назвати та охарактеризувати основні етапи оператора оцінювання Дарбіна.
6. У чому полягають основні концептуальні аспекти практичного застосування алгоритму Уоліса?

Практичне заняття 8

Дослідження автокореляції

Постановка завдання

1 етап	Побудова інформаційного масиву дослідження шляхом збору статистичних даних та розробка моделі парної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), анаіз моделі
2 етап	Побудова множинної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), дослідження моделі
3 етап	Виявлення, дослідження та усунення мультиколінеарності факторів у моделі множинної лінійної регресії методом Фаррара – Глобера
4 етап	Виявлення ефекту гетероскедастичності на основі застосування параметричного тесту Гольдфельда – Квандта та усунення даного ефекту шляхом використання узагальненого методу найменших квадратів (методу Ейткена)
5 етап	Побудова адитивної та мультиплікативної моделі часового ряду на основі проведення декомпозиційного аналізу: визначення структури ряду (тенденції та сезонних (циклічних) коливань), проведення автокореляційного аналізу, визначення якості моделі, розрахунок прогнозних значень
6 етап	Групування елементів економічної системи шляхом кластерного аналізу, побудова деревоподібної дендограми, графічне подання елементів кластеризації
7 етап	Перевірка наявності автокореляції залишків за допомогою критерію Дарбіна – Уотсона, оцінка параметрів рівняння множинної лінійної регресії методом Ейткена
8 етап	**Побудова економіко-математичної моделі методом інструментальних змінних; побудова економетричної моделі на основі даних, що містять помилки вимірювання змінних, оцінка параметрів моделі за допомогою алгоритму Уоліса**
9 етап	Побудова моделі розподіленого лага, оцінка параметрів моделей з лагами в незалежних змінних, оцінювання параметрів авторегресійних моделей; перетворення ARMA і ARIMA
10 етап	Структурне моделювання (моделювання на основі системи структурних рівнянь), оцінка невідомих параметрів моделі, визначення адекватності моделі щодо опису структури вхідних даних

Мета роботи: набуття навичок побудови економетричної моделі методом інструментальних змінних, оцінки її параметрів методом Уоліса. Дослідження моделі.

Початкові дані

За даними 17 рівнів часових рядів вивчається залежність величини активів банків Y (млрд грн) від власного капіталу X_1 млрд грн).

Необхідно:

1. Сформувати масив вхідних статистичних даних шляхом коригування даних на відповідні лаги.
2. Побудувати лінійну модель множинної регресії $\hat{y}_{t-1} = a_0 + a_1 x_{t-1} + a_2 x_{t-2_i}$.
3. Знайти коефіцієнт множинної детермінації. Зробити висновки щодо адекватності побудованої моделі.
4. За допомогою F-критерію Фішера оцінити статистичну значущість рівняння регресії та коефіцієнта детермінації R^2_{yx1x2}.
5. За допомогою критерію Стьюдента оцінити статистичну значимість параметрів, що відповідають факторам x_{t-1} та x_{t-2}.

Приклад виконання

За даними 17 рівнів часових рядів вивчається залежність величини активів банків Y (млрд грн) від власного капіталу X_1 млрд грн).

Таблиця 8.1. Вхідні дані для побудови моделі

Дата	ВК	Усього активів
01.01.2009	119263048	926086498
01.04.2009	117081585	870633535
01.07.2009	112597492	864694968
01.10.2009	117968018	889958533
01.01.2010	120207619	873449574
01.04.2010	126646323	874964709
01.07.2010	127162304	885255711
01.10.2010	132802031	917497465
01.01.2011	137725113	942083994
01.04.2011	138434527	995033185
01.07.2011	147816550	1019811043
01.10.2011	151866259	1029162518
01.01.2012	155486926	1054272287
01.04.2012	162236166,3	1082473105
01.07.2012	163775906,4	1104395259
01.10.2012	165810129,5	1117445882
01.01.2013	170196261,8	1127179379

Розв'язання

Для зручності будемо проводити проміжні розрахунки у вигляді таблиці.

Таблиця 8.2. Проміжні розрахунки

Дата	ВК (X)	Усього активів (Y)	Y_{t-1}	X_{t-1}	X_{t-2}
01.01.2009	1,19E+08	9,26E+08			
01.04.2009	1,17E+08	8,71E+08	926086498	119263048	
01.07.2009	1,13E+08	8,65E+08	870633535	117081585	119263048
01.10.2009	1,18E+08	8,9E+08	864694968	112597492	117081585
01.01.2010	1,2E+08	8,73E+08	889958533	117968018	112597492
01.04.2010	1,27E+08	8,75E+08	873449574	120207619	117968018
01.07.2010	1,27E+08	8,85E+08	874964709	126646323	120207619
01.10.2010	1,33E+08	9,17E+08	885255711	127162304	126646323
01.01.2011	1,38E+08	9,42E+08	917497465	132802031	127162304
01.04.2011	1,38E+08	9,95E+08	942083994	137725113	132802031
01.07.2011	1,48E+08	1,02E+09	995033185	138434527	137725113
01.10.2011	1,52E+08	1,03E+09	1,02E+09	147816550	138434527
01.01.2012	1,55E+08	1,05E+09	1,029E+09	151866259	147816550
01.04.2012	1,62E+08	1,08E+09	1,054E+09	155486926	151866259
01.07.2012	1,64E+08	1,1E+09	1,082E+09	162236166	155486926
01.10.2012	1,66E+08	1,12E+09	1,104E+09	163775906	162236166
01.01.2013	1,7E+08	1,13E+09	1,117E+09	165810130	163775906

1. Побудувати лінійну модель множинної регресії.

Для знаходження параметрів лінійного рівняння множинної регресії $y_{t\text{-}1} = a_0 + a_1 x_{t\text{-}1} + a_2 x_{t\text{-}2t}$, необхідно розв'язати таку систему лінійних рівнянь відносно невідомих параметрів a, b_1, b_2, увівши умовні позначення $x_{t\text{-}1} = x_1 x_{t\text{-}2} = x_2 y_{t\text{-}1} = y$:

$$\begin{cases} na + b_1 \sum x_1 + b_2 \sum x_2 = \sum y \\ a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 = \sum y x_1 \\ a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 = \sum y x_2 \end{cases}$$

або скористатися готовими формулами:

$$b_1 = \frac{\sigma_y}{\sigma_{x_1}} \cdot \frac{r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}$$

$$b_2 = \frac{\sigma_y}{\sigma_{x_2}} \cdot \frac{r_{yx_2} - r_{yx_1} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}$$

$$a = \overline{y} - b_1 \overline{x}_1 - b_2 \overline{x}_2$$

Для знаходження параметрів скористаємося можливостями «Аналіз даних – Регресія» за допомогою MS Excel. Шляхом розрахунків отримаємо такі результати статистичного аналізу, наведені в табл. 8.3.

Таблиця 8.3. Реультати статистичного аналізу залежності величини активів банків *Y* (млрд грн) від власного капіталу X_1, млрд грн) методом інструментальних змінних (алогритму Уоліса)

Регресійна статистика	
Множинний R	0,9812
R-квадрат	0,9628
Нормований R-квадрат	0,9561
Стандартна похибка	19482288,2686
Спостереження	14,0000

Дисперсійний аналіз

	df	*SS*	*MS*	*F*	*Значимість F*
Регресія	2	1,082E+17	5,41E+16	142,53248	1,36468E-08
Залишок	11	4,1752E+15	3,796E+14		
Всього	13	1,1237E+17			

	Коефіцієнти	*Стандартна похибка*	*t-статистика*	*P-значення*	*Нижні 95%*	*Верхні 95%*
Y-перетин	256247237,64	42982113,22	5,96	0,00	161644244,35	350850230,94
1,17E+08	2,53	1,59	1,60	0,14	-0,96	6,03
1,19E+08	2,67	1,65	1,62	0,13	-0,96	6,29

Таким чином, на основі застосування даних табл. 8.3 рівняння $y_{t1} = a_0 + a_1 x_{t\text{-}1} + a_2 x_{t\text{-}2t}$ буде набувати такого вигляду:

$$\hat{y}_{t-1} = 256247237{,}64 + 2{,}53 x_{t-1} + 2{,}67 x_{t-2_t}.$$

2. Коефіцієнт множинної детермінації 0,9628 оцінює частку варіації результату за рахунок поданих у рівнянні факторів у загальній варіації результату. У цьому прикладі ця частка становить 96,28% і вказує на досить високий ступінь обумовленості варіації результату

варіацією факторів, інакше кажучи, на досить тісний зв'язок факторів з результатом.

3. За допомогою *F*-критерію Фішера оцінити статистичну значущість рівняння регресії та коефіцієнта детермінації $R^2_{yx_1x_2}$.

Оцінку надійності рівняння регресії в цілому і показника щільності зв'язку $R^2_{yx_1x_2}$, дає *F*-критерій Фішера:

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

У нашому випадку експериментальне значення *F*-критерію Фішера: 142,53. Отримали, що $F_{факт} > F_{табл} = 3{,}49$ (якщо $n = 20$), тобто, ймовірність отримати таке значення *F*-критерію не перевищує припустимий рівень значущості 5%. Відповідно, отримане значення не випадкове, воно сформувалося під впливом суттєвих факторів, тобто підтверджується статистична значущість усього рівняння і показника щільності зв'язку.

4. За допомогою критерію Стьюдента доведено статистчину значущість параметрів шуканого рівняння, оскільки фактичні значення даного критерію перевищують критично припустимий рівень.

Тести

1. Визначити властивості інструментальної змінної:
   а) повинна корелювати (бажано значною мірою) із заміненою пояснювальною змінною;
   б) повинна корелювати (бажано значною мірою) із заміненою результативною ознакою;
   в) не повинна корелювати з випадковим відхиленням;
   г) має корелювати з випадковим відхиленням.

2. Визначити послідовність етапів оператору оцінювання Вальда:
   а) матриця пояснювальних змінних для цієї моделі запишеться у вигляді:

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

   б) знайти відхилення кожного елемента вектора X від медіани;
   в) використовуючи оператор оцінювання

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

маємо

$$\hat{a} = \begin{pmatrix} n & 0 \\ 0 & \frac{n}{2}(\bar{x}_2 - \bar{x}_1) \end{pmatrix} \begin{pmatrix} n\bar{Y} \\ \frac{n}{2}(\bar{Y}_2 - \bar{Y}_1) \end{pmatrix}, \ \hat{a}_0 = \bar{Y} - a_1\bar{x};$$

   г) величини відхилень, що мають знак «плюс», замінюються на одиниці, величини відхилень, що мають знак «мінус», — на одиниці з цим знаком.

3. За допомогою якого методу ефективність оцінки можна збільшити, якщо розбити впорядковані значення змінної X на три групи однакової величини:
   а) оператор оцінювання Вальда;
   б) оператор оцінювання Бартлета;
   в) оператор оцінювання Дарбіна;
   г) алгоритм Уоліса.

4. Визначити, у межах якого підходу економетрична модель не буде мати вільного члена:
   а) оператор оцінювання Вальда;
   б) оператор оцінювання Бартлета;
   в) оператор оцінювання Дарбіна;
   г) алгоритм Уоліса.

5. Який із підходів дозволяє усунути одну зі складнощів моделі — існування кореляції y_{t-1} з v_t:
   а) оператор оцінювання Вальда;
   б) оператор оцінювання Бартлета;
   в) оператор оцінювання Дарбіна;
   г) алгоритм Уоліса.

6. Як наявність помилок змінних може вплинути на оцінку параметрів моделі?
   а) оцінка параметрів моделей 1 МНК є необґрунтованою;
   б) оцінка параметрів моделей 1 МНК є обґрунтованою;
   в) оцінка параметрів моделей 1 МНК є асимпотично зміщеною;
   г) оцінка параметрів моделей 1 МНК є асимпотично незміщеною.

7. Принципи використання ІЗ передбачають виконання таких умов:
   а) $\operatorname{cov}(Z; X) \neq 0$;
   б) $\operatorname{cov}(Z; X) = 0$;
   в) $M(\varepsilon_t) = 0$;
   г) $M(\varepsilon_t) \neq 0$.

8. Наведена система

$$\begin{cases} b_1^{із} = \dfrac{\sum (z_t - \bar{z})(y_t - \bar{y})}{\sum (z_t - \bar{z})(x_t - \bar{x})} \\ b_0^{із} = \bar{y} - b_1^{із}\bar{x}. \end{cases}$$

відображає математичні співвідношення визначення параметрів економетричної моделі методом інструментальних змінних?
   а) так;
   б) ні.

Т Е М А 9

Моделі розподіленого лага

Поняття лага і лагових змінних • Моделі розподіленого лага. Взаємна кореляційна функція • Оцінка параметрів моделей з лагами в незалежних змінних: метод послідовного збільшення кількості лагів, перетворення Койка (метод геометричної прогресії) • Оцінювання параметрів авторегресійних моделей • Виявлення автокореляції залишків в авторегресійних моделях • Авторегресійне перетворення • Перетворення методом ковзного середнього • Перетворення ARMA і ARIMA

9.1. Поняття лага і лагових змінних

Часовий лаг, або просто лаг – проміжок часу, у який відбувається запізнювання (затримка), що характерно для багатьох економічних процесів, та який проявляється в тому, що ефект від впливу деякого фактора на показник, що характеризує процес, виявляється не відразу, а поступово, через певний час або протягом певного часу.

Наприклад, функція споживання після різкої зміни доходу також змінюється, але не пропорційно до доходу, зокрема, зі збільшенням доходу витрати значно зростають (задовольняються нагальні потреби), а потім можуть зменшуватися шляхом збільшення інвестицій (плануються великі витрати і зростають нагромадження, а споживання зменшується). Вкладання коштів у наукові дослідження не відразу впливає на зростання продуктивності праці – має пройти певний час від виникнення наукової ідеї до її впровадження у виробництво. Капітальне будівництво також не дає миттєвих прибутків і т. ін.

Дистрибутивно-лагова модель – це модель, у якій досліджуваний показник у момент часу t визначається не лише поточними, а й попередніми значеннями незалежних змінних:

$$y_t = a + a_0 x_t + a_1 x_{t-1} + a_2 x_{t-2} + \ldots + \varepsilon_t. \qquad (9.1)$$

Динамічна, авторегресійна, модель – модель, у якій досліджуваний показник у момент часу t визначається своїми попередніми значеннями:

$$y_t = a + a_0 x_t + b_1 y_{t-1} + b_2 y_{t-2} + ... + \varepsilon_t. \quad (9.2)$$

Змінні $x_{t-\tau}$ і $y_{t-\tau}$ у період $t - \tau$ називаються лаговими змінними, а величина τ – періодом зсуву (*лагом*).

Модель з кінцевим лагом – це модель, у якій незалежні змінні використовують за кілька попередніх періодів.

Нескінченна лагова модель – модель, у якій вплив незалежних змінних не обмежується певним періодом.

Звичайно, нескінченна лагова модель більш загальна, однак практичне застосування такої моделі досить проблематичне через велику кількість факторів, складність внутрішньої структури та обмеженість часових рядів – інформаційної бази моделей.

Коефіцієнти a_j, $j = 0, 1, 2 ...$ називаються коефіцієнтами лага, а послідовність $\{a_j, j = 0, 1, 2 ...,\}$ – структурою лага.

Короткостроковий, впливовий мультиплікатор являє собою коефіцієнт a_0 за незалежної змінної x_t, що відбиває її вплив на залежну змінну в поточний період.

Проміжний інтервал – часткові суми коефіцієнтів $(a_0 + a_1)$, $(a_0 + a_1 + b_1)$, ..., що відображають зміну y_t у другий, третій і наступні періоди.

Довгостроковий, загальний, дистрибутивно-лаговий мультиплікатор – загальна сума лагових коефіцієнтів для всієї моделі.

Нормований коефіцієнт лага – відношення кожного коефіцієнта до їх суми.

Нормована структура лага – сукупність нормованих коефіцієнтів лага.

Усі нормовані коефіцієнти менші від одиниці, а їх сума дорівнює одиниці.

9.2. Моделі розподіленого лага

Дистрибутивно-лагові моделі, які ще називають *моделями розподіленого лага*, задовільно описують економічні процеси лише в стабільних (незмінних) умовах. Необхідність враховувати ще й поточні умови функціонування вимагає застосування узагальнених моделей.

Узагальнена модель розподіленого лага – це модель, що містить не лише лагові змінні, а й змінні, які безпосередньо впливають на досліджуваний показник (тобто містить і поточні умови функціонування):

$$y_t = \sum_{\tau \geq 0} a_\tau x_{t-\tau} + \sum_{i=1}^{m} b_i x_{t,i} + \varepsilon_t. \tag{9.3}$$

Оцінювання параметрів таких рівнянь ускладнюється обмеженнями, що накладаються на коефіцієнти при лагових змінних.

Перш ніж будувати економетричну модель з лаговими змінними, необхідно обґрунтувати величину лага.

Взаємна кореляційна функція – послідовність коефіцієнтів кореляції, які визначають ступінь зв'язку кожного елемента вектора залежної змінної з елементом вектора незалежної змінної, зсунутими один відносно одного на часовий лаг τ.

$$r_\tau = \frac{(n-\tau)\sum_{t=1}^{n-\tau} y_t x_{t+\tau} - \sum_{t=1}^{n-\tau} y_t \sum_{t=1}^{n-\tau} x_{t+\tau}}{\sqrt{\left[(n-\tau)\sum_{t=1}^{n-\tau} y_t^2 - \left(\sum_{t=1}^{n-\tau} y_t\right)^2\right]\left[(n-\tau)\sum_{t=1}^{n-\tau} x_{t+\tau}^2 - \left(\sum_{t=1}^{n-\tau} x_{t+\tau}\right)^2\right]}}. \tag{9.4}$$

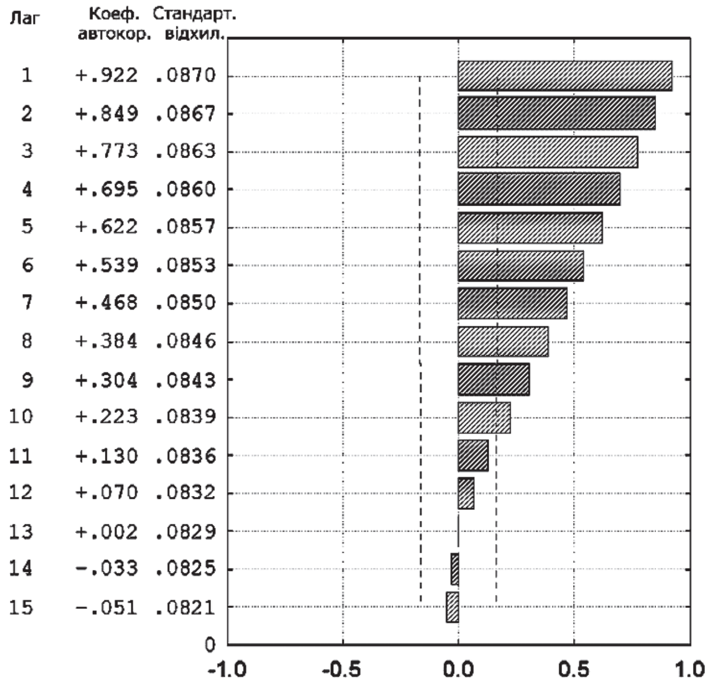


Рис. 9.1. Функція автокореляції

Серед отриманих коефіцієнтів кореляції вибирають найбільший за модулем, а відповідне значення часового зсуву вважають лагом. Якщо таких коефіцієнтів кілька, застосовують модель розподіленого лага.

9.3. Оцінка параметрів моделей з лагами в незалежних змінних: метод послідовного збільшення кількості лагів, перетворення Койка (метод геометричної прогресії)

Для оцінювання параметрів дистрибутивно-лагових моделей звичайно застосовують два можливі підходи – послідовне оцінювання і апріорне оцінювання.

Ідея першого підходу полягає в тому, щоб поступово досліджувати вплив запізнених змінних на залежну змінну. Другий підхід базується на припущенні, що параметри моделі мають певну закономірність, тобто пов'язані між собою деякими співвідношеннями.

Такий метод хоч і повний, однак має певні недоліки. По-перше, те, що невідомою є максимальна тривалість лага, а це не дає можливості передбачити, скільки змінних увійде в модель. По-друге, між послідовними значеннями змінних здебільшого має місце висока кореляція, що породжує проблему мультиколінеарності в моделі. Крім того, через зменшення ступенів свободи в таких моделях оцінки стають дещо непевними, що також знижує їх якість.

Наявність мультиколінеарності між лаговими змінними ускладнює побудову моделі. Щоб усунути мультиколінеарність, на коефіцієнти при лагових змінних накладають додаткові обмеження (*апріорне оцінювання*), а саме: вибирають їх так, щоб вплив лага на досліджуваний показник був "односпрямований" (тобто коефіцієнти були б однакового знака) і зменшувався з кожним наступним кроком у минуле. Такі припущення реалізують, як правило, у моделях, де параметри змінюються в геометричній прогресії. Крім того, нескінченна сума членів спадної геометричної прогресії є скінченною величиною, що

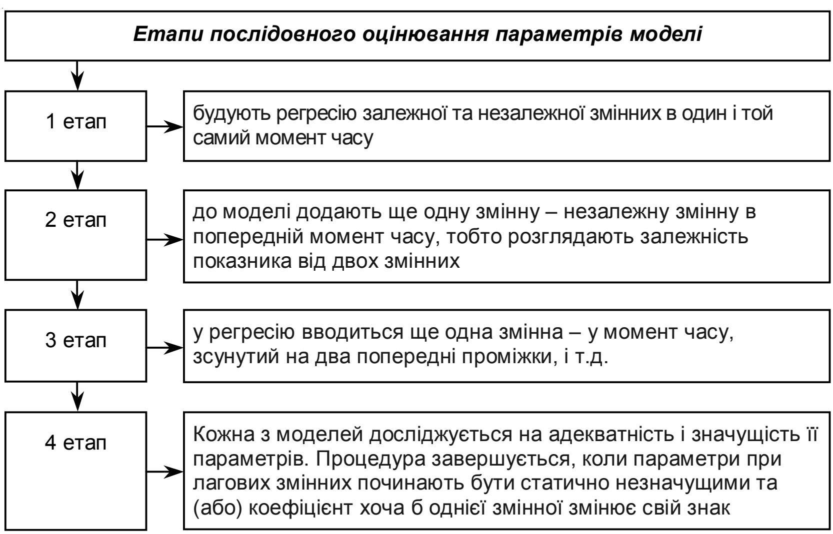


Рис. 9.2. Етапи оцінювання параметрів моделі

дає можливість узагальнити модель з кінцевим лагом і застосовувати однакові методи оцінювання параметрів.

Однак і в цьому разі залишається велика кількість оцінюваних параметрів.

Уведення в модель лагової залежної змінної y_{t-1} (затримка на один період), відоме як перетворення Койка, значно спрощує модель:

$$y_t = w(1-\lambda)x_t + \lambda y_{t-1} + (\varepsilon_t - \lambda\varepsilon_{t-1}), \tag{9.5}$$

де $w = \sum_{j=1}^{\infty} a_j$ (скінченне число) $0 \le \lambda \le 1$;

λ — темп зменшення дистрибутивного лага, тобто деякий постійний темп зменшення в часі лаговагових впливів фактора на результат;

$1 - \lambda$ — швидкість пристосування.

Заміна незалежних лагових змінних $x_{t-1}, x_{t-2}, \ldots$ однією залежною змінною y_{t-1} зменшує кількість оцінюваних параметрів і усуває проблему мультиколінеарності, однак призводить до нових труднощів. Наявність у моделі лагової залежної змінної потребує перевірки передумови про незалежність змінних і залишків при застосуванні звичайного МНК. Крім того, залишки моделі $\nu_t = \varepsilon_t + \lambda\varepsilon_{t-1}$ часто виявляються серійно корельованими, а тому при дослідженні їх на автокореляцію необхідно використати спеціальні тести.

Отримана алгебраїчним способом модель Койка позбавлена теоретичного обгрунтування і фактично є послідовною моделлю.

З певних економічних міркувань можна отримати моделі, що зовні нагадують модель Койка, але з іншою інтерпретацією коефіцієнтів лагових змінних. Такими моделями є модель адаптивних сподівань

$$y_t = \alpha(1-\lambda) + \alpha(1-\lambda)x_t + \lambda y_{t-1} + \varepsilon_t(1 - \lambda\varepsilon_{t-1}) \tag{9.6}$$

та модель часткового коригування

$$y_t = \alpha\lambda + \alpha\gamma x_t + (1-\gamma)y_{t-1} + \varepsilon_t,\ 0 \le \gamma \le 1. \tag{9.7}$$

Ці моделі відрізняються від моделі Койка наявністю вільного члена, але при цьому реалізують різні ідеї щодо економічної діяльності. У першій моделі (9.6) відображено думку про те, що люди навчаються з попереднього досвіду, причому нещодавній досвід має більший вплив, ніж попередній; друга (9.7) базується на тому, що через інертність економічної системи зміна одного економічного показника не відразу впливає на зміну іншого і відповідний рівень залежної змінної досягається через певний час.

9.4. Оцінювання параметрів авторегресійних моделей

Три авторегресійні моделі – Койка (9.5), адаптивних сподівань (9.6) і часткового коригування (9.7) – можна подати в загальній формі:

$$y_t = \alpha_0 + a_1 x_t + a_2 y_{t-1} + \upsilon_t. \quad (9.8)$$

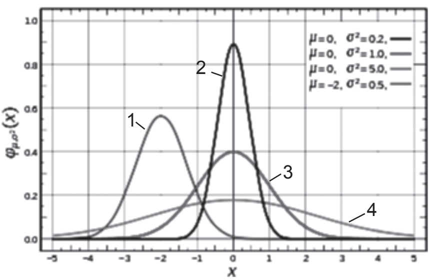
Наявність лагових залежних змінних у динамічних моделях створює певні проблеми при оцінюванні параметрів: серед пояснювальних змінних є стохастичні (залежні лагові змінні), а також існує проблема серійної кореляції залишків моделі та лагових змінних. Залежно від гіпотез щодо залишків таких моделей використовують відповідні методи оцінювання.

Гіпотеза 1. Залишки є нормально розподіленими[1] випадковими величинами з нульовим математичним сподіванням та сталою дисперсією.

---

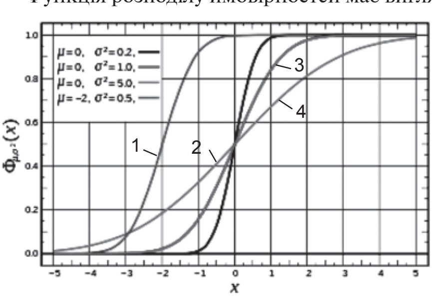
[1] Нормальний розподіл (розподіл Гаусса) — розподіл імовірностей випадкової величини, що характеризується густиною ймовірності.

Функція ймовірностей має вигляд



Крива (3) відповідає *стандартному нормальному розподілу*

Функція розподілу ймовірностей має вигляд



Параметри

Параметри $\mu \in R$ – математичне сподівання; $\sigma^2 > 0$ – дисперсія. Носій функції – $x \in$ R.

Розподіл імовірностей: $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Функція розподілу ймовірностей (cdf):

Гіпотеза 2. Залишки описуються авторегресійною схемою першого порядку:

$$\upsilon_t = \varepsilon_t - \lambda\varepsilon_{y-1},\ 0 \le \lambda \le 1. \tag{9.9}$$

Гіпотеза 3. Залишки автокорельовані та описуються авторегресійною схемою першого порядку:

$$\upsilon_t = \varepsilon_t + \lambda\varepsilon_{t-1}. \tag{9.10}$$

Крім того, $\upsilon_t = \rho\upsilon_{t-1} + \varepsilon_t$.

Перша гіпотеза виконується лише для моделі часткового коригування (9.7); саме для неї можливе застосування звичайного МНК. Однак залежність залишків від лагової змінної $y_{t\text{-}1}$ у цій моделі призводить до зміщення оцінок параметрів. Та хоча оцінки параметрів будуть завищеними, вони матимуть найменшу середньоквадратичну похибку. І після визначення величини зміщення МНК-оцінки будуть найприйнятнішими.

Якщо залишки моделі визначаються через автокорельовані випадкові величини, то МНК-оцінки параметрів моделі також матимуть зміщення, до того ж зміщення матиме також критерій Дарбіна – Уотсона. Тому для перевірки автокореляції залишків застосовують узагальнений критерій Дарбіна – Уотсона. Оцінювання параметрів таких моделей виконують узагальненим методом найменших квадратів (методом Ейткена), в операторі оцінювання якого

$$\hat{a} = \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} Y, \tag{9.11}$$

де $\hat{a}$ – вектор емпіричних оцінок параметрів регресійного рівняння.

$\frac{1}{2}\left[1 + erf\left(\frac{x-\mu}{\sqrt{2\sigma^2}}\right)\right]$. Середнє μ. Медіана μ. Мода μ. Дисперсія σ^2. Коефіцієнт асиметрії 0. Коефіцієнт ексцесу 0. Ентропія: $\frac{1}{2}\ln(2\pi e\sigma^2)$. Твірна функція моментів (mgf): $\exp\left\{\mu t + \frac{1}{2}\sigma^2 t^2\right\}$. Характеристична функція: $\exp\left\{i\mu t - \frac{1}{2}\sigma^2 t^2\right\}$.

$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, де μ – математичне сподівання, σ^2 – дисперсія випадкової величини. Параметр y також відомий як стандартне відхилення. Розподіл із $\mu = 0$ та $\sigma^2 = 1$ називають стандартним нормальним розподілом.

Центральна гранична теорема стверджує, що нормальний розподіл виникає тоді, коли дана випадкова величина являє собою суму великої кількості незалежних випадкових величин, кожна з яких відіграє в утворенні всієї суми незначну роль. Наприклад, відстань від влучення снаряду гармати до цілі за великої кількості пострілів характеризується саме нормальним розподілом.

Коригувальна матриця має вигляд

$$V = \sigma_u^2 \begin{pmatrix} 1+\lambda^2 & -\lambda & 0 & \ldots & 0 \\ -\lambda & 1+\lambda^2 & -\lambda & \ldots & 0 \\ 0 & -\lambda & 1+\lambda^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & 1+\lambda^2 \end{pmatrix}. \qquad (9.12)$$

Якщо лагову модель можна подати у вигляді

$$y_t - \lambda y_{t-1} = a_0 + a_1 x_1 + \upsilon_t, \qquad (9.13)$$

то до перетворених у такий спосіб даних залежної змінної застосовують звичайний МНК. Причому параметр λ вибирають з інтервалу $0 \le \lambda \le 1$ так, щоб мінімізувати суму квадратів залишків $\alpha V^{-1}\varepsilon$.

Якщо відносно залишків моделі приймається третя гіпотеза, то параметри оцінюють за допомогою таких методів:

1) класичного МНК після попереднього перетворення вхідних даних;
2) методу Ейткена (узагальненого МНК);
3) ітераційного методу;
4) методу інструментальних змінних;
5) алгоритму Уолліса.

9.5. Виявлення автокореляції залишків в авторегресійних моделях

Автокореляцію в авторегресійних моделях практично неможливо визначити за допомогою статистики Дарбіна – Уотсона, оскільки для цих моделей значення *DW,* навіть за наявності автокореляції, близьке до 2, що за умовою Дарбіна – Уотсона рівносильне відсутності автокореляції.

Для виявлення автокореляції в авторегресійних моделях Дарбін запропонував використовувати h-статистику, що має вигляд:

$$h = \overline{\rho}\sqrt{\frac{n}{1 - nD(g)}}, \qquad (9.14)$$

де n – обсяг вибірки;

$D(g)$ – дисперсія оцінки коефіцієнта при лаговій змінній y_{t-1};

$\bar{\rho}$ – оцінка коефіцієнта автокореляції першого порядку, яку можна визначити на основі формули:

$$\bar{\rho} \approx 1 - \frac{DW}{2}. \qquad (9.15)$$

Схема використання h-статистики для аналізу автокореляції має такі особливості:

- незалежно від того, скільки лагів змінної y включено в модель, значення h необхідно обчислювати з використанням дисперсії коефіцієнта при y_{t-1};
- статистика h не обчислюється, якщо $nD(g) > 1$ (на практиці такі ситуації майже не зафіксовані);
- застосування h доцільне лише в разі досить великого обсягу вибірки n.

Як зазначалося, автокореляція залишків призводить до отримання зміщених і неспроможних оцінок. Автокореляція може вказувати або на неправильну специфікацію рівняння, або на наявність важливих неврахованих факторів. Однак часто автокореляція викликається наявністю регресійної залежності між відхиленнями, тобто внутрішніми властивостями ряду $\{\nu_t\}$. Існує кілька способів усунення даної проблеми. Зокрема, для авторегресійних моделей пропонуються авторегресійне перетворення, перетворення методом ковзних середніх, моделі ARMA і ARIMA .

9.6. Авторегресійне перетворення

Нехай y – досліджувана величина, і її зміну можна описати за допомогою моделі

$$\left(y_t - \bar{y}\right) = \alpha\left(y_{t-1} - \bar{y}\right) + \varepsilon_t, \qquad (9.16)$$

де $\bar{y}$ – середнє значення Y;

ε_t – некорельовані випадкові відхилення з нульовим математичним очікуванням і постійною дисперсією σ^2 (такі відхилення при розгляді часових рядів іноді називають білим шумом).

Перетворення (9.16) у такому разі називають ***авторегресійним перетворенням першого порядку AR*(1)** (*autoregressive*). При цьому значення

y_t змінної y в момент часу t пропорційне її значенню y_{t-1} у момент часу (t – 1) плюс деяке випадкове відхилення. За аналогією

$$\left(y_t - \overline{y}\right) = \alpha\left(y_{t-1} - \overline{y}\right) + \alpha_2\left(y_{t-2} - \overline{y}\right) + \varepsilon_t \qquad (9.17)$$

називається ***авторегресійним перетворенням другого порядку*** **AR**(2);

$$\left(y_t - \overline{y}\right) = \alpha\left(y_{t-1} - \overline{y}\right) + \alpha_2\left(y_{t-2} - \overline{y}\right) + \ldots + \alpha_p\left(y_{t-p} - \overline{y}\right) + \varepsilon_t \qquad (9.18)$$

називається ***авторегресійним перетворенням порядку P*** **AR** (*P*).

В усіх цих перетвореннях поточне значення y_t змінної Y виражається тільки через її попередні значення і випадкову складову (білий шум) σ_t.

9.7. Перетворення методом ковзного середнього

Нехай модель задається формулою

$$y_t = \gamma + \beta_0\varepsilon_t + \beta_1\varepsilon_{t-1}, \qquad (9.19)$$

де $\gamma = const$, ε_t і ε_{t-1} – білий шум у поточний і попередній моменти часу. У цьому випадку значення змінної y в момент часу t дорівнює сумі константи і ковзної середньої між поточним і попереднім значеннями випадкового відхилення (білого шуму).

Співвідношення (9.19) називають ***перетворенням методом ковзних середніх першого порядку*** **MA(1)** (*moving avarage*).

Співвідношення

$$y_t = \gamma + \beta_0\varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \ldots + \beta_q\varepsilon_{t-q} \qquad (9.20)$$

називають ***перетворенням методом ковзних середніх порядку q MA (q)***.

9.8. Перетворення ARMA і ARIMA

Поєднання перетворень *AR* і *MA* називається ***авторегресійним перетворенням з ковзною середньою ARMA***. Наприклад, для змінної y перетворення ARMA (1,1) буде мати вигляд

$$y_t = \gamma + \alpha_1 y_{t-1} + \beta_0\varepsilon_t + \beta_1\varepsilon_{t-1}. \qquad (9.21)$$

У загальному випадку перетворення ARMA *(p,q)* охоплює *p* авторегресійних членів і *q* ковзних середніх.

Перетворення ARMA в поєднанні з переходом від об'ємних величин до приростних називається ***перетворенням ARIMA***. У деяких випадках такий перехід дозволяє отримати більш точну і явну модель залежності. Тут приростом (кінцевою різницею) першого порядку змінної y називається різниця $y_t - y_{t-1}$. Приростом порядку d змінної y називають різницю $y_t - y_{t-1} - y_{t-2} - \ldots - y_{t-d}$.

У загальному вигляді перетворення ARIMA *(p, d, q)* виражається формулою

$$y_t^* = \alpha_1 y_{t-1}^* + \ldots + \alpha_p y_{t-p}^* + \alpha_0 \varepsilon_t + \alpha_1 \varepsilon_{t-1} + \ldots + \alpha_q \varepsilon_{t-q}, \qquad (9.22)$$

де α_i, $i = 1, 2, .., p$; – невідомі параметри.

Величини y_{t-1}^*, $i = 1, 2, \ldots, p$ являють собою кінцеві різниці порядку d змінної y. ε_{t-1}, $i = 1, 2, \ldots, q$ – незалежні один від одного нормально розподілені випадкові величини з нульовим математичним очікуванням і постійною дисперсією.

Зазначимо, що перетворення AR, MA і ARIMA доцільно використовувати тоді, коли досить ясний набір пояснювальних змінних і загальний вигляд рівняння регресії, але водночас зберігається автокореляція залишків.

Завдання для самоконтролю

1. Дайте визначення часового лага.
2. Наведіть характерні риси дистрибутивно-лагової моделі, визначте особливості її застосування.
3. Записати математичне співвідношення формалізації динамічної або авторегресійної моделі в загальному вигляді.
4. Описати методику визначення взаємної кореляційної функції.
5. Указати гіпотези щодо залишків економетричних моделей, що мають виконуватися при виборі методу оцінювання їх параметрів.
6. Які відхилення при розгляді часових рядів іноді називають білим шумом?
7. Які перетворення називають авторегресійним перетворенням довільного порядку?

Практичне заняття 9

Побудова авторегресійних моделей

Постановка завдання

1 етап	Побудова інформаційного масиву дослідження шляхом збору статистичних даних та розробка моделі парної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), аналіз моделі
2 етап	Побудова множинної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), дослідження моделі
3 етап	Виявлення, дослідження та усунення мультиколінеарності факторів у моделі множинної лінійної регресії методом Фаррара – Глобера
4 етап	Виявлення ефекту гетероскедастичності на основі застосування параметричного тесту Гольдфельда – Квандта та усунення даного ефекту шляхом використання узагальненого методу найменших квадратів (методу Ейткена)
5 етап	Побудова адитивної та мультиплікативної моделі часового ряду на основі проведення декомпозиційного аналізу: визначення структури ряду (тенденції та сезонних (циклічних) коливань), проведення автокореляційного аналізу, визначення якості моделі, розрахунок прогнозних значень
6 етап	Групування елементів економічної системи шляхом кластерного аналізу, побудова деревоподібної дендограми, графічне подання елементів кластеризації
7 етап	Перевірка наявності автокореляції залишків за допомогою критерію Дарбіна – Уотсона, оцінка параметрів рівняння множинної лінійної регресії методом Ейткена
8 етап	Побудова економіко-математичної моделі методом інструментальних змінних; побудова економетричної моделі на основі даних, що містять помилки вимірювання змінних, оцінка параметрів моделі за допомогою алгоритму Уоліса
9 етап	**Побудова моделі розподіленого лага, оцінка параметрів моделей з лагами в незалежних змінних, оцінювання параметрів авторегресійних моделей; перетворення ARMA і ARIMA**
10 етап	Структурне моделювання (моделювання на основі системи структурних рівнянь), оцінка невідомих параметрів моделі, визначення адекватності моделі щодо опису структури вхідних даних

Мета роботи: набуття навичок побудови тренд-сезонних адитивних і мультиплікативних авторегресійних моделей у MS Excel.

Зміст роботи: за допомогою MS Excel провести оцінювання параметрів авторегресійних моделей для банківської системи України на основі вихідних даних.

Початкові дані

На основі самостійно зібраної інформації про банківську систему будь-якої країни (на оцінку 5 балів) необхідно провести:

- формування інформаційної бази дослідження шляхом вибору та проведення збору статистичних даних щодо динаміки релевантних показників кількісної оцінки стабільності банківської системи; комплексне дослідження їх основних закономірностей;
- побудову діаграми динаміки релевантних показників кількісної оцінки стабільності банківської системи;
- декомпозицію розглянутих часових рядів – фільтрацію трендової ($F(t)$), сезонної (циклічної) (S) та випадкової компонент;
- оцінювання параметрів авторегресійних моделей релевантних показників кількісної оцінки стабільності банківської системи;
- розрахунок характеристик (величин c і d) монотонності, згасання чи розхитування кожного із розглянутих часових рядів релевантних показників кількісної оцінки стабільності банківської системи;
- визначення динаміки індикатора стабільності банківської системи;
- визначення напряму та характеру взаємозв'язку між зазначеною результативною ознакою та обсягами власного капіталу, зобов'язань, активів по банківській системі в цілому на основі побудови рівняння нелінійної множинної регресії.

Для вирішення задачі використати: інструктивні матеріали, можливості та вбудовані функції MS Excel.

Вимоги до оформлення звіту

Звіт про проведення даної лабораторної роботи оформлюється, як й інші звіти цього курсу, в окремому зошиті згідно зі встановленими вимогами до оформлення звітів і містить:

- назву, тему, завдання, опис мети лабораторної роботи;
- вихідні дані варіанту;
- результати рішення;
- висновок про отримані результати;
- короткий опис технології вирішення задачі в MS Excel.

Хід виконання

Лібералізація економічних відносин і зростання глобалізації викликані швидким розвитком споживчого та фінансового ринків. Водночас, зазначені фактори негативно позначилися на фінансовому ринку і його сегментах, особливо в банківській системі, тому що банки відіграють найбільш важливу роль на фінансовому ринку і є основними фінансовими посередниками в нашій країні. Крім того, коливання в банківському секторі також може бути причиною обставин, що спричиняють ситуацію економічної нестабільності в країні. Таким чином, виникає питання необхідності проведення дослідження і розробки економічної математичної моделі, яка могла б встановити фактори стабільного розвитку банківської системи і з'ясувати механізм її регулювання на основі регресійного аналізу та аналізу розкладання.

Сучасний літературний аналіз у контексті виявлення і оцінювання стабільності банківської системи [2, 5, 10, 13, 15] дозволяє систематизувати існуючі підходи та розглядати цю категорію як здатність системи вчасно підтримувати стабільність основних характеристик в умовах незначного коливання ринку, адекватно прийняти та протидіяти впливу зовнішніх факторів і підтримувати стан довгострокової динамічної рівноваги [12, 14]. Заснований на сказаному вище підхід пропонує трансформацію стабільності банківської системи для здатності підтримувати стан для довгострокової динамічної рівноваги.

Таким чином, метою даної лабораторної роботи є моделювання динаміки і прогнозування стабільності на прикладі української банківської системи.

Досягнення цієї мети передбачає постановку і вирішення таких завдань:

- запропонувати кількісну оцінку (індикатор) стабільності характеристик банківської системи;
- визначити основні фактори динамічної стабільності формування концепту і дослідження його основних тенденцій та закономірностей;
- формалізувати відносини між сезонністю (циклічністю) факторів стабільності банківської системи та інтегральним показником;
- визначити вплив стабільності індикатора на основні характеристики функціонування банківської системи;
- прогнозувати стабільність банківської системи України протягом 2014–2015 року.

Дослідження стабільності банківської системи вимагає використання бази вхідної інформації, яка складається з даних таких часових рядів [6, 7]: власний капітал, активи, кредити та зобов'язання, процентні доходи і чистий прибуток / (збиток) протягом 2009–2012 років у контексті квартальних даних. Ці статистичні дані наведені в табл. 9.1.

Таблиця 9.1. Вхідна інформаційна база, тис. грн

Період часу	Основна сума	Кредити та заборгованості (позики)	Активи	Чистий прибу-ток/збиток	Процентні доходи
01.01.2009	119263048	741815978	926086498	7304241	88370294
01.04.2009	117081585	737913200	870633535	-7019821	32440838
01.07.2009	112597492	735094597	864694968	-14321422	61849586
01.10.2009	117968018	747775459	889958533	-20943905	91832989
01.01.2010	120207619	726295788	873449574	-31491841	119083052
01.04.2010	126646323	720948800	874964709	-4423405	28204679
01.07.2010	127162304	719569876	885255711	-8306141	55334622
01.10.2010	132802031	744082347	917497465	-9998904	84739672
01.01.2011	137725113	750536379	942083994	-13026584	113334120
01.04.2011	138434527	768065850	995033185	-210766	28123028
01.07.2011	147816550	789549434	1019811043	-1062004	54907633
01.10.2011	151866259	815319860	1029162518	-5641491	82993044
01.01.2012	155486926	813863749	1054272287	-7707548	113352487
01.04.2012	162236166,3	670016569	1082473105	1724830,4	28027985,26
01.07.2012	163775906,4	676956093,8	1104395259	1552033,4	56296229,75
01.10.2012	165810129,5	686335697,4	1117445882	2775150,2	88377226,86
01.01.2013	170196261,8	694381044,8	1127179379	4898805,5	119278015,7

Пропонуємо звернути увагу на діаграму динаміки характеристики показників банківської системи України, подану на рис. 9.3. На цій ілюстрації ми бачимо дві шкали: ліва використовується для відображення даних у вигляді гістограми, права – відповідно, у вигляді динамічних графіків. Аналіз наведених фігур дозволяє робити висновки про тенденції в зоні коливань таких показників, як процентні доходи і чистий прибуток / (збиток), а в контексті інших показників має місце збереження тенденції. Зазначений факт може бути причиною порушення стабільності банківської системи України.

Моделювання динаміки стабільності банківської системи пропонується подати на основі перевірки стаціонарних часових рядів шляхом адаптації методу Forster – Stewart для визначення особливостей предмета дослідження. Таким чином, стаціонарний часовий ряд – це процес, який характеризується постійним математичним очікуванням і дисперсією (без тенденції), автокореляційна функція залежить від

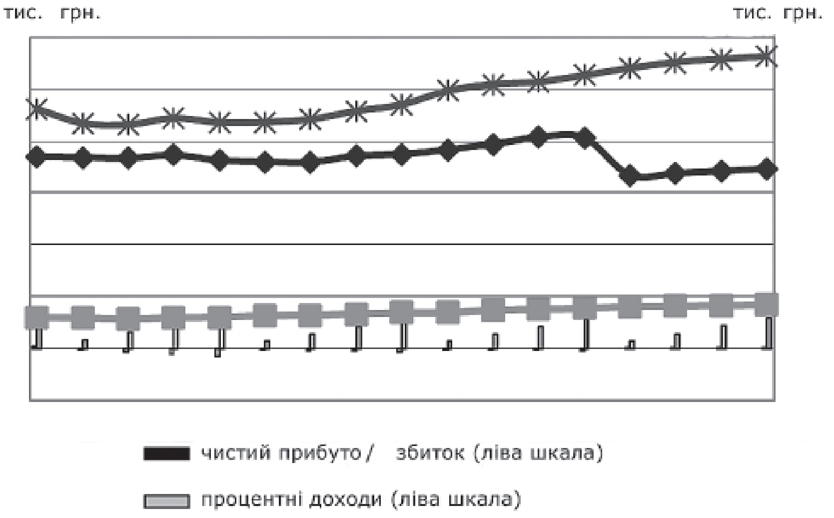


Рис. 9.3. Схема динаміки показників вхідних даних

двох наступних періодів часу, а не від певного періоду часу. Рекомендується більш детально проаналізувати етапи практичної реалізації запропонованого підходу з урахуванням зазначеного визначення.

1 етап. Створення науково-дослідної бази інформації шляхом збору статистичних даних в контексті відповідних показників динаміки кількісної оцінки стабільності банківської системи України (див. табл. 9.2); усебічний аналіз їх основних порушень. Такими показниками є: темпи зростання чистого прибутку / збитку; достатність капіталу банків; процентна маржа; ROA (рентабельність активів); ROE (рентабельність власного капіталу) (рис. 9.4).

Необхідною умовою для застосування методу Forster – Steward для моделювання динаміки стабільності банківської системи є однорідність розглянутих часових рядів даних рівнів.

Гомогенність означає відсутність атипових, аномальних спостережень, а також спотворення тенденції. Аномальний рівень – це окреме значення часового ряду, яке не відповідає потенціалу економічної системи, у процесі вивчення, і яке, залишаючись на рівні часового ряду, справляє істотний вплив на значення основних характеристик часових рядів. Для виявлення аномальних рівнів часових рядів пропонується використовувати метод Ірвіна. Цей метод заснований на порів-

Таблиця 9.2. Показники кількісної оцінки стабільності банківської системи України

Дата	Темпи зростання чистого прибутку / збитку,%	Достатність капіталу банків, частка (10–2)	Процентна маржа, частка (10–2)	ROA Рентабельність активів, частка (10–2)	ROE Рентабельність власного капіталу, частка (10–2)
01.01.2009	–	12,88	11,91	0,79	6,12
01.04.2009	–1,96	13,45	4,40	–0,81	–6,00
01.07.2009	1,04	13,02	8,41	–1,66	–12,72
01.10.2009	0,46	13,26	12,28	–2,35	–17,75
01.01.2010	0,50	13,76	16,40	–3,61	–26,20
01.04.2010	–0,86	14,47	3,91	–0,51	–3,49
01.07.2010	0,88	14,36	7,69	–0,94	–6,53
01.10.2010	0,20	14,47	11,39	–1,09	–7,53
01.01.2011	0,30	14,62	15,10	–1,38	–9,46
01.04.2011	–0,98	13,91	3,66	–0,02	–0,15
01.07.2011	4,04	14,49	6,95	–0,10	–0,72
01.10.2011	4,31	14,76	10,18	–0,55	–3,71
01.01.2012	0,37	14,75	13,93	–0,73	–4,96
01.04.2012	–1,22	14,99	4,18	0,16	1,06
01.07.2012	–0,10	14,83	8,32	0,14	0,95
01.10.2012	0,79	14,84	12,88	0,25	1,67
01.01.2013	0,77	15,10	17,18	0,43	2,88

нянні часових рядів сусідніх значень та розрахунку його характеристик. Розраховані значення порівнюються з критичним значенням і, якщо вони не перевищують критичні значення, відповідні рівні вважаються нормальними. У табл. 9.3 аномальні рівні досліджуваних часових рядів позначені зірочкою, які слід уникати шляхом обчислення середнього значення між попереднім і наступним аномальним рівнем відповідного часового ряду.

Отримані аномальні значення (у табл. 9.3) можуть бути надані в такий спосіб:

- процентна маржа, ROA рентабельність активів, ROE рентабельність власного капіталу на 01.04.2009 і темпи зростання чистого прибутку / збитку на 01.07.2009 – пояснюються початком кризи в банківській системі України, а також одержанням набагато меншого або навіть негативного значення прибутку;

Таблиця 9.3. Аномальні рівні перевірки вивчення часових рядів за допомогою методу Ірвіна

Дата	Темпи зростання чистого прибутку / збитку,%	Достатність капіталу банків, частка	Процентна маржа, частка	ROA Рентабельність активів, частка	ROE Рентабельність власного капіталу, частка
01.01.2009	–	–	–	–	–
01.04.2009	–	0,80	1,69*	1,44*	1,50*
01.07.2009	1,81*	0,60	0,91	0,77	0,83
01.10.2009	0,35	0,33	0,87	0,63	0,62
01.01.2010	0,02	0,71	0,93	1,13	1,04
01.04.2010	0,82	1,00	2,81*	2,81*	2,80*
01.07.2010	1,05	0,15	0,85	0,39	0,38
01.10.2010	0,41	0,15	0,83	0,14	0,12
01.01.2011	0,06	0,20	0,84	0,27	0,24
01.04.2011	0,78	0,99	2,58*	1,23	1,15
01.07.2011	3,03*	0,82	0,74	0,08	0,07
01.10.2011	0,16	0,37	0,73	0,40	0,37
01.01.2012	2,38*	0,01	0,84	0,17	0,15
01.04.2012	0,96	0,34	2,20*	0,81	0,74
01.07.2012	0,68	0,22	0,93	0,02	0,01
01.10.2012	0,54	0,01	1,03	0,10	0,09
01.01.2013	0,01	0,37	0,97	0,17	0,15

- аномальність значення процентної маржі, ROA рентабельність активів, ROE рентабельність власного капіталу на 01.04.2010 можна пояснити певним поліпшенням ситуації в українському банківському секторі, зокрема, значним скороченням банківських витрат.
- процентна маржа на 01.04.2011 пояснюється чотириразовим зниженням процентних доходів у зв'язку зі значним зниженням кредитної активності банків; водночас аномальність даного показника на 01.04.2012 пов'язана з кризою відновлення національної системи і досягненням набагато більшого рівня процентних доходів порівняно з попереднім періодом;
- аномальність усіх значень темпу зростання чистого прибутку / збитку на 01.07.2011 і 01.01.2012 пов'язана, у першу чергу, із збільшенням

обсягу заощаджень, а в другому випадку, навпаки, з переходом вітчизняної банківської системи на корисне функціонування.

Детальний аналіз часових рядів, поданих на рис. 9.4, дозволяє визначити основні тенденції в зміні індексу стабільності національної банківської системи [1]. Таким чином, чітка лінійна тенденція розвитку спостерігається в контексті достатності капіталу і рентабельності активів банківських установ, тоді як сезонні коливання характерні для таких показників, як процентна маржа і рентабельність капіталу. Згадані вище варіації зазначених показників викликають коливання тенденцій у динаміці індикатора стабільності банківської системи.

Формулювання і визначення кількісної оцінки виявлених закономірностей передбачається реалізувати за допомогою використання математичних методів через розкладання розглянутих часових рядів фільтрації тренду ($F(t)$), сезонної (циклічної) (S) і випадкової складових:

- тенденція сезонної авторегресії мультиплікативної моделі рентабельності власного капіталу [3, 4]:

$$ROE_t = \alpha + \beta \cdot ROE_{t-m} + F(t) \cdot S =$$
$$= 0{,}002 + 0{,}31 \cdot ROE_{t-4} + 0{,}60 \cdot (0{,}01t - 0{,}16)(-0{,}24)^{d_1} 0{,}98^{d_2} 1{,}26^{d_3} 1{,}99^{d_4}, \quad (1)$$

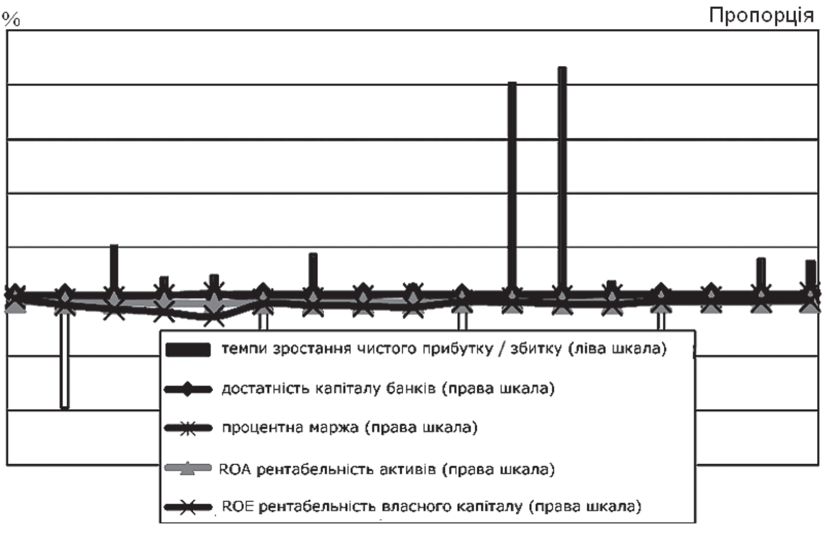


Рис. 9.4. Схема відповідних показників динаміки кількісної оцінки стабільності банківської системи України

де ROE – рентабельність власного капіталу на момент часу T;
d_1 (d_2, d_3, d_4) – індикатор першого (другого, третього, четвертого) кварталу, який набуває значення: «1», якщо момент часу розрахунку відповідає першому (другому, третьому, четвертому) кварталу, «0» – в іншому випадку.

На рис. 9.5 подано графічне зображення формули (1).

Адитивна модель сезонного процентної маржі [8]:

$$PM_t = F(t) + S = (-0{,}01t + 0{,}10) + 0{,}41d_1 + 0{,}81d_2 + 1{,}18d_3 + 1{,}51d_4, \quad (2)$$

де PM_t – процентна маржа в момент часу T.

На рис. 9.6 подано графічне зображення формули (2).

Сезонна мультипликативна модель ROA рентабельності активів:

$$ROA_t = 0{,}001 + 0{,}43 \cdot ROA_{t-4} + \\ + 0{,}22 \cdot (0{,}01t - 0{,}02)(-0{,}24)^{d_1} 0{,}97^{d_2} 1{,}26^{d_3} 2{,}01^{d_4} \quad (3)$$

де ROA_t – рентабельність активів на момент часу T.

На рис. 9.7 ми бачимо графічне зображення формули (3).

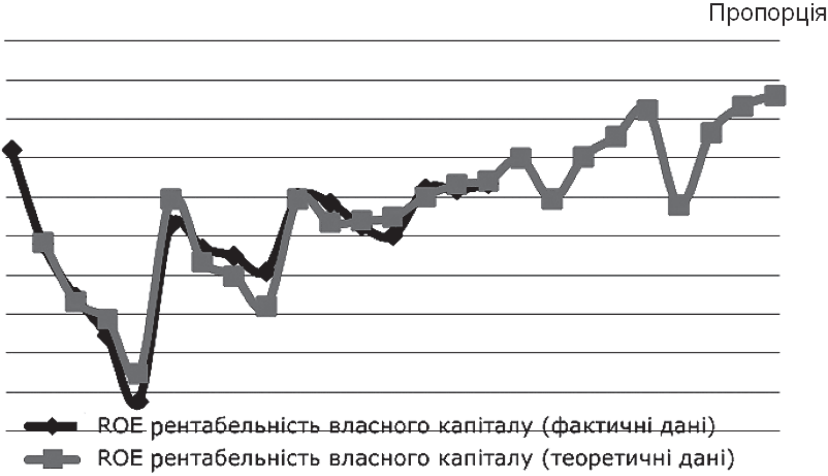


Рис. 9.5. Діаграма порівняння фактичних і теоретичних даних у контексті ROE

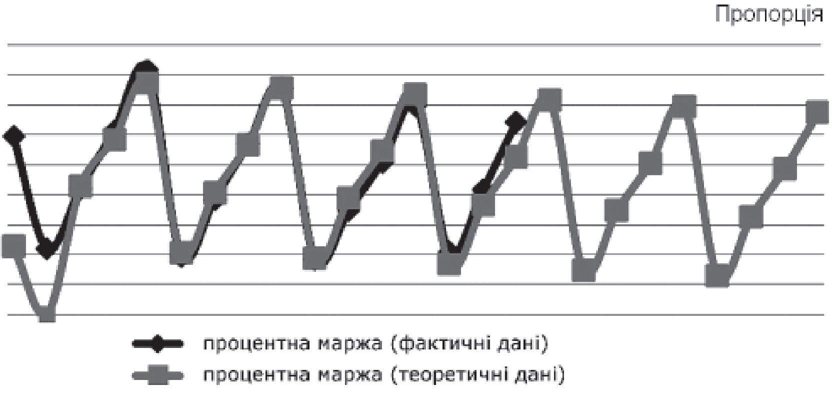


Рис. 9.6. Діаграма порівняння фактичних і теоретичних даних у контексті процентної маржі

Сезонна додаткова модель темпу зростання чистого прибутку / збитку:

$$RGP_t = F(t) + S =$$
$$= (8{,}39t + 4{,}40) - 171{,}52d_1 + 96{,}82d_2 + 91{,}59d_3 - 33{,}581d_4 , \qquad (4)$$

де RGP_t – темп зростання чистого прибутку / збитку на момент часу T.

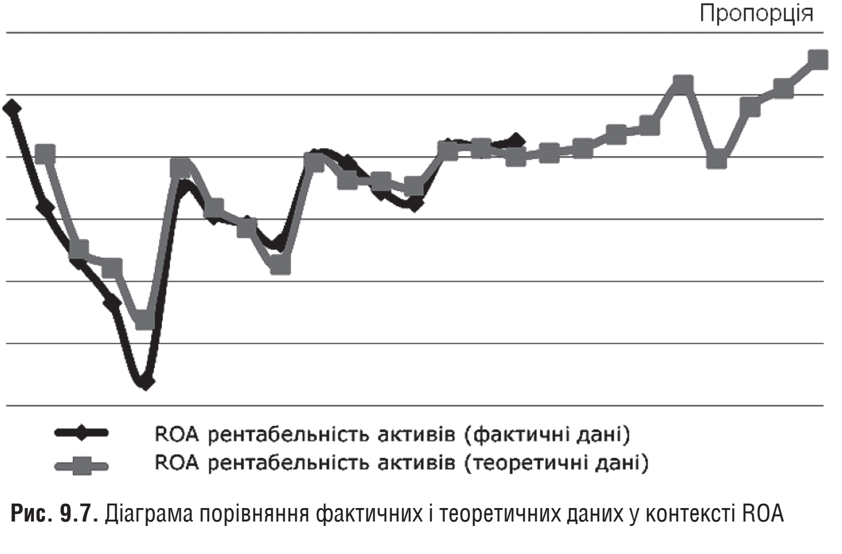


Рис. 9.7. Діаграма порівняння фактичних і теоретичних даних у контексті ROA

На рис. 9.8 подано графічне зображення формули (4).

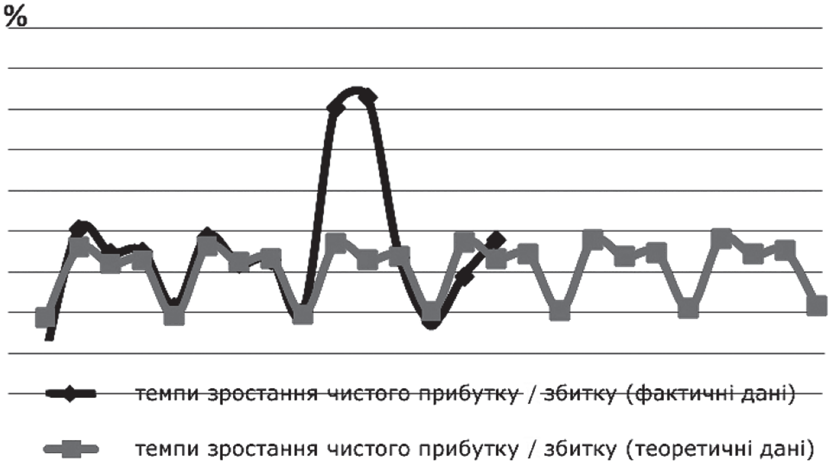


Рис. 9.8. Діаграма порівняння фактичних і теоретичних даних у контексті темпів зростання чистого прибутку / збитку

2 етап (табл. 9.4). Визначення двох числових послідовностей K і L на основі порівняння значення кожного вихідного часового ряду (послідовно один із сукупності окремих показників) з попереднього згідно з таким принципом: у контексті чисельної послідовності K – єдине значення, якщо рівень часового ряду (за винятком першого) більше, ніж усі попередні, і ненульове значення в іншому випадку, у послідовності L – єдине значення, якщо рівень часового ряду (за винятком першого) менше, ніж усі попередні, і ненульове значення в іншому випадку.

3 етап. Розрахунок монотонності характеристик (змінних C і D), згасання або розхитування відповідних показників кожного часового ряду кількісної оцінки стабільності банківської системи України:

$$c = \sum_{t=2}^{n}(k_t + l_t), \quad d = \sum_{t=2}^{n}(k_t - l_t), \qquad (5)$$

де $k_t\,(l_t)$ – значення чисельної послідовності K (відповідно L) у момент часу t.

Таблиця 9.4. Монотонність характеристик часових рядів стійкості відповідних показників банківської системи України

Дата	Темпи зростання чистого прибутку / збитку	Достатність капіталу банків	Процентна маржа	Темпи зростання чистого прибутку / збитку	Достатність капіталу банків	Процентна маржа
01.01.2009						
01.04.2009	0	1	0	0	0	1
01.07.2009	1	0	0	0	1	1
01.10.2009	0	0	1	1	1	0
01.01.2010	0	1	1	1	0	0
01.04.2010	0	1	0	1	0	1
01.07.2010	0	0	0	1	1	1
01.10.2010	0	0	0	1	1	1
01.01.2011	0	1	0	1	0	1
01.04.2011	0	0	0	1	1	1
01.07.2011	1	0	0	0	1	1
01.10.2011	1	1	0	0	0	1
01.01.2012	0	0	0	1	1	1
01.04.2012	0	1	0	1	0	1
01.07.2012	0	0	0	1	1	1
01.10.2012	0	0	0	1	1	1
01.01.2013	0	1	1	1	0	0

1 – ROA рентабельність активів
2 – ROE рентабельність власного капіталу
3 – ROA рентабельність активів
4 – ROE рентабельність власного капіталу

4 етап. Порівняння фактичних значеннь t-відносин з критичними, висновки щодо підтвердження (якщо обчислюване значення перевищує критичну) або відкидання гіпотези про тенденції в середньому і в дисперсії відповідних часових рядів (табл. 9.5). Формула набирає вигляду:

$$t_c = \frac{|c-\mu|}{\sigma_1}, \ \ \sigma_1 = \sqrt{2\ln n - 3{,}43}, \ \ t_d = \frac{|d-0|}{\sigma_2}, \ \ \sigma_2 = \sqrt{2\ln n - 0{,}85}, \qquad (6)$$

де μ – оцінка математичних очікувань часового ряду;
σ_1 – оцінка середньоквадратичного відхилення для значень C;
σ_2 – оцінка середньоквадратичного відхилення для значення D.

Таблиця 9.5. Динаміка показників стабільності банківської системи України

Дата	c	d	t_c	t_d	t_c	t_d	R
01.01.2009							
01.04.2009	4	– 2	0,7982	0,9109	0	0	0
01.07.2009	5	– 3	0,1303	1,3663	0	0	0
01.10.2009	5	– 3	0,1303	1,3663	0	0	0
01.01.2010	5	– 1	0,1303	0,4554	0	0	0
01.04.2010	5	– 3	0,1303	1,3663	0	0	0
01.07.2010	5	– 5	0,1303	2,2772	0	1	1
01.10.2010	5	– 5	0,1303	2,2772	0	1	1
01.01.2011	5	– 3	0,1303	1,3663	0	0	0
01.04.2011	5	– 5	0,1303	2,2772	0	1	1
01.07.2011	5	– 3	0,1303	1,3663	0	0	0
01.10.2011	5	– 1	0,1303	0,4554	0	0	0
01.01.2012	5	– 5	0,1303	2,2772	0	1	1
01.04.2012	5	– 3	0,1303	1,3663	0	0	0
01.07.2012	5	– 5	0,1303	2,2772	0	1	1
01.10.2012	5	– 5	0,1303	2,2772	0	1	1
01.01.2013	5	– 1	0,1303	0,4554	0	0	0

5 етап. Визначення динаміки стабільності індикатора банківської системи (R) у вигляді часових рядів, елементи якого розраховуються як кількість T-відносин у момент часу T.

Практична реалізація зазначених вище підходів дозволяє визначити інтервали часу, протягом яких українська банківська система характеризується недостатнім рівнем стабільності (рис. 9.9). Таким чином, індикатор стабільності набуває значення «1» (факт нестійкого стану) у другій половині 2010 року, у другому кварталі 2011 року і протягом майже всього 2012 року.

Результати, наведені на діаграмі, повністю підтвердили адекватність економічної моделі. Таким чином, розумно сказати, що з 10.05.2010 по 01.09.2012 українська банківська система була нестабі-

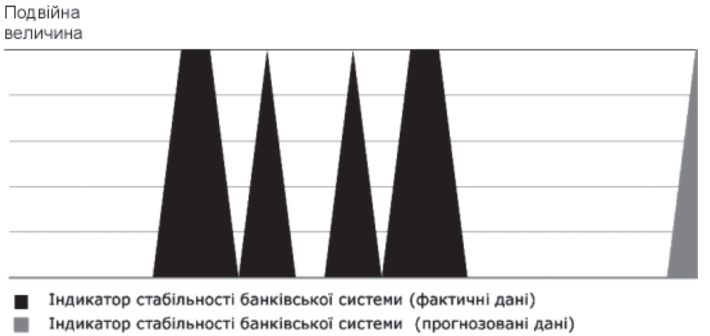


Рис. 9.9. Діаграма порівняння фактичних і теоретичних даних у контексті індикатора стабільності банківської системи

льною. Саме в цей період на фінансову кризу вплинули вражаючі фактори, перш за все, на ліквідність національної банківської системи.

Безперечно, у цей період були квартали, коли системі вдалося повернутися до рівноваги, але продовження цього етапу було коротким. Тільки на початку 2013 року почався відновлювальний період.

Слід зазначити, що на основі прогнозних розрахунків, ми можемо побачити нову хвилю кризи на початку 2015 року. Для нашої країни мають велике значення відносини між фінансовим сектором і політичними подіями. Крім того, очевидно, що до 2015 року НБУ буде тримати ризики валютного обміну під контролем і надавати допомогу проблемним банкам. Таким чином, сезонний характер відповідних показників і циклічного розвитку репресора стабільності банківської системи, створеної на попередніх етапах дослідження, дає можливість отримати цілком адекватне прогнозування результатів формування української банківської системи.

Стабільність банківської системи визначається безліччю факторів і впливами на формування основних показників банківського сектора. У дослідженні доречно визначити напрям і характер зв'язку між регресом і величиною власного капіталу, зобов'язаннями й активами банківської системи в цілому [12, 14]. Таким чином, кількісне визначення регресорів на стабільність індикатора пропонується здійснити на основі економетричного підходу, який передбачає будівництво кількох нелінійних рівнянь регресії в такій формі [9]:

$$R = -91783{,}33 - 44754{,}38 \ln OC - 338405{,}64 \ln E + 385645{,}31 \ln A + \\ + 1260{,}82 \ln^2 OC + 8606{,}86 \ln^2 E - 9744{,}91 \ln^2 A, \quad (9.23)$$

де R – показник стабільності банківської системи;
OC – власний капітал,
E – заборгованість,
A – активи.

Аналіз параметрів рівняння (9.23) дозволяє дійти таких висновків. Зі збільшенням власного капіталу банків на суму 177 480 000 000 (18% ВВП) рівень стабільності банківської системи буде поступово зменшуватися. Збільшення власного капіталу вище встановленого значення забезпечить відродження української стабільності банківської системи. Аналогічна тенденція характерна для заборгованостей (критичне значення 196590000000 – 19,94% ВВП), на відміну від активів, вплив яких на ефективність протилежна. Тому відносини між індикатором стабільності і активами подані у вигляді многочлена другого порядку гілками вниз, збільшення активів до рівня 197 870 000 000 (20,07% ВВП) супроводжується поясненням збільшення змінної. Надлишок фактора вказаного значення поступово призводить до втрати стану стабільності банківської системи.

Точність і адекватність виявлених закономірностей підтверджуються: за критерієм Фішера, фактична вартість якого 3,52 перевищує критичний допустимий рівень 3,41; на коефіцієнт детермінації на рівні 70,10%, до статистично значущих параметрів регресії рівняння (T-критерій) [11].

Таким чином, моделювання динаміки стабільності української банківської системи дозволяє визначити індикатор стабільності, досліджувати його динаміку у вигляді часових рядів, визначити основні чинники його формування, реалізувати розкладання системотвірних компонентів індикатора стабільності.

Практична цінність висновків і рекомендацій таких аспектів: застосування запропонованої моделі підходу дозволяє отримати ранній сигнал про недосконалість існуючого управління фінансовими ресурсами в контексті банківської системи та необхідність корекції; встановлена вище сезонність основних факторів стабільності банківської системи дає можливість отримати основу для підтвердження прийняття рішень; залежності лінійної та нелінійної регресії забезпечують можливість кількісно описати закономірності та перспективи досягнення стабільності системи.

Тести

1. Проміжок часу, у який має місце запізнювання (затримка), що характерно для багатьох економічних процесів, а також проявляється в тому, що ефект від впливу деякого фактора на показник, який характеризує процес, виявляється не відразу, а поступово, через деякий час або протягом деякого часу:

а) часовий лаг;
б) лагова змінна;
в) проміжний інтервал;
г) короткостроковий мультиплікатор;
д) взаємна кореляційна функція.

2. Визначте правильні відповіді:

Динамічна або авторегресійна модель –	модель, у якій досліджуваний показник у момент часу t визначається не лише поточними, а й попередніми значеннями незалежних змінних.
Дистрибутивно-лагова модель –	модель, у якій досліджуваний показник у момент часу t визначається своїми попередніми значеннями.
Узагальнена модель розподіленого лага –	модель, що містить не лише лагові змінні, а й змінні, що безпосередньо впливають на досліджуваний показник (тобто містить і поточні умови функціонування).

3. Визначте правильні відповіді:

Короткостроковий або впливовий мультиплікатор	являє собою коефіцієнт a_0 за незалежної змінної x_t, що відбиває її вплив на залежну змінну в поточний період.
Проміжний інтервал –	часткові суми коефіцієнтів $(a_0 + a_1)$, $(a_0 + a_1 + a_2)$, ..., що відображають зміну y_t в другий, третій і наступні періоди.
Довгостроковий або загальний дистрибутивно-лаговий мультиплікатор –	загальна сума лагових коефіцієнтів для всієї моделлю.

4. Послідовність коефіцієнтів кореляції, які визначають ступінь зв'язку кожного елемента вектора залежної змінної з елементом вектора незалежної змінної, зсунутими один відносно одного на часовий лаг:
   а) часовий лаг;
   б) лагова змінна;
   в) проміжний інтервал;
   г) короткостроковий мультиплікатор;
   д) взаємна кореляційна функція.

5. Визначити послідовність етапів оцінки параметрів моделей з лагами в незалежних змінних:
   а) кожна з моделей досліджується на адекватність і значущість її параметрів;
   б) процедура завершується, коли параметри при лагових змінних починають бути статистично незначущими та (або) коефіцієнт хоча б однієї змінної змінює свій знак;
   в) будують регресію залежної та незалежної змінних в один і той самий момент часу;
   г) у регресію вводиться ще одна змінна — у момент часу, зсунутий на два попередні проміжки, і т.д.;
   д) до моделі додають ще одну змінну — незалежну змінну в попередній момент часу, тобто розглядають залежність показника від двох змінних.

6. Наведена економетрична модель $y_t = w(1-\lambda)x_t + \lambda y_{t-1} + (\varepsilon_t - \lambda\varepsilon_{t-1})$ враховує:
   а) перетворення Койка;
   б) перетворення методом адаптивних сподівань;
   в) перетворення методом часткового коригування.

7. Яка з наведених гіпотез виконується лише для моделі часткового коригування:
   а) залишки є нормально розподіленими випадковими величинами з нульовим математичним сподіванням та сталою дисперсією;
   б) залишки описуються авторегресійною схемою першого порядку.

8. Схема використання *h*-статистики для аналізу автокореляції має такі особливості використання:
   а) незалежно від того, скільки лагів змінної y включено в модель, значення h необхідно обчислювати з використанням дисперсії коефіцієнта при y_{t-1};
   б) незалежно від того, скільки лагів змінної y включено в модель, значення h необхідно обчислювати з використанням дисперсії коефіцієнта при y_{t-2};

в) статистика h не обчислюється, якщо $nD(g) > 1$ (на практиці такі ситуації майже не зафіксовані);
г) статистика h не обчислюється, якщо $nD(g) < 1$ (на практиці такі ситуації майже не мають місця);
д) застосування h доцільне лише в разі досить великого обсягу вибірки n.

9. Співвідношення $y_t = \gamma + \beta_0\varepsilon_t + \beta_1\varepsilon_{t-1} + \beta_2\varepsilon_{t-2} + \ldots + \beta_q\varepsilon_{t-q}$ називають:

а) авторегресійним перетворенням з ковзною середньою ARMA;
б) перетворенням методом ковзних середніх порядку q MA (q);
в) авторегресійним перетворенням другого порядку AR (2).

10. Співвідношення $y_t^* = \alpha_1 y_{t-1}^* + \ldots + \alpha_p y_{t-p}^* + \alpha_0\varepsilon_t + \alpha_1\varepsilon_{t-1} + \ldots + \alpha_q\varepsilon_{t-q}$ називають:

а) авторегресійним перетворенням з ковзною середньою ARMA;
б) перетворенням методом ковзних середніх порядку q MA (q);
в) авторегресійним перетворенням другого порядку AR (2).

ТЕМА 10

Економетричні моделі на основі системи структурних рівнянь

Системи одночасних структурних рівнянь • Структурна і приведена форми моделі • Проблеми ідентифікації. Жорстко ідентифікована, неідентифікована і надідентифікована системи рівнянь • Проблема оцінювання параметрів системи, загальна характеристика методів. Непрямий метод оцінювання параметрів жорстко ідентифікованої системи рівнянь • Двокроковий метод найменших квадратів оцінювання параметрів надідентифікованих систем одночасних рівнянь (2МНК-оцінка) • Трикроковий метод найменших квадратів • Рекурсивні системи одночасних рівнянь, їх характеристика, можливість застосування МНК-оцінки для розрахунку параметрів рекурсивних систем • Прогноз і загальні довірчі інтервали

10.1. Системи одночасних структурних рівнянь

При використанні окремих рівнянь регресії , наприклад для економічних розрахунків, здебільшого передбачається, що аргументи (фактори) можна змінювати незалежно один від одного. Однак це припущення є дуже грубим: практично зміна однієї змінної, як правило, не може відбуватися за абсолютної незмінності інших. Її зміна потягне за собою зміни всієї системи взаємопов'язаних ознак. Отже, окремо взяте рівняння множинної регресії не може характеризувати істинний вплив окремих ознак на варіацію результуючої змінної. Саме тому останніми десятиліттями в економічних дослідженнях важливе місце посідає проблема опису структури зв'язків між змінними системами так званих одночасних рівнянь, званих також структурними рівняннями.

Система рівнянь в економетричних дослідженнях може бути побудована по-різному.

Можлива *система незалежних рівнянь*, коли кожна залежна змінна y розглядається як функція одного й того самого набору факторів x:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\ y_2 = a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_m. \end{cases} \quad (10.1)$$

Набір факторів x_j у кожному рівнянні може варіювати. Кожне рівняння системи незалежних рівнянь може розглядатися самостійно. Для знаходження його параметрів використовується метод найменших квадратів. По суті, кожне рівняння цієї системи є рівнянням регресії. Оскільки фактичні значення залежної змінної відрізняються від теоретичних на величину випадкової помилки, то в кожному рівнянні є величина випадкової помилки ε_i.

Якщо залежна змінна y одного рівняння виступає у вигляді фактора x в іншому рівнянні, то дослідник може будувати модель у вигляді *системи рекурсивних рівнянь*:

$$\begin{cases} y_1 = a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = b_{m1}y_1 + b_{m,m-1}y_{m-1} + a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_m. \end{cases} \quad (10.2)$$

У даній системі залежна змінна y включає в кожне наступне рівняння як фактори всі залежні змінні попередніх рівнянь поряд з набором власне факторів x. Кожне рівняння цієї системи може розглядатися самостійно, і його параметри визначаються методом найменших квадратів.

Найбільше поширення в економетричних дослідженнях набула *система взаємозалежних рівнянь*. У ній одні й ті самі залежні змінні в одних рівняннях входять у ліву частину, а в інших рівняннях – у праву частину системи:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \ldots + b_{1m}y_m + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + \varepsilon_1, \\ y_2 = b_{21}y_1 + b_{23}y_3 + \ldots + b_{2m}y_m + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + \ldots + b_{3m}y_m + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n + \varepsilon_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = b_{m1}y_1 + b_{m2}y_2 + \ldots + b_{m,m-1}y_{m-1} + a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n + \varepsilon_n. \end{cases} \quad (10.3)$$

Система взаємозалежних рівнянь отримала назву системи *спільних, одночасних рівнянь*. Тим самим наголошується, що в системі одні й ті самі змінні одночасно розглядаються як залежні в одних рівняннях і як незалежні в інших. В економетриці ця система рівнянь називається також *структурною формою моделі*. На відміну від попередніх систем кожне рівняння системи одночасних рівнянь не може розглядатися самостійно, і для знаходження його параметрів традиційний МНК непридатний. З цією метою використовуються спеціальні прийоми оцінювання.

10.2. Структурна і приведена форми моделі

Система спільних, одночасних рівнянь (або структурна форма моделі) зазвичай містить ендогенні та екзогенні змінні.

Ендогенні змінні – залежні змінні, кількість яких дорівнює кількості рівнянь у системі і які позначаються через y.

Екзогенні змінні – зумовлені змінні, що впливають на ендогенні змінні, але незалежні від них. Позначаються через x.

Класифікація змінних на ендогенні й екзогенні залежить від теоретичної концепції прийнятої моделі. Економічні змінні можуть в одних моделях бути ендогенними, а в інших – екзогенними змінними. Позаекономічні змінні (наприклад, кліматичні умови, соціальний стан, стать, вікова категорія) входять до системи тільки як екзогенні змінні. Як екзогенні змінні можуть розглядатися значення ендогенних змінних за попередній період часу (*лагові змінні*).

Структурна форма моделі дозволяє побачити вплив змін будь-якої екзогенної змінної на значення ендогенної змінної. Доцільно як екзогенні змінні вибирати такі змінні, які можуть бути об'єктом регулювання. Змінюючи їх і керуючи ними, можна заздалегідь мати цільові значення ендогенних змінних.

Структурна форма моделі в правій частині містить при ендогенних змінних коефіцієнти b_{ik} і екзогенних змінних – коефіцієнти a_{ij}, які називаються ***структурними коефіцієнтами моделі***. Усі змінні в моделі виражені у відхиленнях від середнього рівня, тобто під x мається на увазі $x - \bar{x}$, а під y – відповідно, $y - \bar{y}$. Тому вільний член у кожному рівнянні системи (10.3) відсутній.

Використання МНК для оцінювання структурних коефіцієнтів моделі дає, як прийнято вважати в теорії, зміщені і неспроможні оцінки. Тому зазвичай для визначення структурних коефіцієнтів моделі структурна форма моделі перетворюється в *приведену форму* моделі.

Приведена форма моделі становить собою систему лінійних функцій ендогенних змінних від екзогенних:

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + \ldots + \delta_{1n}x_n + \varepsilon_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + \ldots + \delta_{2n}x_n + \varepsilon_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_m = \delta_{m1}x_1 + \delta_{m2}x_2 + \ldots + \delta_{mn}x_n + \varepsilon_m. \end{cases} \tag{10.4}$$

де δ_{ij} – коефіцієнти приведеної форми моделі;

ε_i – залишкова величина для приведеної форми.

За своїм виглядом приведена форма моделі нічим не відрізняється від системи незалежних рівнянь, параметри якої оцінюються традиційним МНК. Застосовуючи МНК, можна оцінити δ_{ij}, а потім оцінити значення ендогенних змінних через екзогенні.

Коефіцієнти приведеної форми моделі становлять собою нелінійні функції коефіцієнтів структурної форми моделі. Розглянемо це положення на прикладі найпростішої структурної моделі, подавши коефіцієнти приведеної форми моделі через коефіцієнти структурної моделі.

Для структурної моделі виду

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + \varepsilon_2 \end{cases} \tag{10.5}$$

приведена форма моделі має вигляд

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + u_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + u_2. \end{cases} \tag{10.6}$$

З першого рівняння (10.5) можна виразити y_2 таким чином (задля спрощення опускаємо випадкову величину):

$$y_2 = \frac{y_1 - a_{11}x_1}{b_{12}}. \tag{10.7}$$

Підставляючи в друге рівняння (10.5), маємо

$$\frac{y_1 - a_{11}x_1}{b_{12}} = b_{21}y_1 + a_{22}x_2. \tag{10.8}$$

Звідси

$$y_1 = \frac{a_{11}}{1-b_{12}b_{21}}x_1 + \frac{a_{22}b_{12}}{1-b_{12}b_{21}}x_2. \tag{10.9}$$

Діючи аналогічно з другим рівнянням системи (10.5), маємо

$$y_2 = \frac{a_{11}b_{21}}{1-b_{12}b_{21}}x_1 + \frac{a_{22}}{1-b_{12}b_{21}}x_2, \tag{10.10}$$

тобто система (10.5) набуває вигляду

$$\begin{cases} y_1 = \dfrac{a_{11}}{1-b_{12}b_{21}}x_1 + \dfrac{a_{22}b_{12}}{1-b_{12}b_{21}}x_2, \\ y_2 = \dfrac{a_{11}b_{21}}{1-b_{12}b_{21}}x_1 + \dfrac{a_{22}}{1-b_{12}b_{21}}x_2. \end{cases} \tag{10.11}$$

Таким чином, можна дійти висновку про те, що коефіцієнти приведеної форми моделі будуть виражені через коефіцієнти структурної форми таким чином:

$$\begin{aligned} \delta_{11} &= \frac{a_{11}}{1-b_{12}b_{21}}, \quad \delta_{12} = \frac{a_{22}b_{12}}{1-b_{12}b_{21}}, \\ \delta_{21} &= \frac{a_{11}b_{21}}{1-b_{12}b_{21}}, \quad \delta_{22} = \frac{a_{22}}{1-b_{12}b_{21}}. \end{aligned} \tag{10.12}$$

Слід зауважити, що приведена форма моделі хоча й дозволяє отримати значення ендогенної змінної через значення екзогенних змінних, але аналітично вона поступається структурній формі моделі, оскільки в ній відсутні оцінки взаємозв'язку між ендогенними змінними.

10.3. Проблеми ідентифікації. Жорстко ідентифікована, неідентифікована і надідентифікована системи рівнянь

При переході від приведеної форми моделі до структурної економетристи стикаються з проблемою ідентифікації. Ідентифікація – це єдиність відповідності між приведеною та структурною формами моделі.

Структурна модель (10.3) у повному вигляді містить $m \cdot (m + n - 1)$ параметрів, а приведена форма моделі в повному вигляді містить $m \cdot n$ параметрів. Тобто в повному вигляді структурна модель містить більшу кількість параметрів, ніж приведена форма моделі. Відповідно $m \cdot (m + n - 1)$ параметрів структурної моделі не можуть бути однозначно визначені з $m \cdot n$ параметрів приведеної форми моделі.

Щоб отримати єдино можливе рішення для структурної моделі, необхідно припустити, що деякі зі структурних коефіцієнтів моделі через слабкий взаємозв'язок ознак з ендогенною змінною з лівої частини системи дорівнюють нулю. Тим самим зменшиться кількість структурних коефіцієнтів моделі. Зменшення кількості структурних коефіцієнтів моделі можливе й іншим шляхом: наприклад, шляхом прирівнювання деяких коефіцієнтів один до одного, тобто шляхом припущень, що їх вплив на формовану ендогенну змінну однаковий. На структурні коефіцієнти можуть накладатися, наприклад, обмеження виду $b_{ik} + a_{ik} = 0$.

З позиції ідентифікації структурні моделі можна поділити на три групи:

1) ідентифіковані;
2) неідентифіковані;
3) надідентифіковані.

Модель *ідентифікована*, якщо всі її структурні коефіцієнти визначаються однозначно, єдиним чином за коефіцієнтами приведеної форми моделі, тобто якщо кількість параметрів структурної моделі дорівнює кількості параметрів приведеної форми моделі. У такому разі структурні коефіцієнти моделі оцінюються через параметри приведеної форми моделі і модель ідентифікується.

Модель *неідентифікована*, якщо кількість приведених коефіцієнтів менша від кількості структурних коефіцієнтів, і в результаті структурні коефіцієнти не можуть бути оцінені через коефіцієнти приведеної форми моделі .

Модель *надідентифікована*, якщо кількість приведених коефіцієнтів більша від кількості структурних коефіцієнтів. У цьому випадку на основі коефіцієнтів приведеної форми можна отримати два або більше значень одного структурного коефіцієнта. У цій моделі кількість структурних коефіцієнтів менша від кількості коефіцієнтів приведеної форми. Надідентифікована модель на відміну від неідентифікованої моделі практично вирішувана, але вимагає для цього спеціальних методів обчислення параметрів.

Структурна модель завжди становить собою систему спільних рівнянь, кожне з яких потрібно перевіряти на ідентифікацію. Модель вважається ідентифікованою, якщо кожне рівняння системи ідентифіковане. Якщо хоча б одне з рівнянь системи неідентифіковане, то і вся модель вважається неідентифікованою. Надідентифікована модель містить хоча б одне надідентифіковане рівняння.

Виконання умови ідентифікації моделі перевіряється для кожного рівняння системи. Щоб рівняння було ідентифікованим, необхідно, щоб кількість зумовлених змінних, відсутніх у даному рівнянні, але наявних у системі, дорівнювала кількості ендогенних змінних у даному рівнянні без одного.

Якщо позначити кількість ендогенних змінних в i-му рівнянні системи через H, а кількість екзогенних (зумовлених) змінних, які містяться в системі, але не входять до даного рівняння, – через D, то умова ідентифікації моделі може бути записана у вигляді такого правила (табл. 10.1).

Таблиця 10.1. Умова ідентифікації моделі

$D + 1 = H$	рівняння ідентифіковане
$D + 1 < H$	рівняння неідентифіковане
$D + 1 > H$	рівняння надідентифіковане

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

Розглянуте рахункове правило відображає необхідну, але недостатню умову ідентифікації. Більш точно умови ідентифікації визначаються, якщо накладати обмеження на коефіцієнти матриць параметрів структурної моделі. Рівняння ідентифіковане, якщо за відсутніми в ньому змінними (ендогенними і екзогенними) можна з коефіцієнтів при них в інших рівняннях системи отримати матрицю, визначник якої не дорівнює нулю, а ранг матриці не менше, ніж кількість ендогенних змінних у системі без одного.

Доцільність перевірки умови ідентифікації моделі через визначник матриці коефіцієнтів, відсутніх у даному рівнянні, але наявних в інших, пояснюється тим, що можлива ситуація, коли для кожного рівняння системи виконане рахункове правило, а визначник матриці названих коефіцієнтів дорівнює нулю. У такому разі дотримується лише необхідна, але недостатня умова ідентифікації.

В економетричних моделях часто поряд з рівняннями, параметри яких повинні бути статистично оцінені, використовуються балансові тотожності змінних, коефіцієнти при яких рівні. У цьому разі, хоча саме тотожність і не вимагає перевірки на ідентифікацію, бо коефіцієнти при змінних у тотожності відомі, у перевірці на ідентифікацію власне структурних рівнянь системи тотожності беруть участь.

10.4. Проблема оцінювання параметрів системи, загальна характеристика методів. Непрямий метод оцінювання параметрів жорстко ідентифікованої системи рівнянь

Коефіцієнти структурної моделі можуть бути оцінені різними способами залежно від виду системи одночасних рівнянь. Найбільшого поширення в літературі набули такі методи оцінювання коефіцієнтів структурної моделі:

1) непрямий метод найменших квадратів;
2) двокроковий метод найменших квадратів;
3) трикроковий метод найменших квадратів;
4) метод максимальної правдоподібності з повною інформацією;
5) метод максимальної правдоподібності при обмеженій інформації.

Розглянемо коротко сутність кожного з цих методів.

Непрямий метод найменших квадратів (НМНК) застосовується в разі точно ідентифікованої структурної моделі. Процедура застосування НМНК передбачає виконання таких етапів роботи:

1. Структурна модель перетворюється на приведену форму моделі.
2. Для кожного рівняння приведеної форми моделі звичайним МНК оцінюються приведені коефіцієнти.
3. Коефіцієнти приведеної форми моделі трансформуються в параметри структурної моделі.

Якщо система надідентифікована, то НМНК не використовується, бо він не дає однозначних оцінок для параметрів структурної моделі. У цьому разі можуть використовуватися різні методи оцінювання, з-поміж яких найбільш поширеним і простим є *двокроковий метод найменших квадратів* (ДМНК).

Основна ідея ДМНК – на основі приведеної форми моделі отримати для надідентифікованого рівняння теоретичні значення ендогенних змінних, що містяться в правій частині рівняння.

Далі, підставивши їх замість фактичних значень, можна застосувати звичайний МНК до структурної формі надідентифікованого рівняння. Метод дістав назву двокрокового МНК, бо двічі використовується МНК: на першому кроці при визначенні приведеної форми моделі і знаходженні на її основі оцінок теоретичних значень ендогенної змінної $\hat{y}_i = \delta_{i1}x_1 + \delta_{i2}x_2 + \ldots + \delta_{in}x_n$ і на другому кроці стосовно структурного надідентифікованого рівняння при визначенні структурних коефіцієнтів моделі за даними теоретичних (розрахункових) значень ендогенних змінних.

Надідентифікована структурна модель може бути двох типів:

1) усі рівняння системи надідентифіковані;
2) система містить поряд із надідентифікованими точно ідентифіковані рівняння.

Якщо всі рівняння системи надідентифіковані, то для оцінки структурних коефіцієнтів кожного рівняння використовується ДМНК. Якщо в системі є точно ідентифіковані рівняння, то структурні коефіцієнти по них знаходяться з системи приведених рівнянь.

Непрямий і двокроковий методи найменших квадратів докладно описані в літературі і розглядаються як традиційні методи оцінки коефіцієнтів структурної моделі. Ці методи досить легко реалізовуються.

Метод максимальної правдоподібності розглядається як найбільш загальний метод оцінювання, результати якого за нормального розподілу ознак збігаються з МНК. Однак при великій кількості рівнянь системи цей метод призводить до досить складних обчислювальних процедур. Тому як модифікацію використовують *метод максимальної правдоподібності при обмеженій інформації* (метод найменших дисперсійних відносин), розроблений в 1949 р. Т. Андерсоном і Н. Рубіні.

На відміну від методу максимальної правдоподібності в даному методі зняті обмеження на параметри, пов’язані з функціонуванням системи в цілому. Це робить розв’язання більш простим, але трудомісткість обчислення залишається досить високою. Незважаючи на його значну популярність, до середини 1960-х років він був практично витіснений двокроковим методом найменших квадратів (ДМНК) у зв’язку з набагато більшою простотою останнього.

Подальшим розвитком ДМНК є *трикроковий МНК* (ТМНК), запропонований в 1962 р. А. Зельнером і Г. Тейлом. Цей метод оцінювання придатний для всіх видів рівнянь структурної моделі. Однак за певних обмежень на параметри більш ефективним виявляється ДМНК.

10.5. Двокроковий метод найменших квадратів оцінювання параметрів надідентифікованих систем одночасних рівнянь (2МНК-оцінка)

Опис двокрокового методу найменших квадратів (2МНК) супроводимо прикладом його використання для моделі рівноваги на ринках товарів і грошей (*IS-LM*) для закритої економіки при фіксованій податковій ставці *t*.

$$\begin{cases} Y = \alpha_0 + \alpha_{1r} + \alpha_2 G + \alpha_3 t + \varepsilon_1 (\alpha_1 < 0), \\ Y = \beta_0 + \beta_1 r + \beta_2 M + \varepsilon_2 (\beta_1 > 0). \end{cases} \tag{10.13}$$

Перше рівняння системи є перевизначеним (щодо змінної *r*). Щоб оцінити його коефіцієнти, рекомендується скористатися двокроковим методом найменших квадратів, суть якого полягає у використанні як інструментальної змінної оцінки перевизначеної змінної, отриманої на базі екзогенних (чи заздалегідь визначених) змінних моделі.

Крок 1. У першому рівнянні цієї системи перевизначеною змінною є процентна ставка *r*. Її можна оцінити, спираючись лише на екзогенні змінні (наприклад, віднімаючи від другого співвідношення перше):

$$r = \lambda_0 + \lambda_1 M + \lambda_2 G + \lambda_3 t + v\ . \tag{10.14}$$

Застосовуючи для (7.14) МНК, отримуємо оцінку $\hat{r}$ змінної *r:*

$$\hat{r} = \hat{\lambda}_0 + \hat{\lambda}_1 M + \hat{\lambda}_2 G + \hat{\lambda}_3 t, \tag{10.15}$$

де $\hat{r}$– умовна середня при фіксованих значеннях *M*, *G*, *t*.

Крок 2. Підставляючи оцінку (10.15) у друге рівняння системи (10.13), маємо

$$Y = \beta_0 + \beta_1 \hat{r} + \beta_2 M + \varepsilon_2\ . \tag{10.16}$$

Ця заміна дає можливість розв'язати таку істотну проблему перевизначених моделей, як корельованість пояснювальної змінної з випадковим членом (нагадаємо, що така корельованість призводить до отримання зміщених і необґрунтованих оцінок). Дійсно, оцінка $\hat{r}$ виражається лише через екзогенні змінні, а отже, не корелює з випадковим відхиленням. Фактично її можна розглядати як нову екзогенну змінну.

Замінивши в моделі (10.13) друге рівняння на (10.16), отримаємо систему, яку можна оцінити за допомогою МНК.

Якщо модель містить більш ніж одну перевизначену змінну, на першому етапі необхідно оцінити всі такі змінні.

2 МНК має певні властивості, що зумовлюють його широке практичне застосування:

1. У даному методі перший етап (етап побудови зведених рівнянь) виконується для частини перевизначених рівнянь, не зачіпаючи інші рівняння моделі. Це дає можливість мінімізувати обсяг обчислень.
2. За наявності перевизначених рівнянь 2 МНК на відміну від МНК визначає єдині оцінки параметрів моделі.
3. Застосовуючи даний метод, достатньо використовувати лише екзогенні й визначені змінні моделі.
4. Застосування 2 МНК ефективне лише в разі, якщо коефіцієнт детермінації R^2 для зведених рівнянь, побудованих на першому етапі, буде досить великий. При цьому інструментальна змінна (у нашому прикладі це змінна $\hat{r}$) незначною мірою корелює з випадковим відхиленням і наближається до істинного значення ($\hat{r}$) заміненої змінної. При невеликому значенні R^2 використання 2 МНК малопродуктивне, тому що в цьому разі інструментальна змінна мало відповідає істинному значенню заміненої змінної.

Зазначимо, що за допомогою методу інструментальних змінних як складової 2 МНК можна отримувати обґрунтовані оцінки й оцінки стандартних відхилень для вибірок великих обсягів. Однак для малих вибірок висновки будуть не настільки конкретними.

10.6. Трикроковий метод найменших квадратів

Розглянуті методи дають можливість оцінювати параметри окремих рівнянь системи. Кожен з них має переваги та недоліки, однак їх об'єднує спільна риса – значний обсяг розрахунків при роботі із системами великої розмірності, тобто із такими, що містять велику кількість рівнянь, а отже, велику кількість змінних і параметрів. Скорочення обсягу розрахунків стає особливо актуальним у процесі вивчення швидкоплинних процесів, а також у тому разі, якщо змінюється пріоритетність окремих незалежних змінних моделі. У цих випадках краще скористатися методом, що одночасно оцінює параметри всіх рівнянь системи, зокрема трикроковим методом найменших квадратів (3 МНК).

Особливістю 3 МНК є те, що при оцінюванні параметрів системи загалом слід зважати на залежності між окремими рівняннями. Ці залежності виявляються в тому, що залишки окремих рівнянь корелюють між собою, тобто загальна матриця коваріацій системи є недіагональною. У такій ситуації найкращим методом оцінювання є узагальнений метод найменших квадратів (метод Ейткена). Однак у цьому разі необхідно знати перше наближення матриці коваріацій. Для рівнянь множинної регресії з автокорельованими залишками цю матрицю отримують на підставі залишків моделі, параметри якої оцінено за звичайним МНК, і вже після обчислення коефіцієнта кореляції коригують загальний оператор оцінювання параметрів рівняння.

Для систем рівнянь, особливо в разі надідентифікованості окремих рівнянь, краще початкове наближення матриці коваріацій визначають за залишками, які отримано в результаті оцінювання параметрів рівнянь за двокроковим МНК. Отже, саме поєднання 2 МНК і методу Ейткена дало назву цьому методу.

Для практичного застосування 3 МНК потрібно виконати такі вимоги:

1) усі тотожності, які входять до системи рівнянь, виключають з розгляду, тому що вони не містять невідомих параметрів і не параметризуються;
2) кожне неідентифіковане рівняння також виключають із системи, оскільки оцінити їх параметри, в принципі, неможливо;
3) точно ідентифіковані та надідентифіковані рівняння поділяють на дві різні групи, і 3 МНК застосовують до кожної з них окремо;
4) якщо група надідентифікованих рівнянь складається лише з одного рівняння, то 3 МНК перетворюється на 2 МНК;
5) кореляція залишків окремих рівнянь системи призводить до того, що загальна матриця коваріацій системи є недіагональною, однак водночас не між усіма рівняннями системи існує залежність, тому матриця коваріації часто буває блочно-діагональною, тоді оцінювання параметрів на основі 3 МНК виконують окремо для кожної групи рівнянь, що відповідають одному блоку.

10.7. Рекурсивні системи одночасних рівнянь, їх характеристика, можливість застосування МНК-оцінки для розрахунку параметрів рекурсивних систем

Одним із випадків успішного застосування МНК для оцінювання структурних коефіцієнтів моделі є *рекурсивні (трикутні) моделі*, у яких ендогенні змінні послідовно (рекурсивно) пов'язані одна з одною. Перша ендогенна змінна Y_1 залежить лише від екзогенних змінних X_i, $i = 1, 2, ..., m$ і випадкового відхилення ε_1. Друга ендогенна змінна Y_2 визначається лише значеннями екзогенних змінних X_i, $i = 1, 2, ..., m$, випадковим відхиленням ε_2, а також ендогенною змінною. Третя ендогенна змінна Y_3 залежить від тих самих змінних, що й Y_2, випадкового відхилення ε_3, а також від попередніх ендогенних змінних (Y_1, Y_2) і т.д.

У цих моделях структурні рівняння оцінюються поетапно $Y_1 \to Y_2 \to Y_3 \to ... \to Y_N$. Застосовуючи МНК для таких моделей, можна отримати незміщені та обґрунтовані оцінки.

Однак моделі даного типу трапляються досить рідко. У загальному випадку для оцінки структурних коефіцієнтів спочатку необхідно перетворити вихідні рівняння до зведеної форми, а потім застосовувати звичайний МНК.

10.8. Прогноз і загальні довірчі інтервали

Точковий прогноз залежних змінних визначається на підставі зведеної (прогнозної) форми економетричної моделі, заданої системою одночасних рівнянь.

Визначення довірчих інтервалів для цього прогнозу залежить від способу, яким було отримано зведену форму моделі.

У загальному випадку довірчі інтервали для кожної ендогенної змінної задаються співвідношенням

$$(\hat{Y}_i - \sqrt{t_{\alpha/2}^2 S_{ii}^2}, \hat{Y}_i + \sqrt{t_{\alpha/2}^2 S_{ii}^2}), \quad i = 1, 2, .., r, \tag{10.17}$$

де $t_{\alpha/2}^2 = F_\alpha$ – табличне значення критерію Фішера з довірчим рівнем α при $(r, n - k - r + 1)$ ступенях свободи;

r — кількість рівнянь системи;
n — загальна кількість спостережень;
k — кількість екзогенних змінних i-го рівняння системи;
S_2 — дисперсія залишків i-го рівняння моделі.

Довірчі інтервали для всіх ендогенних змінних визначають за формулами

$$\left(\begin{array}{l} \hat{Y}_i - \sqrt{\dfrac{\left[1 + X_j^T \left(X^T X\right)^{-1} X_j\right]\left(n-k\right)^r}{n-k-r+1} S_{ii}^2}\,, \\ \hat{Y}_i + \sqrt{\dfrac{\left[1 + X_j^T \left(X^T X\right)^{-1} X_j\right]\left(n-k\right)^r}{n-k-r+1} S_{ii}^2}\,, \end{array} \right) \tag{10.18}$$

де X_j — матриця спостережень k екзогенних змінних, що ввійшли в i-те рівняння системи, розміром $n \times k$;
X — загальна матриця спостережень розміром $n \times k$;
r — загальна кількість рівнянь системи.

Завдання для самоконтролю

1. Дати визначення системи одночасних структурних рівнянь.
2. Назвати основні відмінності структурної і приведеної форми економетричної моделі.
3. Вказати особливості та описати методику розахунку параметрів економетричної моделі на основі двокрокового МНК.
4. У чому полягає сутність непрямого методу оцінювання параметрів жорстко ідентифікованої системи рівнянь?
5. Вказати доцільність застосування трикрокового МНК при оцінці параметрів економетричної моделі.
6. Навести загальну характеристику рекурсивних систем одночасних рівнянь, надати їх математичну формалізацію.

Практичне заняття 10

Системи структурних рівнянь

Постановка задачі

1 етап	Побудова інформаційного масиву дослідження шляхом збору статистичних даних та розробка моделі парної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), аналіз моделі
2 етап	Побудова множинної лінійної регресії, оцінка її параметрів методом найменших квадратів (1 МНК), дослідження моделі
3 етап	Виявлення, дослідження та усунення мультиколінеарності факторів у моделі множинної лінійної регресії методом Фаррара – Глобера
4 етап	Виявлення ефекту гетероскедастичності на основі застосування параметричного тесту Гольдфельда – Квандта та усунення даного ефекту шляхом використання узагальненого методу найменших квадратів (методу Ейткена)
5 етап	Побудова адитивної та мультиплікативної моделі часового ряду на основі проведення декомпозиційного аналізу: визначення структури ряду (тенденції та сезонних (циклічних) коливань), проведення автокореляційного аналізу, визначення якості моделі, розрахунок прогнозних значень
6 етап	Групування елементів економічної системи шляхом кластерного аналізу, побудова деревоподібної дендограми, графічне подання елементів кластеризації
7 етап	Перевірка наявності автокореляції залишків за допомогою критерію Дарбіна – Уотсона, оцінка параметрів рівняння множинної лінійної регресії методом Ейткена
8 етап	Побудова економіко-математичної моделі методом інструментальних змінних; побудова економетричної моделі на основі даних, що містять помилки вимірювання змінних, оцінка параметрів моделі за допомогою алгоритму Уоліса
9 етап	Побудова моделі розподіленого лага, оцінка параметрів моделей з лагами в незалежних змінних, оцінювання параметрів авторегресійних моделей; перетворення ARMA і ARIMA
10 етап	**Структурне моделювання (моделювання на основі системи структурних рівнянь), оцінка невідомих параметрів моделі, визначення адекватності моделі щодо опису структури вхідних даних**

Мета: побудова математичної моделі соціально-економічного розвитку регіону за допомогою системи структурних рівнянь, що є основою складання науково обґрунтованих програм регіонального розвитку.

Завдання: Основне завдання роботи полягає в тому, щоб скласти модель соціально-економічного розвитку регіону за допомогою системи структурних рівнянь, оцінити невідомі параметри моделі і визначити, наскільки адекватно модель описує структуру вхідних даних. На основі отриманих результатів потрібно визначити перспективні напрямки соціально-економічного розвитку Сумської області на найближчий період.

Основні елементи розв'язання

Визначення чинників та формування моделі

Завдання полягає в тому, щоб скласти модель з групи чинників, а саме: загального фінансового результату від звичайної діяльності до оподаткування, загального обсягу зовнішньоторговельного обороту товарами та послугами, середньомісячної заробітної плати, рівня безробіття, валової продукції сільського господарства та обсягу виробництва основних видів сільськогосподарської продукції, далі оцінити невідомі параметри моделі і визначити, наскільки адекватно модель описує коваріаційну структуру даних.

Вхідними даними є статистична звітність (Вхідні дані.sta). Як файл даних розглянемо матричний файл (рис. 10.1). Дані є коваріаційною матрицею, обчисленою по вибірці. Кожний запис у початковій вибірці відповідає одному показнику, що впливає на соціально-економічний розвиток Сумської області. Необхідно побудувати модель, що зв'язує сільське господарство регіону з промисловістю та соціальною сферою, за отриманими даними за 2004–2008 рр.

У файлі даних прийняті такі позначення: RESULTAT – загальний фінансовий результат від звичайної діяльності до оподаткування; OBOROT – загальний обсяг зовнішньоторгівельного обороту товарами та послугами; ZARPLATA – середньомісячна заробітна плата;

Data: Wheaton.sta (6v by 10c)

	Основні показники соціально-економічного розвитку регіону					
	1 RESULTAT	2 OBOROT	3 ZARPLATA	4 BEZROBITTY	5 VALOVA	6 **VUROBNUCTVO**
RESULTAT	112,834	68,947	62,819	46,783	-35,839	-216,899
OBOROT	68,947	93,364	57,091	51,028	-39,889	-186,831
ZARPLATA	62,819	57,091	121,532	70,495	-32,841	-211,748
BEZROBITTY	46,783	51,028	70,495	94,986	-36,625	-189,875
VALOVA	-35,839	-39,889	-32,841	-36,625	93,61	351,522
VUROBNUCTVO	-216,899	-186,831	-211,748	-189,775	351,522	4500,288

Рис. 10.1. Коваріаційна матриця початкових даних

BEZROBITTY – рівень безробіття; VALOVA – валова продукція сільського господарства; VUROBNUCTVO – обсяг виробництва основних видів сільськогосподарської продукції.

Усі приведені змінні, що характеризують різні параметри регіону, – це спостережувані (явні) змінні, оскільки їх значення приведені у файлі даних. Проте в моделі повинні бути ще й латентні змінні: S/G – сектор сільського господарства; PROM – промисловість, SOC – соціальна сфера. Явні змінні RESULTAT, OBOROT, ZARPLATA, BEZROBITTY, VALOVA, VUROBNUCTVO належать до ендогенних. Латентну змінну S/G можна вважати екзогенною, а латентні змінні PROM, SOC – ендогенними.

Спочатку скористаємося графічними можливостями програми Statistica і побудуємо діаграми шляхів. Діаграми шляхів будуємо, виходячи зі свого розуміння модельованої складної системи, і процедура ця є не формалізованого, тобто для однієї і тієї самої системи можуть бути побудовані різні діаграми. Проблема в тому, щоб побудувати (або вибрати з побудованих) діаграму, що найбільш адекватно описує систему.

Процедура побудови діаграм в Statistica є такою. Спочатку клацаємо по кнопці Graphs у верхньому меню. У вікні (рис. 10.2), що з'явилося, обираємо команду **Multiple Graph Layouts** (множинні шари графіка), а потім вибираємо **Blank Graph** (порожній графік).

На екрані з'явиться графічне вікно Graph1: New Graph, у ньому можна створити найскладнішу діаграму, використовуючи горизонта-

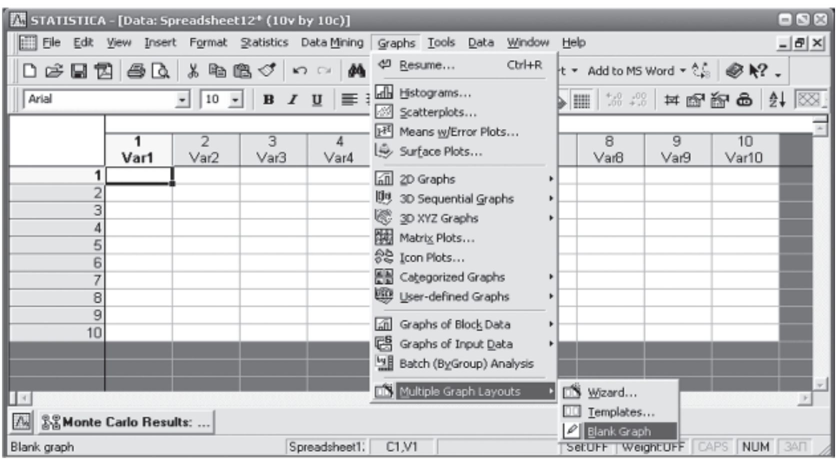


Рис.10.2. Процедура побудови діаграм

льну панель з кнопками, що дозволяють малювати будь-які фігури: прямокутники, стрілки, кола, дуги і т.д. Будувати такі діаграми можна і в текстових редакторах, наприклад, у Word. Побудуємо кілька діаграм (рис. 10.3–10.5).

На рис.10.3 зображена однофакторна модель з одним загальним чинником S/G і двома явними змінними VALOVA і VUROBNUCTVO.

Ця модель відповідає двом регресійним рівнянням:

$$VALOVA = a_1 \cdot S/G + Delta1 \quad (10.19)$$

$$VUROBNUCTVO = a_2 \cdot S/G + Delta2, \quad (10.20)$$

де a_1 та a_2 – невідомі коефіцієнти;
*Delta*1, *Delta*2 – залишкові змінні.

На рис. 10.4 зображена двофакторна модель із загальними чинниками PROM, SOC.

Даній моделі відповідають такі рівняння:

$$RESULTAT = PROM + EPSILON1, OBOROT = a_5 \cdot PROM + EPSILON2 \quad (10.21)$$

$$ZARPLATA = SOC + EPSILON3, BEZROBITTY = a_6 \cdot SOC + EPSILON4 \quad (20.22)$$

На рис. 10.5 приведена регресійна залежність між чинником S/G і чинниками PROM, SOC.

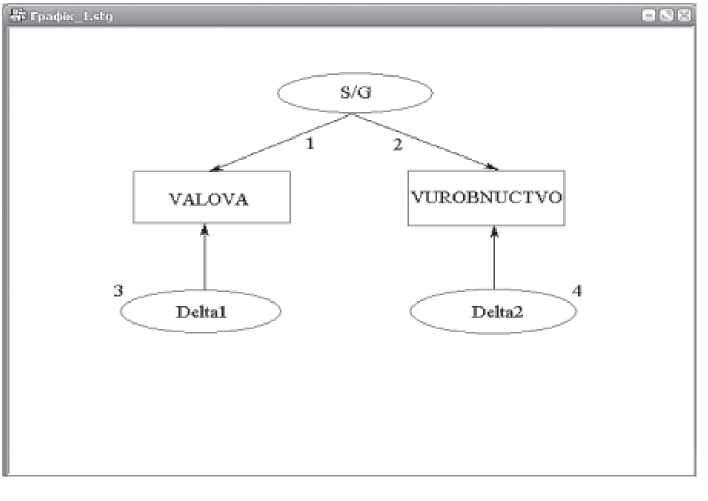


Рис. 10.3. Однофакторна модель

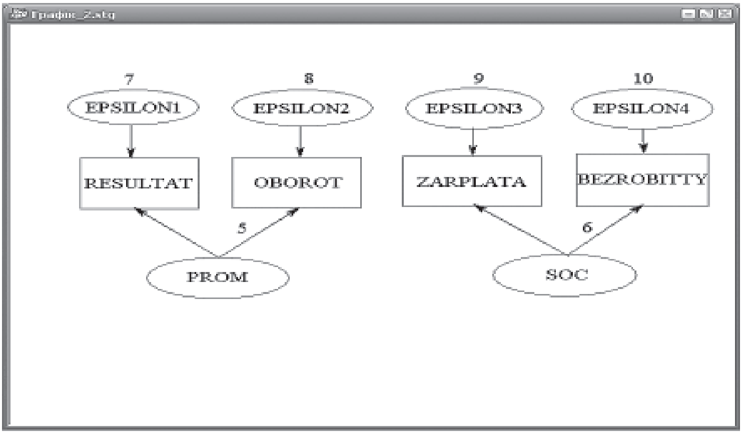


Рис.10.4. Двофакторна модель

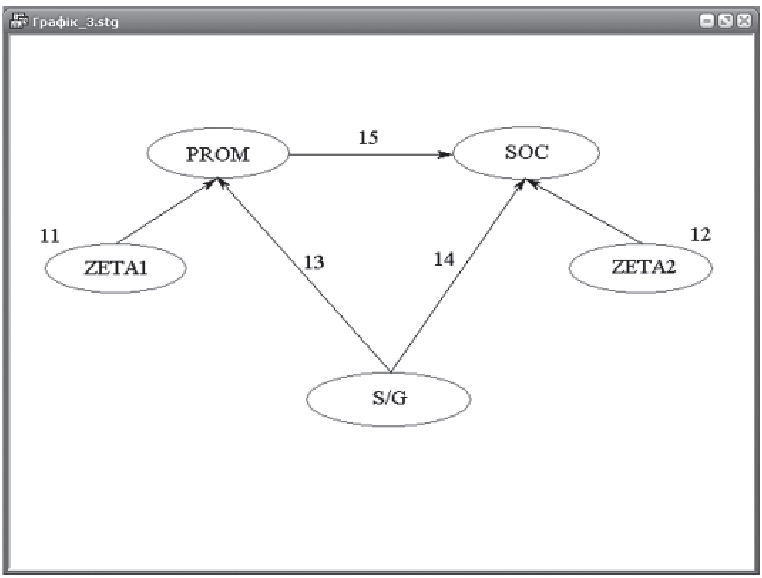


Рис. 10.5. Регресійні залежності

Дана модель описана рівнянням

$$PROM = a_{13} \cdot S / G + ZETA1, \tag{10.23}$$

$$SOC = a_{14} \cdot S / G + a_{15} \cdot PROM + ZETA2, \tag{10.24}$$

де a_{13}, a_{14}, a_{15} – невідомі коефіцієнти,
$ZETA1$, $ZETA2$ – помилки, дисперсії які також є вільними параметрами.

І на рис. 10.6 зображена загальна модель, що охоплює три попередні моделі.

Пояснимо міркування, за якими загальну модель ми подали як сукупність трьох більш простих моделей. При побудові діаграми складної системи бажано в думках поділити її на частини меншого розміру. Як і багато інших моделей структурних рівнянь, наша модель може бути подана у вигляді класичної моделі Lister Karl Joreskog, що складається з трьох менших моделей, дві з яких є загальними чинниками моделей і зазвичай називаються моделями вимірювань (рис. 10.4), і одна модель множинної регресії, що називається структурною моделлю (рис. 10.5). У даному разі нас у першу чергу цікавить регресійна залежність між сільським господарством (S/G), промисловістю

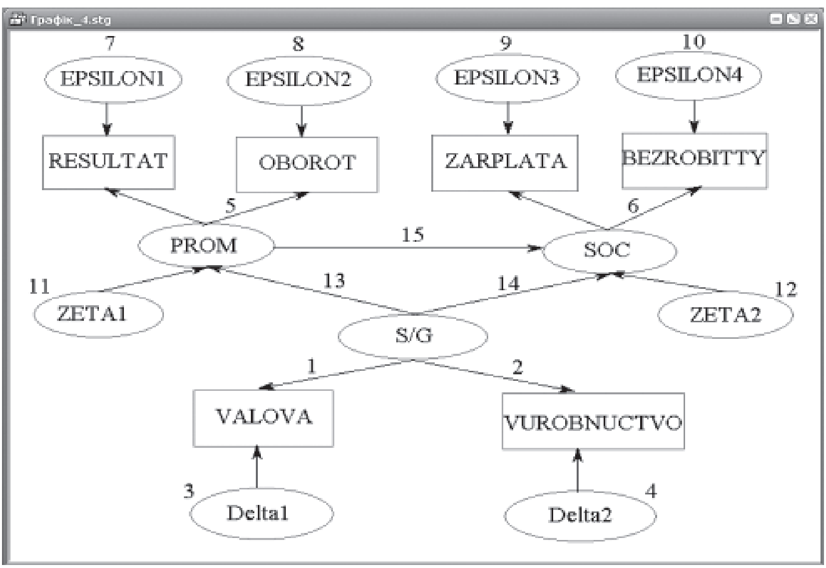


Рис. 10.6. Загальна модель

(PROM) та соціальною сферою (SOC). Як це часто трапляється з даними в економіці, спостережувані змінні, використовувані для оцінювання сільського господарства, промисловості та соціальної сфери, мають різний ступінь надійності. Отже, кореляції між спостережуваними змінними ослаблені можливою невірогідністю даних, тому отримана регресійна залежність між ними може виявитися помилковою. Для того, щоб вирішити цю проблему, у моделях типу Lisrel постулюється регресійна залежність між латентними змінними, які містять помилки вимірювань у вигляді загальних чинників для спостережуваних змінних. Таким чином, маємо дві моделі вимірювань: одна факторна модель для латентних екзогенних змінних – на рис. 10.3, а інша для латентних ендогенних змінних – на рис. 10.4. На рис. 10.6 моделі вимірювань розташовані у верхній і нижній частинах діаграми, а структурна модель знаходиться в центрі.

Розглянемо сенс деяких кількістьвих позначень на діаграмах. Числа 1 і 2, розташовані біля стрілок з S/G до VALOVA і VUROBNUCTVO, представляють номери навантажень (вільні параметри, коефіцієнти a_1 та a_2) чинника S/G на ці змінні, які обчислюються програмою. Дуги, поряд з якими розташовані числа 3 і 4, відповідають дисперсіям змінних залишків Delta1 і Delta2. Ці числа є вільними параметрами, які повинні бути також оцінені програмою. Стрілки з PROM у RESULTAT і з SOC в ZARPLATA не мають поряд розташованих цілих чисел. Це означає, що навантаження (коефіцієнти при PROM, SOC) рівні одиниці. У процесі реалізації процедури Path wizards програма сама визначає, які параметри моделі (коефіцієнти при змінних) є вільними, а які – ні. Тому вид діаграми шляхів завжди повинен бути уточнений після реалізації модуля Sepath.

Програмна реалізація моделі

Після побудови діаграм шляхів перекладемо графічну модель мовою Path1. З урахуванням складності моделі виконаємо це послідовно, крок за кроком, використовуючи діаграми на рис. 10.3–10.6. Після цього клацнемо по кнопці Path wizards (майстер шляхів) (рис. 10.7).

У вікні Sepath, що з'явилося Sepath Wizard – Select Wizard (рис. 10.8), подані дві опції: безпосередньо Structural Modeling (структурне моделювання) і Confirmatory factor analysis (підтверджуючий факторний аналіз).

Моделі підтверджуючого факторного аналізу дуже схожі на звичайні моделі із загальними чинниками, за винятком того, що деякі навантаження чинників і кореляції між чинниками можуть бути

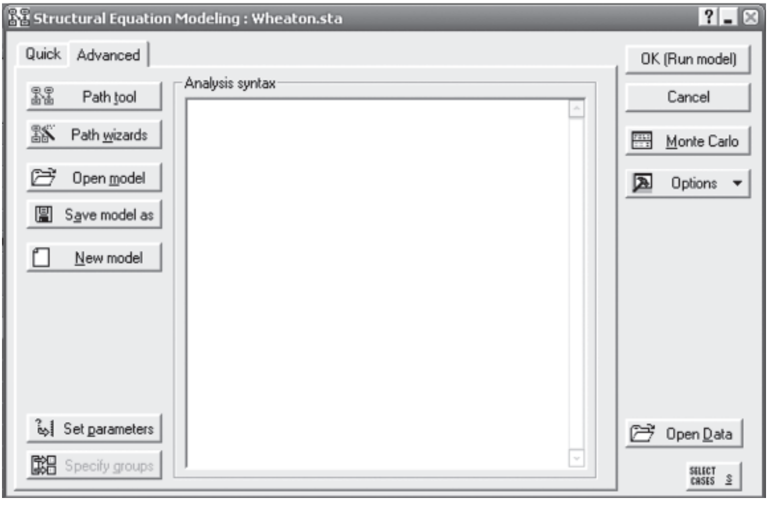


Рис. 10.7. Стартове вікно модуля Sepath

визначені рівними один одному або нулю. Ці умови на структуру чинників легко перевірити за допомогою спеціального запису гіпотез про структуру чинників. Майстер факторного аналізу дозволяє визначити практично будь-яку підтверджуючу факторну модель всього за декілька секунд, якщо ця модель має вісім або менше загальних чинників. Більшість реальних моделей потрапляють до цієї категорії. Визначення

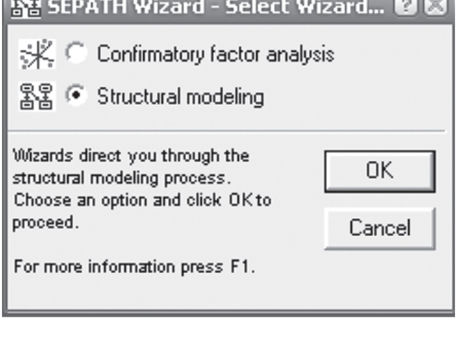


Рис. 10.8. Вибір методу

моделей більшого розміру також не складає особливих труднощів – для цього слід скористатися конструктором шляхів.

Отже, у вікні Sepath Wizard – Select Wizard обираємо опцію Structural Modeling і клацаємо OK. Відкриється вікно Structural Modeling – Exogenous Variables, в якому треба визначити екзогенні змінні. У полі Exogenous variables ми пишемо S/G. Це екзогенний загальний чинник, що визначає рівень сільського господарства регіону. Далі натискаємо кнопку Vars (змінні) і визначаємо явні змінні, пов'язані з екзогенним чинником S/G – VALOVA, VUROBNUCTVO (рис. 10.9). Клацаємо далі OK. Програма повернеться у вікно Structural Modeling – Exogenous Variables. У полі Base name for residual variables (основне ім'я для залишкових змінних) запропонує за умовчанням ім'я для залишкових змінних Delta (тому це ім'я було використане в діаграмі шляхів на рис. 10.6). За бажання можна змінити це ім'я, для цього треба увійти до відповідного рядка і написати інше ім'я.

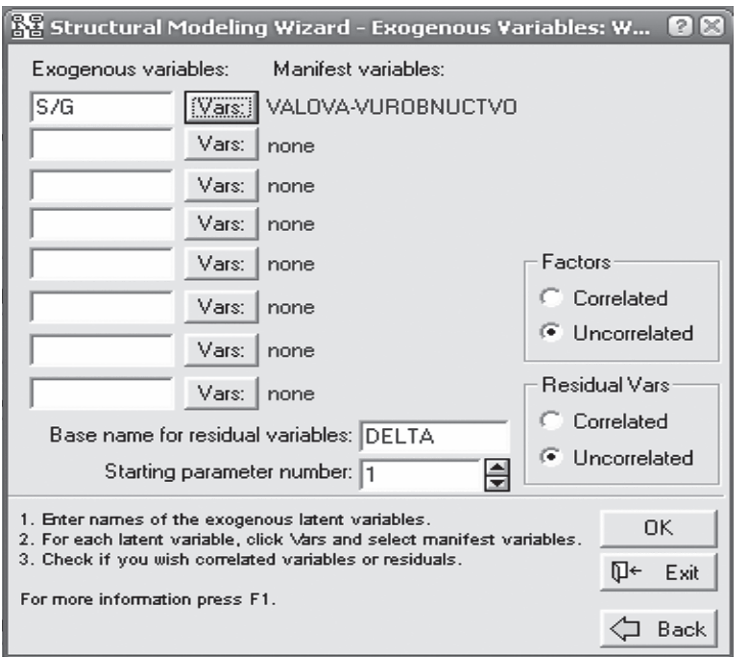


Рис.10.9. Вікно визначення екзогенних змінних

У рамці Residual Vars (залишкові змінні) обираємо опцію Uncorrelated (некорельовані). З діаграми шляхів випливає, що раніше була обрана модель з некорельованими залишковими змінними, оскільки залишкові чинники Delta1, Delta2 некорельовані між собою (між ними немає ніякого шляху). Якщо вибрати Correlated, то розглядатиметься модель з корельованими залишковими змінними. Те саме стосується рамки Factors (чинники) – обираємо опцію Uncorrelated (рис. 10.9).

Далі визначаємо ендогенні змінні, клацнувши кнопку «OK» у вікні Define Exogenous Variables. У вікні, що автоматично з'явилося, Structural Modeling – Endogenous Variables (визначити ендогенні змінні), вибираємо ендогенні змінні відповідно до рис. 10.4, де дана двофакторна модель з двома ендогенними чинниками PROM, SOC. Перш за все записуємо в перше поле Endogenous variables чинник PROM. Далі клацаємо кнопкою Vars і у вікні, що з'явилося, обираємо змінні RESULTAT, OBOROT. Те саме виконуємо і для змінної SOC (рис. 10.10).

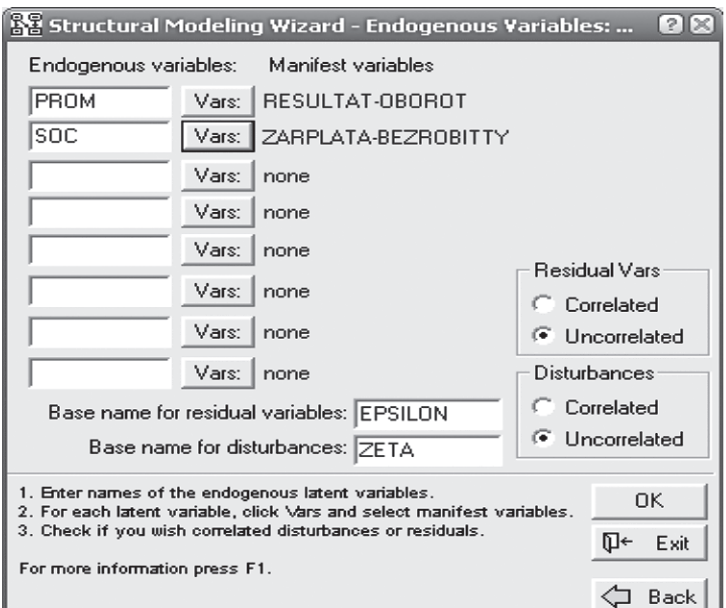


Рис.10.10. Вікно визначення ендогенних змінних

У полі Base name for residual variables автоматично буде прописане ім’я для залишкових змінних – EPSILON (це ім’я також було використане в діаграмі шляхів), у полі Base name for disturbances (основне ім’я для дисперсії) – Zeta. Натискаємо «ОК». З’явиться вікно (рис. 10.11) Define Structural Equation Paths (визначити шляхи структурних рівнянь). Скористаємося діаграмою на рис. 10.5. На цій діаграмі наведена регресійна модель, що зв’язує чинники S/G, PROM, SOC. Перекладемо цю модель мовою Path1.

Спочатку в списку From: (з) висвітлюємо змінну S/G, у списку To (в) виділяємо змінну PROM. Клацаємо кнопку Add (додати). У списку Paths (шляхи) ми побачимо запис (S/G)-13->(PROM).

Таким же чином встановлюємо послідовно зв’язки між чинниками S/G і SOC, між PROM і SOC.

Клацнувши «OK» у вікні, що з’явилося (рис. 10.12) вибираємо опцію Append this model to existing program (приєднати цю модель до існуючої) і знову натискаємо «ОК». Модель визначена, і програма

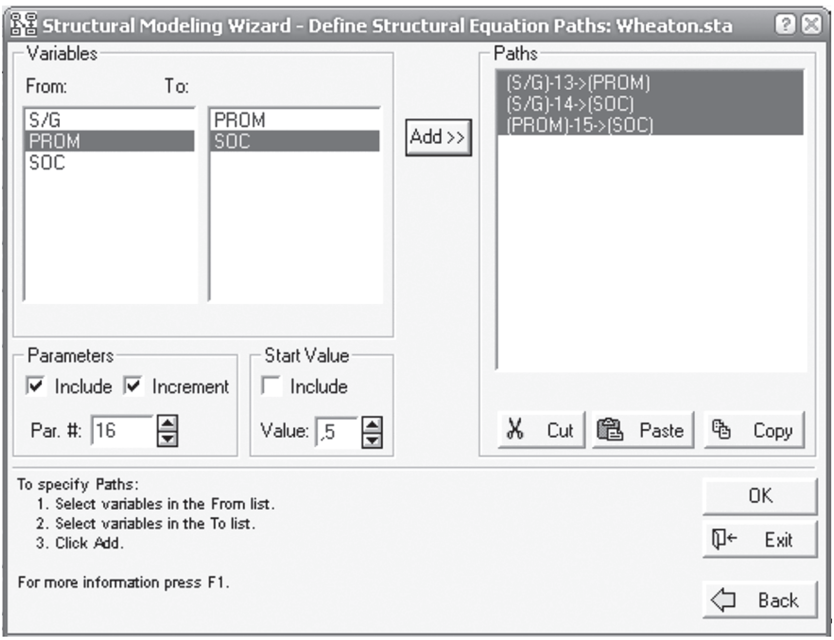


Рис. 10.11. Вікно визначення шляхів структурних рівнянь

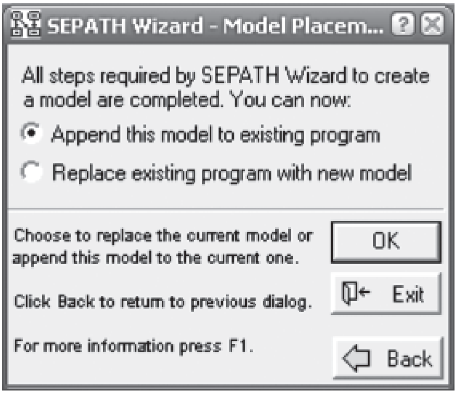


Рис. 10.12. Вибір методу запису моделі

повернулася в стартову панель модуля Structural Equation Modeling (рис. 10.13).

Необхідно звернути увагу на те, що програма самостійно визначила, які параметри моделі (навантаження) вільні, а які – ні. Так, у

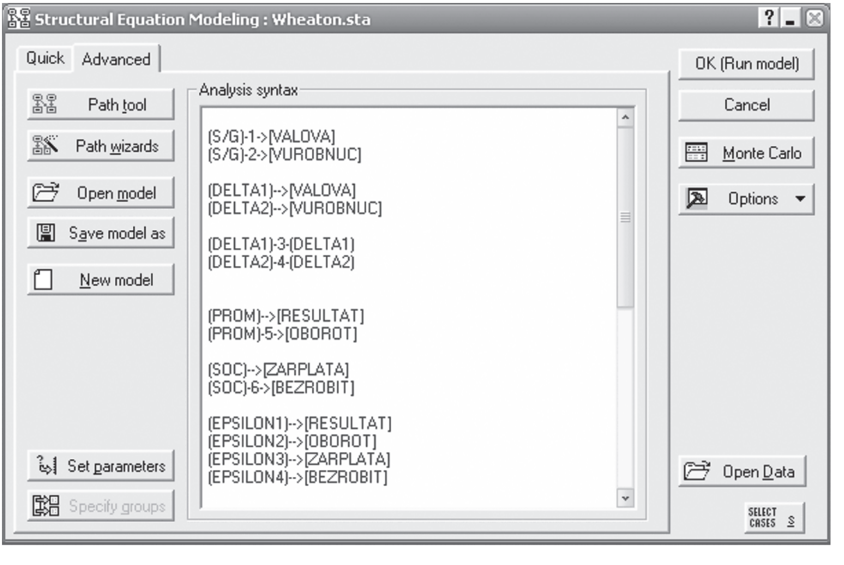


Рис.10.13. Вікно створеної моделі

рядку [PROM]> [RESULTAT] між фактором PROM і змінною RESULTAT відсутній номер вільного параметра, тобто коефіцієнт при PROM дорівнює одиниці, а в рядку [PROM]>[OBOROT] є номер п'ять, що відповідає вільному параметру a_5, що надалі оцінюється програмою.

Для того, щоб змінити, додати нові зв'язки на мові Path1, необхідно натиснути кнопку Path tool (конструктор шляхів), відкриється вікно Path Construction Tool (рис. 10.14).

Таким чином, модель з мови діаграм перекладена мовою Path1. Якщо в стартовій панелі модуля натиснути кнопку Set Parameters (встановити параметри), то на екрані з'явиться вікно Analysis Parameters (параметри аналізу) (рис. 10.15).

Аналіз отриманих результатів

Після того, як модель записана на мові Path1 і параметри аналізу встановлені, можна провести обчислення. Для цього в стартовому вікні

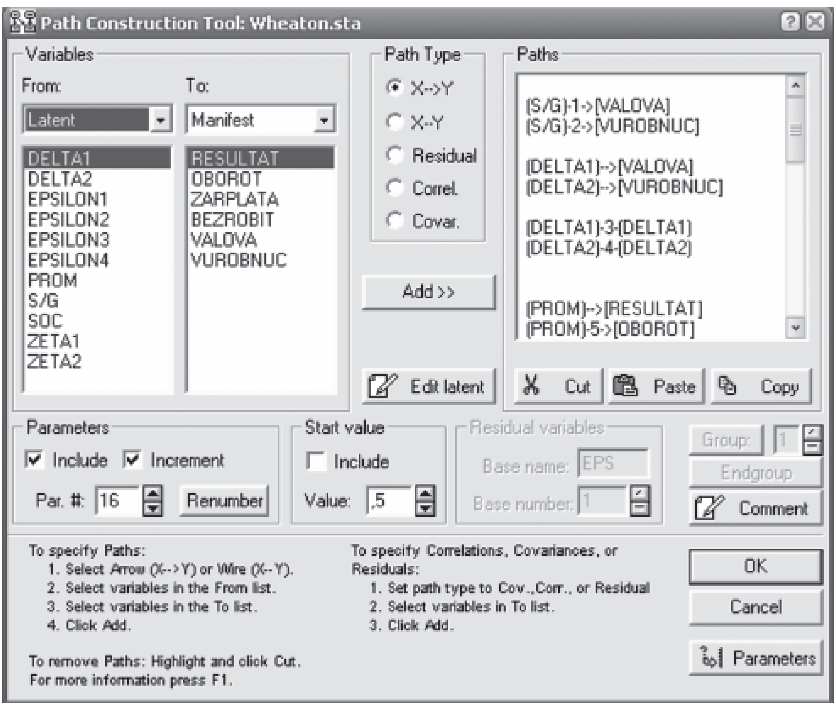


Рис. 10.14. Вікно редагування зв'язків

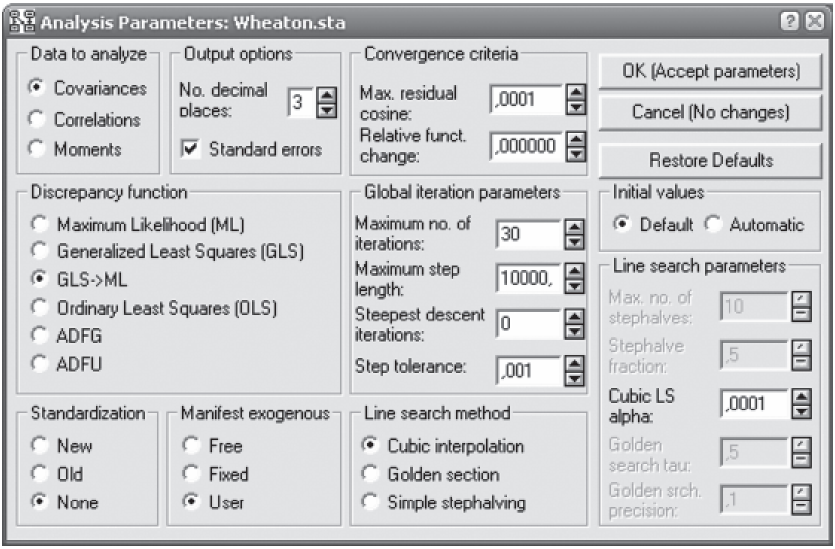


Рис. 10.15. Параметри аналізу

модуля запустимо програму, клацнувши кнопкою «OK», Run model (оцінити модель). Програма почне ітераційну процедуру оцінки невідомих параметрів. На рис. 10.16 наведена таблиця (Iteration Results) результатів ітераційного процесу, що успішно завершився.

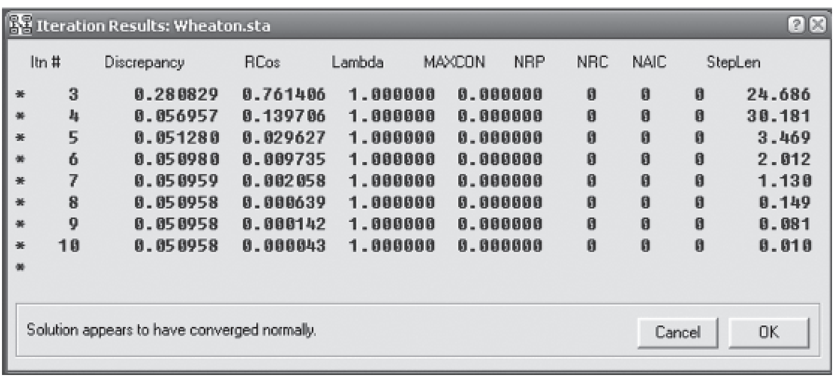


Рис. 10.16. Вікно виведення результатів

У даній таблиці прийняті такі позначення:

Inc # (ітерація) – номер проведеної ітерації.

Discrepancy (незгода) – поточне значення функції незгоди, що мінімізується.

RCos – поточне значення критерію максимуму косинусів залишків.

Lambda (лямбда) – значення множника приросту, використаного на поточному кроці ітерації. Значення 1,0 означає, що перший «повний» крок зменшив значення функції незгоди «достатньо», щоб перейти до наступної ітерації. Значення менш ніж 1,0 означає, що програма використовувала лінійний пошук у вибраному напрямі для вибору конкретної точки, що зменшує значення функції незгоди. Дуже маленькі значення зазвичай говорять про те, що проведення ітерації було практично неможливе.

MAXCON – максимальне значення функцій обмежень. Це значення відмінне від нуля тільки при проведенні обмеженого оцінювання. Обмежене оцінювання використовується, якщо був обраний спосіб New у рамці Standardization або Correlations у рамці Data to analyze. Якщо процес кореляції протікає нормально, то це значення зменшується і наближається до нуля.

NRP – значення NRP показує кількість надмірних параметрів. Програма визначає їх кількість у процесі ітерацій, і, якщо такі параметри є, NRP буде відмінне від нуля.

NRC – кількість надмірних обмежень. Програма визначає їх кількість в процесі ітерацій, і, якщо такі обмеження є, NRS буде відмінний від нуля.

NAIC – кількість активних обмежувальних нерівностей або умов – це те, що використовує програма в процесі ітерації. Програма використовує деякі обмежувальні нерівності, щоб запобігти появі «неможливих» значень параметрів. Наприклад, неприпустимі негативні значення дисперсій. Якщо програма виявляє, що на наступному кроці ітерації дисперсія може прийняти негативне значення, вона встановлює обмежувальну нерівність (ненегативній дисперсії) і після цього мінімізує функцію незгоди, використовуючи решту параметрів. У звичайному факторному аналізі випадком Хейвуда називається ситуація, коли мінімум функції незгоди досягається при негативних значеннях одного або кількох параметрів, відповідних дисперсіям деяких параметрів. Такі значення неприпустимі. При виникненні випадку Хейвуда під час проведення підтверджувального факторного аналізу параметр NAIC буде відмінний від нуля.

StepLen (довжина кроку) – довжина повного кроку при проведенні поточної ітерації. Якщо поряд розташована зірочка, це означає, що був досягнутий максимальний розмір кроку (допустимий розмір кроку можна змінити в рамці Global iteration parameters).

Рядок **Solution appears to have converged normally**, що з'явився внизу вікна (рішення зійшлося нормально) показує, що ітераційний процес зійшовся.

Натискаємо кнопку «ОК». З'явиться діалог аналізу результатів структурного моделювання (рис. 10.17), який складається з двох частин: верхня – інформаційна, де міститься основна інформація про результати оцінювання, і нижня – функціональна, де виділені групи кнопок, що дозволяють всесторонньо проглянути результати, зберегти їх у зручному вигляді, надати графічно і т. д.

У верхній інформаційній частині виводяться такі показники:

Method of estimation (метод оцінювання). Тут виводиться вид використаної функції незгоди.

Discrepancy Function (функція незгоди). Виводиться остаточне значення, що приймається функцією незгоди.

Maximum Residual Cosine (максимум косинуса залишків). Якщо процес ітерації зійшовся успішно, то значення близьке до нуля.

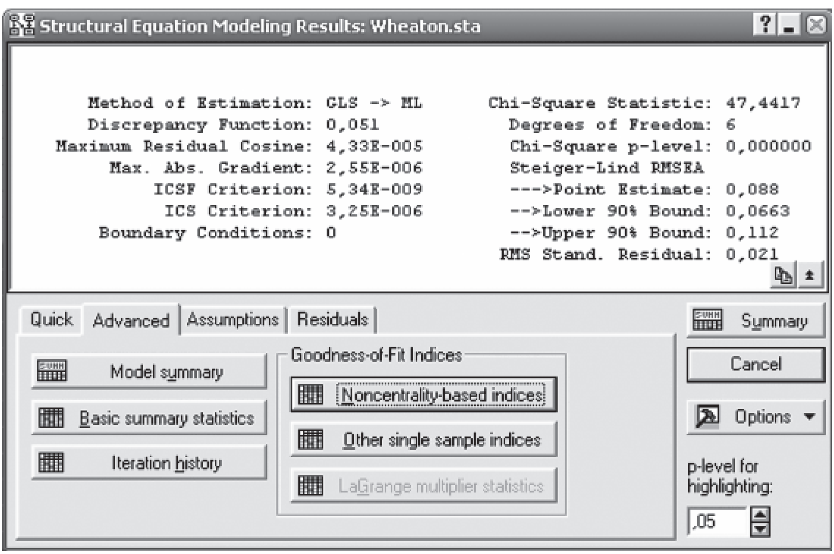


Рис. 10.17. Вікно діалогу аналізу результатів структурного моделювання

Maximum Absolute Gradient (максимальна по модулю компонента градієнта). Це значення дорівнює максимальній по модулю компоненті градієнта.

Структурна модель називається стійкою до множення на постійний множник масштабу (УУПММ), якщо ступінь згоди моделі і даних не змінюється при одночасному множенні всіх змінних на одну й ту саму константу. Більшість, але не всі моделі, що мають місце на практиці, стійкі до множення на постійний множник масштабу. Даний модуль виводить значення двох індексів, що дозволяють судити про ступінь стійкості моделі, що підганяється.

ICSF Criterion – начення цього критерію повинне бути близьким до нуля, якщо структурна модель стійка до множення на постійний множник масштабу.

ICS Criterion – значення даного критерію повинне бути близьким до нуля, якщо дана структурна модель стійка до змін масштабу. При проведенні аналізу кореляції цей індекс повинен бути близьким до нуля.

Boundary Conditions (граничні умови) – виводиться кількість обмежувальних нерівностей, що використовуються при управлінні збіжністю. Воно повинне дорівнювати нулю, якщо тільки в аналізованій моделі не виникла ситуація, подібна випадку Хейвуда. Якщо кількість не дорівнює нулю, статистика X^2 матиме неправильний розподіл, і її використання дає ненадійні результати.

Chi-square Statistic (статистика X^2) – для всіх функцій незгоди, крім LS (МНК), ця статистика, що має асимптотичний розподіл X^2, містить інформацію про істинність нульової гіпотези, тобто про точну відповідність побудованій моделі початковим даним. У моделях з декількома групами ця статистика містить підсумкове значення для всіх груп.

Degrees of Freedom (кількість степенів свободи) – кількість ступенів свободи розрахункової статистики X^2.

Chi-square p-level (p-рівень статистики X^2) – імовірнісний рівень статистики X^2. Нагадаємо, що p – це ймовірність помилкового відхилення нульової гіпотези. Тому якщо p набуває значення, близьке до нуля, то ймовірність помилкового відхилення нульової гіпотези мала і її можна відкинути. Це означає, що чим більше значення p і менше значення критерію X^2, тим достовірніше нульова гіпотеза, а отже, і більш адекватна модель. Якщо $p < 0{,}05$, то відхиляємо нульову гіпотезу при рівні значущості $p = 0{,}05$.

Steiger-Lind RMSEA – тут виводиться Point Estimate (точкова оцінка) і 90% Bound (90% довірчий інтервал) для Steiger-Lind RMSEA.

RMS Stand. Residual (*Root Mean Square Standardized Residual*). Цей індекс також показує якість підгонки моделі. Якщо значення індексу менше ніж 0,05, то підгонка дуже гарна, більше ніж 0,1 – модель неадекватно описує дані.

Розглянемо функціональне призначення кнопок нижньої частини вікна на вкладці Advanced.

Model Summary (підсумкова) модель. Натискаємо цю кнопку, відкриється таблиця з результатами оцінювання. На рис. 10.18 наведений фрагмент таблиці. Рядки таблиці відповідають запису на мові Path1 чергового шляху, у стовпцях наведені оцінки вільного параметра, стандартні помилки, значення t-статистик, *p*-рівні значущості статистик. Значущі *t*-статистики ($p < 0,05$) у програмі STATISTICA виділяються червоним кольором. Якщо *t*-статистика значуща, то правильною є гіпотеза про нерівність нулю оцінки відповідного вільного параметра.

У цій таблиці подані оцінки параметрів регресійної моделі, що зв'язують чинники S/G, PROM, SOC, тобто маючи тільки явні змінні, програма побудувала регресійну модель, що зв'язує приховані загальні чинники. Ні в якому іншому модулі Statistica цього зробити безпосередньо не можна.

Basic Summary Statistics (основні підсумкові статистики). Натиснувши на цю кнопку, з'являться основні статистики, які дані в інфо-

Розрахунки.stw* - Model Estimates (Wheaton.sta)

	Model Estimates (Wheaton.sta)			
	Parameter Estimate	Standard Error	T Statistic	Prob. Level
(EPSILON3)-9-(EPSILON3)	40,290	3,521	11,443	0,000
(EPSILON4)-10-(EPSILON4)	33,817	2,740	12,342	0,000
(ZETA1)-->(PROM)				
(ZETA2)-->(SOC)				
(ZETA1)-11-(ZETA1)	46,766	4,154	11,258	0,000
(ZETA2)-12-(ZETA2)	31,447	3,451	9,113	0,000
(S/G)-13->(PROM)	-4,698	0,360	-13,039	0,000
(S/G)-14->(SOC)	-1,063	0,407	-2,613	0,009
(PROM)-15->(SOC)	0,771	0,054	14,193	0,000

Reproduced Matrix (Wheaton.sta) | Model Estimates (Wheaton.sta)

Рис. 10.18. Фрагмент таблиці обчислених параметрів моделі рмаційній частині вікна. Проте подання їх у таблиці зручне, оскільки далі вони можуть бути збережені, роздруковані, подані графічно і т. д.

Iteration history (звіт про ітерації). У процесі проведення ітерацій програма виводить інформацію про них у вікні Iteration Results. Відкривши цю таблицю, можна знову проглянути отримані результати, а також зберегти їх у файлі даних або провести їх графічний аналіз.

Noncentraility-based indices (індекси нецентральності). Ці індекси показують ступінь адекватності моделі на основі оцінки параметра нецентральності статистики X^2. Клацнувши цією кнопкою, відкриється таблиця (рис. 10.19), у якій послідовно надані: нижня межа 90% довірчого інтервалу, точкова оцінка індексу і верхня межа 90% довірчого інтервалу. У таблиці наведені такі індекси: параметр нецентральності розподілу, Стінгера – Лінда – значення цих індексів менші ніж 0,05 говорять про гарну підгонку моделі, значення цих індексів менші ніж 0,01 говорять про відмінну підгонку моделі; нецентральності МакДональда, гама, скоректований гама-індекс – гарній підгонці відповідають значення цих індексів більші ніж 0,95.

Кнопка **Other Single Sample Indexes** (інші одновибіркові індекси). Якщо натиснути на цю кнопку, з'являються деякі найбільш відомі одновибіркові індекси підгонки, а також деякі пов'язані з ними величини: Джореського, скоректований індекс Джореського, інформаційний критерій Акаїке, байєсівський критерій Шварца.

Для того, щоб перевірити здійснимість умов застосування модуля Sepath, необхідно скористатися вкладкою Assumptions (перевірка припущень, рис. 10.20). Більшість процедур Sepath працює в припущенні нормальності вибірки.

Розрахунки.stw* - Noncentrality Fit Indices (Wheaton.sta)

	Noncentrality Fit Indices (Wheaton.sta)		
	Lower 90% Conf. Bound	Point Estimate	Upper 90% Conf. Bound
Population Noncentrality Parameter	0,026	0,046	0,075
Steiger-Lind RMSEA Index	0,066	0,088	0,112
McDonald Noncentrality Index	0,963	0,977	0,987
Population Gamma Index	0,976	0,985	0,991
Adjusted Population Gamma Index	0,915	0,947	0,970

Iteration History (Wheaton.sta) | Noncentrality Fit Indices (Wheaton.sta)

Рис. 10.19. Таблиця індексів нецентральності

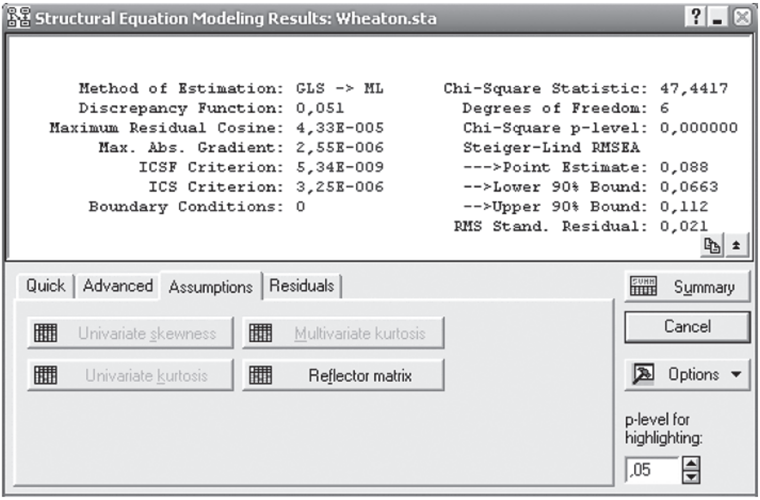


Рис. 10.20. Вікно перевірки припущень

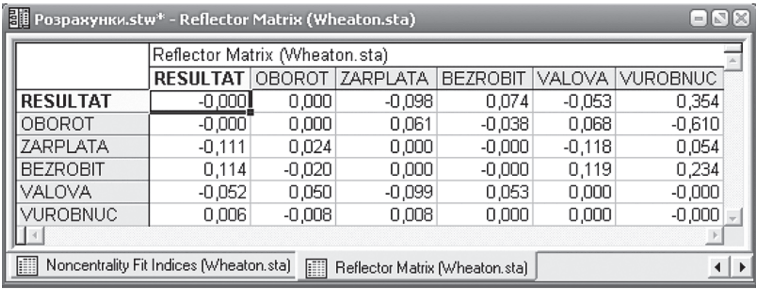
Univariate skewness (одновимірна асиметрія). Натисненням на цю кнопку отримаємо таблицю результатів з різними величинами одновимірної асиметрії.

Univariate kurtosis (одновимірний ексцес). Після натиснення цієї кнопки на екрані відображається таблиця результатів з трьома величинами одновимірного ексцесу: ексцес, уточнений ексцес і нормований ексцес.

Multivariate kurtosis (багатовимірний ексцес). Ця кнопка відкриває таблицю результатів з різними величинами багатовимірного ексцесу, включаючи ті, що забезпечують сумісність з іншими програмами структурного моделювання. Необхідно зазначити, що чим ближче значення асиметрії і ексцесу до нуля, тим більше розподіл відповідає нормальному закону.

Reflector matrix (матриця-рефлектор). Матриця-рефлектор (рис. 10.21) використовується для оцінювання інваріантних властивостей моделі, тобто чи стійка модель до зміни масштабу вимірювання початкових даних. Чим ближчі значення елементів матриці, тим більш стійка модель до зміни масштабу.

Для того, щоб оцінити адекватність моделі, необхідно перейти на вкладку **Residuals** (аналіз залишків, рис. 10.22). Залишки – це різниця між спостережуваними даними і значеннями, що прогнозуються за допомогою моделі. Перегляд залишків дозволяє оцінити якість підгонки.



	Reflector Matrix (Wheaton.sta)					
	RESULTAT	OBOROT	ZARPLATA	BEZROBIT	VALOVA	VUROBNUC
RESULTAT	-0,000	0,000	-0,098	0,074	-0,053	0,354
OBOROT	-0,000	0,000	0,061	-0,038	0,068	-0,610
ZARPLATA	-0,111	0,024	0,000	-0,000	-0,118	0,054
BEZROBIT	0,114	-0,020	0,000	-0,000	0,119	0,234
VALOVA	-0,052	0,050	-0,099	0,053	0,000	-0,000
VUROBNUC	0,006	-0,008	0,008	0,000	0,000	-0,000

Рис. 10.21. Матриця-рефлектор

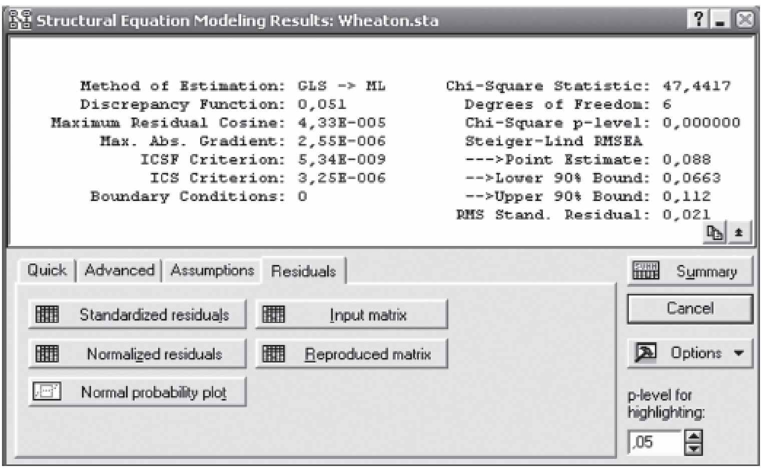


Рис. 10.22. Вікно аналізу залишків

У вікні є такі кнопки для дослідження залишків:

Standardized Residuals (стандартизування залишків) – перегляд стандартизованих залишків;

Normalized Residuals (нормалізовані залишки) – перегляд нормалізованих залишків;

Normal Probability Plot (нормальний імовірнісний графік, рис. 10.23); зауважимо, що однією з ознак адекватності моделі є відповідність закону розподілу залишків нормальному закону; чим щільніше розташовуються точки на прямій, тим більше закон розподілу залишків відповідає нормальному закону.

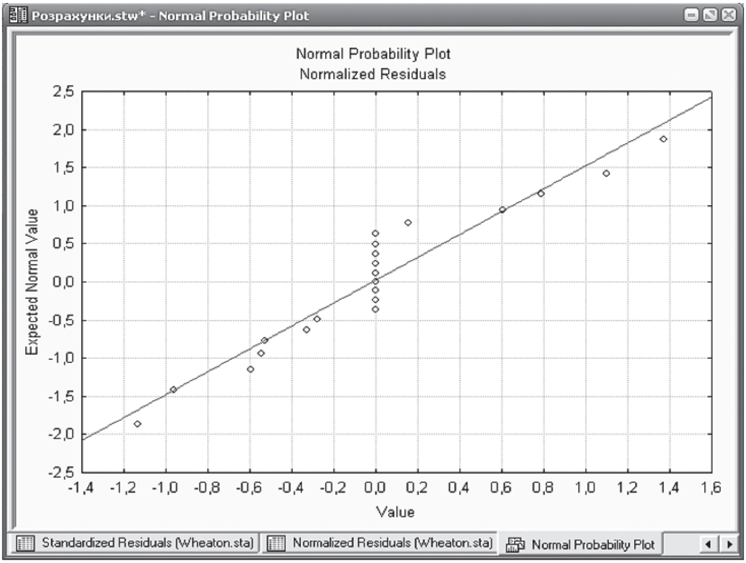


Рис. 10.23. Нормальний імовірнісний графік

Input Matrix (початкова матриця).

Reproduced Matrix (відтворена матриця, рис. 10.24).

Останні дві кнопки дозволяють проглянути і порівняти матриці: початкову (рис. 10.1), яка є початковою для аналізу, і відтворену, яка підрахована на моделі з оціненими параметрами. Чим відмінніші елементи цих матриць, тим менш адекватна модель.

Розрахунки.stw - Reproduced Matrix (Wheaton.sta)

	Reproduced Matrix (Wheaton.sta)					
	RESULTAT	OBOROT	ZARPLATA	BEZROBIT	VALOVA	VUROBNUC
RESULTAT	112,834	68,947	58,094	50,409	-38,021	-204,055
OBOROT	68,947	93,364	58,187	50,489	-38,082	-204,382
ZARPLATA	58,094	58,187	121,532	70,495	-37,931	-203,575
BEZROBIT	50,409	50,489	70,495	94,986	-32,913	-176,645
VALOVA	-38,021	-38,082	-37,931	-32,913	93,610	351,522
VUROBNUC	-204,055	-204,382	-203,575	-176,645	351,522	4500,288

Normal Probability Plot | Reproduced Matrix (Wheaton.sta)

Рис. 10.24. Відтворена матриця

Підсумуємо результати дослідження побудованої моделі. Ітераційний процес (рис. 10.16) зійшовся успішно, значення ICSF Criterion та ICS Criterion близькі до нуля, значення функції незгоди (Discrepancy Functions) мале, значення максимуму косинуса залишків (Maximum Resedual Cosine), максимальна за модулем компонента градієнта (Maximum Absolute Gredient) близькі до нуля. Усі зазначені чинники свідчать про адекватність побудованої моделі.

Використовуючи обчислені параметри моделі (рис. 10.25), запишемо і проаналізуємо співвідношення, що цікавлять нас, між соціально-економічними характеристиками регіону.

Розрахунки.stw - Model Estimates (Wheaton.sta)

	Model Estimates (Wheaton.sta)			
	Parameter Estimate	Standard Error	T Statistic	Prob. Level
(S/G)-1->[VALOVA]	8,093	0,390	20,756	0,000
(S/G)-2->[VUROBNUC]	43,435	2,513	17,282	0,000
(DELTA1)-->[VALOVA]				
(DELTA2)-->[VUROBNUC]				
(DELTA1)-3-(DELTA1)	28,112	4,940	5,691	0,000
(DELTA2)-4-(DELTA2)	2613,690	183,063	14,278	0,000
(PROM)-->[RESULTAT]				
(PROM)-5->[OBOROT]	1,002	0,045	22,371	0,000
(SOC)-->[ZARPLATA]				
(SOC)-6->[BEZROBIT]	0,868	0,041	21,408	0,000
(EPSILON1)-->[RESULTAT]				
(EPSILON2)-->[OBOROT]				
(EPSILON3)-->[ZARPLATA]				
(EPSILON4)-->[BEZROBIT]				
(EPSILON1)-7-(EPSILON1)	43,997	3,119	14,107	0,000
(EPSILON2)-8-(EPSILON2)	24,307	2,622	9,271	0,000
(EPSILON3)-9-(EPSILON3)	40,290	3,521	11,443	0,000
(EPSILON4)-10-(EPSILON4)	33,817	2,740	12,342	0,000
(ZETA1)-->(PROM)				
(ZETA2)-->(SOC)				
(ZETA1)-11-(ZETA1)	46,766	4,154	11,258	0,000
(ZETA2)-12-(ZETA2)	31,447	3,451	9,113	0,000
(S/G)-13->(PROM)	-4,698	0,360	-13,039	0,000
(S/G)-14->(SOC)	-1,063	0,407	-2,613	0,009
(PROM)-15->(SOC)	0,771	0,054	14,193	0,000

Reproduced Matrix (Wheaton.sta) | Model Estimates (Wheaton.sta)

Рис. 10.25. Обчислені параметри моделі

$$VALOVA = 8{,}093 \cdot S/G + 28{,}112 \qquad (10.25)$$

$$VUROBNUCTVO = \ldots S/G + \ldots \qquad (10.26)$$

$$RESULTAT = PROM + \ldots \qquad (10.27)$$

$$OBOROT = \ldots \cdot PROM + \ldots \qquad (10.28)$$

$$ZARPLATA = SOC + \ldots \qquad (10.29)$$

$$BEZROBITTY = \ldots SOC + \ldots \qquad (10.30)$$

З наведених рівнянь випливає, що зі збільшенням рівня розвитку сільського господарства зростають рівень валової продукції сільського господарства та обсяг виробництва основних видів сільськогосподарської продукції; зі збільшенням рівня розвитку промисловості збільшується загальний обсяг зовнішньоторгівельного обороту товарами та послугами й загальний фінансовий результат від звичайної діяльності; а також з підвищенням рівня розвитку соціальної сфери зростає середньомісячна заробітна плата та рівень безробіття.

Лінійні регресійні моделі, що виражають залежність між рівнем розвитку сільського господарства, промисловістю та соціальною сферою, мають такий вигляд:

$$PROM = \ldots \cdot S/G + \ldots \qquad (10.31)$$

$$SOC = \ldots \cdot S/G + \ldots \cdot PROM + \ldots \qquad (10.32)$$

З наведених рівнянь випливає, що чим вищий рівень розвитку сільського господарства, тим нижчий рівень розвитку соціальної сфери і промисловості і чим вищий рівень розвитку промисловості, тим вищий розвиток соціальної сфери.

Тести

1. Система рекурсивних рівнянь передбачає:
   а) якщо залежна змінна y одного рівняння виступає у вигляді фактора x в іншому рівнянні;
   б) якщо незалежна змінна x одного рівняння виступає у вигляді результативної ознаки y в іншому рівнянні.

2. Підставте до кожного визначення в лівій колонці правильну відповідь з правої колонки:

Ідентифікована модель	якщо кількість приведених коефіцієнтів менше числа структурних коефіцієнтів, і в результаті структурні коефіцієнти не можуть бути оцінені через коефіцієнти приведеної форми моделі
Неідентифікована модель	якщо всі структурні її коефіцієнти визначаються однозначно, єдиним чином за коефіцієнтами приведеної форми моделі, тобто якщо кількість параметрів структурної моделі дорівнює числу параметрів приведеної форми моделі
Надідентифікована модель	якщо кількість приведених коефіцієнтів більше числа структурних коефіцієнтів. У цьому випадку на основі коефіцієнтів приведеної форми можна отримати два або більше значень одного структурного коефіцієнта

3. Визначте правильну відповідь відносно умови ідентифікації моделі:

$D + 1 = H$	рівняння ідентифіковане
$D + 1 < H$	рівняння неідентифіковане
$D + 1 > H$	рівняння надідентифіковане

4. Визначте методи оцінювання коефіцієнтів структурної моделі:
   а) непрямий метод найменших квадратів;
   б) двокроковий метод найменших квадратів;
   в) трикроковий метод найменших квадратів;
   г) метод максимальної правдоподібності з повною інформацією;
   д) метод максимальної правдоподібності при обмеженій інформації.

5. Процедура застосування НМНК передбачає виконання таких етапів роботи:
   а) для кожного рівняння приведеної форми моделі звичайним МНК оцінюються приведені коефіцієнти;
   б) структурна модель перетворюється на приведену форму моделі;
   в) коефіцієнти приведеної форми моделі трансформуються в параметри структурної моделі.

6. Для практичного застосування 3 МНК потрібно виконати такі вимоги:
   а) кожне неідентифіковане рівняння також виключають із системи, оскільки оцінити їх параметри, в принципі, неможливо;
   б) усі тотожності, які входять до системи рівнянь, виключають з розгляду, тому що вони не містять невідомих параметрів і не параметризуються;
   в) якщо група надідентифікованих рівнянь складається лише з одного рівняння, то 3 МНК перетворюється на 2 МНК;
   г) точно ідентифіковані та надідентифіковані рівняння поділяють на дві різні групи, і 3 МНК застосовують до кожної з них окремо;
   д) кореляція залишків окремих рівнянь системи призводить до того, що загальна матриця коваріацій системи є недіагональною.

7. Довірчі інтервали для всіх ендогенних змінних визначають за формулами

$$\left(\begin{array}{l}\hat{Y}_i - \sqrt{\dfrac{\left[1 + X_j^T\left(X^T X\right)^{-1} X_j\right]\left(n-k\right)^r}{n-k-r+1} S_{ii}^2}, \\ \hat{Y}_i + \sqrt{\dfrac{\left[1 + X_j^T\left(X^T X\right)^{-1} X_j\right]\left(n-k\right)^r}{n-k-r+1} S_{ii}^2}\end{array}\right),$$

де X_j — матриця спостережень k екзогенних змінних, що ввійшли в i-те рівняння системи, розміром $n \times k$;
X — загальна матриця спостережень розміром $n \times r$;
r — загальна кількість рівнянь системи.

а) так;
б) ні.

2. Система взаємозалежних рівнянь отримала назву:
   а) системи спільних, одночасних рівнянь;
   б) системи незалежних рівнянь;
   в) системи рекурсивних рівнянь.

Питання до іспиту з дисципліни «Економіко-математичні методи та моделі: економетрія»

1. Основні завдання економетрії.
2. Роль економетричних досліджень в економіці.
3. Предмет, цілі, завдання та структура курсу.
4. Місце і значення курсу серед дисциплін фундаментальної підготовки фахівців з економіки.
5. Взаємозв'язки курсу із суміжними дисциплінами.
6. Особливості економетричних моделей.
7. Вибір змінних і структура зв'язків.
8. Роль і місце економетричних моделей у математичному моделюванні.
9. Приклади економетричних моделей: виробнича функція Кобба – Дугласа.
10. Приклади економетричних моделей: моделі пропозиції і попиту на конкурентному ринку.
11. Приклади економетричних моделей: модель Кейнса.
12. Приклади економетричних моделей: модель споживання.
13. Загальна лінійна економетрична модель.
14. Емпірична модель множинної лінійної регресії.
15. Зведення нелінійних економетричних моделей до лінійного вигляду.
16. Метод найменших квадратів.
17. Оператор оцінювання 1МНК.
18. Передумови застосування методу найменших квадратів (1МНК) – умови Гаусса – Маркова.
19. Верифікація моделі.
20. Перевірка значущості та довірчі інтервали.
21. Етапи дослідження загальної лінійної моделі множинної регресії.
22. Прогнозування за лінійною моделлю.
23. Методи побудови багатофакторної регресійної моделі.
24. Поняття мультиколінеарності.
25. Вплив мультиколінеарності на оцінки параметрів моделі.

26. Ознаки мультиколінеарності.
27. Алгоритм Фаррара – Глобера.
28. Методи усунення мультиколінеарності.
29. Метод головних компонент.
30. Моделі з порушенням передумов використання звичайного методу найменших квадратів.
31. Узагальнений метод найменших квадратів.
32. Суть гетероскедастичності.
33. Гетероскедастичність і зважений метод найменших квадратів.
34. Поняття часового ряду.
35. Основні характеристики динаміки часового ряду.
36. Систематичні та випадкові компоненти часового ряду.
37. Перевірка гіпотези про існування тренду.
38. Проблеми аналізу сезонності (та/або циклічності).
39. Фільтрація сезонної компоненти за допомогою індексу сезонності.
40. Метод декомпозиції часового ряду.
41. Методи соціально-економічного прогнозування.
42. Прогнозування тенденцій часового ряду за середніми характеристиками.
43. Прогнозування тенденцій часового ряду за механічними методами.
44. Прогнозування тенденцій часового ряду за аналітичними методами.
45. Вибір форми моделі.
46. Ознаки «гарної» моделі.
47. Види помилок специфікації.
48. Виявлення та коригування помилок специфікації.
49. Дослідження залишкового члена моделі.
50. Проблеми специфікації.
51. Сутність, типологізація та прикладна спрямованість завдань класифікації об'єктів.
52. Класифікація з навчанням.
53. Природа і наслідки автокореляції.
54. Методи визначення автокореляції.
55. Критерій Дарбіна – Уотсона.
56. Критерій фон Неймана.
57. Коефіцієнти автокореляції та їх застосування.
58. Моделі з автокорельованими залишками.
59. Метод оцінювання параметрів Ейткена.
60. Метод Кочрена – Оркатта.
61. Метод перетворення вихідної інформації.
62. Метод Дарбіна.

63. Сутність методу інструментальних змінних.
64. Оператор оцінювання Вальда.
65. Особливості оцінювання методом Бартлета.
66. Оператор оцінювання Дарбіна.
67. Побудова економетричної моделі на основі даних, що містять помилки вимірювання змінних.
68. Оцінка параметрів моделі за допомогою алгоритму Уоліса.
69. Поняття лага і лагових змінних.
70. Моделі розподіленого лага.
71. Взаємна кореляційна функція.
72. Оцінка параметрів моделей з лагами в незалежних змінних: метод послідовного збільшення кількості лагів.
73. Оцінка параметрів моделей з лагами в незалежних змінних: перетворення Койка (метод геометричної прогресії).
74. Оцінювання параметрів авторегресійних моделей.
75. Виявлення автокореляції залишків в авторегресійних моделях.
76. Авторегресійне перетворення.
77. Перетворення методом ковзного середнього.
78. Перетворення ARMA і ARIMA.
79. Системи одночасних структурних рівнянь.
80. Структурна і приведена форма моделі.
81. Проблеми ідентифікації.
82. Жорстко ідентифікована, неідентифікована і надідентифікована системи рівнянь.
83. Проблема оцінювання параметрів системи, загальна характеристика методів.
84. Непрямий метод оцінювання параметрів жорстко ідентифікованої системи рівнянь.
85. Двокроковий метод найменших квадратів оцінювання параметрів надідентифікованих систем одночасних рівнянь (2МНК-оцінка).
86. Трикроковий метод найменших квадратів.
87. Рекурсивні системи одночасних рівнянь, їх характеристика.
88. Можливість застосування МНК-оцінки для розрахунку параметрів рекурсивних систем.
89. Прогноз і загальні довірчі інтервали.

Список рекомендованої літератури

1. Бородич С. А. Эконометрика [Текст]: учебное пособие / С. А. Бородич. – Минск : Новое знание, 2001. – 408 с.
2. Лугінін О. Є. Економетрія [Текст] : навч. посіб. / О. Є. Лугінін. – 2-ге вид., перероб. та доп. – К. : ЦУЛ, 2008. – 278 с.
3. Лук'яненко І. Г. Економетрика [Текст] : підручник / І. Г. Лук'яненко, Л. І. Краснікова. – К. : Знання, 1998. – 494 с.
4. Наконечний С. І. Економетрія [Текст] : навчальний посібник / С. І. Наконечний, Т. О. Терещенко, Т. П. Романюк. – К. : КНЕУ, 1997. – 352 с.
5. Толбатов Ю. А. Економетрика [Текст] : підручник / Ю. А. Толбатов. – К. : Четверта хвиля, 1997. – 320 с.
6. Айвазян С. А. Прикладная статистика и основы эконометрики [Текст] : учебник / С. А. Айвазян, В. С. Мхитарян. – М. : ЮНИТИ, 1998. – 1022 с.
7. Эконометрика [Текст] : учебник / ред. И. И. Елисеева. – М. : Финансы и статистика, 2003. – 344 с.
8. Назаренко О. М. Основи економетрики [Текст] : підручник / О. М. Назаренко. – 2-е вид., випр. – К. : ЦНЛ, 2005. – 392 с.
9. Практикум по эконометрике [Текст] : учебное пособие / ред. И. И. Елисеева. – М. : Финансы и статистика, 2003. – 192 с.
10. Справочник по непараметрической статистике: современный подход [Текст] / пер. с англ. Е. З. Демиденко ; предисл. Ю. Н. Тюрина. – М. : Финансы и статистика, 1982. – 198 с.
11. Anderson T. W. Estimation of dynamic models with error components / T. W. Anderson, C. Hsiao //Journal of the American Statistical Association. – 1981. Vol. 76. – P. 598–606.
12. Andrews D. Tests for parameter stability and structural change with unknown change point / D. Andrews // Econometrica. – 1993. – Vol. 59. – P. 817–858.
13. Bollerslev T. Generalized autoregressive conditional heteroscedasticity / T. Bollerslev // Journal of Econometrics. – 1986. – Vol. 31. – P. 307–327.
14. Box G. E. P. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models / G. E. P. Box and D. A. Pierce // Journal of the American Statistical Association. – 1970. – Vol. 65. – P. 1509–1526.

15. Chen Y. Banking panics: The role of the first-come, first-served rule and information externalities / Y. Chen // Journal of Political Economy. – 1999. – Vol. 107 (5). – P. 946–968.
16. Fama E. F. Efficient capital markets: A review of theory and empirical work / E. F. Fama // Journal of Finance. – 1970. – Vol. 25. – P. 383–417.
17. Filler W. A. Introduction to Statistical Time Series / W. A. Filler. – Wiley, New York, 1976. – P. 366–382.
18. Granger C. W. J. Investigating casual relations by econometric methods and cross-spectral methods / C. W. J. Granger // Econometrica. – 1969. – Vol. 37. – P. 424–438.
19. Granger C. W. J. Spurious regression in econometrics / C. W. J. Granger and P. Newbold // Journal of Econometrics. – 1974. – Vol. 2. – P. 1045–1066.
20. Kaufman H. Debt: the Threat to Financial Stability, in Debt, Financial Stability, and Public Policy / H. Kaufman. – Kansas City, Federal Reserve Bank of Kansas City, 1986.
21. Keuzenkamp H. A. On tests and significance in econometrics / H. A. Keuzenkamp and J. R. Magnus // Journal of Econometrics. – 1995. – Vol. 67. – P. 5–24.
22. MacKie-Mason J. K. Econometric software: A user's view / J. K. MacKie-Mason // Journal of Economic Perspectives. – 1992. – Vol. 6, n. 4. – P. 165–187.
23. Mishkin F. S. Anatomy of a Financial Crisis / F. S. Mishkin. – Cambridge : National bureau of Economic Research. – 1991. – Working Paper No. 3934.
24. Pindyck R. Econometric models and economic forecasts / R. Pindyck, D. Rubinfeld // Mc. Grow- Hill, Inc. USA. – 1991. – 596 p.
25. Timmermans Th. Restructuring of the Belgian Banking Sector and Financial Stability / Th. Timmermans and Ph. Delhez. – BIS – Conference Papers. – 1999. – 7, March. – P. 55–69.

Словник термінів

Автокореляційна функція часового ряду – послідовність коефіцієнтів автокореляції рівнів першого, другого і т.д. порядків.

Автокореляція залишків – одним із припущень класичного регресійного аналізу є припущення про незалежність випадкових величин *ui*, $i = 1, ..., n$, тобто якщо це припущення порушується (незважаючи на те що дисперсія залишків є сталою, наявна гомоскедастичність), то ми маємо справу з явищем, яке називається автокореляцією залишків.

Авторегресійні, або **динамічні, моделі** – моделі, у яких досліджуваний показник у момент часу *t* визначається своїми попередніми значеннями.

Білий шум – некорельовані випадкові відхилення з нульовим математичним очікуванням і постійною дисперсією σ2.

Взаємна кореляційна функція – послідовність коефіцієнтів кореляції, які визначають ступінь зв'язку кожного елемента вектора залежної змінної з елементом вектора незалежної змінної, зсунутими один відносно одного на часовий лаг τ.

Випадкова компонента (залишки, помилки) – складова частина часового ряду, що залишається після вилучення з нього систематичних компонент.

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

Гетероскедастичні моделі – моделі, для яких не виконується умова сталості дисперсії випадкових відхилень $D(\varepsilon_i) = \sigma_\varepsilon^2 \neq const, \;\; i = \overline{1, n}$.

Гомоскедастичні моделі – моделі, для яких виконується умова сталості дисперсії випадкових відхилень.

Динамічна, або **авторегресійна, модель** – модель, у якій досліджуваний показник у момент часу *t* визначається своїми попередніми значеннями.

Динамічний ряд – сукупність спостережень одного показника, впорядкованих залежно від значень іншого показника, що послідовно зростають або спадають.

Дистрибутивно-лагова модель – модель, у якій досліджуваний показник у момент часу *t* визначається не лише поточними, а й попередніми значеннями незалежних змінних.

Дистрибутивно-лаговий мультиплікатор – загальна сума лагових коефіцієнтів для всієї моделі.

Довгостроковий, або **загальний, дистрибутивно-лаговий мультиплікатор** – загальна сума лагових коефіцієнтів для всієї моделі.

Довжина ряду – час, що минув від першого до останнього моменту спостереження. Часто довжиною ряду називають кількість рівнів.

Екзогенні змінні – зумовлені змінні, що впливають на ендогенні змінні, але не залежні від них. Позначаються через x.

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

Економетрія – сукупність різноманітних економічних досліджень, що здійснюються з використанням математичних методів; використання статистичних методів в економічних дослідженнях, а саме: побудова математико-статистичних моделей економічних процесів, оцінка параметрів моделей.

Ендогенні змінні – залежні змінні, кількість яких дорівнює кількості рівнянь у системі і які позначаються через y.

Змінні – економічні величини, що можуть набувати певних значень з деякої множини допустимих величин.

Ідентифікована модель – модель, усі структурні коефіцієнти якої визначаються однозначно, єдиним чином за коефіцієнтами наведеної форми моделі, тобто якщо кількість параметрів структурної моделі дорівнює кількості параметрів наведеної форми моделі.

Індекс сезонності Ij – характеризує ступінь відхилення рівня сезонного часового ряду від ряду середніх u_j (тренду) або, інакше кажучи, ступінь коливань відносно 100%.

Інтервальні часові ряди – рівні часового ряду утворюються шляхом агрегування за певний проміжок (інтервал) часу.

Класифікація – поділ даної сукупності об'єктів або явищ на однорідні, у певному значенні, групи або віднесення кожного із заданої безлічі об'єктів до одного із заздалегідь відомих класів (при цьому «задана множина», що класифікується, може складатися з єдиного об'єкта). Зазначимо, що термін «класифікація» використовується, залежно від контексту, для позначення як самого процесу «поділу-віднесення», так і його результату.

Кластерний аналіз – один із методів багатовимірного аналізу, призначений для групування (кластеризації) сукупності елементів, які характеризуються багатьма чинниками, і отримання однорідних груп (кластерів).

Коефіцієнт детермінації – показує, яка частина варіації залежної змінної описується даним регресійним рівнянням.

Коефіцієнт множинної кореляції – визначає міру зв'язку залежної змінної з усіма незалежними факторами і є коренем квадратним з відповідного коефіцієнта детермінації.

Комбіноване групування – усі ознаки, що характеризують об'єкт, мають дискретний характер або зводяться до таких, а два об'єкти належать до однієї групи тільки при точному збігу зареєстрованих на них градацій одночасно за всіма ознаками, що їх характеризують.

Корелограма – графік залежності значень авто кореляційної функції часового ряду від величини лага (порядку коефіцієнта автокореляції).

Кореляційна залежність – статистична залежність, яка виявляється в тому, що зі змінюванням однієї величини змінюється середнє значення іншої.

Кореляційний зв'язок – статистичний взаємозв'язок двох або кількох випадкових величин. При цьому зміна значень однієї чи декількох із цих величин відповідає систематичній зміні значень іншої/інших величин.

Кореляційно-регресійний аналіз – сукупність методів, за допомогою яких досліджуються та узагальнюються взаємозв'язки кореляційно пов'язаних змінних.

Короткостроковий, або **впливовий, мультиплікатор** – коефіцієнт a_0 при незалежній змінній x_t, що відбиває її вплив на залежну змінну в поточний період.

Лаг – проміжок часу, у який спостерігається запізнювання (затримка), що характерно для багатьох економічних процесів та проявляється в тому, що ефект від впливу деякого фактора на показник, який характеризує процес, виявляється не відразу, а поступово, через певний час або протягом певного часу.

Лінійна регресійна модель – модель, лінійна за своїми параметрами.

Метод інструментальних змінних – сутність цього методу полягає в заміні змінної, що корелює із залишками, інструментальною змінною (ІЗ), яка повинна мати такі властивості: корелювати (бажано значною мірою) із заміненою пояснювальною змінною; не корелювати з випадковим відхиленням.

Модель з кінцевим лагом – модель, у якій незалежні змінні використовують за кілька попередніх періодів.

Моментні часові ряди – часові ряди, утворені показниками, що характеризують економічне явище на певні моменти часу.

Мультиколінеарність – ситуація, коли в багатофакторній регресійній моделі дві або більше незалежні змінні пов'язані між собою лінійною залежністю або, інакше кажучи, мають високий ступінь кореляції.

Надідентифікована модель – модель, у якій кількість наведених коефіцієнтів більша від кількості структурних коефіцієнтів. У цьому разі на основі коефіцієнтів наведеної форми можна отримати два або більше значень одного структурного коефіцієнта.

Неідентифікована модель – модель, у якій кількість наведених коефіцієнтів менша від кількості структурних коефіцієнтів, і в результаті структурні коефіцієнти не можуть бути оцінені через коефіцієнти наведеної форми моделі.

Нескінчена лагова модель – модель, у якій вплив незалежної змінної не обмежується певним періодом.

Параметри – сталі коефіцієнти, що не завжди відомі, але в будь-якій ситуації мають фіксоване значення.

Параметризація рівняння регресії, або **оцінювання параметрів** – визначення значень коефіцієнтів (параметрів) обраної форми статистичного зв'язку змінних на підставі відповідних статистичних даних.

Перехресні дані – 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 належність несуттєва.

Помилки специфікації – неправильний вибір функціональної форми або набору пояснювальних змінних.

Прогноз – науково обґрунтоване судження стосовно можливих станів об'єкта в майбутньому, альтернативні шляхи й терміни їх здійснення.

Прогнозування – процес розроблення прогнозів.

Проміжний інтервал – часткові суми коефіцієнтів $(a_0 + a_1)$, $(a_0 + a_1 + a_2)$, .., що відображають зміну y_t у другий, третій і наступні періоди.

Регресійні рівняння (моделі) – зв'язки між залежною та незалежною (незалежними) змінними, що описуються співвідношеннями $y = f(x) + \varepsilon$; $y = F(x_1, x_2, .., x_m) + \varepsilon$.

Регресійний зв'язок – функція, що описує відношення (залежність) між випадковими змінними величинами.

Сезонні коливання – коливання, які мають періодичний або близький до нього характер упродовж одного року.

Система взаємозалежних рівнянь (система спільних, одночасних рівнянь; структурна форма моделі) – система рівнянь, у якій одні й ті самі залежні змінні в одних рівняннях входять до лівої частини, а в інших рівняннях – до правої частини системи. Тобто, одні й ті самі змінні одночасно розглядаються як залежні в одних рівняннях і як незалежні в інших.

Система незалежних рівнянь – система рівнянь, коли кожна залежна змінна y розглядається як функція одного й того самого набору факторів x.

Система рекурсивних рівнянь – якщо залежна змінна y одного рівняння виступає у вигляді фактора x в іншому рівнянні.

Систематичні компоненти часового ряду – тренд, сезонна і циклічна компоненти.

Специфікація моделі регресії – вибір форми зв'язку змінних.

Статистична залежність – залежність, коли зі змінюванням однієї випадкової величини змінюється закон розподілу ймовірностей іншої.

Статичні моделі – моделі, зв'язки в яких розглядаються у фіксований момент часу, і часові зміни в них ролі не відіграють.

Ступінь гладкості функції – мінімальний ступінь поліному, що найкраще згладжує компоненту U_t.

Тенденція, тренд – невипадкова складова часового ряду, яка змінюється повільно і описується за допомогою певної функції U_t, яку називають функцією тренду.

Узагальнена модель розподіленого лага – економетрична модель, що містить не лише лагові змінні, а й змінні, що безпосередньо впливають на досліджуваний показник (тобто містить й поточні умови функціонування).

Функціональний зв'язок – залежність явищ, за яких зміна одного явища супроводжується зміною іншого.

Циклічні (кон'юнктурні) коливання – коливання схожі на сезонні, але виявляються на більш тривалих інтервалах часу. Циклічні коливання пояснюються дією довготермінових циклів економічної, демографічної або астрофізичної природи.

Часовий ряд – ряд динаміки, упорядкований за часом, або сукупність спостережень економічної величини в різні моменти часу.

Навчальне видання

Козьменко Ольга Володимирівна
Кузьменко Ольга Віталіївна

Економіко-математичні методи та моделі (економетрика)

Навчальний посібник

Головний редактор В.І. Кочубей
Дизайн обкладинки І. Охріменко
Комп'ютерна верстка В.Б. Гайдабрус, А.О. Литвиненко

Підписано до друку 07.07.2014.
Формат 60x84 1/16. Папір офсетний.
Друк цифровий. Ум. друк. арк. 23,6. Обл.-вид. арк. 21,3.
Тираж 300 прим. Замовлення № 46

Відділ реалізації
Тел./факс: (0542) 65-75-85
E-mail: info@book.sumy.ua

ТОВ «ВТД «Університетська книга»
40009, м. Суми, вул. Комсомольська, 27
E-mail: publish@book.sumy.ua
www.book.sumy.ua
Свідоцтво суб'єкта видавничої справи
ДК № 489 від 18.06.2001

Віддруковано на обладнанні ВТД «Університетська книга»
вул. Комсомольська, 27, м. Суми, 40009, Україна
Свідоцтво суб'єкта видавничої справи ДК № 489 від 18.06.2001