МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

Малярець Л. М.

ЕКОНОМІКО-МАТЕМАТИЧНІ МЕТОДИ ТА МОДЕЛІ

Навчальний посібник

Харків. Вид. ХНЕУ, 2014

УДК 330.42(075.8)

ББК 65В6Я73

М21

Рецензенти: докт. екон. наук, професор, завідувач кафедри математики і математичних методів економіки Донецького національного університету *Христіановський В. В.*; докт. екон. наук, проректор з підготовки наукових кадрів Східноєвропейського університету економіки і менеджменту м. Черкаси
Ус Г. О.

Затверджено на засіданні вченої ради Харківського національного економічного університету імені Семена Кузнеця.

Протокол № 9 від 22.04.2014 р.

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

(лист № 1/11-10085 від 02.11.2010 р.)

Малярець Л. М.

М21

Економіко-математичні методи та моделі : навчальний посібник / Л.М. Малярець. – Х. : Вид. ХНЕУ, 2014. – 405 с. (Укр. мов.)

Подано матеріал, який систематизує математичні методи розв’язання оптимізаційних та економетричних задач й економіко-математичні моделі. Наведено реальні економічні задачі, у ході розв’язання яких має місце розвиток економіко-математичних методів.

Рекомендовано для використання студентами у процесі неперервної математичної підготовки, магістрам, аспірантам для проведення наукових досліджень за допомогою економіко-математичних методів і моделей, економістам-практикам для обґрунтування та ухвалення управлінських рішень.

ISBN

УДК 330.42(075.8)

ББК 65ВЯ73

Вступ

Сьогодні характерною особливістю розвитку інструментів дослідження соціально-економічних процесів і явищ на різних рівнях їх управління є інтенсивне поширення математичних методів та моделей. Вони складають фундаментальну основу вирішення реальних аналітичних завдань у всіх сферах діяльності суб'єктів господарювання.

Відомий фахівець з економетрії Е. Маленво так визначив важливість моделей: «Усі науки використовують моделі. Кожна виділяє деякі явища, основні закони яких вона і вивчає. Виходячи з них, вона зменшує складність дійсних ситуацій, спрощує їх, щоб мати можливість застосувати два або три закони, які здаються нам більш важливими для вивчення проблеми. Ізолювання явища і застосування до нього абстрактної формалізації становлять вихідну точку будь-якого наукового методу»[18, с. 49].

Усе це підкреслює важливість і необхідність викладання навчальної дисципліни «Економіко-математичні методи та моделі» для формування компетентностей майбутніх економістів, аналітиків у процесі підготовки бакалаврів і магістрів.

Метою створення даного навчального посібника є допомога студентам, які навчаються на різних напрямах підготовки галузі знань 0305 «Економіка і підприємництво», у вивченні двох нормативних дисциплін: «Оптимізаційні методи та моделі» і «Економетрика», об'єднаних під загальною назвою «Економіко-математичні методи та моделі». У процесі вивчення цих дисциплін студенти набудуть компетентності використання методів оптимізації, а саме методів математичного програмування, економетричних методів для розробки оптимізаційних та економетричних моделей для вирішення реальних економічних завдань на всіх рівнях управління.

Навчальний посібник створено відповідно до чинної міністерської програми навчальної дисципліни «Економіко-математичні методи та моделі» для бакалаврів галузі знань «Економіка і підприємництво».

Відмінною особливістю даної роботи є виклад матеріалу, зміст і обсяг якого відповідає лекціям з нормативних дисциплін «Оптимізаційні методи та моделі» і «Економетрика», при цьому теоретичний матеріал супроводжується великою кількістю демонстраційних прикладів та прикладів з вирішення реальних завдань в економіці за допомогою математичних методів і питаннями для самоперевірки. У деяких темах приклади представлені розв'язання реальних економічних задач з аналізом проблем, які зустрічаються в різних умовах визначеності і невизначеності зовнішнього та внутрішнього середовища діяльності підприємств, фірм.

Навчальний посібник містить узагальнений матеріал із 16 тем навчальної дисципліни «Оптимізаційні методи та моделі» і 16 тем навчальної дисципліни «Економетрика». Основними темами навчального посібника є: концептуальні аспекти математичного моделювання в економіці; задача лінійного програмування і методи її вирішення; симплексний метод розв'язання задач лінійного програмування і деякі його теоретичні аспекти; теорія двоїстості, взаємно двоїсті задачі лінійного програмування; економічна інтерпретація двоїстих невідомих, двоїстий симплекс-метод; аналіз лінійних моделей економічних оптимізаційних задач; транспортна задача; задача дробово-лінійного програмування; цілочислові задачі лінійного програмування; методи нелінійного програмування; квадратичне програмування; теорія ігор, основні методи їх розв'язання та аналізу; динамічне програмування; парна регресія і кореляція в економетричних дослідженнях; перевірка якості рівняння регресії; лінійні моделі множинної регресії; оцінка надійності загальної багатофакторної лінійної моделі; мультиколінеарність, її наслідки та методи усунення; гетероскедастичність та методи її визначення, узагальнений метод найменших квадратів; автокореляція залишків моделі та методи її усунення; проблеми інтерпретації параметрів багатофакторної моделі; узагальнені схеми регресійного аналізу; системи одночасних рівнянь; динамічні економетричні моделі; моделювання одновимірних часових рядів; моделювання тенденції часового ряду.

Таким чином, набуті студентами в результаті вивчення викладеного матеріалу знання, навички, вміння формування оптимізаційних та економетричних задач та моделей в економіці, обґрунтування методів їх розв'язування, аналіз отриманих результатів складають фундамент компетентностей сучасного економіста.

Розділ 1. Оптимізаційні методи та моделі

1. Концептуальні аспекти математичного моделювання в економіці

1.1. Вирішення проблем аналізу даних в економіці на основі економіко-математичного моделювання.

1.2. Загальна технологія визначення величин ознак в економіці.

1.3. Зміст і принципи моделювання.

1.4. Основні типи моделей.

1.5. Етапи економіко-математичного моделювання.

1.1. Вирішення проблем аналізу даних в економіці на основі економіко-математичного моделювання

У сучасних умовах науково-технічного розвитку в усіх сферах діяльності людини стало аксіомою прийняття рішення на основі аналізу даних. Метою аналізу даних є вивчення властивостей об’єктів, явищ та процесів, отримання нових знань про них для більшого підпорядкування.

Сучасний аналіз даних обумовлюється способами отримання величин, методами їх обробки й залежить від розвитку математичних методів і моделювання. Це твердження доведено теорією і практикою.

Наразі методи аналізу даних мають підґрунтям методи математичної статистики, еволюційне моделювання та методи машинного навчання. Сучасний розвиток методів математичної статистики відображається в удосконаленні оцінювання параметрів розподілу величин, перевірці статистичних гіпотез, дисперсійному аналізі, кореляційному аналізі, регресійному аналізі, аналізі часових рядів, багатовимірному статистичному аналізі (БСА). Еволюційне моделювання передбачає використання генетичних алгоритмів, штучних нейронних мереж (ART-мереж, мереж зворотного розповсюдження, мереж зустрічного розповсюдження, мереж Хеммінга, мереж Хопфілда, мереж Кохонена, RBF-мереж). Машинне навчання будується на деревах рішень і використанні ентропійної міри.

Для здійснення автоматичного аналізу даних розроблені сучасні методики, наприклад Data Mining (здобування даних). Методики Data Mining сформовані як синтез методів статистики, теорії інформації, машинного навчання, теорії баз даних. Добування даних визначається як процес аналітичного дослідження великих масивів інформації з метою виявлення закономірностей і систематизації взаємозв'язків між змінними, які потім можна використати для нових даних. Здобування даних як процес включає три основних етапи: дослідження, побудови моделі та її перевірки, що є традиційними для економіко-математичного моделювання.

Проте серед методів аналізу даних пріоритетне місце залишається за методами статистичного аналізу, оскільки вони універсальні, тобто можуть застосовуватися в різних сферах діяльності людини. В економіці пріоритетне місце в аналізі даних займають економетричні методи та методи оптимізації. Але, перш ніж розробляти модель слід отримати величини, тобто їх виміряти. Тут маємо, на перший погляд, протиріччя: величина для моделі, чи модель для отримання величин. Тому й потрібно було уточнити методологічні основи вимірювання величин в економіці. В роботі [19] представлені методологічні основи вимірювання ознак об'єктів в економіці. Підґрунтям для формування цілісних концептуальних основ вимірювання в економіці стали процедури в економіці такі, як: квантифікація, формалізація, моделювання, шкалювання, оцінка, зміст яких цілісно пов'язується зі змістом процесу вимірювання. Конструкція концепції визначення величин ознак зведена на фундаменті триєдиної основи: концепції величин в економіці, умов їх отримання й системи вимірників за допомогою адекватних математичних методів та моделей; схематично це наведено на рис. 1.1 [19]. Названі фундаментальні складові розглядаються як ті, що утворюють базис теорії вимірювання в економіці сьогодні.

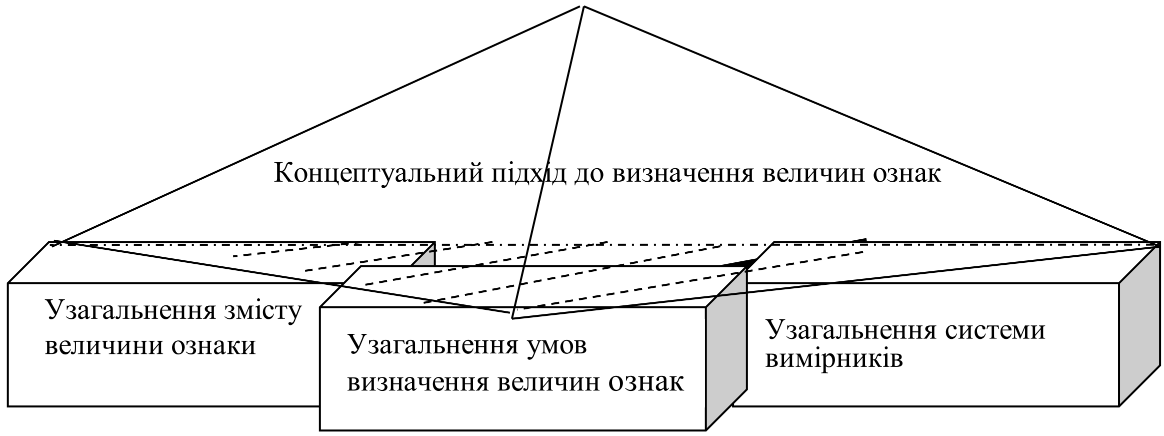


Рис. 1.1. **Базис концепції визначення величин в економіці**

Для реалізації процесу визначення величин потрібні принципи, постулати його здійснення. Сформована система таких принципів, що методологічно забезпечує технологію вимірювання ознак об'єктів в економіці та фундує теорію, наведена в табл. 1.1. Дієвість наведених принципів залежить від дотримання постулатів у вимірюванні в економіці. Наведені постулати та принципи надають об'єктивності та фундаментальності методології моделювання соціально-економічних систем, забезпечують обґрунтованість концепцій та формують змістовність отриманих практичних результатів визначення величин в аналізі даних.

1.2. Загальна технологія визначення величин ознак в економіці

Для визначення величини ознаки потрібно попередньо вивчити дану ознаку. На першому рівні пізнання вивчаються фізичні ознаки об'єкта, оскільки існує можливість емпірично виміряти фізичну величину; на другому рівні пізнання змістовно уточнюється ознака, вимірюється вартісна форма ознаки, аналізується взаємозв'язок її з іншими ознаками об'єкта (фізичними та нефізичними) й вимірюються його складні ознаки; на третьому рівні вимірюється загальна якість ознак об'єкта, яку можна розглядати як загальну якість об'єкта. В табл. 1.1 наведене розширення даного переходу в економіці.

Таблиця 1.1

Конкретизація поняття фізичної величини в економіці

Філософські	Елементи, що	Властивості (ознаки) об'єкта та знання про них

категорії	беруть участь у вимірюванні	Загальні	Форма існування в економіці
Непізнана реальність	Фізичний об'єкт	Апріорні ознаки фізичного об'єкта	Апріорні ознаки фізичного об'єкта в економіці
Пізнана реальність	Об'єкт дослідження	Фізична величина X	Абсолютний показник фізичних ознак об'єкта
Фізичні засоби	Контрольно-вимірювальні прилади	Одиниці фізичної величини	Натуральна форма – кількість – шт.
Знання про реальність	Інформація	Значення фізичної величини $X = \{X\}[X]$	Форма величини фізичної ознаки – вартісний показник. Числове значення показника, одиниці його вимірювання

Існування різних видів величин обумовлює розмежування прямого первинного вимірювання, опосередкованого (непрямого), сумісного та сукупного. Виділені типи вимірювання можна розглядати як процедури технології визначення величин. Первинно вимірюють (за спробами) фізичні й нефізичні величини елементарних ознак об'єкта. Опосередковане (непряме) вимірювання, або його ще називають похідним, маємо, коли вимірюють інтенсивні величини, одержані за допомогою відношення двох екстенсивних та в разі обчислення функції, що залежить від однієї змінної, у ході підстановки екстенсивного аргументу, визначеного в результаті прямого вимірювання.

У цілому ж величини в економіці розділяються на такі групи: основні фізичні величини елементарних ознак, що на даному рівні пізнання можуть квантифікуватися (екстенсивні); похідні фізичні величини, що отримані за допомогою відношення основних фізичних величин. У цій групі можна окремо виділити величини, котрі мають розмірність (у фізичних одиницях), це в основному величини, які отримані як зіставлення результату та витрат, і величини, що не мають розмірності, – це коефіцієнти (інтенсивні); нефізичні величини, а саме метричні величини, що одержані з фізичних за рахунок вартісної форми подання та зіставлення таких величин; утворені, синтезовані метричні величини, що визначають складні ознаки та об'єкт у цілому; статистичні метричні величини, що відображають масові ознаки сукупностей об'єктів в економіці (екстенсивні та інтенсивні); неметричні величини, що відображають якісні ознаки об’єктів в економіці (інтенсивні, якісні); в основному вони вимірюються на неметричних шкалах. У загальному вигляді відповідна система величини наведена на рис. 1.2.

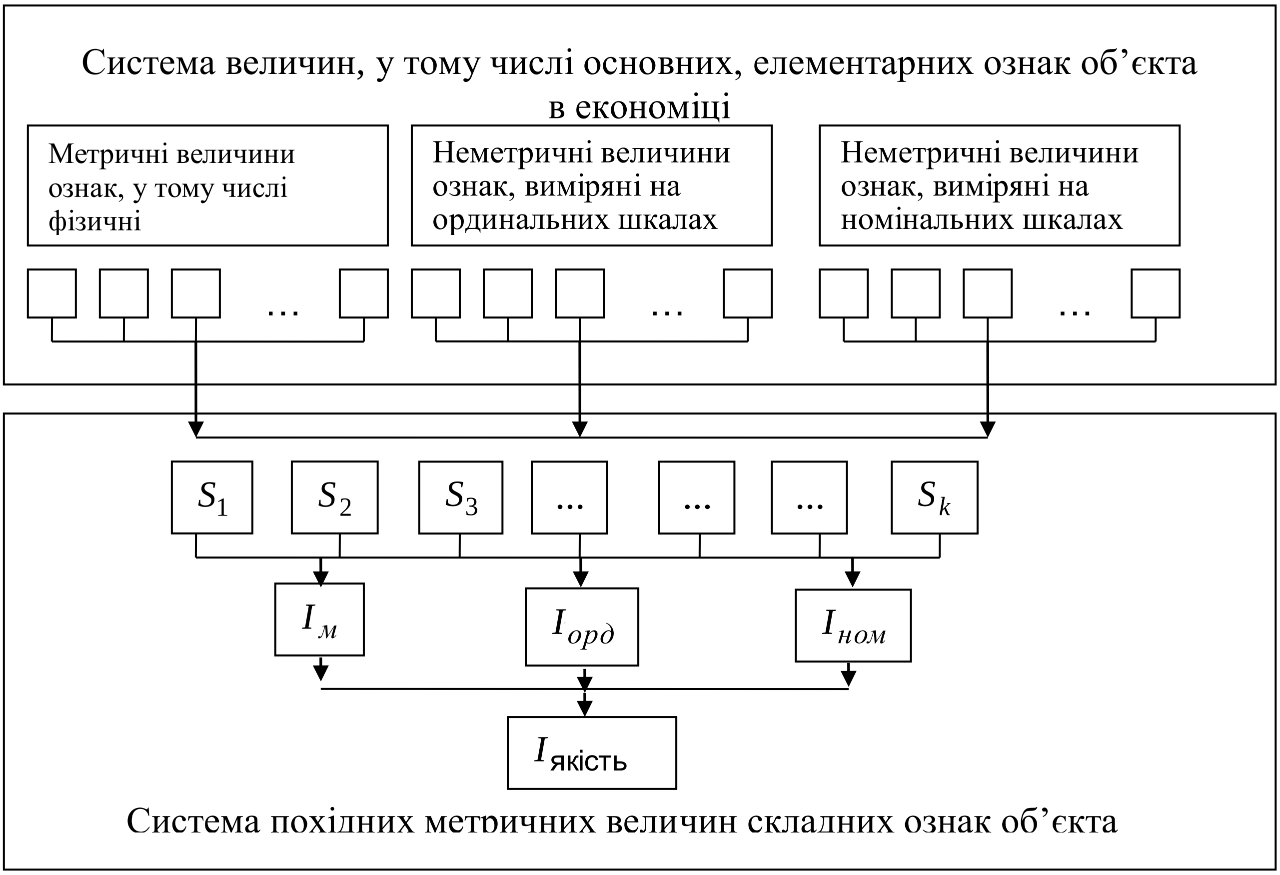


Рис. 1.2. **Система величин в економіці**

Згідно зі змістом величин в економіці доцільно спочатку виділяти етапи визначення величин, що групуються в п’ять процедурних блоків: процедуру постановки; процедуру підготовки; первинне вимірювання; вторинне вимірювання; процедуру контролю за похибками вимірювання (рис. 1.3). Дані процедури різні за змістом та трудомісткістю. Лише третя процедура передбачає безпосереднє вимірювання, оскільки тут відбувається операціоналізація вимірювання. Решта процедур формують умови проведення операціоналізації й одержання результату вимірювання – величини з необхідною точністю. Від правильного виконання даних процедур залежить достовірність визначення величин в аналізі даних, а отже, якість управлінського рішення, що розробляється на його основі.

Схема основних етапів технології визначення величин ознак в економіці	Зміст етапу технології
	1. Аналіз мети визначення величини ознаки об’єкта
	2. Складання когнітивної моделі
	3. Складання змістовної моделі
	4. Розробка концептуальної моделі
	5. Апріорне визначення ознак об’єкта як соціально-економічної системи
	6. Визначення виду ознаки: елементарна чи складна
	7. Визначення виду величини: метрична чи неметрична
	8. Уточнення виду метричної величини
	9. Вибір методу вимірювання, можливих засобів вимірювання
	10. Попередній вибір алгоритму обробки даних
	11. Апріорне оцінювання похибок вимірювання
	12. Вибір системи вимірювання
	13. Визначення типу неметричної величини та шкали вимірювання
	14. Перевірка шкали на досконалість
	15. Реалізація технічної операції вимірювання
	16. Формування метричної величини в показник
	17. Надання неметричній ознаці форми величини
	18. Визначення однорідності величини ознаки
	19. Вибір методу одержання показника
	20. Формування концептуальної моделі складної ознаки
	21. Розробка математичної моделі складної ознаки
	22. Сукупне вимірювання метричних і неметричних величин складної ознаки
	23. Формування величини складної ознаки в узагальнюючий показник
	24. Формування системи вимірників
	25. Аналіз отриманих результатів визначення на предмет методологічних похибок
	26. Аналіз отриманих результатів визначення на предмет методичних похибок
	27. Аналіз отриманих результатів визначення на предмет технічних похибок
	28. Аналіз отриманих результатів визначення на предмет особистісних похибок
	29. Усунення похибок
	30. Звіт про результати визначення величини з показником похибок та перехід до нового, глибшого

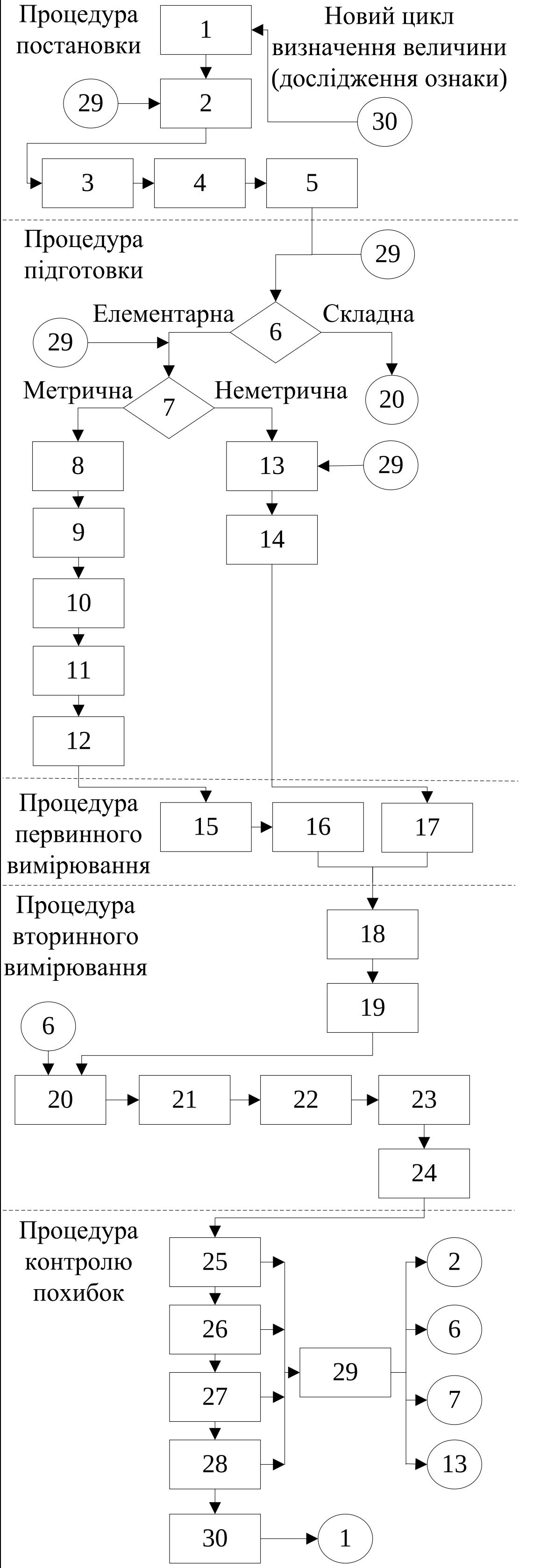
	циклу пізнання

Рис. 1.3. **Схема основних етапів технології визначення величин ознак в економіці**

Запропонована схема етапів визначення величини ознаки організовує даний процес та є єдино можливою. Залежно від мети визначення величини окремі етапи загальної технології можуть бути відсутніми, але логіка в послідовності дій у скороченій схемі зберігається. Комплекс математичних методів може бути різний залежно від визначення конкретного типу ознаки. Таким чином, економіко-математичні методи і моделі створюють умови для визначення величин ознак СЕС, з іншого боку, дотримання технології визначення величин є підґрунтям адекватності побудованих моделей. Маємо суперечність, яка вирішується за допомогою взаємного дотримання означених вимог у реалізації технології і моделювання. Рекомендована структура розширених умов одержання результатів визначення ознак соціально-економічних систем подана в табл. 1.2.

Таблиця 1.2

Складові визначення ознак соціально-економічних систем

Тип визначення	Об’єкт визначення	Спосіб знаходження величини	Форма результату визначення
Первинне вимірювання:			
пряме	Фізичні елементарні ознаки об’єкта, нефізичні величини	Операціоналізація	Фізичні величини, абсолютні натуральні показники. Нефізичні величини, вартісні показники
похідне; опосередковане	Метричні елементарні ознаки; неметричні елементарні ознаки	Безпосереднє обчислення на основі використання законів в економіці; вимірювання в неметричних шкалах	Відносні показники, номінації, порядкові величини, назви
Вторинне вимірювання:			
опосередковане	Складні ознаки об’єкта	Моделі з однією змінною; моделі з багатьма змінними; математичні методи	Показники, вимірники
сукупно-сумісне	Складні ознаки	Моделі з багатьма змінними; математичні методи	Показники, вимірники

Нові величини ознак – вимірники – отримуються за допомогою моделювання й мають принципове значення для економіки; важлива їх роль в самій методології математичного моделювання.

1.3. Зміст і принципи моделювання

Сучасним основним методом дослідження будь-яких систем є метод моделювання, а саме спосіб теоретичного аналізу і практичних дій, спрямований на розробку і використання моделей. Сьогодні багатьма вченими вважається, що розвиток будь-якої науки можна трактувати в загальному сенсі як теоретичне моделювання. Особливе значення набувають моделі у вивченні закономірностей масових процесів, які недоступні безпосередньому спостереженню і не підлягають експериментуванню. Це відноситься, перш за все, до процесів та явищ в економіці, які формуються і розвиваються під впливом нескінченної кількості взаємопов'язаних факторів і за своєю складністю переважають над технічними, хімічними та ін.

У загальному розумінні модель використовується як умовний образ, сконструйований для спрощення реального об'єкта в процесі його вивчення. Під час розробки моделей дотримуються основних методологічних принципів: адекватності, динамізму й евристичності. Принцип адекватності висловлює матеріальна вимога до процесу моделювання: необхідність об'єктивної відповідності моделі оригіналу як умова об'єктивної істинності знання. На відміну від принципу адекватності, принцип динамізму вказує на мінливість всіх основних елементів процесу моделювання: у процесі змінт цілей дослідження, зміни об'єкта, модель повинна пристосовуватися до нових умов і завдань. Принцип заміщення стверджує посередницьку функцію моделі в дослідженні. Особливо важливо це тоді, коли можна провести модельний експеримент. Принцип евристичності націлює на розширене відтворення знань. Цими методологічними принципами керуються під час розробки моделей різних типів.

Математичне моделювання є ідеальним науковим знаковим формалізованим моделюванням, за якого опис об'єкта здійснюється мовою математики, а дослідження моделі виконується за допомогою математичних методів.

1.4. Основні типи моделей

На разі немає загальноприйнятої класифікації моделей. Розрізняють: матеріальні; знакові моделі, а саме графічні та математичні; матеріально-ідеальні. Існує багато визначень математичної моделі. Згідно з математичною енциклопедією «модель – інтерпретація формальної мови» [37, т. 3, с. 770]. Й далі: «математична модель – наближений опис якого-небудь класу явищ зовнішнього світу, виражений за допомогою математичної символіки» [37, т. 3, с. 574].

Також під математичною моделлю прийнято розуміти сукупність співвідношень – рівнянь, нерівностей, логічних умов, операторів і т. д., що визначають характеристики стану об'єкта моделювання, параметри функціонування і розвитку його. Ще математична модель визначається як гомоморфне відображення у вигляді впорядкованої сукупності рівнянь, нерівностей, логічних відносин, графіків, умовний образ об'єкта, складений для спрощення його дослідження.

В економіці найчастіше використовується економічно-математична модель. Вона виражає економічну систему за допомогою формально-математичних термінів, логічна структура якої визначається як об'єктивними властивостями, так і суб'єктивним цільовим чинником дослідження, для якого цей вираз здійснюється. Економіко-математичне моделювання має зворотний вплив на дослідника, вимагаючи від нього чіткості формулювання дослідницьких завдань, суворої логічності у побудові гіпотез і концепцій.

Економіко-математичні моделі класифікуються:

за способом відображення дійсності: аналогові, концептуальні, структурні, інформаційні, функціональні;

за ознакою цільового призначення: теоретико-аналітичні, прикладні;

щодо практичного призначення: балансові, дескриптивні, імітаційні, рівноваги, нерівноваги, прескриптивні, оптимізаційні;

за способом логіко-математичного опису економічних систем: аналітичні, імовірнісні, детерміновані, лінійні, нелінійні, математико-статистичні, матричні, економетричні;

за тимчасовими і просторовими ознаками: динамічні, статичні, точкові, трендові;

за внутрішньою структурою модельного опису системи: автономні, глобальні, закриті, відкриті, комплекс моделей, макроекономічні, мікроекономічні, багатосекторні, одно-, багатопродуктові;

за областю використання: за типом економічних завдань, за видом математичного методу, застосованого під час розробки моделей.

На мікрорівні, наприклад, на рівні підприємств, у вирішенні реальних економічних завдань найчастіше в економіко-математичному моделюванні використовуються економетричні, оптимізаційні та балансові моделі. Математико-статистичні моделі передбачають застосування інструментів математичної статистики. Оптимізаційні моделі охоплюють деяку кількість варіантів виробництва, розподілу і споживання і призначена для вибору таких значень змінних, що характеризують ці варіанти, щоб був знайдений кращий з них [35, с. 356]. Відмінною рисою цих моделей є те, що вони містять як рівняння, що описують взаємозв’язки між змінними, так і критерії для вибору – функціонал або цільову функцію. Дані моделі відносяться до класу екстремальних завдань і описують умови функціонування економічної системи.

Опис кількісних взаємозв'язків між ознаками здійснюють економетричні моделі. Згідно з економіко-математичним енциклопедичним словником економетрика (економетрія) − наукова дисципліна, що дозволяє на основі положень економічної теорії та результатів економічного вимірювання надавати конкретні кількісні вирази загальним (якісним) закономірностям, обумовленим економічною теорією. Економетричні моделі визначаються як економіко-математичні моделі, параметри яких оцінюються за допомогою методів математичної статистики. Економетрична модель як засіб аналізу і прогнозування конкретних економічних процесів як на макро-, так і на мікроекономічному рівні на основі реальної статистичної інформації.

Типовими економетричними моделями є виробничі функції, що виражають взаємозв'язок між витратами та результатами діяльності економічних систем; моделі функціонування національної економіки; типологізація об'єктів і поведінки агентів (країн, регіонів, фірм, споживачів); цільові функції споживчої переваги і функції попиту; моделі розподільних відносин у суспільстві; моделі ринку й економічної рівноваги; моделі інтернаціоналізації національних економік; моделі міжрегіонального аналізу та між державами та ін. Спираючись на класифікацію завдань, що вирішуються за допомогою економетрики, а саме виділення за кінцевими прикладними цілями, за рівнем ієрархії і за профілем аналізованої економічної системи, розрізняють прогнозні економетричні моделі, імітаційні, моделі макро-, мезо- та мікрорівня, моделі, в яких зосереджено увагу на окремих проблемах економічних систем.

Балансові моделі представляють систему балансів виробництва і розподілу продукції і записуються у формі квадратних матриць. Балансові моделі служать для встановлення пропорцій і залежностей у процесі планування різних галузей економіки.

У цілому економіко-математичні моделі розробляються для досягнення трьох цілей; кращого розуміння об'єктивної реальності; розробки раціонального варіанту дій, а також для вибору оптимальних рішень в практичній діяльності.

1.5. Етапи економіко-математичного моделювання

Якщо в моделюванні керуватися правильною методикою, то розроблені моделі будуть адекватні реальному об'єкту, допустимі з точки зору обчислювальних процедур, а сам процес моделювання – ефективним і обгрунтованим. У такому випадку моделювання досягає мети.

Процес математичного моделювання узагальнено представляється чотирма етапами. За першим етапом формулюються закони, що зв’язують основні об’єкти моделювання. На другому етапі відбувається дослідження математичних задач, до яких приводить математична модель. Третій етап полягає у з’ясуванні, чи задовольняє прийнята (гіпотетична) модель критерію практики, тобто вияснення питання про те, чи узгоджуються результати спостережень з теоретичними наслідками моделі в межах точності спостережень. На четвертому етапі виконується наступний аналіз моделі у зв’язку з накопиченням даних про явища, що вивчаються, та модернізація моделі [37, т. 3, с. 574].

Економіко-математичне моделювання можна розглядати як послідовну логіку взаємопов'язаних етапів: визначення мети моделювання, аналіз сформульованої проблеми і розробка концептуальної моделі; побудова математичної моделі та її аналіз; підготовка висхідних даних; чисельне рішення моделі; аналіз числових результатів на несуперечність з концептуальною моделлю та за необхідністю удосконалення моделі; використання моделі для обгрунтування управлінського рішення [23; 32].

Мету моделювання слід формулювати виходячи з сутності проблеми дослідження. Тому спочатку формуються гіпотези, що пояснюють поведінку і розвиток модельованого об'єкта, а також прогнозують цілі, які будуть змінювати в часі.

На етапі аналізу сформульованої проблеми і розробки концептуальної моделі передбачається виділення меж економічної системи, її структуризація. Під структурою економічної системи розуміють її статичне уявлення в розрізі матеріальних і нематеріальних елементів, які забезпечують її форму й організованість. Процес логічного поділу великої проблеми на окремі елементи передбачає отримання об’єктивного управлінського рішення, прийнятого на основі економіко-математичної моделі. Призначення концептуальних моделей – змістовно представляти суттєві властивості об'єкта і головні зв'язки між цими властивостями.

На етапі побудови математичної моделі здійснюється формалізація концептуальної моделі. Математична формалізація позначає, що відпрацьовані конкретні правила дій, концептуальні положення, адекватні цілям дослідження і прийнятій системі гіпотез, здійснюється глибинний зв'язок між математичним інструментом, предметом дослідження і дослідником. Етап математичного аналізу моделі пов'язаний з тим, що математичними прийомами дослідження виявляються загальні властивості моделі та її рішень, при цьому важливим моментом є доказ існування рішення сформульованої задачі.

На етапі підготовки висхідних даних передбачається вимірювання ознак об'єкта, які є його основними властивостями і відображення величин в системі показників. Інформаційною моделлю в економіко-математичному моделюванні вважають ієрархічну систему показників, що відображають ознаки об'єкта.

На етапі чисельного рішення моделі використовують існуючі програмні засоби, прикладні пакети, розробляються спеціальні обчислювальні програми для реалізації моделі. Обчислення можуть мати різноманітний характер для вивчення поведінки моделі у різних умовах і обмеженнях.

Аналіз числових результатів та їх використання містять перевірку правильності та повноти результатів моделювання та використання в практичній діяльності, а також для вдосконалення самої моделі. На цьому етапі слід виконати верифікацію (перевірку правильності структури, логіки моделі) і валідацію моделі (перевірку відповідності даних, отриманих на основі моделі, реальному процесу). У випадку виявлення помилки, неточностей слід виявснити причину та повернутися на попередні етапи для удосконалення.

Етап використання результатів моделювання для прийняття управлінського рішення складається з якісного аналізу отриманих результатів, які не тільки представляються формулами, але і для наочності зображуються у вигляді графіків, таблиць, схем.

Окремо слід вказати, що кожен етап моделювання доцільно супроводити оцінкою отриманих результатів для своєчасного усунення виявлених помилок. Найтиповішими помилками є включення в модель несуттєвих для даної задачі змінних, не включення в модель істотних змінних, низька точність оцінок параметрів моделі, недоліки в структурі моделі, що призводить до неправильної специфікації моделі. Процес моделювання має циклічний характер і, починаючи моделювання об'єкта з розробки простої моделі переходять до розробки складних моделей, вдосконалення їх за допомогою урахування нових умов і уточнення математичних залежностей.

У математичному моделюванні розрізняють такі види контролю: контроль характеру залежностей, контроль екстремальних ситуацій, контроль граничних умов, контроль математичної замкнутості. За допомогою контролю характеру залежностей перевіряється напрямок і швидкість змін одних ознак у разі зміни інших.

У цілому процес економіко-математичного моделювання можна розглядати як послідовність розроблення моделей: когнітивної, змістовної моделі (описової, що пояснює, передбачувальної), концептуальної моделі (логіко – семантичної, структурно – функціональної, причинно – наслідкової), формалізованої моделі (математичної та інформаційної). Когнітивною прийнято називати модель, що сформована в голові дослідника як деякий уявний образ. Представлення когнітивної моделі звичайною мовою є змістовною моделлю. Концептуальною моделлю називають змістовну модель, у процесі формування якої використовують поняття і уявлення предметної галузі знань, до якої належить об'єкт моделювання. Формалізована модель є представленням концептуальної моделі за допомогою однієї або декількох формальних мов.

У процесі розроблення різних економіко-математичних моделей дотримуються загальної технології економіко-математичного моделювання. Так процес розроблення оптимізаційної моделі в економіці в загальному вигляді може складатися з чотирьох етапів: формулювання проблеми; безпосереднє побудова моделі; знаходження оптимального модельного розв'язку; перевірка адекватності моделі [30, с. 18 – 19]. До першого етапу відноситься формулювання мети дослідження, виявлення можливих альтернатив розв'язання, визначення властивих системі, що досліджується, вимог, умов, обмежень. На другому етапі встановлюються параметри, що управляються, кількісні співвідношення для виразу цільової функції і обмежень у вигляді функцій для параметрів, що управляються. На третьому етапі розв'язується сформульована задача, знаходиться оптимальний розв'язок, проводиться аналіз на чуттєвість, який покаже можливість змінени розв'язку при зміненнюванні числових значень параметрів системи. На четвертому етапі проводиться перевірка адекватності моделі.

Процес побудови економетричної моделі об'єкта в економіці, що є економічною системою, має свої особливості, що виражаються в послідовності таких етапів [20, с. 17-23; 27, с. 13-14]:

1) аналіз характерних ознак економічної системи і процесів, обгрунтування відповідної системи показників і її структури взаємозв'язків, визначення класу моделей, найбільш відповідних реальному об'єкту для його опису;

2) формування сукупності спостережень – вихідних даних для моделювання, які є тимчасовими, просторовими і просторово-часовими вибірками;

3) перевірка однорідності і точності вихідних даних;

4) оцінка параметрів обраної форми моделі на основі вихідних даних;

5) перевірка якості обчисленої моделі на основі статистичних критеріїв та висновки щодо її використання для подальшого економетричного дослідження;

6) прогнозування показників в економіці на основі моделі.

Запитання для самоперевірки:

1. Які основні методологічні принципи моделювання?
2. Яка основна суть кожного методологічного принципу моделювання?
3. Дайте визначення математичної моделі.
4. Які основні типи моделей?
5. Як класифікуються економіко-математичні моделі?
6. Які особливості оптимізаційних моделей?
7. Які особливості економетричних моделей?
8. З яких етапів складається побудова економетричних моделей?

9. З яких етапів складається технологія економіко-математичного моделювання?

10. Яка модель називається когнітивною?

11. Яка модель називається змістовною?

12. Яка модель називається концептуальною?

13. Яка модель називається формалізованою?

Тема 2. Оптимізаційні економіко-математичні моделі

2.1. Класифікація оптимізаційних задач.

2.2. Основи класичної теорії оптимізації.

2.3. Загальна постановка задачі оптимізації.

2.4. Класична задача умовної оптимізації. Формулювання задачі.

2.5. Метод множників Лагранжа.

Велику групу математичних методів в економіці утворюють оптимізаційні методи, оскільки перед менеджерами, економістами, працівниками системи управління, інженерами різного рівня виникають проблеми прийняття рішення, які вимагають оптимізації результатів різних видів діяльності з урахуванням наявних ресурсів. Існує спеціальна теорія прийняття рішення, що має економічні, психологічні, політичні, соціальні, фінансові та інші аспекти, враховуючи які слід шукати найкращі рішення. Алгоритмізація процесу розробки моделі передбачає використання різних математичних методів для пошуку найкращого рішення. Ці методи поділяються на методи лінійної та нелінійної оптимізації.

Типовими оптимізаційними завданнями є задачі оптимального планування, в яких виділяють змінні і параметри (кількість купованих продуктів, кількість виробленої продукції, кількість перевезеного вантажу), ціль, яку бажають досягти (функція цілі) і яку слід оптимізувати (мінімізувати витрати на споживання, максимізувати прибуток підприємства, мінімізувати вартість перевезень) і обмеження, тобто умови, що обмежують можливості досягнення бажаної цілі (в раціоні повинні бути визначені компоненти, обмежені ресурси підприємства, кількість перевезеного товару). Цільова функція має сенс очікуваної «цінності» або «корисності». Задача оптимального планування також називається оптимізаційною або екстремальною задачею.

У задачах оптимізації можуть бути виділені характеристики об'єкта (об'єктів), якими можна або треба варіювати для досягнення цілі. Такі характеристики називаються керованими змінними або керованими параметрами. Набір значень керованих змінних у задачах оптимізації називаються розв’язком. Значення керованих змінних можуть бути обмеженими. Розв’язок, що задовольняє висунуті обмеження називають допустимим розв’язком. Оптимальним називається допустимий розв’язок, який з деяких причин переважає над іншим розв’язком, наприклад, розв’язок, за якого цільова функція екстремальна. Існують також некеровані змінні, тобто зміна їх значень не залежить від керуючого суб'єкта.

Задачі оптимізації в економіці об'єднуються в розділі «Дослідження операцій», який передбачає використання математичних методів для моделювання та аналізу ситуацій в економіці. Серед цих методів окремо виділяють лінійне і нелінійне програмування, називаючи математичним програмуванням, що є одним з основних методів прийняття виробничо-економічних рішень.

2.1. Класифікація оптимізаційних задач

Будь-яка модель в оптимізаційній задачі включає шукані змінні, обмеження, що накладаються на них і формулювання цілі. Ціль визначає цільову функцію, що задається на безлічі припустимих розв’язків D. Безліч D виражає міру здійснення цілі: якщо D порожнє, то розв’язку не існує; якщо D містить одну точку, то ця точка є єдиним допустимим розв’язком задачі і таке завдання не представляє інтересу для дослідження; якщо D містить більш ніж один розв’язок, тоді задача оптимізації полягає в знаходженні оптимального розв’язку на безлічі допустимих розв’язків. При цьому, якщо D скінченне, то оптимальний розв’язок може бути знайдений у результаті простого перебирання всіх точок D і обчислення в них функції цілі. Якщо D перелічене або D є континуумом, то оптимальний розв’язок слід шукати в нескінченній множині допустимих розв’язок.

Математична модель задачі оптимізації передбачає, що всі змінні, параметри, обмеження і цільова функція моделі кількісно вимірні. Якщо змінні

$X=(x_1, x_2, ..., x_n)$ є керованими змінними, а $Z=(z_1, z_2, ..., z_k)$ – не керованими параметрами й умова функціонування досліджуваної системи визначається m обмеженнями, то математичну модель записують так:

знайти точку $Y=(y_1, y_2, ..., y_n)$, в якій досягається екстремум, мінімум або максимум, цільової функції $f(X, Z)$:

$$f(Y, Z)=extrf(X, Z)$$

за обмежень

$$\varphi_i(X, Z)\le(\ge,=)b_i, i=\overline{1,m},$$
$$x_j\ge 0, \ j=\overline{1,n}.$$

Таким чином, задача на умовний екстремум зазвичай має сенс, якщо $m<n$.

Оптимізаційні задачі за видом цільової функції і за виглядом обмежень класифікуються на групи. Якщо функція цілі і система обмежень є лінійними, то говорять про лінійне програмування, в іншому випадку виникає задача нелінійного програмування. У разі квадратичної функції цілі і лінійної системи обмежень задачу оптимізації називають задачею квадратичного програмування. Якщо функцію цілі можна представити у вигляді суми таких функцій, що кожна з них залежить тільки від однієї змінної, то розглядають задачу сепарабельного програмування. Якщо керовані змінні принципово можуть бути тільки цілими, то така задача оптимізації називається цілочисельною.

Якщо функція цілі є опуклою функцією, то така задача оптимізації називається задачею опуклого програмування. Якщо функція φ_i, визначають обмеження $\varphi_i(X, Z)\le(\ge,=)b_i$ є опуклими вгору (опуклими вниз) функціями, то вони породжують опуклу множину допустимих розв'язків.

За інформаційними властивостями оптимізаційні задачі діляться на статичні і динамічні. Якщо суб'єкт у процесі прийняття рішення не змінює свій інформаційний стан, то ухвалення рішення розглядається як миттєвий факт і такі задачі називаються статичними. Якщо суб'єкт у процесі прийняття рішення змінює свій інформаційний стан, то рішення приймається поетапно і задача називається динамічною.

Крім наведених моделей до оптимізаційних іноді відносять імітаційні моделі та евристичні. Імітаційні моделі дозволяють імітувати поведінку дуже складних систем, для яких не можливо побудувати математичні моделі та отримати розв'язок. Тут виникають складнощі, пов'язані з розробкою експерименту, перевіркою статистичної значущості його результатів, а також сам процес оптимізації викликає ускладнення. Якщо наявна задача оптимізації настільки складна, але існує припущення, що рішення є, то використовують різні евристичні методи, які засновані на інтуїції дослідника або на емпіричних даних. Слід вказати, що точність моделі також залежить від обсягу і складу наявних вихідних даних.

2.2. Основи класичної теорії оптимізації

Нехай функція $f(X)$ визначена в області O n-мірного евклідового простору $\Re^n$, $X=(x_1, x_2, ..., x_n) \in O \subseteq \Re^n$ і приймає значення на деякому інтервалі O_f дійсній осі, $f=(X) \in O_f \subseteq \Re$. Точка $Y=(y_1, y_2, ..., y_n) \in O$ називається точкою локального мінімуму або локального максимуму функції $f(X)$, якщо існує таке $\varepsilon > 0$, що нерівність

$$f(Y) \leq f(Y + \delta X) \text{ або } f(Y) \geq f(Y + \delta X) \qquad (2.1)$$

виконується для всіх $\delta X = (\delta x_1, \delta x_2, ..., \delta x_n)$, таких, що $0 < |\delta x| < \varepsilon$. Значення функції $f(Y)$ в цій точці називається локальним мінімумом або локальним максимумом. Множина точок $U_\varepsilon(Y) = \{Y + \delta X\}$ утворює відкриту сферу радіуса ε с центром у точці $У$ або $\varepsilon -$ окіл. Якщо нерівність (2.1) виконується строго:

$$f(Y) < f(Y + \delta X) \text{ або } f(Y) > f(Y + \delta X) \qquad (2.2)$$

то такий мінімум (максимум) називається строгим, в іншому випадку мінімум (максимум) − нестрогий. Пошук точок мінімуму або максимуму це пошук точок екстремуму.

Точка $Y \in O$ називається точкою глобального мінімуму (максимуму) функції $f(X)$, якщо нерівність (2.1) або (2.2) виконується для всіх точок δX області визначення O ($\delta X \in O$):

$$f(Y) \le f(X) \text{ або } f(Y) \ge f(X).$$

Відомо, що точка глобального екстремуму є точкою локального екстремуму, а зворотне твердження неправильне.

Якщо точка $Y \in O$ є точкою, в якій функція $f(X)$ досягає глобального мінімуму або максимуму, то з визначення екстремуму випливає, що число $f^* = f(Y)$ є відповідно точною нижньою межею або точною верхньою межею множини значень функції $f(X)$:

$$f^* = \inf_{X \in O} f(X) = f(Y) \text{ або } f^* = \sup_{X \in O} f(X) = f(Y).$$

Зворотне може бути непрвильно, якщо функція $f(X)$ не має екстремуму, то:

$$f^* = \inf_{X \in O} f(X) < f(X) \text{ або } f(X) = \sup_{X \in O} f(X) = f^*.$$

Від локального або глобального мінімуму (максимуму) функції $f(X)$ завжди можна перейти до локального або глобального максимуму (мінімуму) функції $f(X)$ за допомогою перетворення:

$$F(X) = -f(X).$$

Функції, що мають єдиний мінімум (максимум), називаються унімодальні.

2.3. Загальна постановка задачі оптимізації

У загальному вигляді задача оптимізації або задача визначення екстремуму ставиться таким чином.

Нехай задані:

Функція $f(X)$, визначена на множині $O \subseteq \Re^n$;

множина $D \subseteq \Re^n$.

Знайти точку $Y = (y_1, y_2, ..., y_n) \in D$, в якій функція $f(X)$ досягає екстремального (мінімального чи максимального) значення, тобто:

$$f(Y) = extr\ f(X) \text{ і } Y \in D.$$

Функція $f(X)$ називається цільовою функцією, змінні X – керованими змінними, D – допустимою множиною, а будь-який набір значень $У$ керованих змінних, що належить D ($Y \in D$) – допустимим розв'язком задачі оптимізації.

Точка $У$, в якій $f(X)$ досягає свого екстремуму, повинна належати перетинанню області визначення O функції $f(X)$ і допустимій множині D ($Y \in O \bigcap D$). Якщо множина O і D співпадають з усім простором $\Re^n$ $\left(O = D = \Re^n\right)$, то така задача називається задачею на безумовний екстремум. Якщо хоча б одна з множин O або D є власною підмножиною простору $\Re^n$ $\left(O \in \Re^m, D \in \Re^n\right)$ або множини O і D перетинаються $\left(O \bigcap D \neq \varnothing\right)$, то така задача називається задачею на умовний екстремум, в іншому випадку $\left(O \bigcap D = \varnothing\right)$ точка екстремуму $У$ не існує. Якщо множина O і D перетинаються в одній точці $У$, то ця точка $У$ є єдиним допустимим розв'язком і обговорювати проблему пошуку екстремуму безглуздо.

Зазвичай у задачі умовного екстремуму задається не сама допустима множина розв'язків D, а система співвідношень, її визначальна,

$\varphi_i(x_1, x_2, ..., x_n) \leq (=, \geq)\, 0, i = \overline{1,m}$

тобто $D = \left\{\varphi_i(x_1, x_2, ..., x_n) \leq (=, \geq)\, 0, i = \overline{1,m}\right\} \in \Re^n$,

або безліч D може одночасно задаватися як в явному вигляді, тобто допустимий розв'язок X повинен належати деякій області $P \in \Re$, так і системою обмежень.

Таким чином, задача на безумовний екстремум цільової функції $f(X)$ полягає в знаходженні точки $Y = (y_1, y_2, ..., y_n) \in \Re^n$, в якій досягається екстремум $f(X)$:

$$f(Y) = extr\ f(X). \tag{2.3}$$

Задача на умовний екстремум цільової функції $f(X)$ полягає в знаходженні точки $Y=(y_1, y_2, ..., y_n) \in P \subseteq \Re^n$, в якій досягається екстремум $f(X)$:

$$f(Y) = extr\ f(X), \quad (2.4)$$

за умови, що точка *У* задовольняє обмеженням

$$\varphi_i(x_1, x_2, ..., x_n) \le (=, \ge)\, 0, i = \overline{1,m}, \quad (2.5)$$

де $\varphi_i(X)$ задані на деякій множині $O' \in \Re^m$, тобто

$Y \in O \bigcap O' \bigcap D$

$D = \{X : X \in P\ і\ \varphi_i(X) \le (=, \ge)\, 0, i = \overline{1,m}\} \subseteq \Re^n$.

Тут $O \bigcap O'$ – область визначення функцій $f(X)$ і $\varphi_i(X)$.

Як правило, в задачі умовної оптимізації $D \subseteq O \bigcap O'$ і тому область визначення функцій $f(X)$ і $\varphi_i(X)$ явно не вказують, а розглядають тільки множину *P*, завжди маючи на увазі, що функція цілі $f(X)$ і функції $\varphi_i(X)$, що визначають обмеження, принаймні, задані на *P*. Природно, що за необхідності обмеження, які обумовлені множинами *O* і O', включаються до *P*.

У вирішенні реальних економічних проблем задачі знаходження безумовного екстремуму, як правило, зустрічаються надзвичайно рідко.

Слід зазначити, що якщо *У* – точка безумовного локального екстремуму функції $f(X)$ знаходиться всередині безлічі допустимих значень *D*, то обмеження виконуються автоматично, якщо точка *У* – за межами безлічі допустимих значень *D*, то точка умовного локального екстремуму $f(X)$ знаходиться на границі *D*, тобто обмеження-нерівності перетворюються на обмеження-рівності. У задачі тільки з обмеженнями-рівностями кількість обмежень *m* повинно бути менше числа змінних *n*, *m*<*n*, так як при *m*>*n* принаймні (*m*-*n*) обмежень виявляються надлишковими, або система обмежень (2.5), або система обмежень (2.5) перевизначена і, отже, не має припустимого рішення. При *m*=*n* допустимий розв'язок, швидше за все, буде єдиним, і така задача оптимізації також не представляє інтересу. Задача пошуку оптимального розв'язку має сенс, коли *m*<*n*.

Якщо $m<n$, то завжди з n змінних $X=(x_1,x_2,...,x_n)$ можна вибрати $n-m$ змінних, які називаються незалежними (вільними) змінними; інші змінні називаються залежними змінними.

2.4. Класична задача умовної оптимізації. Формулювання задачі

Розглянемо задачу умовної оптимізації (2.4), (2.5), яка полягає в знаходженні точки $Y=(y_1,y_2,...,y_n)\in P\subseteq\Re^n$, в якій досягається екстремум $f(X)$:

$$f(Y)=extr\ f(X), \tag{2.4}$$

за умови, що точка Y задовольняє обмеженням:

$$\varphi_i(x_1,x_2,...,x_n)\le(=,\ge)\,0,\ i=\overline{1,m}. \tag{2.5}$$

Сформулюємо необхідні і достатні умови існування екстремуму в задачі на умовний екстремум.

Нехай вектор $Z\in D\subseteq\Re^n$ задає можливий напрямок на множині допустимих значень D, якщо при всіх достатньо малих $\alpha>0$ вектор $Y+\alpha Z\in D$. Всі можливі напрямки на множині допустимих значень D функції $f(X)$ в точці Y утворюють безліч можливих напрямків $v(Y,f)\subseteq\Re^n$.

Теорема. Якщо точка Y є точкою локального мінімуму (максимуму) задачі (2.4), (2.5), то $w(Y,f)\cap v(Y,f)\neq\varnothing$ де $w(Y,f)$ – множина напрямків спадання (зростання) $f(X)$ в точці Y.

Дана теорема показує, що неможливо зміщення з точки локального екстремуму Y, яке призводить до зменшення (збільшення) цільової функції і при цьому не виходить за межі допустимої області. Автор книги «Введення в теорію і методи оптимізації для економістів» Фролькіс В. А. [30] вважає, що дана теорема тривіальна і її слід доповнити теоремою Вейєрштрасса.

Теорема Вейєрштрасса. Нехай $D\subseteq\Re^n$ – замкнута обмежена множина і $f(X)$ – неперервна функція, визначена на D. Тоді існує точка $Y\in D$ глобального мінімуму (максимуму) функції $f(X)$, $f(Y)=min(max)f(X)$.

Наведені дві теореми не вказують алгоритм знаходження екстремальної точки. Припустимо, функція $f(X)$ визначена і неперервна в обмеженій замкненій області *D*. У цій області (згідно з теоремою Вейєрштрасса) знайдеться точка, в якій функція буде екстремальною. Насправді своє екстремальне значення функція $f(X)$ може досягти як у внутрішній точці області *D*, так і на її межі. Якщо додатково функція $f(X)$ має кінцеві частинні похідні першого і другого порядку, то для отримання необхідних і достатніх умов існування локального екстремуму цільової функції $f(X)$ у внутрішніх точках *D* можна використовувати теореми, що визначають необхідні і достатні умови існування локального безумовного екстремуму, але вони не діагностують екстремальні точки, що лежать на межі *D*. Тому для того, щоб знайти глобальний екстремум $f(X)$ в області *D*, необхідно знайти всі внутрішні точки локального екстремуму функції $f(X)$, обчислити в них значення $f(X)$ і порівняти зі значеннями функції $f(X)$ на межі області. Найменше (найбільше) з них і буде шуканим глобальним значенням. Якщо внутрішніх стаціонарних точок немає, то екстремальне значення може досягатися тільки на межі області. У багатьох задачах умовної оптимізації екстремум досягається саме на межі області *D*, тому для таких задач результати теорем класичного аналізу неприйнятні.

Для того щоб не обчислювати значення цільової функції вздовж межі області допустимих значень *D*, а сформулювати необхідні і достатні умови екстремуму як для внутрішніх точок області *D*, так і для точок на її межі, необхідно розглянути окремий випадок задачі (2.4), (2.5) у вигляді, який називається класична задача оптимізації.

Знайти точку локального екстремуму $Y=(y_1, y_2, ..., y_n)$ цільової функції $f(X)$,

$$f(Y)=extr\ f(X), \tag{2.6}$$

за умови, що точка *У* задовольняє обмеженням-рівностям:

$$\varphi_i(X)=0, i=\overline{1,m},\ m<n, \tag{2.7}$$

тобто $Y\in D$, де допустима множина визначається як:

$$D=\{X:\ \varphi_i(X)=0, i=\overline{1,m}\}\subset\Re^n \tag{2.8}$$

$$\text{і } D \subseteq O \bigcap O'.$$

У теорії умовної оптимізації існує кілька методів, що дозволяють звести задачу умовного екстремуму до задачі безумовного екстремуму і завдяки такому прийому використовувати необхідні і достатні умови екстремуму. До таких найбільш поширених методів відноситься метод множників Лагранжа і метод Якобі.

2.5. Метод множників Лагранжа

Метод множників Лагранжа дозволяє звести класичну задачу на умовний екстремум до задачі на безумовний екстремум. Ідея методу Лагранжа полягає в побудові допоміжної функції, так званої функції Лагранжа, точка безумовного локального екстремуму *У* збігається з точкою умовного локального екстремуму функції задачі (2.6) – (2.8).

Такими властивостями володіє функція виду:

$$L(\Lambda, \lambda_0, X) = \lambda_0 f(X) + \sum_{i=1}^{m} \lambda_i \varphi_i(X), \qquad (2.9)$$

параметри λ_0 і $\Lambda = (\lambda_1, \lambda_2, ..., \lambda_m) \in \Re^m$ якої називаються множниками Лагранжа. Функцію Лагранжа можна записати у вигляді:

$$L(\Lambda, \lambda_0, X) = \lambda_0 f(X) - \sum_{i=1}^{m} \lambda_i \varphi_i(X).$$

Якщо функції $f(X)$ і $\varphi_i(X)$ неперервні і мають частинні похідні першого порядку, то частинні похідні функції $L(\Lambda, \lambda_0, X)$ за змінними x_j мають вигляд:

$$\frac{\partial L(\Lambda, \lambda_0, X)}{\partial x_j} = \lambda_0 \frac{\partial f(X)}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial \varphi_i(X)}{\partial x_j}. \qquad (2.10)$$

Складений з них вектор позначається як:

$$L'_x(\Lambda, \lambda_0, X) = \lambda_0 f'(X) + \sum_{i=1}^{m} \lambda_i \varphi'_i(X).$$

Частинні похідні функції $L(\Lambda, \lambda_0, X)$ за змінними $\lambda_i, i = \overline{1,m}$ завжди існують і рівні:

$$\frac{\partial L(\Lambda, \lambda_0, X)}{\partial \lambda_i} = \varphi_i(X). \qquad (2.11)$$

Необхідні умови екстремуму першого порядку в задачі безумовної оптимізації функції Лагранжа формулюються у вигляді наступної теореми.

Теорема (правило множників Лагранжа). Якщо точка $Y \in D \subseteq \Re^n$ – точка локального оптимуму задачі на умовний екстремум (2.6) − (2.8):

$$f(Y) = extr\ f(X) \text{ і } \varphi_i(X) = 0, i = \overline{1,m},$$

в околі деякої функції $f(X)$ і $\varphi_i(X)$ безперервно диференційовані, а в точці *У* ранг матриці Якобі дорівнює *m*, то існують такі одночасно нерівні нулю вектор Λ' і параметр λ_0', що точка $(\Lambda', \lambda_0', Y)$ є стаціонарною точкою задачі на безумовний екстремум функції Лагранжа (2.9), відповідної задачі (2.6) – (2.8):

$$grad\ L(\Lambda', 1, Y) = 0.$$

Враховуючи вид похідних (2.10) – (2.11) стаціонарну точку *У* функції Лагранжа можна визначити як точку, що задовольняє умови:

$$L_x'(\Lambda, \lambda_0, X) = 0, \varphi_i(X) = 0, i = \overline{1,m}.$$

На відміну від задачі на безумовний екстремум стаціонарна точка функції Лагранжа може перебувати на межі області *D*.

Приклад. Знайти точки умовного екстремуму функції;

$$f(X) = xyz + 8x^2 + 12y^2 + 25z^2$$

за обмежень

$$x + y + 7z = 12,$$
$$6x + 3y - 4z = 25.$$

Функція Лагранжа задачі, що розглядається:

$$L(\Lambda, X) = xyz + 8x^2 + 12y^2 + 25z^2 + \lambda_1(x + y + 7z - 12) + \lambda_2(6x + 3y + 4z - 25).$$

Для того щоб точка (Λ', Y) була стаціонарною точкою $L(\Lambda, X)$, повинні виконуватися необхідні умови екстремуму:

$$\frac{\partial L(\Lambda, X)}{\partial x} = 16x + xy + \lambda_1 + 6\lambda_2 = 0,$$
$$\frac{\partial L(\Lambda, X)}{\partial y} = xy + 24y + \lambda_1 + 3\lambda_2 = 0,$$
$$\frac{\partial L(\Lambda, X)}{\partial z} = xy + 50z + 7\lambda_1 + 4\lambda_2 = 0,$$
$$\varphi_1(X) = x + y + 7z - 12 = 0, \quad \varphi_2(X) = 6x + 3y + 4z - 25 = 0.$$

Розв'язком системи є дві стаціонарні точки (Λ^1, Y^1) і (Λ^2, Y^2) з координатами:

$$Y^1 = (2{,}917;\ 0{,}95;\ 1{,}162), \quad \Lambda^1 = (-4{,}581;\ -7{,}2),$$
$$Y^2 = (113036;\ -245{,}198;\ 20{,}595), \quad \Lambda^2 = (3872;\ -105{,}217),$$
$$f(Y^1) = 115{,}877; \quad f(Y^2) = 263500\ .$$

За допомогою частинних похідних другого порядку доводиться, що точка Y^1 є точкою мінімуму функції $f(X)$, а точка Y^2 – точка максимуму.

Необхідно розглянути економічну інтерпретацію множників Лагранжа. Для задачі на умовний екстремум:

знайти точку локального екстремуму $Y = (y_1, y_2, ..., y_n)$ цільової функції $f(X)$:

$$f(Y) = extr\ f(X),$$

за умови, що керовані змінні $X = (x_1, x_2, ..., x_n)$ задовольняють обмеженням-рівностям:

$$\varphi_i(X) = b_i,\ i = \overline{1, m}.$$

Якщо цільову функцію $f(Y)$ в точці екстремуму $У$ інтерпретувати як оптимальний дохід (або вартість), а постійні b_k – як обсяги будь-яких ресурсів, то множники Лагранжа λ_k показують, як зміниться оптимальний дохід (або вартість) при збільшенні кількості k-го ресурсу на одиницю. Іншими словами, множники Лагранжа вказують, як зміниться оптимальне значення цільової функції $f(Y)$ при зміні параметра b_k в правій частині обмежень на одиницю, тобто визначають цінність k-го ресурсу.

Запитання для самоперевірки:

1. Наведіть класифікацію оптимізаційних задач.
2. Наведіть постановку загальної оптимізаційної моделі.
3. Яка точка називається точкою екстремуму?
4. Як визначається допустима множина в оптимізаційній моделі?
5. Як формулюється задача на безумовний екстремум?
6. Як формулюється задача на умовний екстремум?
7. Яка точка називається точкою локального мінімуму (максимуму)?
8. Які методи дозволяють привести задачу умовного екстремуму до задачі безумовного екстремуму?
9. Як будується функція Лагранжа?
10. Сформулюйте правило множників Лагранжа.
11. Яка економічна інтерпретація множників Лагранжа?

3. Задача лінійного програмування і методи її розв’язання

3.1. Постановка задачі лінійного програмування.

3.2. Канонічна форма задачі лінійного програмуванняю.

3.3. Графічний метод розв’язання задачі лінійного програмування.

3.4. Властивості можливих розв’язків задачі лінійного програмування.

3.1. Постановка задачі лінійного програмування

У щоденній практиці економісти, менеджери ухвалюють управлінські рішення, від обґрунтованості яких залежить ефективне функціонування підприємства, його розвиток. Якщо при пошуку управлінського рішення враховуються основні наявні умови, критерії дії, тобто раціональність і доцільність ґрунтуються на оптимальності, то такі рішення рідко призводять до збитків та втрати. Тому більшість задач в економіці є оптимізаційними. Однією з найпоширеніших оптимізаційних задач є задачі, пов’язані з плануванням виробництва на підприємстві: розробка плану випуску максимального обсягу продукції при обмежених сировинних ресурсах або в межах конкретного плану виробництва визначення мінімально необхідного обсягу виробничих витрат. До цієї групи належить задача про оптимальну суміш, наприклад, задача про дієту, де необхідно враховувати калорійність і споживчу цінність продуктів, а також їх вартість; задачі про визначення оптимального складу сумішей у процесі відгодівлі тварин. До цієї ж групи, також, належить задача про розкрій матеріалів, суть якої полягає в розробці таких технологічно допустимих планів розкрою, за яких виходить необхідний комплект заготовок, а відходи зводяться до мінімуму. Іншим найбільш поширеним класом задач є задача про мінімізацію транспортних витрат, коли мова йде про раціональне перевезення деякого однорідного продукту від виробників до споживачів.

Розглянемо типові задачі лінійного програмування (ЗЛП).

Задача про оптимальне використання ресурсів

Нехай для виробництва деякої продукції A_1, A_2, A_3 використовуються ресурси B_1, B_2, B_3, B_4 (сировина, напівфабрикати, обладнання, фінанси), запаси яких складають b_1, b_2, b_3, b_4. Відомі норми витрат a_{ij} кожного виду ресурсу B_i на виготовлення одиниці продукції A_j, а також прибуток c_j від виробництва і реалізації кожного виду продукції (табл. 3.1).

Таблиця 3.1

Умова задачі

Ресурси	Продукція			Запаси ресурсів
	A_1	A_2	A_2	
B_1	a_{11}	a_{12}	a_{13}	b_1
B_2	a_{21}	a_{22}	a_{23}	b_2
B_3	a_{31}	a_{32}	a_{33}	b_3
B_4	a_{41}	a_{42}	a_{43}	b_4
Прибуток	c_1	c_2	c_3	

Необхідно знайти план виробництва, тобто необхідно виготовити таку кількість x_1, x_2, x_3 продукції кожного виду, щоб максимізувати загальний прибуток. Оскільки прибуток від виробництва одиниці A_1 дорівнює c_1, то прибуток від планової кількості x_1 буде дорівнювати $c_1 x_1$; загальний прибуток буде дорівнювати сумі прибутків від виробництва всіх видів продукції:

$$Z = c_1 x_1 + c_2 x_2 + c_3 x_3 \to max.$$

Витрати ресурсу B_1 на виробництва запланованих обсягів продукції x_1, x_2, x_3 будуть дорівнювати добуткам $a_{11}x_1$, $a_{12}x_2$, $a_{13}x_3$. Необхідно врахувати, що загальні витрати цього ресурсу $a_{11}x_1 + a_{12}x_2 + a_{13}x_3$ не можуть перевищувати його запаси b_1, тобто: $a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \le b_1$. Запишемо аналогічні умови за усіма видами ресурсів і отримаємо систему нерівностей:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \le b_1; \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \le b_2; \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \le b_3; \\ a_{41}x_1 + a_{42}x_2 + a_{43}x_3 \le b_4. \end{cases}$$

Додамо ще вимогу невід'ємності невідомих $x_1, x_2, x_3 \ge 0$.

Математична модель цієї задачі має вигляд:

знайти максимум лінійної функції цілі

$$Z = \sum_{j=1}^{n} c_j x_j \to \max \qquad (3.1)$$

за наявності лінійної системи обмежень-нерівностей:

$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j \le b_i, \\ (i = \overline{1, m}), \end{cases} \qquad (3.2)$$

де усі $x_j \ge 0$; n – кількість видів продукції; m – кількість видів ресурсів.

Це є задача лінійного програмування (ЗЛП).

Модель, де потрібно знайти максимум лінійної функції цілі при лінійних обмеженнях типу $\le$ є ***першою стандартною формою*** ЗЛП. Якщо продукція вимірюється в штуках, то в модель ще потрібно включити умову, щоб усі невідомі були цілими.

Приклад. Для виготовлення двох типів в'язальних станків підприємство використовує три основних види сировини: метал, пластик, алюміній. На виробництво одного в'язального верстату першого типу необхідно витратити металу 10 кг, пластику 5 кг і алюмінію 4 кг. На виробництво одного в'язального станка другого типу необхідно витратити металу 9 кг, пластику 11 кг і алюмінію 15 кг. Виробництво забезпечено металом в обсязі 1870 кг, пластиком

– 1455 кг, алюмінієм – 1815 кг. Прибуток від реалізації одиниці в’язального верстату першого типу становить 700 грн, а від реалізації одиниці в’язального с верстату другого типу становить 900 грн. Скласти план виробництва в’язальних с верстатів на підприємстві, який забезпечує максимальний прибуток від їх реалізації.

Умова даної задачі може бути записана в табличній формі (табл. 3.2)

Таблиця 3.2

Умова задачі

Сировина (кг)	В’язальні верстати		Запаси сировини (кг)
	1-го типу	2-го типу	
Метал	10	9	1 870
Пластик	5	11	1 455
Алюміній	4	15	1 815
Прибуток (грн)	700	900	

Для складання економіко-математичної моделі позначимо: x_1 – кількість в’язальних верстатів першого типу; x_2 – кількість в’язальних верстатів другого типу, які може виробити підприємство, враховуючи обмежені обсяги сировини та максимізуючи прибуток. Критерієм оптимальності плану є забезпечення максимуму прибутку.

Математична модель цієї задачі має вигляд:

знайти максимум лінійної функції цілі

$$Z = 700x_1 + 900x_2 \rightarrow max,$$

за обмежень:

$$\begin{cases} 10x_1 + 9x_2 \le 1870, \\ 5x_1 + 11x_2 \le 1455, \\ 4x_1 + 15x_2 \le 1815, \\ x_1,\ x_2 \ge 0. \end{cases}$$

Задача складання найдешевшого раціону (кормової суміші) або задача про суміші (задача про дієту)

З наявних у продажу продуктів необхідно скласти такий раціон харчування, щоб його загальна вартість була найменшою, при цьому повинні бути забезпечені усі потреби організму в поживних речовинах (білки, жири, вуглеводи). У табл. 3.3 наведено дані про зміст поживних речовин B_1, B_2, B_3 у кожному виді продуктів A_1, A_2, A_3, A_4, вартості продуктів і мінімальні потреби організму в поживних речовинах.

Таблиця 3.3

Умова задачі

Продукти	Поживні речовини			Вартість продуктів
	B_1	B_2	B_3	
A_1	a_{11}	a_{12}	a_{13}	c_1
A_2	a_{21}	a_{22}	a_{23}	c_2
A_3	a_{31}	a_{32}	a_{33}	c_3
A_4	a_{41}	a_{42}	a_{43}	c_3
Щоденний попит	b_1	b_2	b_3	

Необхідно визначити, з якої кількості кожного виду продуктів повинен складатися найбільш дешевий раціон. Позначимо ці невідомі кількості (склад раціону) через x_1, x_2, x_3, x_4. Вартість одиниці продукту A_1 дорівнює c_1, в раціон входить x_1 одиниць продукту загальною вартістю $c_1 x_1$. Враховуючи витрати на інші види продуктів, отримаємо вартість кожного раціону:

$$Z = c_1 x_1 + c_2 x_2 + c_3 x_3 + c_4 x_4 \to min.$$

Далі потрібно визначити, яка кількість першого виду поживної речовини B_1 міститься в раціоні. У табл. 3.3 наведено дані a_{i1} про зміст речовини B_1 в одиниці кожного виду продуктів A_i. Ці величини необхідно помножити на x_i і скласти суму: $a_{11} x_1 + a_{21} x_2 + a_{31} x_3 + a_{41} x_4$. Потрібно, щоб у раціоні сумарна кількість цього виду поживної речовини була не менше, ніж мінімальна добова потреба: $a_{11} x_1 + a_{21} x_2 + a_{31} x_3 + a_{41} x_4 \geq b_1$.

Уся система обмежень за видами поживних речовин має вигляд:

$$\begin{cases} a_{11} x_1 + a_{21} x_2 + a_{31} x_3 + a_{41} x_4 \geq b_1; \\ a_{12} x_1 + a_{22} x_2 + a_{32} x_3 + a_{42} x_4 \geq b_2; \\ a_{13} x_1 + a_{23} x_2 + a_{33} x_3 + a_{43} x_4 \geq b_3. \end{cases}$$

До цієї системи слід додати умову невід'ємності невідомих:

$$x_1, x_2, x_3, x_4 \geq 0.$$

Математична модель задачі представлена в такому вигляді:

знайти мінімум лінійної функції цілі

$$Z = \sum_{i=1}^{m} c_i x_i \to min \tag{3.3}$$

за обмежень:

$$\begin{cases} \sum_{i=1}^{m} a_{ij} x_i \geq b_j \\ (j = \overline{1, n}) \end{cases} \tag{3.4}$$

$$x_i \geq 0,$$

де n – кількість видів поживних речовин; m – кількість продуктів, які входять в раціон.

Ця задача є задачею лінійного програмування. Порівняно з попередньою задачею моделі в деякому змісті взаємні – зараз потрібно знайти мінімум лінійної функції цілі при лінійних обмеженнях типу $\geq$. Така форма моделі є ***другою стандартною формою*** задачі ЗЛП.

Приклад. До деякого сплаву може входити принаймні 4% нікелю і не більше 80% заліза. Для складання сплаву використовують три види сировини (наприклад, відходи, лом), що містять залізо, нікель та інші речовини; крім того можуть використовуватись добавки з чистого нікелю, заліза, та інших речовин. Вартість кожного виду сировини, місткість в них відповідних речовин і їх запаси наведені в табл. 3.4. Скільки кілограмів першого, другого, третього видів сировини, чистого заліза, нікелю та інших елементів слід включити до шихти, щоб собівартість 1 т сплаву була мінімальною?

Таблиця 3.4

Умова задачі

	Вид сировини			Компонент сплаву		
залізо	70 %	90 %	85 %	100 %		
нікель	5 %	2 %	7 %		100 %	
інші	25 %	8 %	8 %			100 %
Вартість 1 кг, грн	6	4	5	2,5	67	20

Запаси, кг	800	700	1 100	не обмежено	не обмежено	не обмежено

Позначимо: x_1, x_2, x_3 – обсяги сировини першого, другого, третього видів, які необхідно знайти (кг); x_4 – обсяг чистого заліза (кг); x_5 – обсяг нікелю (кг); x_6 – обсяг інших елементів, що входять в 1 т шихти (кг).

Математична модель даної задачі має вигляд:

знайти мінімум лінійної функції цілі (мінімум вартості 1 т (1 000 кг) шихти):

$$Z = 6x_1 + 4x_2 + 5x_3 + 2{,}5x_4 + 67x_5 + 20x_6 \to min,$$

за обмежень

$$\begin{cases} 0{,}7x_1 + 0{,}9x_2 + 0{,}85x_3 + x_4 \le 800, \\ 0{,}05x_1 + 0{,}02x_2 + 0{,}07x_3 + x_5 = 40, \\ x_1 \le 800, \\ x_2 \le 700, \\ x_3 \le 1100, \\ x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 1000, \\ x_{1,\dots,6} \ge 0. \end{cases}$$

Транспортна задача

У транспортній задачі потрібно скласти найбільш дешевий план перевезення однорідного продукту з пунктів виробництва A_1, A_2 в пункти споживання B_1, B_2, B_3. У табл. 3.5 представлені відомості про запаси a_i у пунктах виробництва, попит b_j у пунктах споживання і тарифи c_{ij} – вартості перевезення одиниці товару з кожного пункту відправлення в кожний пункт призначення.

Таблиця 3.5

Умова задачі

Постачальники	Споживачі			Запаси
	B_1	B_2	B_3	
A_1	c_{11}	c_{12}	c_{13}	a_1
A_2	c_{21}	c_{22}	c_{23}	a_2
Попит	b_1	b_2	b_3	

Невідомі x_{ij} – кількість вантажу, перевезеного від кожного постачальника A_i до кожного споживача B_j. Як правило, в збалансованій задачі загальний попит $b_1 + b_2 + b_3$ дорівнює загальному запасу $a_1 + a_2$. Необхідно вивезти всі запаси і задовольнити весь попит з найменшими транспортними витратами.

Складаємо функцію цілі, що описує загальні транспортні витрати. Тариф c_{11} показує вартість перевезення одиниці продукції за маршрутом A_1–B_1; витрати на перевезення за цим маршрутом кількості товару x_{11} будуть дорівнювати добутку $c_{11}\,x_{11}$. Загальні витрати будуть рівні сумі таких величин:

$$Z = (c_{11}x_{11} + c_{12}x_{12} + c_{13}x_{13}) + (c_{21}x_{21} + c_{22}x_{22} + c_{23}x_{23}) \to min.$$

Кількість товару, вивезеного з A_i, має бути рівним запасу a_i:

$$\begin{cases} x_{11} + x_{12} + x_{13} = a_1\,; \\ x_{21} + x_{22} + x_{23} = a_2\,. \end{cases}$$

Кількість товару, вивезеного з B_j, має бути рівним запасу b_j:

$$\begin{cases} x_{11} + x_{12} = b_1\,; \\ x_{21} + x_{22} = b_2\,; \\ x_{31} + x_{32} = b_3\,. \end{cases}$$

Усі невідомі в цій задачі − невід'ємні: $x_{ij} \geq 0$.

Математична модель транспортної задачі має вигляд:

знайти мінімум лінійної функції цілі

$$Z = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \to \min \qquad (3.5)$$

за обмежень-рівностей:

$$\begin{cases} \sum\limits_{j=1}^{n} x_{ij} = a_i \\ (i = \overline{1, m}) \end{cases}; \qquad \begin{cases} \sum\limits_{i=1}^{m} x_{ij} = b_j \\ (j = \overline{1, n}) \end{cases} \qquad (3.6)$$

$$x_{ij} \geq 0.$$

Транспортна задача також є задачею ЛП.

Приклад. Підприємство має три цеха (A_1, A_2, A_3) і чотири склади (B_1, B_2, B_3, B_4). Перший цех виробляє 30 тис. шт. виробів, другий – 40 тис. шт. виробів, третій – 20 тис. шт. виробів. Пропускна здатність складів характеризується показниками: склад 1 – 20 тис. шт., склад 2 – 30 тис. шт., склад 3 – 30 тис. шт., склад 4 – 10 тис. шт. Вартість перевезення з першого цеху до складів 1, 2, 3, 4 за одну тисячу штук виробів відповідно 2, 3, 2, 4 грн., з другого – 3, 2, 5,1 грн., з третього – 4, 3, 2, 6 грн. Скласти такий план перевезення виробів, за якого витрати на перевезення були б найменшими.

Умова задачі в табличній формі представлена в табл. 3.6.

Таблиця 3.6

Умова задачі

Цех	Склад				Обсяги виробництва, тис. шт.
	B_1	B_2	B_3	B_4	
A_1	1	2	3	4	30
A_2	3	2	5	1	40
A_3	4	3	2	6	20
Місткість складів	20	30	30	30	90

Позначимо: x_{ij} – кількість вантажу, який необхідно перевезти з цеху A_i до складу B_j, щоб витрати перевезення були б найменшими.

Математична модель даної транспортної задачі:

знайти мінімум лінійної функції цілі:

$$Z = x_{11} + 2x_{12} + 3x_{13} + 4x_{14} + 3x_{21} + 2x_{22} + 5x_{23} + x_{24} +$$
$$4x_{31} + 3x_{32} + 2x_{33} + 6x_{34} \to \min,$$

за обмежень-рівностей:

$$\begin{cases} x_{11}+x_{12}+x_{13}+x_{14}=30, \\ x_{21}+x_{22}+x_{23}+x_{24}=40, \\ x_{31}+x_{32}+x_{33}+x_{34}=20, \\ x_{11}+x_{21}+x_{31}=30, \\ x_{13}+x_{23}+x_{33}=30, \\ x_{14}+x_{24}+x_{34}=10, \\ x_{ij}\geq 0, \quad i=\overline{1,3}, \quad j=\overline{1,4}. \end{cases}$$

Задача обліку понаднормових робіт

Дотепер всі невідомі в умовах задач були невід'ємними $x_j \geq 0$, і в традиційній постановці задач ЛП ця властивість вважається навіть обов'язковою. Але у більш складних задачах з'являються невідомі без обмеження на знак $x_j \gtrless 0$, які залежно від виробничої ситуації можуть приймати як додатні, так і від'ємні значення.

Приклад. У процесі виготовлення виробів A і B здійснюється послідовна обробка на двох різних верстатах. Час роботи кожного верстата – 8 годин на добу, але його можна збільшити на 4 години за рахунок понаднормових робіт з погодинною додатковою оплатою 5 грн за годину. Продуктивності верстатів і прибуток від виготовлення та реалізації кожного виробу наведені в табл. 3.7.

Таблиця 3.7

Умова задачі

Верстати	Продуктивність, шт /година		Фонд часу, години
	A	B	
I	8	4	8
II	6	5	8
Прибуток, грн /шт	4	6	

Потрібно визначити кількості виробів кожного виду (x_1, x_2), які потрібно виготовити, щоб дістати найбільший прибуток (Z), за умови, що час роботи обладнання може бути збільшений тільки за рахунок понаднормових робіт.

Якщо понаднормові роботи не допускаються, то ця задача відноситься до вже розібраного вище класу завдань оптимального використання ресурсів:

$$\begin{cases} \dfrac{x_1}{8}+\dfrac{x_2}{4}\leq 8 & \text{(завантаження першого верстату не перевищує 8 год)} \\ \dfrac{x_1}{6}+\dfrac{x_2}{5}\leq 8 & \text{(завантаження другого верстату не перевищує 8 год)} \end{cases}$$

$z = 4x_1 + 6x_2 \to max$ (прибуток)

$x_1, x_2 \geq 0.$

Оптимальний розв'язок цієї задачі: $x_1 = 24$, $x_2 = 20$, $z = 216$.

Устаткування повинне бути завантажене повністю по 8 годин на добу. Тепер врахуємо можливість понаднормового завантаження устаткування. Позначимо через $y_1, y_2 \leq 4$ – понаднормове завантаження, що за умовою завдання не може перевищувати 4 години на добу. Тепер завантаження устаткування буде описуватися у вигляді наступних рівностей:

$$\begin{cases} \dfrac{x_1}{8} + \dfrac{x_2}{4} = 8 + y_1, \\ \dfrac{x_1}{6} + \dfrac{x_2}{5} = 8 + y_2. \end{cases}$$

Ця система рівнянь правильно описує обидві виробничі ситуації – коли понаднормові роботи допускаються (тоді $y_1, y_2 \geq 0$), і коли понаднормові роботи не допускаються (тоді $y_1, y_2 \leq 0$). Виходить, що в цієї задачі обидві нові невідомі не мають обмеження на знак $y_1, y_2 \gtrless 0$.

Якщо врахувати додаткові витрати в функції цілі у випадку допустимості понаднормових робіт, то функція цілі має вигляд:

$$Z = 4x_1 + 6x_2 - 5w_1 - 5w_2 \to max,$$

де невідомі w_1, w_2 мають такий зміст:

$$\begin{cases} \text{ЯКЩО } y_1 > 0 \text{ ТО } w_1 = y_1 \text{ ІНАКШЕ } w_1 = 0; \\ \text{ЯКЩО } y_2 > 0 \text{ ТО } w_2 = y_2 \text{ ІНАКШЕ } w_2 = 0. \end{cases}$$

Логічна функція ЯКЩО цілком припустима. Можна спростити модель:

$$Z = 4x_1 + 6x_2 - 5 \cdot \text{MAX}(0, y_1) - 5 \cdot \text{MAX}(0, y_2) \to max.$$

Функції MAX, MIN, ЯКЩО і багато інших – цілком припустимі і передбачені практично в будь-якому програмному забезпеченні. Запис функції цілі z за допомогою функцій MAX значно компактніший попереднього запису через функції ЯКЩО і повністю йому еквівалентний. Дійсно, якщо $y_i > 0$, то $w_i = \max(0, y_i) = y_i$; інакше (якщо $y_i \leq 0$) $w_i = \max(0, y_i) = 0$.

Функція MAX – нелінійна, отже, задача зведена до такої задачі нелінійного програмування:

Потрібно знайти найбільше значення нелінійної функції цілі:

$$Z = 4x_1 + 6x_2 - 5\cdot\max(0, y_1) - 5\cdot\max(0, y_2)$$

за обмежень:

$$\begin{cases} \dfrac{x_1}{8} + \dfrac{x_2}{4} - y_1 = 8, \\ \dfrac{x_1}{6} + \dfrac{x_2}{5} - y_2 = 8, \\ y_1 \le 4, \\ y_2 \le 4, \end{cases}$$

$$\text{де} \qquad x_1, x_2 \ge 0;$$

$$y_1, y_2 \gtrless 0.$$

Якщо врахувати, що функція $w_i = \max(0, y_i)$ еквівалентна системі *лінійних* нерівностей, то задача зводиться до задачі лінійного програмування:

$$\begin{cases} w_i \ge y_i; \\ w_i \ge 0, \\ w_i \to min. \end{cases}$$

Остаточно одержуємо задачу лінійного програмування у вигляді:

знайти максимальне значення лінійної функції цілі

$$Z = 4x_1 + 6x_2 - 5w_1 - 5w_2 \to max,$$

за лінійних обмежень:

$$\begin{cases} \dfrac{x_1}{8} + \dfrac{x_2}{4} - y_1 = 8 \\ \dfrac{x_1}{6} + \dfrac{x_2}{5} - y_2 = 8 \end{cases},$$

$$y_1 \le 4,\ y_2 \le 4,$$

$$y_1 - w_1 \le 0,\ y_2 - w_2 \le 0,$$

$$\text{де} \qquad x_1, x_2, w_1, w_2 \ge 0$$

$$y_1, y_2 \gtrless 0.$$

Умови $w_i \to min$ виконуються автоматично, оскільки у цільову функцію $Z \to max$ невідомі w_i входять з від'ємними коефіцієнтами і найбільшому значенню z відповідають найменші w_1, w_2.

Оптимальний розв'язок цієї задачі: $x_1 = 36$, $x_2 = 30$, $y_1 = y_2 = 4$, $z = 284$.

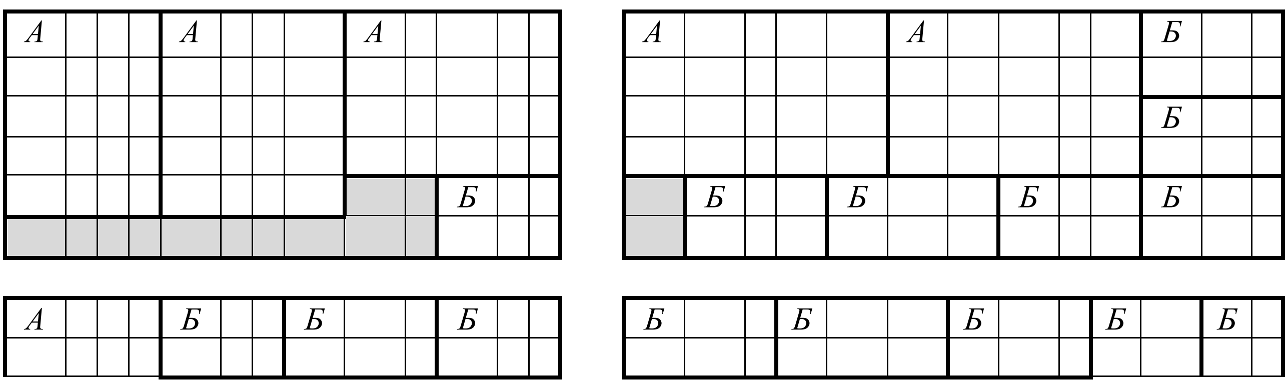
Завдяки понаднормовим роботам прибуток збільшився на 68 одиниць.

Задача про розкрій матеріалу

Згідно з технологією виготовлення багатьох видів продукції починається з розкрою листового матеріалу. Викроюють не тільки одяг і взуття, але також деталі корпусу корабля, кузову автомобіля, фюзеляжу літака. Розкроюють тканини, шкіру, скло, метал, пластмасу. Під час масового виробництва або виготовлення продукції з дефіцитного матеріалу проблема оптимізації розкрою стає досить актуальною.

Приклад. Зі стандартних листів дефіцитного матеріалу розміром 6×13 м^2 необхідно викроїти заготівки типу $А$ розміром 5×4 м^2 і типу $Б$ розміром 2×3 м^2. Для виготовлення готових виробів потрібно забезпечити умову комплектності заготівок – на одну заготівку $А$ повинно припадати 5 заготівок $Б$. Треба скласти оптимальний план розкрою стандартних листів матеріалу, що забезпечує одержання: максимальної кількості комплектів; мінімальної кількості відходів дефіцитного матеріалу.

Розв'язування задачі про розкрій починається з переліку всіх способів розміщення заготівок на одному листі матеріалу, причому рекомендується починати з розміщення найбільш великих заготівок. Виявляється, що в цій задачі є тільки 4 основні способи розкрою, які зображені на рис. 3.1. Тут крім явно невигідних варіантів були також відкинуті всі еквівалентні варіанти розміщення заготівок.



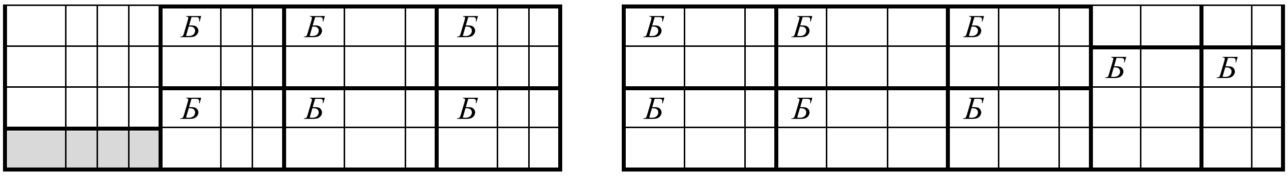


Рис. 3.1. **Основні варіанти розміщення заготівок**

Характеристики основних варіантів розкрою зводимо в табл. 3.8:

Таблиця 3.8

Характеристики способів розкрою

Способи розкрою	Заготівки		Обрізки, $м^2$	План розкрою
	А	*Б*		
I	3	1	12	x_1
II	2	6	2	x_2
III	1	9	4	x_3
IV	0	13	0	x_4
Усього	y_1	y_2		

Зміст плану розкрою x_1, x_2, x_3, x_4: скільки листів матеріалу треба розкроювати кожним способом. Крім цих невідомих корисно ввести позначення y_1, y_2 для загальної кількості заготівок *А* і *Б*, що будуть отримані при відомому плані розкрою.

В умовах задачі може бути задана загальна кількість листів матеріалу (N), у такому випадку всі x_j, y_i повинні бути цілими. Якщо ж загальну кількість листів матеріалу не задано, його приймають рівним 100 %, тоді план розкрою x_j буде також виражатися у відсотках – у якому співвідношенні слід розкроювати стандартні листи різними способами.

Складемо математичну модель задачі:

$$\begin{cases} N = x_1 + x_2 + x_3 + x_4 = 100, \\ y_1 = 3x_1 + 2x_2 + x_3 + 0x_4, \\ y_2 = x_1 + 6x_2 + 9x_3 + 13x_4, \\ y_2 = 5y_1, \end{cases}$$

$$Z = y_1 \to max$$

$$\text{усі } x_j, y_i \geq 0.$$

Маємо, що задача про розкрій матеріалу є також задачею ЛП.

У складеній моделі задано знайти найбільше число комплектів, що дорівнює загальному числу заготівок типу *А*. Якщо ж ціллю є мінімізація відходів, то треба врахувати, що у відходи слід включати не тільки обрізки матеріалу, площі яких наведені в таблиці варіантів розкрою (рис. 3.1), але також усі заготівки, що не складають повного комплекту.

3.2. Канонічна форма задачі лінійного програмування

У загальній задачі ЛП розшукується максимум або мінімум лінійної функції цілі за наявності системи лінійних обмежень у вигляді рівностей і нерівностей будь-якого типу ($\le$ або $\ge$). Невідомі можуть бути невід'ємними, недодатними, не мати ніяких обмежень на знак, мати двосторонні обмеження типу $d_j \le x_j \le g_j$.

Існує три стандартних форми завдання, першу і другу стандартні форми були розглянуті вище. Третя (основна) стандартна форма названа «канонічною».

У канонічній формі відшукується максимум лінійної функції цілі:

$$Z = \sum_{j=1}^{n} c_j x_j \to max \qquad (3.7)$$

за наявності системи лінійних обмежень-рівностей:

$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j = b_i, \\ (i = \overline{1, m}). \end{cases} \qquad (3.8)$$

Усі невідомі повинні бути невід'ємними: $x_j \ge 0$.

Будь-яку задачу ЛП завжди можна привести до канонічної форми:

а) пошук мінімуму функції f завжди можна замінити пошуком максимуму функції $Z = -f$, тому що $f_{min}(x^0) = -Z_{max}(x^0)$, де x^0 – оптимальне значення невідомої (див. рис. 3.2);

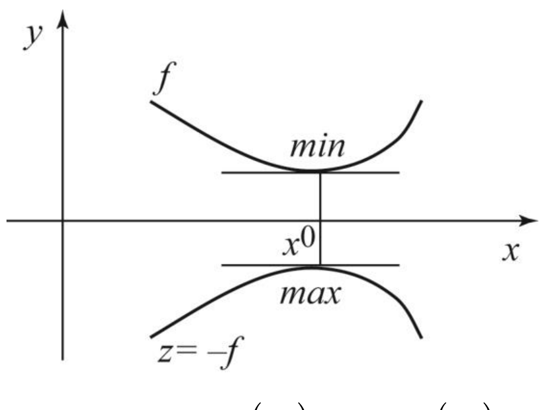


Рис. 3.2 $f_{\min}(x^0) = -Z_{\max}(x^0)$

б) обмеження-нерівності типу $\Sigma a_{ij}x_j \le b_i$ можна замінити на рівності $\Sigma a_{ij}x_j + x_{i+k} = b_i$, а нерівності $\Sigma a_{ij}x_j \ge b_i$ – на $\Sigma a_{ij}x_j - x_{i+k} = b_i$. В обох цих випадках уводяться додаткові невід'ємні балансові невідомі $x_{i+k} \ge 0$, де k – число вихідних невідомих;

в) як правило, всі невідомі – невід'ємні за своїм фізичним (економічним) сенсом. Проте необхідно передбачити наявність декількох невідомих без будь-якого обмеження на знак – такі невідомі часто з'являються у ході постановки задач нелінійного програмування. Невідомі без обмеження на знак (якщо такі є) завжди можна замінити різницею двох невід'ємних невідомих $x_j = x'_j - x''_j$, де $x'_j, x''_j \ge 0$. Цей прийом досить простий, але приводить до збільшення загального числа невідомих у канонічній формі задачі. Інший спосіб: можна виразити невідомі без обмеження на знак через інші невід'ємні невідомі і, таким чином, виключити їх з подальшого аналізу аж до закінчення процесу оптимізації. Розмір задачі при цьому не збільшується, а зменшується.

При однобічних обмеженнях типу $x_j \ge d_j$ уводимо нові невід'ємні невідомі $x'_j = x_j - d_j$. Обмеження типу $x_j \le g_j$ включаємо в загальне число лінійних обмежень, які перетворюються до виду рівностей шляхом введення додаткових балансових невідомих.

Крім невідомих без обмеження на знак іноді доводиться вводити фіктивні невідомі, які тотожно дорівнюють нулю $x_j \equiv 0$. Як буде показано далі, обмеження-рівності, фіктивні невідомі і невідомі без обмеження на знак тісно пов'язані між собою.

У канонічній формі загальне число обмежень-рівностей m, загальне число (невід'ємних) невідомих n, включаючи балансові.

Приклад. Привести до канонічної форми умови задачі про облік понаднормового часу. Вихідна модель має вигляд:

$$Z = 4x_1 + 6x_2 - 5w_1 - 5w_2 \to max,$$

$$\begin{cases} \dfrac{x_1}{8} + \dfrac{x_2}{4} - y_1 = 8, \\ \dfrac{x_1}{6} + \dfrac{x_2}{5} - y_2 = 8, \end{cases}$$

$$y_1 \le 4,\ y_2 \le 4,$$

$$y_1 - w_1 \le 0,\ y_2 - w_2 \le 0,$$
$$\text{де} \quad x_1, x_2, w_1, w_2 \ge 0,$$
$$y_1, y_2 \gtrless 0.$$

Для кожного обмеження-нерівності введемо невід'ємну балансову невідому як різницю між більшою і меншою частинами нерівності:

$$u_1 = 0 - y_1 + w_1; \qquad v_1 = 4 - y_1;$$
$$u_2 = 0 - y_2 + w_2; \qquad v_2 = 4 - y_2.$$

Невідомі y_1, y_2 не мають обмежень на знак, тому за допомогою останніх двох рівностей виражаємо їх через невід'ємні балансові невідомі

$$y_1 = 4 - v_1, \qquad y_2 = 4 - v_2$$

і виключаємо із всіх інших обмежень, а також з функції цілі:

$$Z = 4x_1 + 6x_2 - 5w_1 - 5w_2 \to max,$$
$$\begin{cases} \frac{x_1}{8} + \frac{x_2}{4} + v_1 = 12, \\ \frac{x_1}{6} + \frac{x_2}{5} + v_2 = 12, \\ w_1 - u_1 + v_1 = 4, \\ w_2 - u_2 + v_2 = 4. \end{cases}$$

Усі невідомі тепер – невід'ємні.

До речі, рівності в канонічній формі прийнято записувати так, щоб вільні члени обмежень-рівнянь були невід'ємними.

3.3. Графічний метод розв'язання задачі лінійного програмування

Графічний метод застосовується, в основному, для розв'язання задач двовимірного простору і ґрунтується на геометричній інтерпретації задач лінійного програмування. У разі тривимірного простору задача ускладнюється, а для чотиривимірного графічне розв'язування стає неможливим.

Нехай задача лінійного програмування задана в двовимірному просторі, тобто обмеження містять дві змінні. Знайти максимальне значення функції цілі:

$$Z = c_1 x_1 + c_2 x_2 \to \max, \tag{3.9}$$

за обмежень:

$$
\begin{cases}
a_{11}x_1 + a_{12}x_2 \le b_1, \\
a_{21}x_1 + a_{22}x_2 \le b_2, \\
\dots\dots\dots\dots\dots\dots \\
a_{m1}x_1 + a_{m2}x_2 \le b_m,
\end{cases} \tag{3.10}
$$

$$
x_1 \ge 0, \quad x_2 \ge 0. \tag{3.11}
$$

Кожна нерівність згідно з рівнянням прямої $a_{i1}x_1 + a_{i2}x_2 = b_i$ поділяє площину $0X_1X_2$ на дві півплощини і визначає ту півплощину, в якій виконується нерівність. Обмеження за знаком (3.3) також визначають півплощини, обмежені прямими: $x_1 = 0$ і $x_2 = 0$; таким чином, область визначення даної задачі розміщена в І-й чверті координатної площини. Якщо система обмежень сумісна, півплощини, де виконуються відповідні нерівності, перетинаються і утворюють багатокутник можливих розв'язків ЗЛП, тобто багатокутник планів (рис. 3.3).

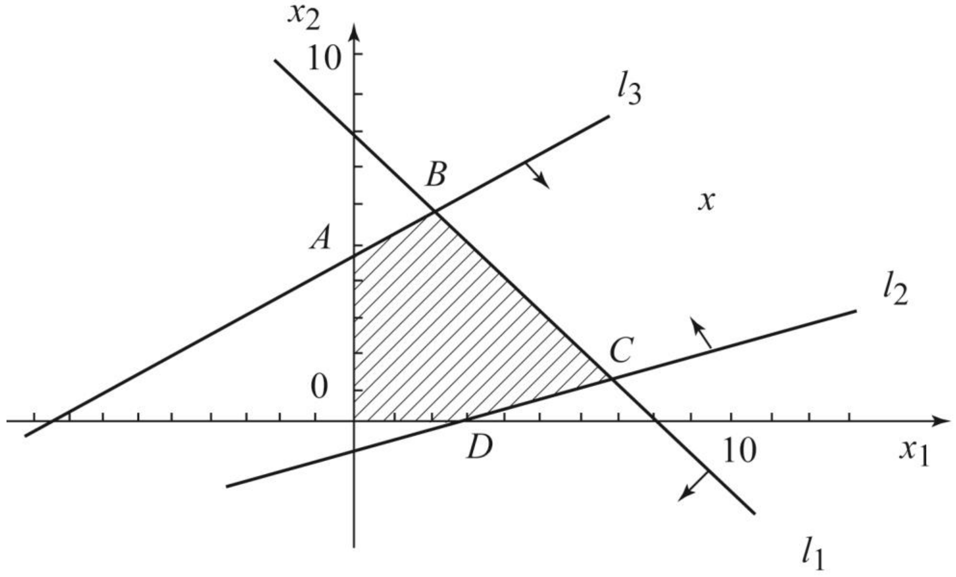


Рис. 3.3. **Багатокутник планів, який відповідає початковій системі**

Розглянемо графічну інтерпретацію цільової функції. Оскільки цільова функція є функцією двох змінних, на площині $0X_1X_2$ зображується лінією рівня, тобто лінією, відповідної сталому значенню цільової функції $Z = const$. Оскільки лінія рівня, яка відповідає нульовому значенню цільової функції, має рівняння: $c_1x_1 + c_2x_2 = 0$, її графіком є пряма, яка проходить через початок координат. Коефіцієнти цільової функції мають значення проекцій нормалі до лінії рівня $\boldsymbol{N} = (c_1; c_2)$. Для лінійної функції градієнт, тобто вектор, який вказує напрямок найшвидшого зростання функції, збігається з вектором нормалі: $grad\ Z = \frac{\partial z}{\partial x_1} i + \frac{\partial z}{\partial x_2} j = c_1 i + c_2 j = N$. Якщо значення цільової функції зростає, лінія рівня пересувається в напрямку нормалі (рис. 3.4).

Із геометричної інтерпретації задачі лінійного програмування слідує алгоритм її графічного розв'язання: 1) на основі системи обмежень будують область допустимих розв'язків; 2) будують вектор градієнтного напрямку $N = (c_1; c_2)$; 3) проводять довільну лінію рівня $Z = Z_0$ в напрямку вектора N так, щоб вона дотикалась області допустимих розв'язків в її крайньому положенні (крайній точці). У випадку розв'язування задачі на мінімум лінію рівня $Z = Z_0$ переміщують в антиградієнтному напряму; 4) визначають оптимальний план $X^* = \left(x_1^*, x_2^*\right)$ та екстремальне значення цільової функції $Z^* = Z\left(X^*\right)$.

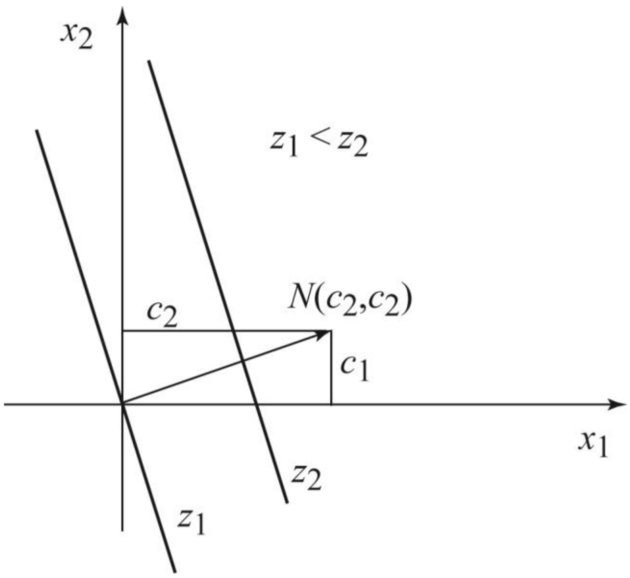


Рис. 3.4. **Побудова лінії рівня**

З графічного розв'язання слідують висновки: 1) система лінійних обмежень визначає опуклу багатокутну область планів; 2) оптимальне рішення ЗЛП досягається в одній з вершин багатокутної області планів. При цьому мають місце такі випадки:

1. Система обмежень може виявитися суперечливою, область планів – порожньою ($D = \varnothing$); у цьому випадку задача розв'язку не має (рис. 3.5).

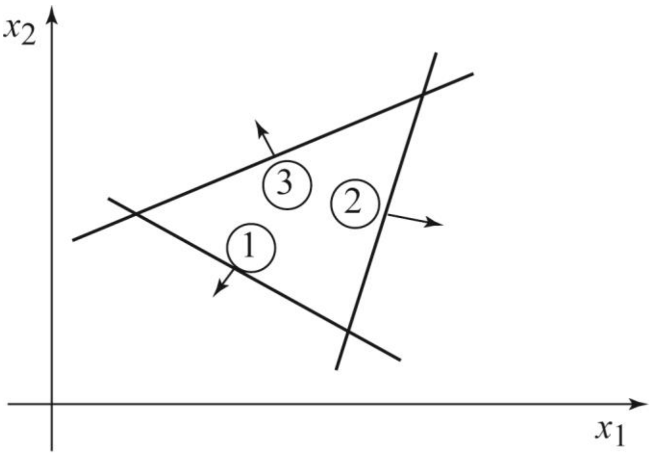


Рис. 3.5. **Немає розв’язків**

2. Область планів може бути відкритою, а максимальне значення функції цілі – необмеженим (рис. 3.6).

3. Як правило, оптимальний розв’язок – єдиний, але бувають випадки, коли є незліченна безліч еквівалентних оптимальних планів. На рис. 3.7 лінія рівня функції цілі (яка перпендикулярна градієнту) проходить відразу через дві вершини A і B багатокутної області планів. Оптимальними планами тут є всі точки відрізку [AB]. Рівняння відрізка можна записати як комбінацію його вершин: $X = (1 - \lambda)A + \lambda B$, де $0 \le \lambda \le 1$.

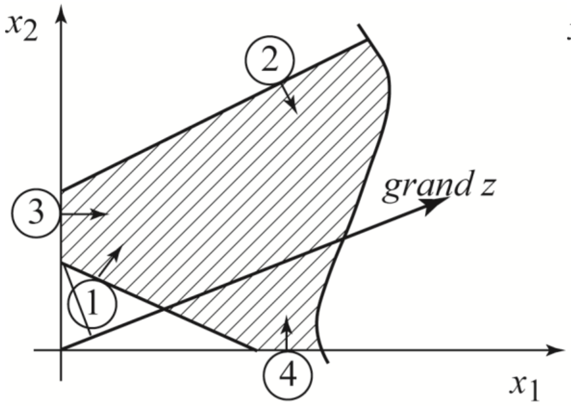


Рис. 3.6. **Необмежена область планів**

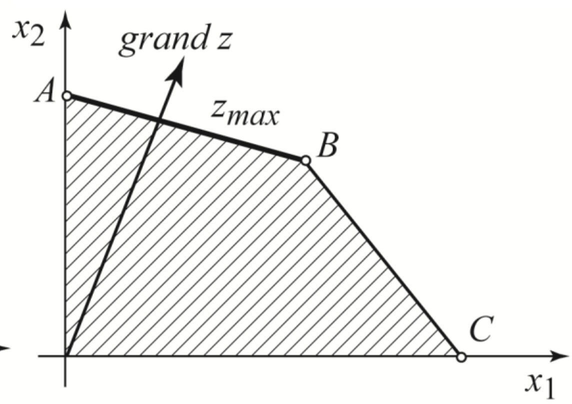
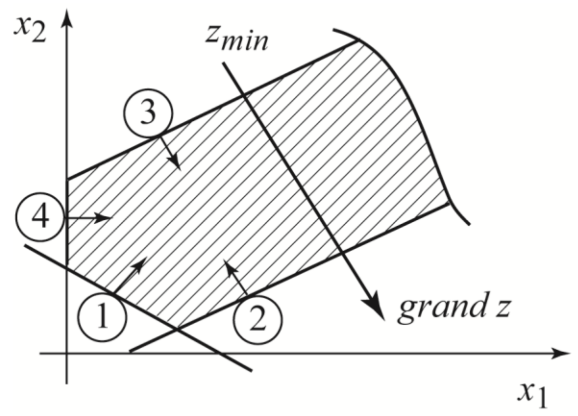


Рис. 3.7. **Неєдиність оптимального розв’язку**

При $\lambda = 0$ маємо першу вершину A; при $\lambda = 1$ маємо другу вершину B; при $\lambda = {}^1\!/_2$ маємо середину відрізка. Якщо максимальне значення функції цілі



досягається відразу в трьох вершинах *A*, *B*, *C*, то загальний оптимальний розв'язок (рівняння грані *ABC*) можна записати аналогічним чином: $X = \alpha A + \beta B + \gamma C$, де $\alpha + \beta + \gamma = 1$, $\alpha, \beta, \gamma \geq 0$. Знання альтернативних оптимальних розв'язків дозволяє вибрати серед них найбільш прийнятний план з урахуванням другорядних особливостей, які не відображені в моделі. Існують специфічні випадки, що зустрічаються значно рідше.

4. Область планів може бути відкритою, але граничні значення функції цілі – обмеженими (рис. 3.8).

Рис. 3.8. **Область планів відкрита, функція цілі – обмежена**

5. Область планів може виродитися в єдину точку. Це єдиний розв'язок за відсутністю вибору автоматично буде оптимальним (рис. 3.9).

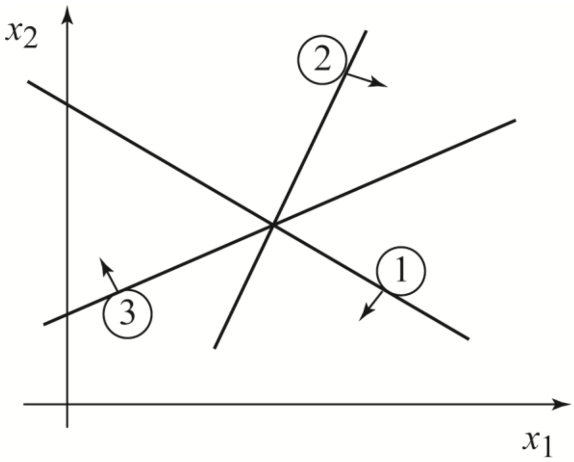


Рис. 3.9. **Єдиний розв'язок**

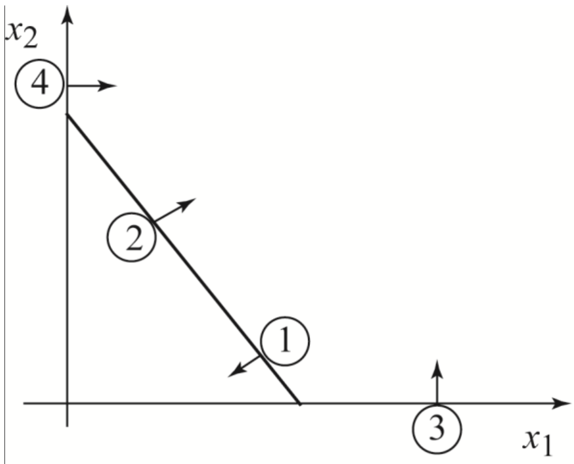


Рис. 3.10. **Область розв'язків – відрізок**

6. Область планів може становити лінію (відрізок). Цей випадок має місце, якщо серед обмежень є обмеження-рівність, що завжди можна замінити системою двох протилежних нерівностей (рис. 3.10).

7. Наведемо ще один, на вид малопомітний, але необхідний особливий випадок – випадок виродження. На рис. 3.11 обмеження 4 – явно надлишкове. Максимальне значення функції цілі тут досягається у вершині *B*.

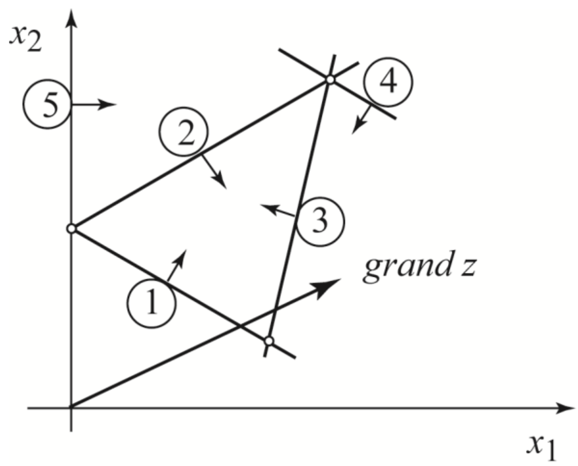


Рис. 3.11. **Виродження плану**

У двовимірному випадку вершина визначається перетинанням двох ліній, але в цій вершині перетинаються відразу три лінії. Тому у вершині B фактично злилися разом три вершини B_1, B_2 і B_3, кожна з яких визначається різними парами ліній. Як говорять, вершина B – кратна. Для графічного методу кратність вершин не має значення, але при аналітичному вирішенні це може привести до деяких утруднень. До речі, мінімальне значення функції цілі в цьому прикладі також виходить у кратній вершині, в утворенні якої беруть участь не тільки лінії 1 і 2, але також умова невід'ємності $x_1 \geq 0$.

Приклад. У задачі про оптимальне завантаження устаткування при відсутності понаднормових робіт наведений розв'язок. Одержимо цей розв'язок графічно. Математична модель задачі має вигляд:

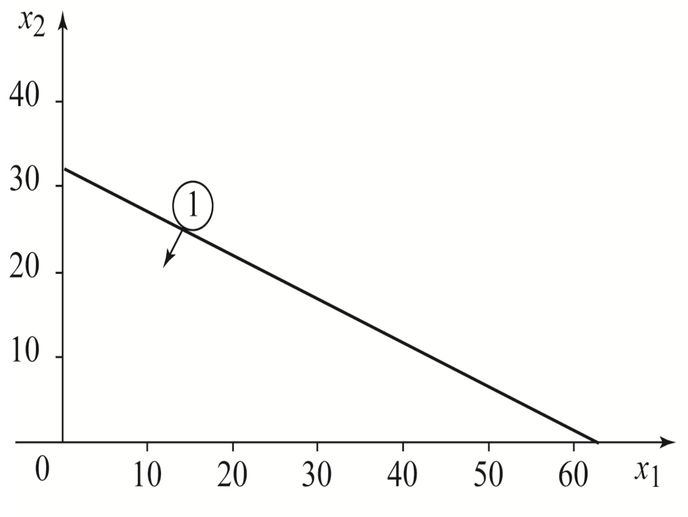
$$Z = 4x_1 + 6x_2 \rightarrow max$$

$$\begin{cases} \dfrac{x_1}{8} + \dfrac{x_2}{4} \leq 8 \quad , \; 1 \\ \dfrac{x_1}{6} + \dfrac{x_2}{5} \leq 8 \quad , \; 2 \end{cases}$$

$$x_1 \geq 0 \, , \; 3$$

$$x_2 \geq 0 \, . \; 4$$

Знайдемо графічно розв'язок 1-ї нерівності $x_1 + 2x_2 \leq 64$, що визначає на площині (x_1, x_2) деяку область. Границя цієї області (активне обмеження) $x_1 + 2x_2 = 64$ становить пряму лінію; будуємо її графік за двома точками, наприклад, точок перетину з осями координат (64; 0) і (0; 32). Ця лінія ділить площину на дві частини, в одній з яких дотримується зазначена нерівність. Досить перевірити виконання нерівності в довільній точці, що не лежить на границі (звичайно для цього вибирається початок координат). Оскільки для $x_1 = x_2 = 0$ нерівність задовольняється ($0 \leq 64$), то областю розв'язку 1-ї нерівності є частина площини убік початку координат (рис. 3.12).



Таким же чином знаходять розв'язки інших нерівностей і визначають загальну область, у якій справедливі всі обмеження. Для лінійних обмежень завжди виходить опукла багатокутна область *D*, що має назву «область планів».
Рис. 3.12. **Область розв’язання 1-ї нерівності**

На рис. 3.13 зображена багатокутна область планів *OACD* для нашої задачі.

Тепер будуємо сімейство ліній рівня функції цілі $Z = const$. На рис. 3.14 побудоване сімейство $Z_0 = 6x_2 + 4x_2$ для $z_0 = 0, 60, 120, 180, 240$.

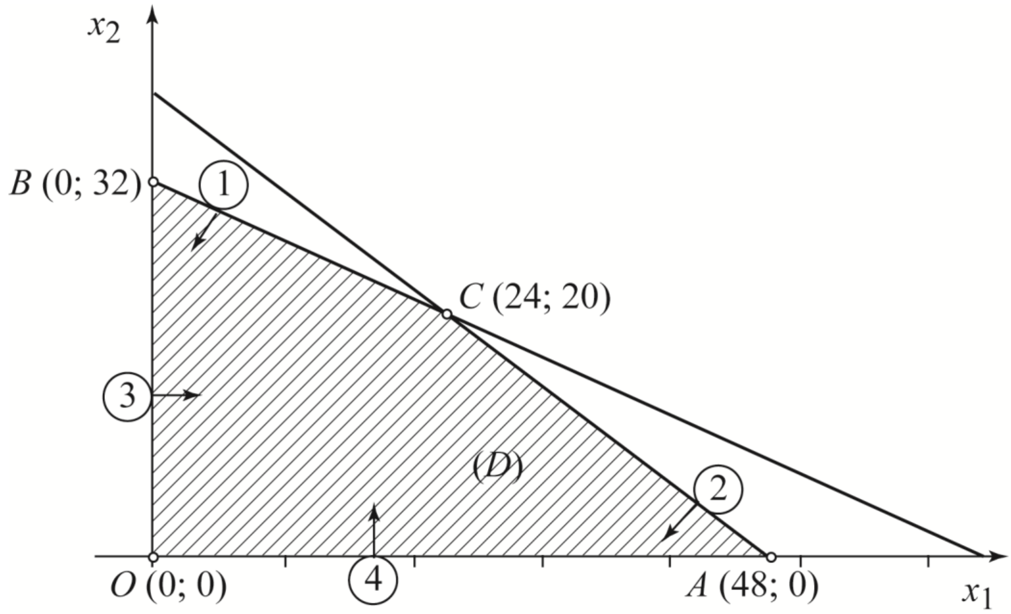
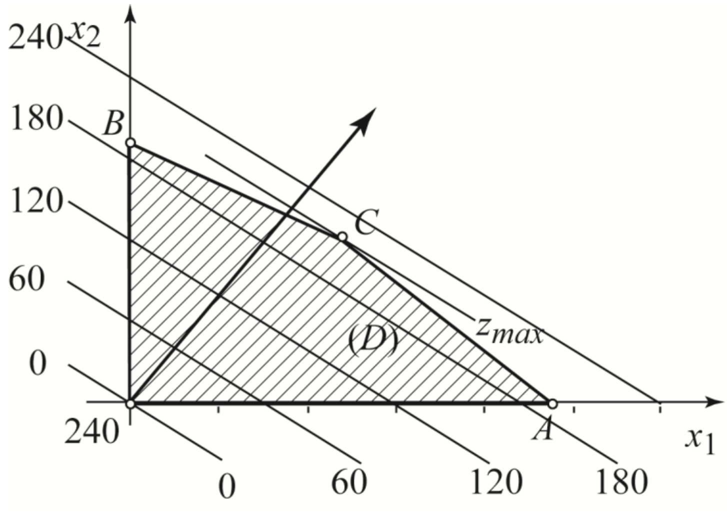


Рис. 3.13. **Область розв’язків системи нерівностей**

У будь-якій точці лінії рівня значення функції цілі залишається тим самим $Z = Z_0$. Якщо лінія рівня перетинає область планів, то є реальні плани з таким значенням функції цілі.



Відомо, що в будь-якій точці лінія рівня перпендикулярна вектору-градієнту $gradZ$, напрямок якого є напрямком найшвидшого зростання функції (це напрямок «крутого сходження»).

Рис. 3.14 **Сімейство ліній рівня $z = z_0$**

Проекціями градієнта є частинні похідні функції цілі за кожною невідомою. Для лінійної функції цілі $Z = c_1 x_1 + c_2 x_2$ ці похідні сталі і всі лінії рівня перпендикулярні одному вектору-градієнту $grad\, Z = (c_1, c_2)$. Для досягнення максимуму лінію рівня потрібно пересувати паралельно самій собі в напрямку градієнта доти, поки лінія рівня має хоча б одну загальну точку з областю планів. У нашому прикладі із градієнтом $grad\, Z = (4, 6)$ такою граничною точкою є вершина C.

У задачах ЛП використання градієнта спрощує побудову сімейства ліній рівня і проблему визначення оптимального плану – для цього досить побудувати вектор-градієнт і спроектувати на нього крайні точки області планів (звичайно проектують всі вершини багатокутної області). Точка області D з найбільшою проекцією на градієнт є оптимальним планом з максимальним значенням функції цілі, а точка з найменшою проекцією – оптимальним планом з мінімальним значенням функції цілі.

У даному прикладі мінімум досягається на початку координат, а максимум – у точці C – загальній точці активних обмежень ① і ②. Розв’язуємо систему обмежень-рівностей:

$$\begin{cases} \dfrac{x_1}{8} + \dfrac{x_2}{4} = 8 \\ \dfrac{x_1}{6} + \dfrac{x_2}{5} = 8 \end{cases}$$

і одержуємо оптимальний розв'язок: $x_1 = 24$; $x_2 = 20$; $Z_{max} = 4 \cdot 24 + 6 \cdot 20 = 216$.

Приклад. Розв’язати графічним методом задачу планування виробництва в’язальних верстатів на підприємстві. Математична модель цієї задачі має вигляд (рис.3.15):

знайти максимум лінійної функції цілі

$$Z = 700 x_1 + 900 x_2 \rightarrow max,$$

за обмежень

$$\begin{cases} 10x_1 + 9x_2 \le 1870, \\ 5x_1 + 11x_2 \le 1455, \\ 4x_1 + 15x_2 \le 1815, \\ x_1, x_2 \ge 0. \end{cases}$$

Побудуємо багатокутник розв'язків. Для цього в системі координат x_1Ox_2 на площині побудуємо граничні прямі:

$$10x_1 + 9x_2 = 1870, \quad (L_1)$$
$$5x_1 + 11x_2 = 1455, \quad (L_2)$$
$$4x_1 + 15x_2 = 1815. \quad (L_3)$$

Багатокутником розв'язків даної задачі є п'ятикутник *OABCD* (рис. 3.15).

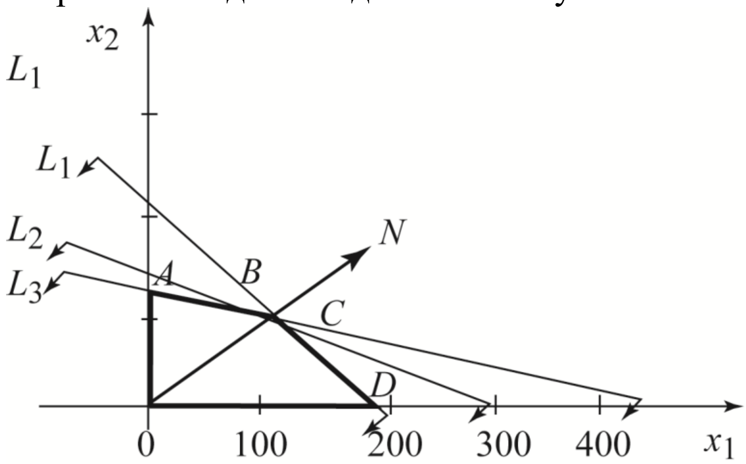


Рис. 3.15. **Область допустимих розв'язків задачі**

Для побудови прямої $700x_1 + 900x_2 = 0$ побудуємо вектор $\boldsymbol{N} = (700; 900)$ і через точку *O* проводимо пряму, яка перпендикулярна йому. Побудовану пряму $Z = 0$ переміщуємо паралельно самій собі за напрямом вектора *N*. Згідно з рис. 3.15 маємо, що опорною стосовно багатокутника розв'язків ця пряма стає в точці *C*, де функція Z приймає максимальне значення. Точка *C* є перетином прямих:

$$\begin{cases} 10x_1 + 9x_2 = 1870, \\ 5x_1 + 11x_2 = 1455. \end{cases}$$

Таким чином, оптимальний розв'язок $x_1 = 115, \quad x_2 = 80$. При цьому

$$Z_{\max} = 700 \cdot 115 + 900 \cdot 80 = 152\,500 \text{ грн.}$$

Отже, підприємство отримає максимальний прибуток у розмірі 152 500 грн, якщо буде виробляти в’язальних станків першого типу 115 штук, а в’язальних станків другого типу − 80 штук.

Зауваження. Аналогічну графічну інтерпретацію може мати задача лінійного програмування і для більшої кількості змінних, якщо різниця між кількістю змінних і кількістю рівнянь основної системи обмежень дорівнює двом: $n - m = 2$. Розглянемо цей випадок.

Нехай маємо цільову функцію $Z = c_1 x_1 + c_2 x_2 + ... + c_n x_n \to \min$

за обмежень:

$$\begin{cases} a_{11}x_1 + a_{12}x_1 + ... + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_1 + ... + a_{2n}x_n = b_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_1 + ... + a_{mn}x_n = b_m. \end{cases}$$

У даній системі обмежень всі рівняння лінійно незалежні і виконується умова $n - m = 2$. Виокремимо m перших невідомих в базис, а решта – дві останні незалежні x_{m+1} і x_n є вільними. Система обмежень має вигляд:

$$\begin{cases} x_1 + a'_{1\,m+1}x_{m+1} + a'_{1n}x_n = b'_1, \\ x_2 + a'_{2\,m+1}x_{m+1} + a'_{2n}x_n = b'_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ x_m + a'_{m\,m+1}x_{m+1} + a'_{mn}x_n = b'_m. \end{cases}$$

Використавши цю систему обмежень, виразимо цільову функцію тільки через вільні невідомі та враховуючи, що всі базисні невідомі невід’ємні ($x_j \geq 0, \quad j = \overline{1, m}$), відкинемо їх та перейдемо до системи нерівностей:

$$\begin{cases} a'_{1\,m+1}x_{m+1} + a'_{1n}x_n \leq b'_1, \\ a'_{2\,m+1}x_{m+1} + a'_{2n}x_n \leq b'_2, \\ \dots\dots\dots\dots\dots\dots \\ a'_{m\,m+1}x_{m+1} + a'_{mn}x_n \leq b'_m, \\ x_{m+1} \geq 0, x_n \geq 0. \end{cases}$$

При цьому цільова функція: $Z = c'_{m+1}x_{m+1} + c'_n x_n \to \min$.

Отримана задача має дві невідомі і її можна розв’язати графічним методом. Знайдене оптимальне значення x_{m+1} і x_n підставляємо в систему з m невідомими та знаходимо оптимальні значення $x_1, x_2, ..., x_m$.

3.4. Властивості можливих розв’язків задачі лінійного програмування

Розглянемо основну систему обмежень ЗЛП в канонічній формі (3.8):

$$\sum_{j=1}^{n} a_{ij} x_j = b_i \quad i = \overline{1,m} .$$

Постановка ЗЛП має сенс тільки коли система (3.8) має безліч розв’язків. У цьому випадку серед безлічі розв’язків необхідно вибрати той, за якого цільова функція досягає екстремуму. Згідно з теоремою Кронекера – Капеллі, система m рівнянь з n невідомими сумісна і має безліч розв’язків, якщо ранг матриці коефіцієнтів системи $\mathbf{A} = (a_{ij})_{m\,x\,n}$ дорівнює рангу розширеної матриці $A|B$ і менше кількості невідомих:

$$r_A = r_{A|B} = k$$
$$r_A = r_{A|B} < n$$

Невідомі, коефіцієнти при яких утворюють матрицю порядку k, і визначник якої не дорівнює нулю, є основними (або базисними). Решта $n - k$ невідомих є вільними. Якщо в загальному рішенні системи, де основні невідомі визначаються через вільні, вільним невідомим присвоїти значення рівне нулю, то такий розв’язок називається базисним. Серед безлічі розв’язків основної системи обмежень (3.4) кількість базисних розв’язків не перевищує кількості способів, якими можна відібрати m змінних серед n їх загальної кількості, тобто C_n^m – кількість сполучень з n по m. Якщо взяти до уваги обмеження на знак:

$$x_j \geq 0, \quad j = \overline{1,m},$$

то серед базисних розв’язків планами ЗЛП будуть тільки ті, для яких всі базисні змінні невід'ємні. Такі плани називаються опорними. Якщо всі k базисних компонентів є додатними, відповідний план називається невиродженим. В іншому випадку (тобто, коли кількість додатних компонентів опорного плану менше ніж k) план є виродженим.

Теорема. Безліч планів задачі лінійного програмування є опуклим.

Варто відмітити, що опуклою є множина, для якої виконується таке твердження: якщо точки M_1 і M_2 належать цій множині, то будь-яка внутрішня точка M відрізка M_1M_2 також належить цій множині. Тобто $M = \lambda M_1 + (1-\lambda)M_2$, $0 \le \lambda \le 1$, де M_1, M_2, M − радіус-вектори, що відповідають цим точкам. Кажуть, що вектор M є опуклою лінійною комбінацією векторів M_1 і M_2.

Доведення. Розглянемо задачу лінійного програмування в канонічній формі:

$$Z = C \cdot X \quad \rightarrow \quad \max$$

$$A \cdot X = B, \tag{3.12}$$

$$X \ge 0. \tag{3.13}$$

Нехай вектори X_1 і X_2 – можливі розв'язки задачі, тобто задовольняють основній системі обмежень (3.9):

$$A \cdot X_1 = B \quad i \quad A \cdot X_2 = B.$$

Необхідно розглянути вектор X який є лінійною комбінацією векторів X_1 і X_2:

$$X = \lambda X_1 + (1-\lambda)X_2, \quad 0 \le \lambda \le 1,$$

і підставити його в систему обмежень (3.9):

$$A \cdot X = A \cdot (\lambda X_1 + (1-\lambda)X_2) = \lambda \cdot AX_1 + AX_2 - \lambda \cdot AX_2 =$$
$$= \lambda B + B - \lambda B = B.$$

Тобто, вектор X також є розв'язком основної системи обмежень і належить до безлічі можливих розв'язків задачі лінійного програмування – що й потрібно було довести. ■

Якщо розглянути графічну інтерпретацію опуклої множини (рис. 3.12), то опуклим є п'ятикутник *ABCDE*, а п'ятикутник *FGHPQ* не є опуклим, оскільки не всі точки відрізка M_1M_2 належать цій множині.

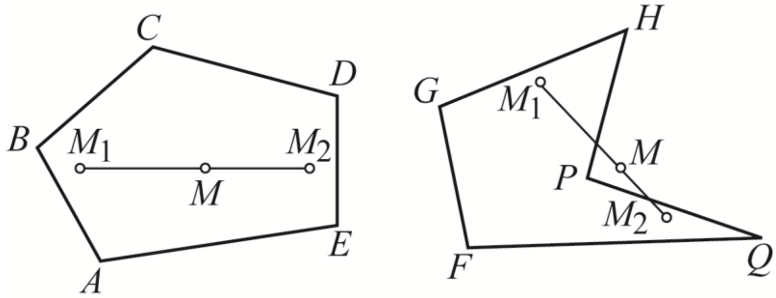


Рис. 3.16. **Опукла (*ABCDE*) і неопукла (*FGYPQ*) множини**

Згідно з теоремою 3.1 тільки п'ятикутник *ABCDE* може бути багатокутником планів задачі лінійного програмування.

Теорема. Опорний розв’язок задачі лінійного програмування відповідає допустимій екстремальній точці простору розв’язків (*без доведення*).

Іншими словами, кожному опорному плану задачі лінійного програмування відповідає вершина (кутова точка) багатокутника планів і, навпаки, кожній вершині багатокутника планів відповідає опорний план.

Запитання для самоперевірки

1. Які типові задачі лінійного програмування розрізняють?
2. Запишіть задачу лінійного програмування в канонічній формі.
3. Які задачі лінійного програмування можна розв’язати графічно?
4. Які основні етапи графічного методу розв'язання задач лінійного програмування?
5. Які випадки графічного розв'язання задачі лінійного програмування розрізняють?
6. Наведіть властивості можливих розв’язків задачі лінійного програмування.

4. Симплексний метод розв'язання задач лінійного програмування і деякі його теоретичні аспекти

4.1. Пошук оптимального плану. Умова оптимальності.

4.2. Алгоритм симплексного методу.

4.3. Метод штучного базису. Розширена *M*-задача.

4.4. Проблема виродження.

4.1. Пошук оптимального плану. Умова оптимальності

Розглянемо задачу лінійного програмування в канонічній формі:

$$Z = \sum_{j=1}^{n} c_j x_j \to \min ,$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \quad i = \overline{1,m}$$

$$x_j \geq 0, \quad j = \overline{1,n}.$$

Щоб задача мала розв'язок, система її обмежень повинна бути сумісною. Це можливо, якщо ранг r системи не більше числа невідомих n: $r \leq n$. Випадок $r > n$ неможливий. При $r = n$ система лінійних рівнянь має єдиний розв'язок. Якщо ж цільова функція досягає екстремального значення більш ніж в одній крайній точці, то вона досягає того ж значення в будь-якій точці, яка є їх опуклою лінійної комбінації. Для розв'язання задачі лінійного програмування (ЗЛП) необхідно знайти всі крайні точки багатогранника планів і порівняти в них значення цільової функції. Отже, для відшукання оптимального плану необхідно дослідити всі опорні плани, максимальна кількість яких дорівнює C_n^m. Перебирати всі опорні плани не доцільно, а слід скористатись упорядкованим переходом від одного опорного плану до іншого. Таку можливість надає симплексний метод, який дозволяє від відомого опорного плану за скінчене число ітерацій отримати оптимальний план. Загальна ідея симплекс-методу (методу послідовного покращення плану) полягає у 1) знаходженні початкового опорного плану; 2) наявності ознаки оптимальності опорного плану; 3) вміння переходити до не гіршого опорного плану.

Розглянемо теоретичні основи симплекс-методу. За допомогою методу Жордана−Гаусса виразимо перші невідомі у вигляді:

$$\begin{cases} x_1 + a_{1m+1}x_{m+1} + \ldots + a_{1n}x_n = b_1, \\ x_2 + a_{2m+1}x_{m+1} + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_m + a_{m\,m+1}x_{m+1} + \ldots + a_{m\,n}x_n = b_1. \end{cases} \tag{4.1}$$

Дану систему (4.1) можна записати у векторній формі:

$$x_1A_1 + x_2A_2 + \ldots + x_mA_m + x_{m+1}A_{m+1} + \ldots + x_nA_n = A_0, \tag{4.2}$$

де $$A_1 = \begin{pmatrix} 1 \\ 0 \\ \ldots \\ 0 \end{pmatrix}, \ A_2 = \begin{pmatrix} 0 \\ 1 \\ \ldots \\ 0 \end{pmatrix}, \ldots, A_m = \begin{pmatrix} 0 \\ 0 \\ \ldots \\ 1 \end{pmatrix},$$

$$A_{m+1}=\begin{pmatrix}a_{1\,m+1}\\a_{2\,m+1}\\\ldots\\a_{m\,m+1}\end{pmatrix},\ A_{m+2}=\begin{pmatrix}a_{1\,m+2}\\a_{2\,m+2}\\\ldots\\a_{m\,m+2}\end{pmatrix},\ \ldots,A_n=\begin{pmatrix}a_{1\,n}\\a_{2\,n}\\\ldots\\a_{m\,n}\end{pmatrix},\ A_0=\begin{pmatrix}b_1\\b_2\\\ldots\\b_m\end{pmatrix}.$$

Базис складають m векторів $A_1, A_2, ..., A_m$.

Маємо m базисних невідомих: x_1, x_2, …, x_m і їх потрібно виразити через вільні невідомі x_{m+1}, x_{m+2}, …, x_n:

$$x_i=b_i-\sum_{k=1}^{n-m}a_{ik}x_{m+k},\quad i=\overline{1,m}.$$

Отримаємо висхідний опорний план:

$$X_0=\left(b_1,b_2,...,b_m,\underbrace{0,0,...,0}_{n-m}\right),$$

або $X_0=\left(x_1=b_1,x_2=b_2,...,x_m=b_m,\underbrace{x_{m+1}=0,x_{m+2}=0,...,x_n=0}_{n-m}\right).$

Значення функції для висхідного опорного плану:

$$Z_0=Z(X_0)=c_1x_1+c_2x_2+...+c_mx_m.$$

Виразимо:

$$\begin{cases}A_{m+1}=a_{1\,m+1}A_1+a_{2\,m+1}A_2+...+a_{m\,m+1}A_m,\\A_{m+2}=a_{1\,m+2}A_1+a_{2\,m+2}A_2+...+a_{m\,m+2}A_m,\\\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\A_n=a_{1\,n}A_1+a_{2\,n}A_2+...+a_{m\,n}A_m,\end{cases}\tag{4.3}$$

де $A_1, A_2, ..., A_m$ – лінійно незалежні вектори. Для даного плану відповідає розклад:

$$A_0=x_1A_1+x_2A_2+...+x_mA_m.\tag{4.4}$$

Для зменшення значення цільової функції треба побудувати інший опорний план, тобто перейти до нового базису.

Кожний з n векторів співвідношення (4.2) розкладається за векторами базису єдиним чином:

$$A_j = \sum_{i=1}^{m} x_{ij} A_i \ .$$

Введемо в базис вектор A_{m+1}, який має хоча б один додатний коефіцієнт $x_{i\,m+1}$ в розкладанні

$$A_{m+1} = x_{1\,m+1}A_1 + x_{2\,m+1}A_2 + \ldots + x_{m\,m+1}A_m \,. \qquad (4.5)$$

Помножимо даний вираз на $\theta > 0$ та віднімемо від (4.4), отримаємо:

$$\left(x_1 - \theta x_{1\,m+1}\right)A_1 + \left(x_2 - \theta x_{2\,m+1}\right)A_2 + \ldots + \left(x_m - \theta x_{m\,m+1}\right)A_m + \theta A_{m+1} = A_0 \,. \qquad (4.6)$$

Вектор $X_1 = \left(x_1 - \theta x_{1\,m+1}; x_2 - \theta x_{2\,m+1}; \ldots; x_m - \theta x_{m\,m+1}; \theta; 0; \ldots; 0\right)$ є планом, якщо його компоненти невід'ємні. Оскільки $\theta > 0$, то всі компоненти вектора X_1, в які входять недодатні $x_{i\,m+1}$, невід'ємні. Тому будемо розглядати тільки компоненти, до яких входять додатні $x_{i\,m+1}$. Іншими словами слід визначити таке $\theta > 0$, при якому для всіх $x_{i\,m+1} > 0$:

$$x_i - \theta x_{i\,m+1} \geq 0. \qquad (4.7)$$

Звідки маємо $\dfrac{x_i}{x_{i\,m+1}} \geq \theta$. Вектор X_1 є планом задачі для будь-якого θ, що задовольняє умові $\min \dfrac{x_i}{x_{i\,m+1}} \geq \theta > 0$ при $x_{i\,m+1} > 0$.

Вираз (4.6) можна розглядати як розкладання вектора A_0 за новим базисом. Один з коефіцієнтів розкладання повинен дорівнювати нулю. Нехай $\theta = \theta_0 = \min\limits_i \dfrac{x_i}{x_{i\,m+1}} \geq \theta > 0$, тоді відповідна компонента, в якій досягається мінімум, дорівнює нулю. Якщо ця компонента є першою, тобто:

$$\theta = \theta_0 = \min_i \frac{x_i}{x_{i\,m+1}} = \frac{x_1}{x_{1\,m+1}},$$

то підставивши величину θ в (4.6) отримаємо:

$$\left(x_1 - \frac{x_1}{x_{1\,m+1}} x_{1\,m+1}\right)A_1 + \left(x_2 - \frac{x_1}{x_{1\,m+1}} x_{2\,m+1}\right)A_2 + \ldots + \left(x_m - \frac{x_1}{x_{1\,m+1}} x_{m\,m+1}\right)A_m +$$

$$+\frac{x_1}{x_{1\,m+1}} A_{m+1} = A_0,$$

і відповідно розкладання

$$A_0 = x_2' A_2 + x_3' A_3 + \ldots + x_m' A_m + x_{m+1}' A_{m+1},$$

якому відповідає новий опорний план:

$$X_1 = \left(0;\, x_2';\, x_3';\ldots;\, x_m';\, x_{m+1}';\, 0;\ldots;0\right),$$

де $x_i' = x_i - \theta_0 x_{i\,m+1}$, $x_{i\,m+1}' = \theta_0$, а система векторів $A_2, A_3, \ldots, A_m, A_{m+1}$ утворює новий базис. Число θ називають симплексним відношенням.

Отже, перехід до нового опорного плану здійснюється через виключення одного вектора і включення до базису іншого. Критерій, що використовується для визначення вектора, який має включатись до базису є одним із основних складових симплексного методу. Якщо вектор A_{m+1} має бути включеним до базису, а в його розкладі всі $x_{i\,m+1} \le 0$ і не має змоги вибрати таке $\theta > 0$, що виключало б один з векторів розкладу і план X_1 містить $m+1$ додатних компонент, то говорять, що система векторів $A_1, A_2, A_3, \ldots, A_m, A_{m+1}$ лінійно залежна і визначає внутрішню точку багатогранника розв'язків. У цьому випадку лінійна функція не може досягати мінімального значення і гіперплощина, що відповідає лінійній функції, не може стати опорною до багатогранника розв'язків і лінійна функція не обмежена на багатограннику розв'язків.

Умова оптимальності

Нехай задача лінійного програмування має плани і кожний її опорний план не вироджений. Для висхідного опорного плану X_0 маємо:

$$A_0 = x_1 A_1 + x_2 A_2 + \ldots + x_m A_m,$$

значення лінійної функції цілі дорівнює:

$$x_1 C_1 + x_2 C_2 + \ldots + x_m C_m = Z(X_0). \tag{4.8}$$

Розклад будь-якого вектора A_j за векторами даного базису $A_1, A_2, \ldots, A_m$, єдиний:

$$x_{1j}A_1 + x_{2j}A_2 + \ldots + x_{mj}A_m = A_j, \ \left(j = \overline{1,m}\right), \qquad (4.9)$$

тому розкладу вектора A_j в базисі відповідає і єдине значення лінійної функції:

$$x_{1j}C_1 + x_{2j}C_2 + \ldots + x_{mj}C_m = Z_j, \ \left(j = \overline{1,m}\right). \qquad (4.10)$$

Нехай C_j − коефіцієнт лінійної функції, який відповідає вектору A_j.

Теорема. Якщо для деякого вектора A_j виконується умова $Z_j - C_j > 0$, то план X_0 не є оптимальним, і можна побудувати такий план X, для якого виконується нерівність $Z(X) < Z(X_0)$.

Доведення. Помножимо вираз (4.9) і вираз (4.10) на додатне число $\theta > 0$ і віднімемо відповідно з:

$$A_0 = x_1A_1 + x_2A_2 + \ldots + x_mA_m \text{ та } x_1C_1 + x_2C_2 + \ldots + x_mC_m = Z(X_0)$$

відповідно, отримаємо

$$(x_1 - \theta x_{1j})A_1 + (x_2 - \theta x_{2j})A_2 + \ldots + (x_m - \theta x_{mj})A_m + \theta A_j = A_0,$$

$$(x_1 - \theta x_{1j})C_1 + (x_2 - \theta x_{2j})C_2 + \ldots + (x_m - \theta x_{mj})C_m + \theta C_j = Z(X_0) - \theta(Z_j - C_j).$$

Оскільки всі $x_1, x_2, \ldots, x_m$ додатні, то завжди можна вибрати таке число $\theta > 0$, щоб всі коефіцієнти при векторах $A_1, A_2, \ldots, A_m, A_j$ були невід'ємними, а отже отримати новий план задачі:

$$X = (x_1 - \theta x_{1j}; x_2 - \theta x_{2j}; \ldots; x_m - \theta x_{mj}; \theta; 0; \ldots; 0),$$

якому відповідає значення лінійної функції $Z(X) = Z(X_0) - \theta(Z_j - C_j)$. Оскільки за умовою теореми $Z_j - C_j > 0$ і $\theta > 0$, то $Z(X) < Z(X_0)$. ■

Наслідок. Якщо для деякого плану X_0 розклад всіх векторів A_j в даному базисі задовольняє умові $Z_j - C_j \le 0$, то план X_0 є оптимальним.

Нерівність $Z_j - C_j \le 0$ є ***умовою оптимальності оптимізаційної задачі***, в якій цільова функція відшукується на мінімум. Величину $Z_j - C_j$ називають ***оцінками плану***.

Для задачі лінійного програмування:

$$Z = \sum_{j=1}^{n} c_j x_j \to \max,$$

$$\begin{cases} x_1 + a_{1m+1}x_{m+1} + \ldots + a_{1n}x_n = b_1, \\ x_2 + a_{2m+1}x_{m+1} + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_m + a_{mm+1}x_{m+1} + \ldots + a_{mn}x_n = b_m, \end{cases}$$

$$x_j \ge 0, \ \ j = \overline{1,n}.$$

Справедлива така теорема:

Теорема. Якщо для деякого вектора A_j виконується умова $Z_j - C_j < 0$, то план X_0 не є оптимальним, і можна побудувати такий план X, для якого виконується нерівність $Z(X) > Z(X_0)$.

Теорема без доведення.

Наслідок. Якщо для деякого плану X_0 розклад всіх векторів A_j в даному базисі задовольняє умові $Z_j - C_j \ge 0$, то план X_0 є оптимальним.

4.2. Алгоритм симплексного методу

Розглянемо задачу лінійного програмування:

$$Z = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \min,$$

$$\begin{cases} x_1 + a_{1m+1}x_{m+1} + \ldots + a_{1n}x_n = b_1, \\ x_2 + a_{2m+1}x_{m+1} + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_m + a_{mm+1}x_{m+1} + \ldots + a_{mn}x_n = b_m, \end{cases}$$

$$x_j \ge 0, \quad j = \overline{1,n}.$$

Нехай задача має опорний план:

$$X_0 = \left(x_1 = b_1, x_2 = b_2, \ldots, x_m = b_m, \underbrace{x_{m+1} = 0, x_{m+2} = 0, \ldots, x_n = 0}_{n-m} \right),$$

який визначається системою m-вимірних одиничних векторів $A_1, A_2, \ldots, A_m$. Для визначення оптимальності даного опорного плану необхідно вектори A_j системи розкласти за векторами базису $A_1, A_2, \ldots, A_m$ та обчислити $Z_j - C_j$. Оскільки базис утворюють одиничні вектори, тому коефіцієнтами розкладу вектора A_j за базисом є його компонентами $x_{ij} = a_{ij}$ $\left(i = \overline{1,m},\ j = \overline{1,n}\right)$. Подальші обчислення доцільно проводити в симплексній таблиці (табл. 4.1), де коефіцієнти цільової функції, які відповідають базисним змінним, записують в колонці $C_б$, висхідний опорний план X_0, проміжні опорні плани та оптимальний план знаходимо в колонці A_0, коефіцієнти розкладу вектора A_j за базисом записуємо в колонці A_j.

Таблиця 4.1

i	Базис	$C_б$	C	c_1	c_2	$\dots$	c_l	$\dots$	c_m	c_{m+1}	$\dots$	c_j	$\dots$	c_k	$\dots$	c_n
			A_0	A_1	A_2	$\dots$	A_l	$\dots$	A_m	A_{m+1}	$\dots$	A_j	$\dots$	A_k	$\dots$	A_n
1	A_1	c_1	x_1	1	0	$\dots$	0	$\dots$	0	x_{1m+1}	$\dots$	x_{1j}	$\dots$	x_{1k}	$\dots$	x_{1n}
2	A_2	c_2	x_2	0	1	$\dots$	0	$\dots$	0	x_{2m+1}	$\dots$	x_{2j}	$\dots$	x_{2k}	$\dots$	x_{2n}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$		$\dots$		$\dots$		$\dots$	
l	A_l	c_l	x_l	0	0	$\dots$	1	$\dots$	0	$x_{l\,m+1}$	$\dots$	$x_{l\,j}$	$\dots$	$x_{l\,k}$	$\dots$	$x_{l\,n}$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$		$\dots$		$\dots$		$\dots$	
m	A_m	c_m	x_m	0	0	$\dots$	0	$\dots$	1	$x_{m\,m+1}$	$\dots$	$x_{m\,j}$	$\dots$	$x_{m\,k}$	$\dots$	$x_{m\,n}$
$m+1$	$\Delta_j = Z_j -$ $-C_j$		Z_0	0	0	$\dots$	0	$\dots$	0	$Z_{m+1} - C_{m+1}$		$Z_j - C_j$		$Z_k - C_k$	$\dots$	$Z_n - C_n$

В $(m+1)$ рядку в колонці A_0 записуємо значення лінійної функції $Z(X_0)$, а в колонках A_j записуємо значення оцінок $\Delta_j = Z_j - C_j$. Значення $Z(X_0)$ знаходимо як $Z(X_0) = C_б X_0 = \sum_{i=1}^{m} c_i x_i$, $Z_j = C_б X_j = \sum_{i=1}^{m} c_i x_{ij}$. Аналізуючи значення $\Delta_j = Z_j - C_j$ маємо: якщо $Z_j - C_j \le 0$, то опорний план X_0 оптимальний і мінімальне значення лінійної функції цілі дорівнює $Z(X_0)$. В противному випадку, тобто якщо $Z_j - C_j > 0$, то план X_0 не є оптимальним, можна знайти менше значення лінійної функції. Якщо додатних оцінок $\Delta_j = Z_j - C_j$ декілька, то в базис має бути включений вектор, якому відповідає $\max(\theta_{0j}(Z_j - C_j))$, за умови $Z_j - C_j > 0$ і θ_{0j} обчислюється для кожного j. Якщо значень $\max(\theta_{0j}(Z_j - C_j))$ декілька, то до базису включається той вектор, у якого $\min C_j$.

Якщо хоча б для однієї $Z_j - C_j > 0$ коефіцієнти розкладу x_{ij} відповідного вектора недодатні, то лінійна функція не обмежена в багатограннику розв'язків або багатогранник розв'язків в даному випадку є необмеженою багатогранною областю.

Нехай $\max\left(\theta_{0j}\left(Z_j - C_j\right)\right) = \theta_{0k}\left(Z_k - C_k\right)$, де $m < k \le n$. Тоді в базис включається вектор A_k і виключається вектор, якому відповідає $\theta_{0k} = \min\left\{\dfrac{x_i}{x_{ik}}\right\} \left(x_{ik} > 0\right)$.

Нехай $\theta_{0k} = \min\left\{\dfrac{x_i}{x_{ik}}\right\} = \dfrac{x_l}{x_{lk}}$, отже вектор A_l виключається з базису.

Елемент x_{lk} називають ***направляючим***, а відповідний рядок і стовпець ***направляючими***. Тепер новий базис утворюють вектори $A_1, \ldots, A_{l-1}, A_k, A_{l+1} \ldots, A_m$. Новий опорний план $X_0 = \left(x'_1, x'_2, \ldots, x'_k, x'_m\right)$ обчислюється за формулами:

$$x'_i = x_i - \frac{x_l}{x_{lk}} x_{ik} \quad \left(i \ne l\right),$$

$$x'_k = \frac{x_l}{x_{lk}} \quad \left(i = l\right).$$

Розкладання векторів у новому базисі визначається за формулами:

$$x'_{ij} = x_{ij} - \frac{x_{lj}}{x_{lk}} x_{ik} \quad \left(i \ne l\right),$$

$$x'_{lj} = \frac{x_{lj}}{x_{lk}} \quad \left(i = l\right).$$

Таблиця 4.2

i	Базис	$C_б$	C	c_1	c_2	$\cdots$	c_l	$\cdots$	c_m	c_{m+1}	$\cdots$	c_j	$\cdots$	c_k	$\cdots$	c_n
			A_0	A_1	A_2	$\cdots$	A_l	$\cdots$	A_m	A_{m+1}	$\cdots$	A_j	$\cdots$	A_k	$\cdots$	A_n
1	A_1	c_1	x'_1	1	0	$\cdots$	x'_{1l}	$\cdots$	0	x'_{1m+1}	$\cdots$	x'_{1j}	$\cdots$	0	$\cdots$	x'_{1n}
2	A_2	c_2	x'_2	0	1	$\cdots$	x'_{2l}	$\cdots$	0	x'_{2m+1}	$\cdots$	x'_{2j}	$\cdots$	0	$\cdots$	x'_{2n}
…	…	…	…	…	…	…		…	…		…		…		…	
l	A_k	c_l	x'_k	0	0	$\cdots$	x'_{kl}	$\cdots$	0	$x'_{k\,m+1}$	$\cdots$	$x'_{k\,j}$	$\cdots$	1	$\cdots$	$x'_{k\,n}$
…	…	…	…	…	…	…		…	…		…		…		…	
m	A_m	c_m	x'_m	0	0	$\cdots$	$x'_{m\,l}$	$\cdots$	1	$x'_{m\,m+1}$	$\cdots$	$x'_{m\,j}$	$\cdots$	0	$\cdots$	x'_{mn}
$m+1$	$\Delta_j = Z_j - C_j$		Z'_0	0	0	$\cdots$	$Z'_l - C_l$		0	$Z'_{m+1} - C_{m+1}$		$Z'_j - C_j$		0	$\cdots$	$Z'_n - C_n$

Дані формули є формулами повного виключення Жордана−Гаусса. Всі обчислення за наведеними формулами доцільно проводити в симплексній таблиці (табл. 4.2).

Якщо в табл. 4.2 в $(m+1)$ рядку всі оцінки $Z_j - C_j \le 0$, то отриманий план X_0є оптимальним. У випадку наявності додатних оцінок відшукуємо наступний опорний план. Процес продовжується до отримання оптимального плану або встановлення необмеженості лінійної функції.

Якщо оцінками оптимального плану є тільки нульові оцінки, що відповідають базисним векторам, то це підтверджує єдиність оптимального плану. Якщо нульова оцінка відповідає вектору, що не входить до базису, то оптимальний план не єдиний. У даному випадку задача лінійного програмування має нескінченну множину оптимальних планів.

Для задачі лінійного програмування, в якої відшукується максимальне значення в разі невиконання умов оптимальності $Z_j - C_j \ge 0$ в базис включається вектор, якому відповідає $\min\left(\theta_{0j}\left(Z_j - C_j\right)\right)$, за умови $Z_j - C_j < 0$. Якщо мінімальних оцінок декілька, то в базис включається той вектор, якому відповідає $\max C_j$. Решта всіх дій і обчислень аналогічно задачі лінійному програмуванню, в якій мінімізується цільова функція.

Приклад. На підприємстві можна організувати виробництво продукції двома способами, при цьому необхідно мати три види ресурсів: сировину, обладнання і електроенергію. Витрати ресурсів за один місяць і наявний загальний їх обсяг представлено в табл. 4.3.

Таблиця 4.3

Норми витрат ресурсів

Ресурси	Витрати ресурсів за 1 місяць при організації виробництва		Запаси ресурсів
	1-м способом	2-м способом	
Сировина	1	2	4
Обладнання	1	1	3
Електроенергія	2	1	8

При першому способі виробництва підприємство випускає за один місяць 3 тисячі виробів, при другому – 4 тисячі виробів. Скільки місяців має пропрацювати підприємство кожним із способів, щоб за наявності ресурсів забезпечити максимальний випуск продукції?

Складемо математичну модель задачі. Позначимо x_1- період роботи підприємства при організації виробництва продукції першим способом, x_2- період роботи підприємства при організації виробництва продукції другим способом. Математична модель має вигляд:

$$Z = 3x_1 + 4x_2 \to \max,$$

за обмежень:

$$\begin{cases} x_1 + 2x_2 \le 4, \\ x_1 + x_2 \le 3, \\ 2x_1 + x_2 \le 8, \\ x_1 \ge 0,\ x_2 \ge 0. \end{cases}$$

Приведемо задачу до канонічного вигляду:

$$Z = 3x_1 + 4x_2 \to \max,$$

$$\begin{cases} x_1 + 2x_2 + x_3 = 4, \\ x_1 + x_2 + x_4 = 3, \\ 2x_1 + x_2 + x_5 = 8, \\ x_j \ge 0,\ j = \overline{1,5}. \end{cases}$$

Розв'яжемо задачу симплексним методом у симплексній таблиці (табл. 4.4).

Таблиця 4.4

i	Базис	$C_б$	C	3	4	0	0	0
			A_0	A_1	A_2	A_3	A_4	A_5
1	A_3	0	4	1	2	1	0	0
2	A_4	0	3	1	1	0	1	0
3	A_5	0	8	2	1	0	0	1
4		Δ	0	-3	-4	0	0	0

Маємо початковий опорний план: $X_0 = (0, 0, 4, 3, 8)$. Значення цільової функції дорівнює $Z(X_0) = 0$. Опорний план не оптимальний, оскільки в індексному рядку є дві оцінки: Δ_1 та Δ_2 від'ємні. Оскільки

$min\left(\theta_{0j}\left(Z_j - C_j\right)\right) = (-3)\cdot\frac{3}{1} = -9$, то $x_{21} = 1$ є направляючим, і, отже, вектор A_4 слід вивести з базису, а вектор A_1 ввести в базис. Перетворення зробимо в симплексній таблиці за методом Жордана−Гаусса (табл. 4.5)

Таблиця 4.5

i	Базис	$C_б$	C	3	4	0	0	0
			A_0	A_1	A_2	A_3	A_4	A_5
1	A_3	0	1	0	1	1	-1	0
2	A_1	3	3	1	1	0	1	0
3	A_5	0	2	0	-1	0	-2	1
4		Δ	9	0	-1	0	3	0

Маємо новий опорний план: $X_0 = (3,0,1,0,2)$. Значення цільової функції дорівнює $Z(X_0) = 9$. Опорний план не є оптимальний, оскільки в індексному рядку є оцінка Δ_2 від'ємна. Згідно з $\theta_{02} = min\left\{\frac{1}{1},\frac{3}{1}\right\} = 1$, то $x_{12} = 1$ є направляючим, і, отже, вектор A_3 слідує вивести з базису, а вектор A_2 ввести в базис. Перетворення зробимо в симплексній таблиці за методом Жордана-Гаусса (табл. 4.6)

Таблиця 4.6

i	Базис	$C_б$	C	3	4	0	0	0
			A_0	A_1	A_2	A_3	A_4	A_5
1	A_2	4	1	0	1	1	-1	0
2	A_1	3	2	1	0	-1	2	0
3	A_5	0	3	0	0	1	-3	1
4		Δ	10	0	0	1	2	0

Маємо опорний план: $X_0 = (3,1,0,0,3)$. Він є оптимальний, оскільки в індексному рядку всі оцінки Δ_0 невід'ємні. Максимальне значення цільової функції дорівнює $Z_{max}(X_0)=10$. Отже, для забезпечення максимального випуску продукції у розмірі 10 тисяч одиниць підприємство має випускати продукцію першим способом два місяці, а другим способом – один місяць.

4.3. Метод штучного базису. Розширена *М*-задача

Якщо обмеження задачі лінійного програмування містять одиничну матрицю порядку m, то за умови невід’ємних значень вільних членів рівнянь початковий план визначений та за допомогою симплексного методу знаходять оптимальний план.

Якщо систему обмежень задачі лінійного програмування можна перетворити до вигляду $AX \le A_0, A_0 \ge 0$, то вона завжди містить одиничну матрицю. Але є багато задач лінійного програмування, які не містять одиничної матриці і не перетворюються до вказаного вигляду, проте мають розв’язки.

В такому випадку використовується метод штучного базису.

Нехай потрібно дослідити функцію на мінімум $Z=\sum_{j=1}^{n} c_j x_j \to min$, за умови:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots a_{mn}x_n = b_m, \end{cases}$$

$$x_j \ge 0, \quad j=\overline{1,n}.$$

Система обмежень не містить одиничної матриці. Тому для отримання одиничної матриці добавимо до кожної рівності по одній змінній $x_{n+i} \ge 0,\ i=\overline{1,m}$, які називають штучними і розв’язують розширену задачу:

$$Z = c_1x_1 + c_2x_2 + \ldots + c_nx_n + Mx_{n+1} + \ldots + Mx_{n+m} \to min,$$

за обмежень:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots a_{1n}x_n + x_{n+1} = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots a_{2n}x_n + x_{n+2} = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots a_{mn}x_n + x_{n+m} = b_m, \end{cases}$$

$$x_j \ge 0, \quad j = \overline{1, n+m}.$$

Величину M беруть достатньо велике додатне число, якщо цільова функція в задачі досліджується на мінімум, і достатньо мале від'ємне число, якщо цільова функція в задачі досліджується на максимум. Вектори $A_{n+1}, A_{n+2}, \ldots, A_{n+m}$ утворюють штучний базис.

Для відшукання оптимального плану даної задачі використовують таку теорему.

Теорема. Якщо в оптимальному плані $X_0 = (x_1, x_2, \ldots, x_n, 0, \ldots, 0)$ розширеної задачі штучні змінні $x_{n+i} = 0, \quad i = \overline{1, n}$, то план $X = (x_1, x_2, \ldots, x_n)$ є оптимальним планом початкової задачі.

Доведення. Якщо план X_0 є оптимальним планом розширеної задачі, то план X є планом висхідної задачі, при цьому $Z(X_0) = Z(X)$, оскільки останні m компонент плану X_0 дорівнюють нулю.

Доведення того, що план X є оптимальним планом висхідної задачі виконаємо від супротивного. Тобто нехай X не є оптимальним планом. Тоді існує такий оптимальний план $X^* = (x_1^*, x_2^*, \ldots, x_n^*)$, для якого $Z(X^*) < Z(X)$. Звідси для вектора $X_0^* = (x_1^*, x_2^*, \ldots, x_n^*, 0, \ldots, 0)$, який є планом розширеної задачі, маємо:

$$Z(X_0^*) = Z(X^*) < Z(X) = Z(X_0),$$

тобто $Z(X_0^*) < Z(X_0)$. Таким чином, план X_0 розширеної задачі не є оптимальним, що суперечить умові теореми. ■

Використовуючи симплексний метод для розв'язування розширеної задачі, отримуємо план, в якому всі штучні змінні дорівнюють нулю. Якщо початкова задача не має планів, тобто вона не сумісна, то оптимальний розв'язок розширеної задачі має хоча б одну $x_{n+i} > 0$. Висхідна задача буде мати оптимальний розв'язок, якщо вдалось виключити з базису всі штучні змінні.

Таким чином, якщо маємо задачу зі змішаними обмеженнями, наприклад:

$$Z = \sum_{j=1}^{n} c_j x_j \to extr: \max \text{ або } \min$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n \le b_k, \\ a_{k+11}x_1 + a_{k+12}x_2 + \ldots + a_{k+1n}x_n \ge b_{k+1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{l1}x_1 + a_{l2}x_2 + \ldots + a_{ln}x_n \ge b_l, \\ a_{l+11}x_1 + a_{l+12}x_2 + \ldots + a_{l+1n}x_n = b_{l+1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m, \\ x_j \ge 0, \quad j = \overline{1,n}. \end{cases}$$

додамо в перші k нерівності $x_{n+i}, \ldots, x_{n+k}$, $A_{n+1}, \ldots, A_{n+k}$ складуть базис

$$Z = c_1x_1 + c_2x_2 + c_nx_n + 0 \cdot x_{n+1} + \ldots + 0 \cdot x_{n+l} \quad \to \quad extr$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n + x_{n+1} = b_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{k1}x_1 + a_{k2}x_2 + \ldots + a_{kn}x_n + x_{n+k} = b_k, \\ a_{k+11}x_1 + a_{k+12}x_2 + \ldots + a_{k+1n}x_n - x_{n+k+1} = b_{k+1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{l1}x_1 + a_{l2}x_2 + \ldots + a_{ln}x_n - x_{n+l} = b_l, \\ a_{l+11}x_1 + a_{l+12}x_2 + \ldots + a_{l+1n}x_n = b_{l+1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m, \\ x_j \ge 0, \quad j = \overline{1,n+l}. \end{cases}$$

Вектори $A_{n+1}, \ldots, A_{n+l}$ не утворюють базис.

Необхідно використовувати метод штучного базису.

Приклад. Відомо, що за агротехнічними нормами в 1 га чорнозему треба вносити протягом сезону основні корисні речовини в кількостях не менше 6 одиниць фосфору, 8 одиниць азоту, 12 одиниць калію. Фермерське господарство має можливість придбати добрива виду A_1, A_2 і A_3. В 1 кг добрив залежно від складу міститься основних корисних речовин в кількостях:

Добрива	Фосфор	Азот	Калій
A_1	2	1	3

A_2	1	2	4
A_3	3	1,5	2

Ціни за 1 кг добрив: A_1 − 2 грн, A_2 − 3,5 грн, A_3 − 2,5 грн. Скласти найбільш економний план закупівлі добрив.

Згідно з умовою задачі складемо цільову функцію:

$$Z = 2x_1 + 3{,}5x_2 + 2{,}5x_3 \quad \to \min,$$

обмеження:

$$\begin{cases} 2x_1 + x_2 + 3x_3 \ge 6, \\ x_1 + 2x_2 + 1{,}5x_3 \ge 8, \\ 3x_1 + 4x_2 + 2x_3 \ge 12, \\ x_j \ge 0, \quad j = \overline{1,3}. \end{cases}$$

Перетворимо задачу до канонічної форми:

$$Z = 2x_1 + 3{,}5x_2 + 2{,}5x_3 + 0 \cdot y_1 + 0 \cdot y_2 + 0 \cdot y_3 \quad \to \min$$

$$\begin{cases} 2x_1 + x_2 + 3x_3 - y_1 = 6, \\ x_1 + 2x_2 + 1{,}5x_3 - y_2 = 8, \\ 3x_1 + 4x_2 + 2x_3 - y_3 = 12, \\ x_j \ge 0, \quad j = \overline{1,6}. \end{cases}$$

Виконаємо перетворення, щоб ввести в базис змінні y_1, y_2:

$$\begin{cases} x_1 + 3x_2 - x_3 + y_1 - y_3 = 6, \\ 2x_1 + 2x_2 + 0{,}5x_3 + y_2 - y_3 = 4, \\ 3x_1 + 4x_2 + 2x_3 - y_3 = 12, \\ x_j \ge 0, \quad j = \overline{1,6}. \end{cases}$$

Для того щоб отримати базис слід ввести штучну змінну y_4 та розв'язати розширену *M*-задачу:

$$Z = 2x_1 + 3{,}5x_2 + 2{,}5x_3 + 0 \cdot \mathrm{y}_1 + 0 \cdot y_2 + 0 \cdot y_3 + M \cdot y_4 \quad \to \min$$

$$\begin{cases} x_1 + 3x_2 - x_3 + y_1 - y_3 = 6, \\ 2x_1 + 2x_2 + 0{,}5x_3 + y_2 - y_3 = 4, \\ 3x_1 + 4x_2 + 2x_3 - y_3 + y_4 = 12, \\ x_j \ge 0, \quad j = \overline{1{,}7}. \end{cases}$$

i	Базис	$C_б$	C	2	3,5	2,5	0	0	0	M
			A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7
1	A_4	0	6	1	3	-1	1	0	-1	0
2	A_5	0	4	2	**2**	1/2	0	1	-1	0
3	A_7	M	12	3	4	2	0	0	-1	1
4	$\Delta_j = Z_j - c_j$		12$\boldsymbol{M}$	3M-2	4M- -3,5	2M- -2,5	0	0	-M	0

$$\theta_2 = \min\left\{6/3;\ 4/2;\ 12/4\right\} = 2$$

$X_0 = (0, 0, 0, 6, 4, 0, 12),\ Z(X_0) = 12M$.

i	Базис	$C_б$	C	2	3,5	2,5	0	0	0	M
			A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7
5	A_4	0	0	-2	0	-7/4	1	-3/2	1/2	0
6	A_2	3,5	2	1	1	1/4	0	1/2	-1/2	0
7	A_7	M	4	-1	0	**1**	0	-2	1	1
8	$\Delta_j = Z_j - c_j$		4$\boldsymbol{M}$+ +7	3/2- -M	0	***M*-** **-13/8**	0	7/4- -2M	M-7/4	0

$X_0 = (0, 2, 0, 0, 0, 0, 4),\ Z(X_0) = 4M + 7$.

i	Базис	$C_б$	C	2	3,5	2,5	0	0	0	M
			A_0	A_1	A_2	A_3	A_4	A_5	A_6	A_7
9	A_4	0	7	-15/4	0	0	1	-5	9/4	7/4

10	A_2	3,5	1	5/4	1	0	0	1	-3/4	-1/4
11	A_3	2,5	4	-1	0	1	0	-2	1	1
12	$\Delta_j = Z_j - c_j$		13,5	-1/8	0	0	0	-3/2	-1/8	13/8- -M

Оптимальний план задачі: X^*= (0; 1; 4; 7; 0; 0; 0), $Z(X^*) = 13{,}5$.

Таким чином, для 1 га чорнозему фермерському господарству слід закупити 1 кг добрив складу A_2 і 4 кг добрив складу A_3 (добриво складу A_1 зовсім не закуповувати). Цим повністю задовольнятимуться вимоги внесення до чорнозему азоту і калію, а кількість фосфору на 1 га перевищить необхідний мінімум на 7 одиниць. Загальна вартість добрив для цього плану мінімальна і складе 13,5 грн з розрахунку 1 га.

4.4. Проблема виродження

У процесі використання симплексного методу монотонне зростання цільової функції у ході дослідження на максимум (або монотонне зменшення у ході дослідження на мінімум) має місце за умови, що на кожній ітерації маємо невироджений опорний план. Мають місце задачі, коли основна система обмежень в правій частині містить один або кілька нулів. Така ситуація може виникнути під час розв'язання задачі. Так, в рядку 5 симплексної таблиці прикладу пункту 4.3 міститься рівняння вільний член якого дорівнює нулю, отже опорний план вироджений $X_0 = (0, 2, 0, 0, 0, 0, 4)$, оскільки він містить одну базисну невідому, значення якої дорівнює нулю. При цьому, якби довелось вводити в базис A_6, то найменше значення в симплексному відношенні було б нуль.

Розглянемо, як зміниться цільова функція, якщо при черговому перетворенні за методом Жордана−Гаусса елемент розв'язування міститься в рядку з нульовою правою частиною. Відомо, що при кожній ітерації цільова функція змінюється на величину $\Delta Z = |\theta \cdot \Delta|$. Оскільки в цьому випадку $\theta = 0$, то введення в базис вектора, якому відповідає нульове симплексне відношення, не призводить до зміни цільової функції, отже, таке перетворення не має сенсу. Як правило, після декількох перетворень з використанням нульового симплексного відношення можна отримати план, який був раніше, тобто має місце зациклення.

На практиці такі випадки зустрічаються дуже рідко, і мають, як правило, штучний характер, але теоретично такий випадок можливий.

Запитання для самоперевірки

1. Для розв’язання яких математичних задач використовується симплексний метод?

2. У чому суть симплекс-методу?

3. На яких властивостях задач лінійного програмування заснований симплекс-метод?

4. Як визначити початковий опорний план задачі лінійного програмування?

5. Сформулюйте послідовність етапів практичної реалізації симплекс-методу у ході розв’язування задач лінійного програмування.

6. Які ознаки оптимальності базисного плану?

7. Коли виникає необхідність використання симплекс-методу з штучним базисом?

8. У чому суть проблеми виродження?

5. Теорія двоїстості. Взаємно двоїсті задачі лінійного програмування

5.1. Правила складання умов взаємно двоїстих задач.

5.2. Теореми двоїстості.

5.1. Правила складання умов взаємно двоїстих задач

У різних розділах математики часто зустрічаються так звані *теореми двоїстості*. Кожна з них дозволяє для будь-якого твердження даної теорії побудувати – за певними стандартними правилами – інше твердження таким чином, що із справедливості першого автоматично випливає справедливість другого. Стосовно лінійного програмування це виражається в тому, що, розв’язуючи одну оптимізаційну задачу, одержуємо розв’язок ще однієї оптимізаційної задачі, *двоїстої до вихідної*.

Можна вказати на ряд практичних використань теорії двоїстості. У ході розв'язування задачі ЛП симплекс-методом розв'язок парної задачі виходить автоматично без додаткових зусиль з нашого боку, тому з двох зв'язаних між собою задач розв'язувати слід ту, яка простіше. Розглядаючи за кроками процес розв'язування вихідної задачі стандартним симплекс-методом і спостерігаючи за пов'язаним з ним процесом автоматичного розв'язування двоїстої задачі, був відкритий алгоритм так званого двоїстого *симплекс-методу*. Новий метод істотно розширив обчислювальні можливості і виявився дуже корисним під часрозв'язання цілочисельних задач ЛП і задач ЛП з параметрами. Цікаво, що, записавши умови парних задач разом, виявилося можливим звести задачу оптимізації до звичайного розв'язування системи алгебраїчних рівнянь (з подвійним числом невідомих, з яких половина невідомих повинна дорівнювати нулю). Такий прийом використовується у процесі розв'язання транспортної задачі ЛП і є основою для розв'язання задач квадратичного програмування. Розв'язання двоїстої задачі корисне само по собі – воно дає можливість провести аналіз стійкості розв'язку вихідної задачі щодо малих змін в її умовах. Так, у задачі оптимального використання обмежених ресурсів можливо перевірити чутливість знайденого оптимального плану до малих варіацій у запасах ресурсів. Економічна інтерпретація двоїстих оцінок дозволяє визначити корисність кожного ресурсу і вказати найбільш прийнятний шлях зміни їх запасів для досягнення максимального прибутку.

Перша задача. Підприємство має m видів ресурсів у кількостях b_i $(i=\overline{1,m})$ з яких виготовляється n видів продукції. Відома матриця технологічних коефіцієнтів $A=(a_{ij})_{m\times n}$, а також відома вартість одиниці кожної продукції c_j ($j=\overline{1,n}$). Необхідно скласти оптимальний план виробництва, щоб загальний прибуток був максимальним. Тобто висхідна задача формулюється так: скільки і якої продукції x_j $\left(j=\overline{1,n}\right)$ необхідно виробити, щоб при заданих вартостях c_j одиниці продукції і обсягах наявних ресурсів b_i максимізувати випуск продукції у вартісному вираженні. Тобто знайти вектор $X=(x_1, x_2, ..., x_n)$, який задовольняє обмеженням:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots a_{1n}x_n \le b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots a_{2n}x_n \le b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots a_{mn}x_n \le b_m, \end{cases}$$

$$x_j \ge 0, \quad j = \overline{1,n},$$

і доставляє максимальне значення функції:

$$Z = c_1x_1 + c_2x_2 + \ldots + c_nx_n \to \max.$$

Це є висхідна задача.

Друга задача. За одиницю вартості ресурсів приймають одиницю вартості випущеної продукції. Позначимо через y_i $\left(i = \overline{1,m}\right)$ вартість одиниці i-го ресурсу. Тоді вартість всіх потрачених ресурсів, використаних для виготовлення одиниці j-го виду продукції, дорівнює $\sum\limits_{i=1}^{m} a_{ij}y_i$, що не може бути менше вартості кінцевої продукції, тому $\sum\limits_{i=1}^{m} a_{ij}y_i \ge c_j$. Вартість всіх наявних ресурсів буде дорівнювати $\sum\limits_{i=1}^{m} b_iy_i$.

Отже, двоїста задача формулюється так. Яка повинна бути ціна одиниці кожного з ресурсів, щоб при заданих кількостях ресурсів b_i і величинах вартості одиниці продукції c_j мінімізувати загальну вартість витрат.

Знайти вектор $Y = \left(y_1, y_2, \ldots, y_m\right)$, який задовольняє обмеженням:

$$\begin{cases} a_{11}y_1 + a_{21}y_2 + \ldots a_{m1}y_m \ge c_1, \\ a_{12}y_1 + a_{22}y_2 + \ldots a_{m2}x_n \ge c_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{1n}y_1 + a_{2n}y_2 + \ldots a_{mn}y_m \ge c_n, \end{cases}$$

$$y_i \ge 0, \quad i = \overline{1,m},$$

і доставляє мінімальне значення функції:

$$F = b_1y_1 + b_2y_2 + \ldots + b_my_m \to \min.$$

Це є двоїста задача.

Змінні y_i називаються оцінками або обліковими, неявними цінами.

Обидві задачі утворюють пару сполучених двоїстих задач.

Правила складання умов взаємно двоїстих задач:

1. Основним невідомим однієї задачі відповідають балансові невідомі іншої задачі і навпаки:

Задача 1	Основні змінні $x_1, x_2, \dots, x_n$	Балансові змінні $u_1, u_2, \dots, u_m$
Задача 2	$v_1, v_2, \dots, v_n$ Балансові змінні	$y_1, y_2, \dots, y_m$ Основні змінні

Число основних невідомих однієї задачі дорівнює числу обмежень іншої задачі. Між невідомими взаємно двоїстих задач існують зв'язки: $x_j \leftrightarrow v_j,\ y_i \leftrightarrow u_i$.

2. Невід'ємним невідомим однієї задачі відповідають невід'ємні невідомі іншої задачі.

3. Фіктивним невідомим (тотожно рівним нулю) однієї задачі відповідають в іншій задачі невідомі без обмеження на знак і навпаки.

4. Якщо умови однієї задачі приведені до 1-ї стандартної форми, то умови іншої задачі будуть представлені в 2-й стандартній формі і навпаки.

5. Вільні члени в обмеженнях однієї задачі є коефіцієнтами цільової функції іншої задачі і навпаки.

6. Матриці коефіцієнтів при невідомих в обмеженнях обох задач – взаємно транспоновані.

Види математичних моделей двоїстих задач

I. Симетричні задачі

Висхідна задача	Двоїста задача
I.1. $Z = CX \to \min$	$F = YA_0^T \to \max$
$AX \ge A_0,$	$YA \le C^T,$
$X \ge 0,$	$Y \ge 0.$

Висхідна задача	Двоїста задача
I.2. $Z = CX \to \max$	$F = YA_0^T \to \min$
$AX \le A_0,$	$YA \ge C^T,$
$X \ge 0,$	

$$Y \geq 0 .$$

II. Несиметричні задачі

Висхідна задача	Двоїста задача
II.1. $Z = CX \to \min$	$F = YA_0^T \to \max$
$AX = A_0,$	$YA \leq C^T .$
$X \geq 0,$	

Висхідна задача	Двоїста задача
II.2. $Z = CX \to \max$	$F = YA_0^T \to \min$
$AX = A_0,$	$YA \geq C^T .$
$X \geq 0,$	

5.2. Теореми двоїстості

Властивості розв'язків взаємно двоїстих задач зазвичай групуються у два блоки, які іменуються 1-ю і 2-ю теоремами двоїстості.

У теорему 1 увійшли наступні твердження:

***Теорема* 1.1.** Якщо одна задача має оптимальний розв'язок, то інша задача також має оптимальний розв'язок, причому $Z_{\max}\left(X^*\right) = F_{\min}\left(Y^*\right)$ або $Z_{\min}\left(X^*\right) = F_{\max}\left(Y^*\right)$.

Доведення. Розглянемо несиметричну пару спряжених задач. Візьмемо математичну модель висхідної задачі у вигляді II.1:

$$\begin{gathered} Z = CX \to \min, \\ AX = A_0, \\ X \geq 0. \end{gathered} \tag{5.1}$$

Нехай ця задача має розв'язок. Тоді її оптимальний план можна визначити симплексним методом. Необхідно вважати, що базисними векторами є вектори $A_1, A_2, ..., A_m$ і остання ітерація симплексного методу має вигляд, представлений в табл. 5.1.

Таблиця 5.1

		C	c_1	c_2	$\dots$	c_m	c_{m+1}	$\dots$	c_n

Базис	$C_б$	A_0	A_1	A_2	…	A_m	A_{m+1}	…	A_n
A_1	c_1	$x^*{}_1$	1	0	…	0	x_{1m+1}	…	x_{1n}
A_2	c_2	$x^*{}_2$	0	1	…	0	x_{2m+1}	…	x_{2n}
…	…	…	…	…	…	…	…	…	…
A_m	c_m	$x^*{}_m$	0	0	…	1	x_{mm+1}	…	x_{mn}
$\Delta_j = Z_j - C_j$		$Z^*{}_0$	$Z_1 - c_1$	$Z_2 - c_2$	…	$Z_m - c_m$	$Z_{m+1} - c_{m+1}$	…	$Z_n - c_n$

Позначимо через D матрицю, що складається з компонентів векторів $A_1, A_2, ..., A_m$ (базисні вектори останньої ітерації), тоді табл. 5.1 містить коефіцієнти розкладання векторів $A_1, A_2, ..., A_n$ вихідної системи за векторами базису, тобто кожному вектору A_j в цій таблиці відповідає вектор X_j, такий що:

$$A_j = DX_j \;\; j = \overline{1,n} \; . \qquad (5.2)$$

Для оптимального плану маємо:

$$A_0 = DX^*, \qquad \text{де} \qquad X^* = \left(x_1^*, x_2^*, ..., x_n^*\right)^T . \qquad (5.3)$$

Позначимо $\overline{X}$ матрицю, що складається з коефіцієнтів у розкладанні векторів A_j в табл. 5.1. З урахуванням співвідношень (5.2), (5.3) маємо:

$$A = D\overline{X} \text{ , звідки } D^{-1}A = \overline{X} \; , \qquad (5.4)$$

$$A_0 = DX^* \text{ , звідки } D^{-1}A_0 = X^* . \qquad (5.5)$$

За таблицею мінімальне значення цільової функції дорівнює:

$$Z_0^* = C_б X^* . \qquad (5.6)$$

Розглянемо вектор $\overline{Z}$**:**

$$\overline{Z} = C_б X^* - C = \left(C_б X_1 - c_1, C_б X_2 - c_2, ..., C_б X_n - c_n\right) = \\ = \left(Z_1 - c_1, Z_2 - c_2, ..., Z_n - c_n\right). \qquad (5.7)$$

Оскільки його компоненти є відповідними оцінками $\Delta_j = Z_j - c_j \le 0 (j = \overline{1,n})$ оптимального плану, то вектор $\overline{Z}$ не додатний: $\overline{Z} = C_б \overline{X} - C \le 0$.

Оптимальний план вихідної задачі має вигляд: $X^* = D^{-1}A_0$, тому оптимальний план двоїстої задачі:

$$F = YA_0^T \to \max \tag{5.8}$$

$$YA \le C^T,$$

відшукаємо в аналогічній формі: $Y^* = C_б D^{-1}$. Покажемо, що вектор Y^* дійсно є планом двоїстої задачі (5.8). Для цього її систему обмежень запишемо у вигляді $YA - C^T \le 0$ і підставимо в ліву частину нерівності вектор Y^*. Враховуючи співвідношення (5.4), (5.7) і (5.8), маємо:

$$YA - C^T = C^* D^{-1} A - C^T = C^* D^{-1} D\overline{X} - C^T = C^* \overline{X} - C^T \le 0.$$

Таким чином, $Y^* A \le C^T$. Оскільки вектор Y^* задовольняє систему обмежень задачі (5.8), то Y^* є планом двоїстої задачі. Значення цільової функції двоїстої задачі з цим планом буде $F(Y^*) = Y^* A_0^T$. З іншого боку,

$$F(Y^*) = Y^* A_0^T = C_б^* D^{-1} A_0^T = C^* X^* = Z(X^*) = Z_{\min}.$$

Таким чином, доведено, що цільова функція двоїстої задачі для плану Y^* дорівнює мінімальному значенню цільової функції вихідної задачі. Залишається довести, що Y^* є оптимальним планом двоїстої задачі.

Слід помножити рівняння $AX = A_0$ на будь-який довільний план Y двоїстої задачі, а нерівність $YA \le C^T$ помножимо на будь-який план вихідної задачі X. Маємо:

$$YAX = YA_0 = F(Y),$$

$$YAX \le C^T X = Z(X).$$

Звідси випливає, що для будь-яких довільних планів X і Y виконується нерівність $F(Y) \le Z(X)$. Це співвідношення поширюється і на екстремальні значення цільової функції, тобто:

$$F_{\max}(Y) \leq Z_{\min}(X).$$

З останньої нерівності випливає, що цільова функція досягає максимального значення в тому випадку, якщо $F_{\max}(Y) = Z_{\min}(X)$, але це має місце, якщо $Y = Y^*$, отже, Y^* і є оптимальний план двоїстої задачі.

Аналогічно можна довести, що якщо двоїста задача має розв'язок, то й висхідна також має розв'язок, причому $Z_{\min}(X) = F_{\max}(Y)$.

Для доведення другої частини теореми, припустимо, що цільова функція вихідної задачі не обмежена знизу. Тоді випливає, що $F(Y) \leq -\infty$. Це означає, що двоїста задача має суперечливу систему обмежень, а отже, не має розв'язків.

Таким же чином припустимо, що цільова функція двоїстої задачі не обмежена зверху, тобто $Z(X) \geq +\infty$. Це означає, що висхідна задача має необмежену цільову функцію на множині своїх планів або її система обмежень суперечлива, а отже, вихідна задача не має розв'язків. ■

Отже, сформулюємо без доведення ще три теореми.

***Теорема* 1.2.** Для довільних розв'язків двох задач $Z_1 \leq F_2$.

***Теорема* 1.3.** Якщо в одній задачі функція цілі необмежена, то інша задача розв'язків не має.

***Теорема* 1.4.** Якщо одна задача не має розв'язків, то в іншій задачі функція цілі необмежена, або ж інша задача також не має розв'язків.

***Теорема* 2**. Для оптимальності двох планів пари спряжених задач необхідно і достатньо, щоб ці плани задовольняли умову:

$$y_i^* \cdot (\sum_{j=1}^{n} a_{ij} x_j^* - b_i) = 0, \quad i = \overline{1,m}, \tag{5.9}$$

$$x_j^* \cdot (\sum_{i=1}^{m} a_{ij} y_i^* - c_j) = 0, \quad j = \overline{1,n}, \tag{5.10}$$

де X^* та Y^* − оптимальні плани відповідно висхідної та двоїстої задач.

Доведення. Для доведення звернемо увагу на співмножники, що стоять у дужках у виразах (5.9) - (5.10). Це різниці між лівими та правими частинами основних обмежень кожної з пари спряжених задач. Отже, другу теорему двоїстості можна сформулювати таким чином. Два плани пари спряжених задач лінійного програмування будуть оптимальними тоді і лише тоді, коли всім відмінним від нуля компонентам плану кожної з задач відповідають обмеження спряженої задачі, які виконуються у вигляді точних рівнянь, а всім обмеженням, які залишаються у вигляді нерівностей, відповідають змінні спряженої задачі, що обов'язково дорівнюють нулю. Іншими словами, для того, щоб опорний план був оптимальним, необхідно і достатньо, щоб всім додатнім базисним змінним відповідали обмеження спряженої задачі, які задовольняються як точні рівняння, а обмеження, що залишаються у вигляді нерівностей, відповідали б вільним змінним, які дорівнюють нулю.

Доведемо необхідність виконання умов (5.9) − (5.10). Нехай X^* та Y^* − оптимальні плани симетричної пари спряжених задач.

Висхідна задача

$$Z = CX \to \max$$

$$\sum_{j=1}^{n} a_{ij} x_j \le b_i, \quad i = \overline{1,m}, \qquad (5.11)$$

$$x_j \ge 0, \quad j = \overline{1,n}.$$

Двоїста задача

$$F = YA_0^T \to \min$$

$$\sum_{i=1}^{m} a_{ij} y_i \ge c_j, \quad j = \overline{1,n}, \qquad (5.12)$$

$$y_i \ge 0, \quad i = \overline{1,m}.$$

Покажемо, що умови (5.11) та (5.12) задовольняються.

Оскільки X^* є планом задачі (5.12), то виконується співвідношення:

$$\sum_{j=1}^{n} a_{ij} x_j^* \le b_i, \quad i = \overline{1,m}, \tag{5.13}$$

а оскільки Y^* є планом задачі (5.12), то виконується співвідношення:

$$\sum_{i=1}^{m} a_{ij} y_i^* \ge c_j, \quad j = \overline{1,n}. \tag{5.14}$$

Помножимо нерівності системи (5.13) на відповідні їм змінні оптимального плану двоїстої задачі y_i^* $(i = \overline{1,m})$, а нерівності системи (5.14) – на змінні x_j^* $(j = \overline{1,n})$ оптимального плану вихідної задачі. Отримаємо співвідношення:

$$\sum_{i=1}^{m} y_i^* \sum_{j=1}^{n} a_{ij} x_j^* \le \sum_{i=1}^{m} y_i^* b_i = F^*, \tag{5.15}$$

$$\sum_{j=1}^{n} x_j^* \sum_{i=1}^{m} a_{ij} y_i^* \ge \sum_{j=1}^{n} x_j^* c_j = Z^*. \tag{5.16}$$

Згідно з першою теоремою двоїстості $F^* = Z^*$, отже:

$$\sum_{i=1}^{m} y_i^* \sum_{j=1}^{n} a_{ij} x_j^* \le \sum_{i=1}^{m} y_i^* b_i = \sum_{j=1}^{n} x_j^* c_j \le \sum_{j=1}^{n} x_j^* \sum_{i=1}^{m} a_{ij} y_i^* . \qquad (5.17)$$

Неважко помітити, що крайні члени останнього виразу – це та ж сама величина, а саме: $\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} x_j^* y_i^*$, отже, у виразі (5.17) знаки нестрогої нерівності треба замінити знаками рівняння, звідки дістанемо:

$$\sum_{i=1}^{m} y_i^* \sum_{j=1}^{n} a_{ij} x_j^* = \sum_{i=1}^{m} y_i^* b_i \Rightarrow \quad \sum_{i=1}^{m} y_i^* (\sum_{j=1}^{n} a_{ij} x_j^* - b_i) = 0, \qquad (5.18)$$

$$\sum_{j=1}^{n} x_j^* \sum_{i=1}^{m} a_{ij} y_i^* = \sum_{j=1}^{n} x_j^* c_j \Rightarrow \quad \sum_{j=1}^{n} x_j^* (\sum_{i=1}^{m} a_{ij} y_i^* - c_j) = 0. \qquad (5.19)$$

Кожний з доданків у рівнянні (5.18) буде недодатним, бо $y_i^* \ge 0$, а $\sum_{j=1}^{n} a_{ij} x_j^* - b_i \le 0$, що в добутку дає недодатну величину. Оскільки сума доданків одного знаку (у даному випадку недодатних) дорівнює нулю, це можливо лише тоді, коли кожен з доданків цієї суми дорівнює нулю:

$$y_i^* \cdot (\sum_{j=1}^{n} a_{ij} x_j^* - b_i) = 0, \quad i = \overline{1,m} .$$

Застосувавши аналогічні міркування до визначення знаків доданків у виразі (5.19), отримаємо:

$$x_j^* \cdot (\sum_{i=1}^{m} a_{ij} y_i^* - c_j) = 0, \quad j = \overline{1,n} .$$

Таким чином, умову необхідності доведено.

Доведемо тепер ознаку достатності. Нехай для деяких планів X' та Y' спряженої пари задач (5.11) та (5.12) виконуються умови:

$$y_i' \cdot (\sum_{j=1}^{n} a_{ij} x_j' - b_i) = 0 \quad \Rightarrow \quad y_i' \sum_{j=1}^{n} a_{ij} x_j' = y_i' b_i, \quad i = \overline{1,m}, \qquad (5.20)$$

$$x_j' \cdot (\sum_{i=1}^{m} a_{ij} y_i' - c_j) = 0 \quad \Rightarrow \quad x_j' \sum_{i=1}^{m} a_{ij} y_i' = c_j x_j', \quad j = \overline{1,n} . \qquad (5.21)$$

Шляхом послідовного додавання лівих та правих частин рівнянь у системі (5.20) визначимо суму всіх m рівнянь системи:

$$\sum_{i=1}^{m} y_i' \sum_{j=1}^{n} a_{ij} x_j' = \sum_{i=1}^{m} y_i' b_i = F(\mathbf{Y}'). \tag{5.22}$$

Аналогічно визначимо суму всіх n рівнянь системи (5.21):

$$\sum_{j=1}^{n} x_j' \sum_{i=1}^{m} a_{ij} y_i' = \sum_{j=1}^{n} c_j x_j' = Z(\mathbf{X}'). \tag{5.23}$$

Як бачимо, ліві частини виразів (5.22) та (5.23) однакові, отже, для планів X' та Y' маємо, що $Z(X') = F(Y')$. За першою теоремою двоїстості це є критерієм оптимальності планів пари спряжених задач. Звідси $X' = X^*$ та $Y' = Y^*$. Теорему доведено. ■

Теорема. Якщо при підстановці компонент оптимального плану в систему обмежень висхідної задачі i-те обмеження перетворюється в нерівність, то i-та компонента оптимального плану двоїстої задачі дорівнює нулю.

Якщо i-та компонента оптимального плану двоїстої задачі додатна, то i-те обмеження висхідної задачі задовольняється її оптимальним розв'язком як строга рівність.

Зауваження. У ході розв'язання ЗЛП симплекс-методом розв'язок двоїстої задачі знаходиться в рядку цільової функції.

Приклад. Побудувати і розв'язати двоїсту задачу для висхідної задачі:

$$Z = x_1 - x_2 \to \max,$$

$$\begin{cases} -2x_1 + x_2 \le 2, \\ x_1 - 2x_2 \le 2, \\ x_1 + x_2 \le 5, \\ x_1 \ge 0,\ \ x_2 \ge 0. \end{cases}$$

Необхідно розглянути розв'язки задач з використанням теорем двоїстості.

Висхідна задача:

$$Z = x_1 - x_2 \to \max,$$

$$\begin{cases} -2x_1 + x_2 \le 2, \\ x_1 - 2x_2 \le 2, \\ x_1 + x_2 \le 5, \\ x_1 \ge 0,\ \ x_2 \ge 0. \end{cases}$$

Двоїста задача:

$$F = 2y_1 + 2y_2 + y_3 \to \min,$$

$$\begin{cases} -2y_1 + y_2 + y_3 \ge 1, \\ y_1 - 2y_2 + y_3 \ge -1, \\ y_i \ge 0,\ \ i = \overline{1,3}. \end{cases}$$

Розв'яжемо вихідну задачу симплексним методом (табл. 5.2), і перетворимо нерівності в рівності:

$$\begin{cases} -2x_1 + x_2 + x_3 = 2, \\ x_1 - 2x_2 + x_4 = 2, \\ x_1 + x_2 + x_5 = 5, \\ x_i \ge 0, \ \ j = \overline{1,5}. \end{cases}$$

Таблиця 5.2

Базис	$C_б$	C	1	-1	0	0	0
		A_0	A_1	A_2	A_3	A_4	A_5
A_3	0	2	-2	1	1	0	0
A_4	0	2	**1**	-2	0	1	0
A_5	0	5	1	1	0	0	1
Цільова функція		0	0	0	0	0	0
$\Delta_j = Z_j - c_j$			-1	1	0	0	0
A_3	0	6	0	-3	1	2	0
A_1	1	2	1	-2	0	1	0
A_5	0	3	0	**3**	0	-1	1
Цільова функція		2	1	-2	0	1	0

Закінчення табл. 5.2

$\Delta_j = Z_j - c_j$		2	0	-1	0	1	0
A_3	0	9	0	0	1	1	1
A_1	1	4	**1**	0	0	$\frac{1}{3}$	$\frac{2}{3}$
A_2	-1	1	0	1	0	$-\frac{1}{3}$	$\frac{1}{3}$
Цільова функція		3	1	-1	0	$\frac{2}{3}$	$\frac{1}{3}$
$\Delta_j = Z_j - c_j$			0	0	0	$\frac{2}{3}$	$\frac{1}{3}$

З табл. 5.2 випливає, що $X^* = (4,1), \quad Z(X^*) = 3$.

На підставі 1-й теореми двоїстості маємо: $Z(X^*) = F(Y^*) = 3$.

Оскільки $x_1 \geq 0, x_2 \geq 0$, то за 2-ю теоремою двоїстості систему обмежень двоїстої задачі запишемо у вигляді:

$$\begin{cases} -2y_1 + y_2 + y_3 = 1; \\ y_1 - 2y_2 + y_3 = -1. \end{cases}$$

Підставивши $X^* = (4,1)$ у систему обмежень висхідної задачі:

$$\begin{cases} -2 \cdot 4 + 1 \leq 2; \\ 4 - 2 \cdot 1 \leq 2; \\ 4 + 1 \leq 5; \end{cases} \quad \begin{cases} -7 < 2; & \Rightarrow y_1 = 0; \\ 2 = 2; & \Rightarrow y_2 > 0; \\ 5 = 5; & \Rightarrow y_3 > 0. \end{cases}$$

Тоді система обмежень двоїстої задачі прийме вигляд:

$$\begin{cases} y_2 + y_3 = 1; \\ -2y_2 + y_3 = -1. \end{cases}$$

Звідки $Y^* = \left(0, \frac{2}{3}, \frac{1}{3}\right), \quad F(Y^*) = 3$.

Нехай відомо розв'язок двоїстої задачі $Y^* = \left(0, \frac{2}{3}, \frac{1}{3}\right), \quad F(Y^*) = 3$, знайдемо розв'язок висхідної задачі. За 1-ю теоремою двоїстості $F(Y^*) = Z(X^*) = 3$. Оскільки $y_2 > 0, y_3 > 0$, то за 2-ю теоремою двоїстості друга і третя нерівності висхідної задачі перетворюються в рівності:

$$\begin{cases} x_1 - 2x_2 = 2; \\ x_1 + x_2 = 5. \end{cases}$$

Звідки, $X^* = (4,1), \quad Z(X^*) = 3$.

Розглянемо розв'язок задач методом, заснованим на взаємно однозначній відповідності між змінними: основним змінним висхідної задачі відповідають балансові змінні двоїстої, і навпаки.

Нехай двоїста задача розв'язана симплексним методом (табл. 5.3):

$$F = 2y_1 + 2y_2 + 5y_3 \quad \rightarrow \quad \min$$

$$\begin{cases} -2y_1 + y_2 + y_3 - y_4 = 1; \\ -y_1 + 2y_2 - y_3 + y_5 = 1; \\ y_i \geq 0, i = \overline{1,3}. \end{cases}$$

Таблиця 5.3

Базис	$C_б$	C A_0	2 A_1	2 A_2	5 A_3	0 A_4	0 A_5
	0	1	-2	1	**1**	-1	0
A_5	0	1	-1	2	-1	0	1
Цільова функція		0	0	0	0	0	0
$\Delta_j = Z_j - c_j$							0
A_3	5	1	-2	1	1	-1	0
A_5	0	2	-3	**3**	0	-1	1
Цільова функція		5	-12	5	5	-5	0
$\Delta_j = Z_j - c_j$			-8	3	0	-5	0
A_3	5	$\frac{1}{3}$	-1	0	1	$-\frac{2}{3}$	$-\frac{1}{3}$
A_2	2	$\frac{2}{3}$	-1	1	0	$-\frac{1}{3}$	$\frac{1}{3}$
Цільова функція		3	-7	2	5	-4	-1
$\Delta_j = Z_j - c_j$			-9	0	0	-4	-1

Розв'язок іншої задачі знайдемо за відповідністю між змінними:

Задача 1	Основні змінні x_1 x_2 x_3	Балансові змінні x_4 x_5
Задача 2	y_4 y_5 y_1 Балансові змінні	y_2 y_3 Основні змінні

Значення x_j визначаємо за симплексною таблицею (табл. 5.3) у рядку Δ_i у відповідному стовпці, причому значення x_j слід брати за модулем:

$$x_1 \to y_4, \quad x_1 = |\Delta_4| = 4,$$
$$x_2 \to y_5, \quad x_2 = |\Delta_5| = 1.$$

Таким чином, розв'язок висхідної задачі: $X^* = (4,1)$, $\quad Z(X^*) = 3$.

Якщо висхідна задача розв’язана симплексним методом, то розв'язок двоїстої задачі може бути знайдений за формулою: $Y^* = \boldsymbol{C} \cdot \boldsymbol{A}^{-1}$, де $\boldsymbol{C}$ – матриця-рядок коефіцієнтів при базисних змінних цільової функції в оптимальному розв’язанні висхідної задачі; $\boldsymbol{A}^{-1}$ – обернена матриця для матриці $\boldsymbol{A}$, що є матрицею коефіцієнтів базисних змінних системи обмежень висхідної задачі в оптимальному розв'язку. З табл. 5.3 маємо:

$$\boldsymbol{C} = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix},\ \boldsymbol{A} = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -2 & 0 \\ 1 & 1 & 0 \end{pmatrix},\ \boldsymbol{A}^{-1} = \begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & -\frac{1}{3} & \frac{1}{3} \\ 1 & 1 & 1 \end{pmatrix},$$

$$Y^* = \boldsymbol{C} \cdot \boldsymbol{A}^{-1} = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ 0 & -\frac{1}{3} & \frac{1}{3} \\ 1 & 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & \frac{2}{3} & \frac{1}{3} \end{pmatrix}.$$

Таким чином, розв'язок двоїстої задачі: $Y^* = \left(0, \frac{2}{3}, \frac{1}{3}\right)$, $\quad F(Y^*) = 3$.

Запитання для самоперевірки

1. Наведіть приклади практичного використання теорії двоїстості.
2. Сформулюйте правила складання умов взаємно двоїстих задач.
3. Сформулюйте основні теореми теорії двоїстості.
4. Поясніть економічний зміст теорем двоїстості.
5. Як за розв'язком прямої задачі знайти розв'язок двоїстої?

Тема 6. Економічна інтерпретація двоїстих невідомих. Двоїстий симплекс-метод

6.1. Економічна інтерпретація двоїстих невідомих.

6.2. Аналіз стійкості двоїстих оцінок.

6.3. Двоїстий симплекс-метод.

6.1. Економічна інтерпретація двоїстих невідомих

Зазвичай вільні члени обмежень типу $\sum a_{ij}x_j \le b_i$ трактуються як запаси ресурсів, коефіцієнти перед невідомими – як норми витрат ресурсів на виробництво одиниці продукції даного виду, балансові невідомі – як залишки ресурсів. При цьому самі нерівності показують, що загальні витрати кожного ресурсу не можуть перевищувати його запаси. Функція цілі, Z_1 максимум якої розшукують, зазвичай трактується як прибуток від реалізації виробленої продукції. На підставі теореми 1.1 прибуток в оптимальному плані можна виразити через запаси ресурсів: $Z_{1max} = F_{2min} = \sum y_i b_i$.

Звідси випливає, що невідома (двоїста до залишку ресурсів u_i) дорівнює першій похідній від максимального прибутку за запасом *i*-го ресурсу $\dfrac{\partial Z_{1max}}{\partial b_i} = 0$ і тому маємо таку інтерпретацію: невідома показує, на скільки збільшиться прибуток в оптимальному плані, якщо збільшити запас ресурсу b_i на одиницю. Іншими словами прибуток у разі збільшення ресурсу b_i на Δb_i збільшиться на $y_i\Delta b_i$, при цьому варіація ресурсу Δb_i не повинна перевищувати деякого допустимого рівня.

За наявності обмеження типу $\ge$, коли вільні члени обмежень не можна трактувати як запаси ресурсів, двоїсті невідомі інтерпретуються так: двоїста невідома y_i показує інтенсивність збільшення функції цілі в оптимальному розв'язку при зміні вільного члена b_i на одиницю убік ослаблення відповідного обмеження.

Таким чином, невідомі y_i є певними вартісними оцінками корисності ресурсів. Це не означає, що така реальна ціна ресурсів, за яку їх можна придбати; тіньові ціни характеризують корисність кожного ресурсу для подальшого збільшення прибутку під час розширеного виробництва.

Якщо якийсь ресурс є в надлишку, залишок цього ресурсу відмінний від нуля $u_k > 0$, отже, його двоїста невідома дорівнює нулю $y_i = 0$ і подальше збільшення запасу цього ресурсу не буде приводити до зміни прибутку (тіньова ціна даного ресурсу дорівнює нулю).

Якщо виявилося, що в оптимальному плані вичерпані запаси відразу декількох ресурсів (балансові невідомі дорівнюють нулю, відповідні обмеження реалізуються як строгі рівності), то тіньові ціни y_i будуть показувати найбільш бажаний шлях розширення ресурсів. Якщо в задачі враховано витрати на придбання ресурсів, то найбільш вигідним буде збільшення запасів ресурсів пропорційно їх тіньовим цінам:

$$\Delta b_1 : \Delta b_2 : \Delta b_3 = y_1 : y_2 : y_3 .$$

Економічну інтерпретацію двоїстих задач і оцінки дефіцитності ресурсів в околі оптимального плану розглянемо на прикладі.

Приклад. Для виготовлення двох видів продукції *A* і *B* використовують три види сировини, яку підприємство може щомісяця закуповувати в обмежених обсягах. Щомісячне знаходження необхідної сировини, витрати його на виготовлення одиниці кожного виду продукції, а також прибуток від реалізації одиниці продукції представлені у табл. 6.1.

Таблиця 6.1

Види сировини	Норми витрат		Запаси, кг
	A	*B*	
I	9	6	540
II	5	10	500
III	14	7	980
Прибуток, грн	5	6	

Визначити, яку кількість продукції кожного виду підприємство повинно випускати щомісячно, щоб прибуток від її реалізації був максимальним.

Позначимо через x_1 і x_2 відповідно кількість продукції *A* і *B*, яку підприємство випускає протягом місяця. Критерієм ефективності *Z* є прибуток від реалізації продукції. Маємо математичну модель задачі:

$$Z = 5x_1 + 6x_2 \quad \to \quad \max \tag{6.1}$$

$$\begin{cases} 9x_1 + 6x_2 \le 540 \\ 5x_1 + 10x_2 \le 500 \\ 14x_1 + 7x_2 \le 980 \\ x_1 \ge 0, \quad x_2 \ge 0 \end{cases}. \tag{6.2}$$

Складемо двоїсту задачу. Нехай кожному виду сировини поставлена у відповідність оцінка y_1, y_2 та y_3. Тоді загальна оцінка сировини *F*, яка використовується під час виготовлення продукції, становить:

$$F = 540y_1 + 500y_2 + 980y_3 \quad \rightarrow \quad \min, \qquad (6.3)$$

а система обмежень:

$$\begin{cases} 9y_1 + 5y_2 + 14y_3 \geq 5, \\ 6y_1 + 10y_2 + 7y_3 \geq 6, \\ y_i \geq 0, \ i = \overline{1,3}. \end{cases} \qquad (6.4)$$

Задачі (6.1), (6.2) та (6.3), (6.4) утворюють симетричну пару взаємно двоїстих задач. Рішенням вихідної задачі є оптимальний план виробництва продукції *A* і *B*, а двоїстої − оптимальна система оцінок. Щоб розв'язати ці задачі, досить розв’язати одну з них. Оскільки висхідна задача містить тільки дві невідомі, то її можна розв'язати графічно. Потім, за теоремою двоїстості, розв'яжемо і другу задачу.

Розв’яжемо висхідну задачу симплексним методом (табл. 6.2).

Таблиця 6.2

Базис	$C_б$	C	5	6	0	0	0
		A_0	A_1	A_2	A_3	A_4	A_5
A_3	0	540	9	6	1	0	0
A_4	0	500	5	**10**	0	1	0
A_5	0	980	14	7	0	0	1
Цільова функція		0	0	0	0	0	0
$\Delta_j = Z_j - c_j$		0	-5	**-6**	0	0	0

Закінчення табл. 6.2

A_3	0	240	**6**	0	1	-0,6	0
A_2	6	50	1/2	1	0	0,1	0
A_5	0	630	21/2	0	0	-0,7	1
Цільова функція		300	3	6	0	0,6	0
$\Delta_j = Z_j - c_j$			**-2**	0	0	0,6	0
A_1	5	40	1	0	1/6	-0,1	0
A_2	6	30	0	1	-1/12	0,15	0

A_5	0	60	0	0	-7/4	0,25	1
Цільова функція		300	380	5	6	1/3	0,4
$\Delta_j = Z_j - c_j$			0	0	1/3	0,4	0

$$\mathbf{X}^* = (40;\quad 30;\quad 0;\quad 0;\quad 60),\qquad \mathbf{Y}^* = (1/3;\quad 2/5;\quad 0).$$

$$F_{\min} = 540 \cdot 1/3 + 500 \cdot 0{,}4 = 380$$

Оскільки $F_{\min} = Z_{\max}$, то обидві задачі розв'язані правильно.

Підставимо значення двоїстих оцінок $y_1^*= 1/3$, $y_2^*= 2/5$ і $y_3^*= 0$ до системи обмежень двоїстої задачі:

$$\begin{cases} 9 \cdot \dfrac{1}{3} + 5 \cdot 0{,}4 = 5, \\ 6 \cdot \dfrac{1}{3} + 10 \cdot 0{,}4 = 6. \end{cases}$$

Обидва обмеження двоїстої задачі виконуються як рівняння, тобто двоїсті оцінки сировини не перевищують прибуток від реалізації кожного виду продукції. Тому продукцію доцільно виробляти у кількостях, які визначаються оптимальним планом вихідної задачі. Будь-яка зміна початкових умов вихідної задачі може вплинути як на її оптимальний план, так і на систему двоїстих оцінок, тому економічний аналіз з використанням двоїстих оцінок передбачає визначення інтервалу їх стійкості.

Змінні $y_1^*= 1/3$ і $y_2^*= 2/5$ визначають умовні двоїсті оцінки одиниці сировини відповідно І та ІІ типів. Ці дві оцінки відмінні від нуля і тому використовуються цілком. У цей час двоїста оцінка сировини ІІІ типу дорівнює нулю $y_3^* = 0$, тому при оптимальному плані виробництва $\mathbf{X}^*$ залишається певний запас сировини ІІІ типу. Таким чином, позитивні умовні оцінки мають тільки ті типи сировини, які повністю використовуються при оптимальному плані виробництва. Отже, двоїсті оцінки характеризують дефіцит сировини, що використовується.

Розглянемо числа, які містяться у стовпці вектора A_3 останньої ітерації симплексної таблиці. Отже, зростання прибутку від реалізації у разі збільшення обсягу сировини І типу на одиницю обумовлено переходом до нового оптимального плану, за яким випуск продукції A збільшується на 1/6 одиниць, випуск продукції У зменшується на 1/12 одиниць, а використання сировини III типу зростає на 7/4 кг. Відповідно, числа у стовпці вектора A_4 показують, як зміниться оптимальний план, якщо збільшити кількість сировини II типу на 0,4 кг.

Величина двоїстої оцінки показує, наскільки зросте максимальне значення цільової функції вихідної задачі у разі збільшення кількості сировини відповідного типу на 1 кг. Так, збільшення кількості сировини I типу на 1 кг призведе до того, що з'явиться можливість знайти новий план виробництва, за яким щомісячний загальний прибуток від реалізації виробів може збільшитися на $\Delta Z = 1/3$ *грн.*

6.2. Аналіз стійкості двоїстих оцінок

Змінні двоїстої задачі визначають темп зміни цільової функції вихідної задачі, якщо змінюється запас *i*-го виду сировини. Це справедливо в околі точки b_i, а саме поки базис *D* залишається незмінним. Якщо $y_i^* > 0$, то значення цільової функції висхідної задачі можна збільшити за рахунок збільшення запасів *i*-ї сировини, оскільки ця сировина буде використовуватися для збільшення випуску продукції. Якщо $y_i^* = 0$, то збільшення *i*-го виду сировини не супроводжується збільшенням цільової функції, так як цієї сировини достатньо. Наведені співвідношення відповідають другій теоремі двоїстості і є основою для теорії доповнюючої нежорсткості. Система рівнянь (6.5) показує, що зміна величини b_i супроводжується зміною екстремального значення цільової функції $Z_{\max}$ на величину $\left|y_i^*\right|$ і може бути охарактеризована лише тоді, коли при зміні b_i значення змінних y_i^* в оптимальному плані не змінюється. Це справедливо для тих $b_i + \Delta b_i$, для яких немає невід'ємних чисел серед компонент вектора *B*, де матриця D^{-1} є оберненою до матриці D, утвореної з компонент базисних векторів, що визначали початковий план вихідної задачі:

$$G = D^{-1} = \begin{pmatrix} b_1 + \Delta b_1 \\ b_2 + \Delta b_2 \\ \dots\dots\dots \\ b_m + \Delta b_m \end{pmatrix}. \qquad (6.5)$$

Отже, якщо знайдено рішення висхідної задачі, то без особливих труднощів можна провести аналіз стійкості двоїстих оцінок щодо зміни величин b_i. Це дозволяє оцінити стійкість оптимального плану двоїстої задачі щодо зміни b_i, визначити ступінь впливу зміни b_i на максимальне значення цільової функції висхідної задачі, що дає можливість знайти найбільш доцільне спрямування можливих змін b_i.

Приклад. Необхідно знайти матрицю D^{-1}, яка є оберненою до матриці D, утвореної з компонентів базисних векторів A_3, A_4 і A_5. Матрицю D^{-1} можна записати за останньою ітерацією симплексної таблиці (табл. 6.2):

$$D^{-1} = \begin{pmatrix} \frac{1}{6} & -0{,}1 & 0 \\ -\frac{1}{12} & 0{,}15 & 0 \\ -\frac{7}{4} & 0{,}25 & 1 \end{pmatrix}.$$

Тоді компоненти вектора G (6.5) варто записати таким чином:

$$D^{-1} \cdot \begin{pmatrix} b_1 + \Delta b_1 \\ b_2 + \Delta b_2 \\ \dots\dots\dots \\ b_m + \Delta b_m \end{pmatrix} = \begin{pmatrix} \frac{1}{6} & -0{,}1 & 0 \\ -\frac{1}{12} & 0{,}15 & 0 \\ -\frac{7}{4} & 0{,}25 & 1 \end{pmatrix} \cdot \begin{pmatrix} 540 + \Delta b_1 \\ 500 + \Delta b_2 \\ 980 + \Delta b_3 \end{pmatrix} = \begin{pmatrix} 40 + \frac{1}{6}\Delta b_1 - 0{,}1\Delta b_2 \\ 30 + \frac{1}{2}\Delta b_1 + 0{,}15\Delta b_2 \\ 60 - \frac{7}{4}\Delta b_1 + 0{,}25\Delta b_2 + \Delta b_3 \end{pmatrix}.$$

Слід визначити, за яких значень Δb_i $(i = \overline{1,3})$ ці компоненти невід'ємні:

$$\begin{cases} 40 + \frac{1}{6}\Delta b_1 - 0{,}1\Delta b_2 \ge 0, \\ 30 + \frac{1}{2}\Delta b_1 + 0{,}15\Delta b_2 \ge 0, \\ 60 - \frac{7}{4}\Delta b_1 + 0{,}25\Delta b_2 + \Delta b_3 \ge 0. \end{cases} \tag{6.6}$$

Якщо запаси сировини I та II типів не змінюються, система має рішення:

$$\begin{cases} \Delta b_1 = 0 \\ \Delta b_2 = 0 \\ \Delta b_3 \geq -60 \end{cases} \Rightarrow \quad \Delta b_3 \in [-60;+\infty).$$

Тобто запаси сировини III типу можуть необмежено зростати або зменшуватися в межах 60 кг, що не супроводжується зміною плану двоїстої задачі $\mathbf{Y}^* = (1/3; \quad 2/5; \quad 0)$. Оскільки за цим планом оцінка, що відповідає сировині III типу, дорівнює нулю, екстремальне значення цільової функції у разі зміни кількості сировини III типу не змінюється.

Якщо залишаються незмінними запаси сировини I і III типів, система нерівностей (6.7) дає рішення:

$$\begin{cases} \Delta b_1 = 0 \\ \Delta b_3 = 0 \\ -240 \leq \Delta b_2 \leq 400 \end{cases} \Rightarrow \quad \Delta b_2 \in [-240;400].$$

У разі зміни сировини II типу в межах $\Delta b_2 \in [-240;400]$ екстремальне значення цільової функції змінюється на величину $\Delta Z_{\max} = 2/5 \cdot \Delta b_2$, що пов'язано із зміною оптимального плану виробництва. При постійних запасах сировини II та III типів зміна запасів сировини I типу в межах $\Delta b_1 \in [-240; 240/7]$ не призведе до зміни оптимального плану двоїстої задачі, а екстремальне значення цільової функції (у разі відповідної зміни плану виробництва) зміниться на величину $\Delta Z_{\max} = 1/3 \cdot \Delta b_1$.

Отже, якщо змінюються запаси сировини тільки одного з типів, інтервал стійкості двоїстих оцінок наступний:

$$\Delta b_3 \in [-60;+\infty), \text{ якщо } \Delta b_1 = 0, \quad \Delta b_2 = 0;$$

$$\Delta b_2 \in [-240;400], \text{ якщо } \Delta b_1 = 0, \quad \Delta b_3 = 0;$$

$$\Delta b_1 \in [-240; 240/7], \text{ якщо } \Delta b_2 = 0, \quad \Delta b_3 = 0.$$

Слід перевірити, чи змінюється оптимальний план двоїстої задачі, якщо збільшити запаси всіх видів сировини на 20, 30 і 10 кг відповідно. Для цього необхідно підставити дані, наведені в системі нерівностей (6.6):

$$\begin{cases} 40 + 1/6 \cdot 20 - 0{,}1 \cdot 30 > 0 \\ 30 - 1/2 \cdot 20 + 0{,}15 \cdot 30 > 0 \\ 60 - 7/4 \cdot 20 + 0{,}25 \cdot 30 + 10 > 0 \end{cases}.$$

Оскільки система нерівностей виконується, то оптимальний план двоїстої задачі залишається незмінним. Збільшення сировини призведе до збільшення максимального значення цільової функції вихідної задачі на величину:

$$\Delta Z_{\max} = \Delta b_1 \cdot y_1^* + \Delta b_2 \cdot y_2^* + \Delta b_3 \cdot y_3^* = 20 \cdot {}^1\!/_3 + 30 \cdot {}^2\!/_5 + 10 \cdot 0 = {}^{56}\!/_3.$$

Слід зауважити, що доцільно розглянути економічний зміст вимог позитивності оцінок вихідної задачі, якщо її цільова функція досліджується на максимум. Якщо $\Delta_j = Z_j - c_j < 0$, то питома вага вартості ресурсів Z_j, які необхідні для збільшення рівня виробничої діяльності на одиницю, менше ніж питомий прибуток, тому необхідно ввести x_j в базис, що еквівалентно збільшенню небазисної змінної x_j від нуля до певного позитивного значення. Отже, потрібно перейти до нового опорного плану. Процес продовжується до тих пір, поки всі оцінки стануть позитивними, тобто для всіх видів сировини питома вага дорівнюватиме або стане меншою ніж питомий прибуток.

6.3. Двоїстий симплекс-метод

У теорії подвійності вихідна задача вирішується симплекс-методом, двоїста задача вирішується двоїстим симплекс-методом. Існує цілий клас задач, наприклад задача про дієту, умови якої природним чином формулюються відразу в другій стандартній формі. Для вирішення таких задач двоїстий симплекс-метод більш зручний, ніж звичайний алгоритм симплекс-методу.

Якщо розглядати першу симплексну таблицю з одиничним додатковим базисом, то маємо, що в стовпцях записана висхідна задача, а в рядках – двоїста. Оцінками плану висхідної задачі є c_j, а оцінками плана двоїстої задачі $-b_i$. Розв'язавши двоїсту задачу в симплексній таблиці, в якій записана висхідна задача, можна знайти оптимальний план двоїстої задачі, разом з оптимальним планом висхідної задачі. Цей метод називається ***двоїстим симплексним методом***.

Розглянемо висхідну задачу у вигляді:

$$Z = CX \to \min$$

$$AX = A_0,$$

$$X \geq 0,$$

тоді двоїста задача така:

$$F = YA_0^T \to \max$$

$$YA \leq C^T.$$

Нехай базис складають вектори $A_1, A_2, ..., A_l, ..., A_m$, при якому хоча б одна з компонент вектора $X = D^{-1}A_0 = (x_1, x_2, ..., x_l, ..., x_m)$ від'ємна, наприклад, $x_l < 0$. Для всіх векторів A_j виконується $Z_j - C_j \leq 0$. За теоремами двоїстості маємо план двоїстої задачі $Y = C_б D^{-1}$, але він не є оптимальним, оскільки базис містить від'ємну компоненту і не є планом висхідної задачі, при цьому оцінки оптимального плану двоїстої задачі мають бути невід'ємними. Тут вектор A_l слід виключити з базису висхідної задачі, а вектор, якому відповідає від'ємна оцінка, включити в базис двоїстої задачі. Якщо в $l-$му рядку не міститься $x_{lj} < 0$, то лінійна функція двоїстої задачі не обмежена в багатокутнику розв'язків, а висхідна задача не має розв'язків. Якщо ж в $l-$му рядку міститься $x_{lj} < 0$, то обчислюємо $\theta_{0j} = \min\left\{\frac{x_i}{x_{ik}}\right\} \geq 0$ та визначаємо вектор, що відповідає $\max(\theta_{0j}(Z_j - C_j))$ у ході розв'язування висхідної задачі на мінімум і $\min(\theta_{0j}(Z_j - C_j))$ у ході розв'язування висхідної задачі на максимум. Цей вектор включають в базис висхідної задачі. Вектор, що слід виключити з базису висхідної задачі визначається направляючим рядком.

Якщо $\theta_{0j} = \min\left\{\frac{x_i}{x_{ik}}\right\} = 0$, тобто $x_i = 0$, то направляючого вибирають x_{ij}, якщо $x_{ij} > 0$, то процес продовжуємо до отримання $X \geq 0$. Таким чином знаходимо оптимальний план двоїстої задачі, а отже, і оптимальний план висхідної задачі. За алгоритмом двоїстого симплексного методу умову $Z_j - C_j \leq 0$ можна не враховувати до виключення всіх $x_i < 0$, і далі оптимальний план знаходимо звичайним симплексним методом. Це зручно використовувати, якщо всі $x_i < 0$, тоді для переходу до плану висхідної задачі за одну ітерацію необхідно θ_{0j} визначити не по мінімуму, а по максимуму, тобто $\theta_{0j} = \max\left\{\frac{x_i}{x_{ik}}\right\} \geq 0$.

Двоїстий симплексний метод зручно використовувати в задачах лінійного програмування, в яких система обмежень містить вільні члени будь-якого знаку.

Приклад. Знайти $F = 2x_1 + 3x_2 \to \min$ за обмежень:

$$\begin{cases} 3x_1 + 2x_2 \geq 6 \\ x_1 + 4x_2 \geq 4 \\ x_1 \geq 0,\ x_2 \geq 0 \end{cases}$$

Приведемо задачу до канонічної форми.

$$\begin{cases} 3x_1 + 2x_2 - x_3 = 6 \\ x_1 + 4x_2 - x_4 = 4 \\ x_1 \geq 0,\ x_2 \geq 0 \end{cases} \Rightarrow \begin{cases} -3x_1 - 2x_2 + x_3 = -6 \\ -x_1 - 4x_2 + x_4 = -4 \\ x_1 \geq 0,\ x_2 \geq 0 \end{cases}.$$

Розв’язування прикладу здійснимо в симплекс-таблиці.

№ п/п	Базис	$c_{баз}$	c_j	2	3	0	0	Примітки
			A_0	A_1	A_2	A_3	A_4	
1	A_3	0	-6	**-3**	-2	1	0	$\theta = \min\left\{\frac{-b_i}{-a_{ik}}\right\}$

2	A_4	0	-4	-1	-4	0	1	
	Δ_1		0	-2	-3	0	0	
3	A_1	2	2	1	2/3	-1/3	0	(3)=(1):[-3]
4	A_4	0	-2	0	**-10/3**	-1/3	1	(4)=(2)+(3)
	Δ_2		4	0	-5/3	-2/3	0	
5	A_1	2	8/5	1	0	-2/5	1/5	(5)=(3)+(6)[-2/3]
6	A_2	3	3/5	0	1	1/10	-3/10	(6)=(4):[-10/3]
	Δ_3		5	0	0	-1/2	-1/2	

$F_{\min} = Z_{\max} = 5, \quad X^* = (8/5, 3/5, 0, 0).$

Запитання для самоперевірки

1. Дайте економічну інтерпретацію властивостей двоїстих оцінок.
2. Яка економічна інтерпретація двоїстих невідомих?
3. У чому полягає аналіз стійкості двоїстих оцінок?
4. У чому суть двоїстого симплекс-методу?

7. Аналіз лінійних моделей економічних оптимізаційних задач

7.1. Огляд задач в економіці, які зводяться до розв’язання задач лінійного параметричного програмування.

7.2. Задачі ЛП з параметрами у вільних членах обмежень.

7.1. Огляд задач в економіці, які зводяться до розв’язання задач лінійного параметричного програмування

Під час обґрунтування управлінського рішення в економічній задачі реального підприємства стикаються з проблемою, що значення обсягів ресурсів, ціни, прибуток, собівартість можуть змінюватись у відповідних межах. В даній ситуації важливо знати поведінку оптимального плану у разі зміни висхідних даних залежно від параметрів, які вводять в задачу для її розв’язання. Розроблено спеціальний розділ – параметричне програмування, в якому розглядаються оптимізаційні задачі з цільовими функціями й обмеженнями, що залежать від параметрів.

Розглянемо декілька прикладів типових оптимізаційних задач параметричного програмування. В задачі планування виробництва (3.1) – (3.2):

знайти максимум лінійної функції цілі

$$Z = \sum_{j=1}^{n} c_j x_j \to \max(\min),$$

за наявності лінійної системи обмежень-нерівностей:

$$\begin{cases} \sum\limits_{j=1}^{n} a_{ij} x_j \le b_i, \\ \quad (i = \overline{1, m}), \end{cases}$$

де усі $x_j \ge 0$; n – кількість видів продукції; m – кількість видів ресурсів; коефіцієнти c_j цільової функції виражають ціну одиниці продукції j-го виду, x_j – кількість одиниць j-ї продукції, що планується реалізовувати. Якщо ціна на продукцію залежить від сезону, то вона є функцією часу t, а для кожного виду продукції вона змінюється по-різному, що враховується коефіцієнтом α_j. Тому вартість продукції, що описується цільовою функцією, складається зі сталої частини – $\sum_{j=1}^{n} c_j x_j$ і змінної $\sum_{j=1}^{n} \alpha_j t\, x_j$, яка залежить від терміну зберігання t, де $t \in I_1$, I_1 – вибраний інтервал.

Нехай залежність ціни від параметра t лінійна, а обмеження задачі залишились тими ж, де a_{ij}, b_i – сталі, тоді ЗЛП буде мати вигляд:

$$Z = \sum_{j=1}^{n} (c_j + \alpha_j t) x_j \to \max(\min), \tag{7.1}$$

$$\begin{cases} \sum\limits_{j=1}^{n} a_{ij} x_j \le b_i, \\ x_j \ge 0, \quad (i = \overline{1, m}), t \in I_1. \end{cases} \tag{7.2}$$

До задачі (7.1) – (7.2) зводить і задача, в якій c_j означають кількість деякої продукції, яка виробляється за j-ю технологією, а x_j – відрізок терміну роботи за цією технологією за умови, що змінення інтенсивності виробництва за кожною технологією лінійно залежить від деякого параметру t. Тоді функція цілі виражає загальну кількість випущеної продукції за планом $(x_1,...,x_n)$ використання різних технологій залежно від параметра t з виконанням системи обмежень (7.2).

Якщо в задачі (3.1) – (3.2) можуть змінюватись обсяги ресурсів b_i залежно від параметра t, а решта коефіцієнтів задачі a_{ij}, c_j сталі, тоді маємо ЗЛП у вигляді:

$$Z=\sum_{j=1}^{n} c_j x_j \to \max(\min),$$

$$\begin{cases} \sum_{j=1}^{n} a_{ij} x_j \le b_i + \gamma_i\, t, \\ x_j \ge 0, \quad (i=\overline{1,m}), t \in I_2, \end{cases}$$

$(\gamma_1,...,\gamma_m)$ – задані значення, I_2 – заданий інтервал.

Існують також ситуації в економці, коли коефіцієнти a_{ij}, що є нормами витрат ресурсів на виробництво одиниці продукції також змінюються і залежать від деякого параметра t і при цьому і решта коефіцієнтів задачі c_j, b_i також залежать від одного параметра. Тоді задача має вигляд:

$$Z=\sum_{j=1}^{n} (c_j+\alpha_j t) x_j \to \max(\min),$$

$$\begin{cases} \sum_{j=1}^{n} (a_{ij}+\beta_{ij} t) x_j \le b_i + \gamma_i\, t, \\ x_j \ge 0, \quad (i=\overline{1,m}), t \in I_2, \end{cases}$$

при цьому допустимі області змінення параметра t, а також сталі $\alpha_j, \beta_{ij}, \gamma_i$ визначають, виходячи з реальних виробничих умов.

7.2. Задачі ЛП з параметрами у вільних членах обмежень

Володіння двоїстих невідомих y_i дозволяє встановити відносну цінність кожного дефіцитного ресурсу і з’ясувати, запаси яких ресурсів вигідно змінювати, і в якій пропорції. Однак неясно, до яких меж можна збільшувати запас одного, нехай навіть найбільш дефіцитного ресурсу, оскільки з його збільшенням будуть вичерпані запаси інших ресурсів і вже недостача інших ресурсів буде стримувати можливий подальший ріст прибутку. Для того щоб з'ясувати, в яких межах варіації ресурсів знайдений план залишається оптимальним і допустимим, потрібно вирішувати задачу з параметрами. У ході вирішення задач ЛП з параметрами у вільних членах $b = b_0 + b_1 \cdot t_1 + b_2 \cdot t_2$ в симплекс-таблицю вводимо кілька стовпців вільних членів b_0, t_1, t_2. Після визначення оптимального рішення по першому (основному) стовпцю b_0 слід виписати комбіноване рішення з урахуванням інших стовпців вільних членів і провести аналіз залежності рішення від параметрів. Якщо у разі зміни параметрів комбінований вільний член у будь-якому обмеженні змінює знак і стає від'ємним, то отримане рішення стає неприпустимим. У цьому випадку зручно скористатися ітерацією двоїстого симплекс-методу, в результаті якої умова оптимальності не буде порушена і буде отримане оптимальне рішення для іншої області зміни параметра.

Облік параметрів необхідний для перевірки стійкості знайденого статичного рішення, для визначення його чутливості до неминучих малих порушень умов задачі. Для ілюстрації цього розглянемо задачу.

Приклад. У табл. 7.1 наведені норми витрат ресурсів (сталь, кольорові метали, завантаження токарних верстатів) для виробництва двох видів виробів *A* і *B*. У цій же таблиці вказаний прибуток від виробництва і реалізації кожного виробу.

Таблиця 7.1

Види ресурсів	*A*	*B*	Запаси
Сталь, кг	10	70	570
Кольорові метали, кг	20	50	420
токарні верстати, верстато-г.	300	400	5 600
Прибуток, тис. грн	3	8	

Необхідно знайти оптимальний план виробництва продукції і перевірити його на стійкість стосовно варіацій запасів ресурсів.

Позначимо через x_1, x_2 – кількість виробів кожного виду і сформулюємо умови задачі у вигляді.

Слід знайти максимальне значення функції цілі.

$$Z = 3x_1 + 8x_2 \to \max,$$

за обмежень

$$\begin{cases} 10x_1 + 70x_2 \le 570, \\ 20x_1 + 50x_2 \le 420, \\ 300x_1 + 400x_2 \le 5600, \end{cases}$$

усі $x_j \ge 0$.

Можна скоротити перші дві нерівності на 10, а третя нерівність – на 100; це призведе до того, що норми витрат a_{ij}, запаси b_i, залишки y_i і варіації $t_i = \Delta b_i$ ресурсів будуть вимірюватися у десятках кг і сотнях верстато-годин відповідно.

У цій задачі необхідно знайти не тільки одне статичне рішення для конкретних значень вільних членів, але також дослідити оптимальне рішення на стійкість при можливих варіаціях вільних членів. При економічній інтерпретації потрібно не тільки отримати оптимальний план виробництва продукції при заданих запасах ресурсів, але також вирішити деякі виробничі питання, наприклад: які ресурси можна скоротити і на скільки, які ресурси вигідно збільшити, в якій пропорції і на скільки. Тому загальний стовпець вільних членів необхідно представити у вигляді чотирьох доданків $b = b_0 + b_1 t_1 + b_2 t_2 + b_3 t_3$ і ввести 3 параметри – по одному на кожен вільний член. Тут b_0 – вихідний стовпець вільних членів, а коефіцієнти стовпців параметрів t_2, t_3 утворюють в симплекс-таблиці одиничну матрицю.

Слід переписати систему обмежень з урахуванням параметрів:

$$\begin{cases} x_1 + 7x_2 \le 57 + t_1, \\ 2x_1 + 5x_2 \le 42 + t_2, \\ 3x_1 + 4x_2 \le 56 + t_3, \end{cases}$$

привести задачу до канонічної форми і заповнити симплекс-таблицю (табл. 7.2):

Таблиця 7.2

		C				3	8	0	0	0	
№	Базис	A_0	t_1	t_2	t_3	A_1	A_2	A_3	A_4	A_5	*Примітки*
1	A_3	57	1	0	0	1	7	1	0	0	
2	A_4	42	0	1	0	2	5	0	1	0	
3	A_5	56	0	0	1	3	4	0	0	1	
4	Δ_1	$Z_1 = 0$				-3	-8	0	0	0	

Вирішуємо задачу, орієнтуючись тільки на основний стовпець вільних членів b_0, не звертаючи поки уваги на стовпці параметрів t_1 , t_2 , t_3 . Спочатку потрібно ввести в базис x_2 (табл. 7.3)

Таблиця 7.3

		C				3	8	0	0	0	
№	Базис	A_0	t_1	t_2	t_3	A_1	A_2	A_3	A_4	A_5	*Примітки*
5	A_2	$^{57}/_7$	$^1/_7$	0	0	$^1/_7$	1	$^1/_7$	0	0	[1] : 7
6	A_4	$^9/_7$	$-^5/_7$	1	0	$^9/_7$	0	$-^5/_7$	1	0	[2] – [5]×5
7	A_5	$^{164}/_7$	$-^4/_7$	0	1	$^{17}/_7$	0	$-^4/_7$	0	1	[3] – [5]×4
8	Δ_2	$Z_2 =$ $^{456}/_7$	$^8/_7$	0	0	$-^{13}/_7$	0	$^8/_7$	0	0	

Тепер ввести в базис x_1, оскільки $\Delta_2 < 0$ (табл. 7.4).

Таблиця 7.4

		C				3	8	0	0	0	
№	Базис	A_0	t_1	t_2	t_3	A_1	A_2	A_3	A_4	A_5	*Примітки*
9	A_1	1	$-^5/_9$	$^7/_9$	0	1	0	$-^5/_9$	$^7/_9$	0	[6] : $^9/_7$
10	A_2	8	$^2/_9$	$-^1/_9$	0	0	1	$^2/_9$	$-^1/_9$	0	[5] – [9]×$^1/_7$
11	A_5	21	$^7/_9$	$-^{17}/_9$	1	0	0	$^7/_9$	$-^{17}/_9$	1	[7] – [9]×$^{17}/_7$
12	Δ_3	$Z_3 =$ 67	$^1/_9$	$^{13}/_9$	0	0	0	$^1/_9$	$^{13}/_9$	0	

Після цих двох ітерацій у рядку 12 (рядку функції цілі) більше немає від'ємних елементів. Треба виписувати отримане оптимальне рішення разом з інформацією в стовпцях t_1 , t_2 , t_3 :

$$x_1 = 1 - {}^5/_9\, t_1 + {}^7/_9\, t_2 ;$$

$$x_2 = 8 + {}^2\!/_9\, t_1 - {}^1\!/_9\, t_2\,;$$
$$\boldsymbol{x_5} = 21 + {}^7\!/_9\, t_1 - {}^{17}\!/_9\, t_2 + t_3\,;$$
$$\boldsymbol{x_3} = \boldsymbol{x_4} = 0\,;$$
$$z_{max} = 67 + {}^1\!/_9\, t_1 + {}^{13}\!/_9\, t_2\,.$$

Слід нагадати, що вільні члени функції цілі записані в симплекс-таблиці із зворотними знаками.

Потрібно зауважити, що стовпці параметрів $t_1 = \Delta b_1$, $t_2 = \Delta b_2$, $t_3 = \Delta b_3$ в таблицях повністю дублюються стовпцями залишків ресурсів $\boldsymbol{x_3}$, $\boldsymbol{x_4}$, $\boldsymbol{x_5}$, тому немає особливої необхідності їх спеціально обчислювати. У даному прикладі всі обмеження були обмеженнями-нерівностями одного типу (типу $\leq$), тому стовпці при балансових невідомих у вихідній таблиці повністю збігалися зі стовпцями параметрів. При обмеженнях-нерівностях протилежного типу (типу $\geq$) стовпці параметрів відрізняються від стовпців балансових невідомих тільки знаком. У всякому разі, у ході вирішення більшості задач ЛП із стандартної симплекс-таблиці можна виписати не один статичний розв'язок, а більш загальний розв'язок з урахуванням можливих варіацій запасів ресурсів (стовпці параметрів потрібно вводити тільки в разі обмежень-рівностей, для яких балансові невідомі тотожно рівні нулю). У даній задачі вийшло, що прибуток в оптимальному плані не залежить від $t_3 = \Delta b_3$.Це природно, оскільки в цьому оптимальному плані третій ресурс не витрачений повністю, залишок цього ресурсу відмінний від нуля $\boldsymbol{x_5} = 21$ (тобто можна скоротити надлишковий запас $b_3 = 56$ на 21 одиницю без скорочення виробництва). Перші два ресурси витрачені повністю $\boldsymbol{x_3} = \boldsymbol{x_4} = 0$, і саме це стримує подальше зростання прибутку. Розглядаючи залежність прибутку від параметрів $z_{max} = 67 + {}^1\!/_9\, t_1 + {}^{13}\!/_9\, t_2$, варто зауважити, що збільшення запасу 1-го ресурсу на одиницю (на 10 кг) призводить до збільшення прибутку на ${}^1\!/_9$, а збільшення запасу 2-го ресурсу призводить до збільшення прибутку на ${}^{13}\!/_9$ вартісних одиниць (тис. грн). З цих перших двох ресурсів другий є більш дефіцитним, більш корисним для отримання максимального прибутку. У процесі розширення виробництва треба в першу чергу вкладати кошти у збільшення запасу другого ресурсу.

Найбільша віддача буде при збільшенні ресурсів у пропорції

$$\Delta b_1 : \Delta b_2 : \Delta b_3 = {}^1\!/_9 : {}^{13}\!/_9 : 0 = 1 : 13 : 0.$$

Але як дізнатися, до яких меж можна змінювати запаси ресурсів?

Необхідно спочатку варіювати запас нпйбільш дефіцитного другого ресурсу (кольорові метали) і розглянути, в яких межах зміни параметра t_2 знайдений план залишається допустимим. За будь-якими можливими значеннями параметра компоненти плану повинні залишатися невід'ємними:

$x_1 = 1 + {}^7/_9\, t_2 \geq 0;$ звідки $t_2 \geq -{}^9/_7 = -1{,}286;$

$x_2 = 8 - {}^1/_9\, t_2 \geq 0;$ звідки $t_2 \leq 72;$

$x_5 = 21 - {}^{17}/_9\, t_2 \geq 0;$ звідки $t_2 \leq {}^{189}/_{17} = 11{,}118;$

$\boldsymbol{x_3} = \boldsymbol{x_4} = 0;$

$z_{max} = 67 + {}^{13}/_9\, t_2\,.$

Виявляється, знайдений план залишається оптимальним і допустимим тільки в інтервалі варіювання параметра $-1{,}286 \leq t_2 \leq 11{,}118$, або ж для таких значень запасу другого ресурсу: $40{,}714 \leq b_2 \leq 53{,}118$. Занепокоєння тут викликає нижня межа – варто тільки зменшити заданий запас $b_2 = 43$ всього на 1,3 одиниці (тобто на 3 %), як план стає неприпустимим (компонента x_1 буде негативною). Як кажуть, знайдений план нестійкий щодо малих змін вільних членів. Природно, щоб не було неприємної несподіванки, треба заздалегідь знайти оптимальне рішення для $b_2 \leq 40{,}714$.

Варто продублювати далі попередню табл. 7.4 (рядки 9 – 12):

№	Базис	A_0	t_1	t_2	t_3	A_1	A_2	A_3	A_4	A_5	*Примітки*
9	A_1	1	$-{}^5/_9$	${}^7/_9$	0	1	0	$-{}^5/_9$	${}^7/_9$	0	[1] : ${}^9/_7$
10	A_2	8	${}^2/_9$	$-{}^1/_9$	0	0	1	${}^2/_9$	$-{}^1/_9$	0	[5] – [9]×${}^1/_7$
11	A_5	21	${}^7/_9$	$-{}^{17}/_9$	1	0	0	${}^7/_9$	$-{}^{17}/_9$	1	[7] – [9]×${}^{17}/_7$
12	Δ_3	$Z_3 =$ 67	${}^1/_9$	${}^{13}/_9$	*0*	*0*	*0*	${}^1/_9$	${}^{13}/_9$	*0*	

При $b_2 < 40{,}714$ ($t_2 < -1{,}286$) комбінований вільний член в 1-му обмеженні (9-й рядок таблиці) $b_2 = 1 + {}^7/_9\, t_2$ стає негативним.

Потрібно виправити цю ситуацію однією ітерацією двоїстого симплекс-методу, для чого в рядку 9 знаходимо негативні елементи (якщо таких немає, то задача для цих значень параметра рішення не має); в рядку 9 є тільки один негативний елемент (виділений рамочкою і кольором), варто взяти його за розв'язний і провести заміну базису ітерацією Гаусса – Жордана (табл. 7.5):

Таблиця 7.5

№	Базис	A_0	t_1	t_2	t_3	A_1	A_2	A_3	A_4	A_5	*Примітки*
13	A_3	$-9/5$	1	$-7/5$	*0*	$-9/5$	0	1	$-7/5$	0	[9] : ($-5/9$)
14	A_2	$42/5$	0	$1/5$	*0*	$2/5$	1	0	$1/5$	0	[10] – [13]$\times 1/7$
15	A_5	$112/5$	0	$-4/5$	*1*	$7/5$	0	0	$-4/5$	1	[11] – [13]$\times 17/7$
16	Δ	$Z = 336/5$	0	$8/5$	*0*	$1/5$	0	0	$8/5$	0	

Необхідно виписувати новий оптимальний розв'язок і визначити припустимі межі варіювання параметра t2, для яких справедливий знайдений розв'язок:

$x_3 = -9/5 - 7/5\, t_2 \geq 0;$ звідки $t_2 \leq -9/7 = -1{,}286;$

$x_2 = 42/5 + 1/5\, t_2 \geq 0;$ звідки $t_2 \geq -42;$

$\mathbf{x_5} = 112/5 - 4/5\, t_2 \geq 0;$ звідки $t_2 \leq 28;$

$x_1 = x_4 = 0;$

$z_{max} = 336/5 + 8/5\, t_2 .$

Новий розв'язок справедливий на інтервалі $-42 \leq t_2 \leq -1{,}286$, або $0 \leq b_2 \leq 40{,}714$. З фізичних міркувань очевидно, що параметр t_2 не може бути менше –42, оскільки запас другого ресурсу не може бути негативним (до речі, з рядка 14 випливає, що при $t_2 < -42$ невід'ємних розв'язків немає, так як у цьому рядку з від'ємним вільним членом немає невід'ємних коефіцієнтів при вільних невідомих).

Виключно в навчальних цілях необхідно знайти оптимальні розв'язки даної задачі для всіх значень параметра t_2 ($-42 \leq t_2 < \infty$) при $t_1 = t_3 = 0$. Вже знайдені розв'язки в діапазонах $-42 \leq t_2 \leq -1{,}286$ і $-1{,}286 \leq t_2 \leq 11{,}118$. Для аналізу діапазону $t_2 > 11{,}118$ потрібно повернутися до попередньої табл. 7.4 (рядки 9 – 12).

№	Базис	A_0	t_1	t_2	t_3	A_1	A_2	A_3	A_4	A_5	*Примітки*
9	A_1	1	$-5/9$	$7/9$	0	1	0	$-5/9$	$7/9$	0	
10	A_2	8	$2/9$	$-1/9$	0	0	1	$2/9$	$-1/9$	0	
11	A_5	21	$7/9$	$-17/9$	1	0	0	$7/9$	$-17/9$	1	
12	Δ	$Z = 67$	$1/9$	$13/9$	0	0	0	$1/9$	$13/9$	0	

Знову аналізуючи розв'язок на інтервалі $-1{,}286 \le t_2 \le 11{,}118$, переконуємося, що при $t_2 > 11{,}118$ комбінований вільний член $b_3 = 21 - {}^{17}/_9\, t_2$ у 3-му обмеженні (рядок 11) змінює знак. У цьому рядку є один від'ємний елемент, варто приймати його за розв'язний; у результаті ітерації двоїстого симплекс-методу з цим розв'язним елементом отримуємо (в рядках 17 – 20) черговий розв'язок, який справедливий для всіх $t_2 > 11{,}118$ (табл. 7.6):

Таблиця 7.6

№	Базис	A_0	t_1	t_2	t_3	A_1	A_2	A_3	A_4	A_5	*Примітки*
17	A_4	$-189/17$	$-7/17$	1	$-9/17$	0	0	$-7/17$	1	$-9/17$	[11] : $(-17/9)$
18	A_1	$164/17$	$-4/17$	0	$7/17$	1	0	$-4/17$	0	$7/17$	[9] – [17]$\times 7/9$
19	A_2	$115/17$	$3/17$	0	$-1/17$	0	1	$3/17$	0	$-1/17$	[10] + [17]$\times 1/9$
20	Δ	$Z = 1412/17$	$12/17$	0	$13/17$	0	0	$12/17$	0	$13/17$	

Виписуємо останній розв'язок:

$\boldsymbol{x_4} = -{}^{189}/_{17} + t_2 \ge 0$; звідки $t_2 \ge {}^{189}/_{17} = 11{,}118$;

$x_1 = {}^{164}/_{17} \ge 0$; $x_2 = {}^{115}/_{17} \ge 0$; $z_{max} = {}^{1412}/_{17}$.

Отже, знайдені оптимальні розв'язки для всіх значень параметра t_2. Далі наведена зведена таблиця розв'язків для всіх значень запасу кольорових металів $b_2 = 42 + t_2$, десятки кг (табл. 7.7):

Таблиця 7.7

$b_2 < 0$	Розв'язків немає		Немає
$0 \le b_2 \le 40{,}714$	$x_1 = 0$	$x_2 = 0{,}2\ b_2$	$z_{max} = 1{,}6\ b_2$
$40{,}714 \le b_2 \le 53{,}118$	$x_1 = 0{,}778\ b_2 - 31{,}667$	$x_2 = 12{,}667 - 0{,}111\ b_2$	$z_{max} = 1{,}444\ b_2 + 6{,}333$
$b_2 > 53{,}118$	$x_1 = 9{,}647$	$x_2 = 8{,}765$	$z_{max} = 83{,}059$

Варто раз відзначити, що стовпці параметрів t_i повністю дублюються в таблицях стовпцями балансових невідомих y_i, тому не було особливої необхідності заздалегідь передбачати їх у вихідній симплекс-таблиці.

Запитання для самоперевірки

1. Сформулюйте задачу лінійного програмування з параметрами у вільних членах.

2. Для яких цілей потрібно врахування параметрів у вільних членах?

3. Приведіть етапи розв'язання задачі лінійного програмування з параметрами у вільних членах.

8. Транспортна задача. Методи розв'язання транспортної задачі

8.1. Загальна постановка транспортної задачі.

8.2. Способи складання першого базисного плану.

8.3. Критерій оптимальності. Метод потенціалів.

8.4. Виродження плану транспортної задачі.

8.1. Загальна постановка транспортної задачі

Маємо m пунктів відправлення $A_1, A_2, ..., A_m$ і n пунктів призначення $B_1, B_2, ..., B_n$. Відомі запаси a_i продукції у кожного постачальника A_i, потреби b_j кожного споживача B_j і тарифи c_{ij} (вартість перевезення одиниці товару від пункту A_i в пункт B_j). Потрібно скласти оптимальний план перевезення, тобто визначити, яка кількість вантажу x_{ij} має бути відправлена з кожного пункту відправлення в кожний пункт призначення, щоб загальна вартість всіх перевезень була найменшою.

Дано: тарифи, запаси і потреби

Поста-чальники	Споживачі				Запа-си
	B_1	B_2		B_n	
A_1	c_{11}	c_{12}		c_{1n}	a_1
A_2	c_{21}	c_{22}		c_{2n}	a_2
…	…	…		…	…
A_m	c_{m1}	c_{m2}		c_{mn}	a_m
Потреби	b_1	b_2		b_n	

Знайти: оптимальний план перевезення вантажів

Поста-чальники	Споживачі				Запа-си
	B_1	B_2		B_n	
A_1	x_{11}	x_{12}		x_{1n}	a_1
A_2	x_{21}	x_{22}		x_{2n}	a_2
…	…	…		…	…
A_m	x_{m1}	x_{m2}		x_{mn}	a_m
Потреби	b_1	b_2		b_n	

Якщо з кожного пункту відправлення в кожний пункт призначення заплановано перевезення x_{ij} кількість вантажу, то вартість кожного перевезення буде $c_{ij}x_{ij}$, а вартість всього плану перевезень – $\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}x_{ij}$. Кількість вивезеного товару від A_i не може перевищувати його запасу a_i:

$$\sum_{j=1}^{n}x_{ij} \le a_i :$$

кількість привезеного товару в B_j не може перевищувати потреби b_j:

$$\sum_{i=1}^{m}x_{ij} \le b_j .$$

Якщо $\sum_{i=1}^{m}a_i = \sum_{j=1}^{n}b_j$, то транспортна задача називається збалансованою або закритою. У цьому випадку весь вироблений товар повинен бути вивезений, усі потреби повинні бути задоволені:

$$\begin{cases} \sum_{j=1}^{n}x_{ij} = a_i, \\ \sum_{i=1}^{m}x_{ij} = b_j. \end{cases}$$

Математична модель транспортної задачі:

Знайти найменше значення лінійної функції:

$$F = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \to min, \qquad (8.1)$$

при обмеженнях

$$\begin{cases} \sum_{j=1}^{n} x_{ij} = a_i,\ i = \overline{1,m}, \\ \sum_{i=1}^{m} x_{ij} = b_j,\ j = \overline{1,n}, \\ x_{ij} \geq 0. \end{cases} \qquad (8.2)$$

Теорема. Будь-яка транспортна задача, у якій сумарний обсяг запасів дорівнює сумарному обсягу потреб, має розв'язок.

Доведення. Для доведення теореми слід показати, що при заданих умовах існує хоча б один план задачі і лінійна функція на множині планів обмежена.

Нехай $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j = M > 0$. Тоді величини $x_{ij} = \frac{a_i b_j}{M}$ є планом, оскільки вони задовольняють системі обмежень (8.2), а саме підставляючи x_{ij} в дану систему обмежень маємо:

$$\sum_{j=1}^{n} x_{ij} = \sum_{j=1}^{n} \frac{a_i b_j}{M} = \frac{a_i}{M}\sum_{j=1}^{n} b_j = \frac{a_i}{M}M = a_i,$$

$$\sum_{j=1}^{n} x_{ij} = \sum_{j=1}^{n} \frac{a_i b_j}{M} = \frac{b_j}{M}\sum_{j=1}^{n} a_i = \frac{b_j}{M}M = b_j.$$

Виберемо зі значень c_{ij} найбільше $C' = max c_{ij}$ і замінимо в лінійній функції (8.1) всі коефіцієнти на C', тоді, враховуючи $\sum_{j=1}^{n} x_{ij} = a_i$, отримаємо

$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \leq C'\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} = C'\sum_{i=1}^{m} a_i = C'M.$$

Виберемо зі значень c_{ij} найменше $C'' = min c_{ij}$ і замінимо в лінійній функції (8.1) всі коефіцієнти на C'', тоді, враховуючи $\sum_{j=1}^{n} x_{ij} = a_i$, отримаємо

$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} \geq C''\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} = C''\sum_{i=1}^{m} a_i = C''M.$$

Об'єднуючи розглянуті дві нерівності в одну, отримаємо

$$C'M \leq Z \leq C''M,$$

тобто лінійна функція обмежена на множині планів транспортної задачі. ■

Якщо транспортна задача незбалансована, то її можна закрити введенням фіктивного постачальника або фіктивного споживача:

якщо $\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j$ вводиться фіктивний постачальник $A_ф$, запаси якого дорівнюють $a_ф = \sum_{j=1}^{n} b_j - \sum_{i=1}^{m} a_i$, тарифи при цьому $c_ф = 0$;

якщо $\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$ вводиться фіктивний споживач $B_ф$, потреби якого дорівнюють $b_ф = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j$, тарифи при цьому $c_ф = 0$.

8.2. Способи складання першого базисного плану

Оптимальний план перевезень треба шукати тільки серед базисних планів, в яких число відкритих маршрутів (ненульових x_{ij}) найменше.

Система обмежень транспортної задачі містить mn невідомих і $m+n$ рівнянь, але так як $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$, то одне з рівнянь виявляється зайвим, і система обмежень має містити $r = m + n - 1$ лінійно незалежних рівнянь, а отже не вироджений опорний план має включати $m + n - 1$ додатних компонент або перевезень.

Якщо транспортна задача розв'язується у таблиці, то клітинки, які мають відмінні від нуля перевезення, називають зайнятими, решту клітинок – незайнятими. Зайняті клітинки відповідають базисним невідомим, і для невиродженого опорного плану їх кількість дорівнює $m + n - 1$. Опорність плану обумовлюється його ациклічністю, тобто в таблиці не можна побудувати замкнутий цикл, всі вершини якого знаходяться в зайнятих клітинках.

Циклом називають послідовну сукупність клітинок вигляду $(i_1 j_1)(i_1 j_2)\times(j_2 i_2)$ ... $(j_1 i_m)$, в якому дві і тільки дві сусідні клітинки розміщені в одному стовпці або одному рядку таблиці, причому остання клітинка знаходиться в тому ж рядку або стовпці, що і перша. Побудова циклів розпочинається з будь-якої зайнятої клітинки і потім переходять по стовпчику або по рядку в іншу зайняту клітинку, при цьому роблять повороти під прямим кутом у зайнятих клітинках і рухаються до іншої зайнятої клітинки по рядку або по стовпцю для повернення в першопочаткову клітинку. Якщо таке повернення можливе, то отриманий цикл і план не є опорним, в противному випадку – опорний.

Для складання першого опорного плану розроблено декілька методів, з яких найпростішими є діагональний метод (північно-західного кута) і метод найменшої вартості.

Приклад. Маємо 3 пункти відправлення і 4 пункти призначення. Відомі запаси a_i продукції у кожного постачальника A_i, потреби b_j кожного споживача B_j і тарифи c_{ij} (вартість перевезення одиниці товару від пункту A_i в пункт B_j). Потрібно скласти оптимальний план перевезення, тобто визначити, яка кількість вантажу x_{ij} має бути відправлена з кожного пункту відправлення в кожний пункт призначення, щоб загальна вартість всіх перевезень була найменшою.

Постачальники	Споживачі				Запаси
	B_1	B_2	B_3	B_4	
A_1	4	3	2	7	46
A_2	1	1	6	4	34
A_3	3	5	9	4	40
Потреби	40	35	30	45	120 150

Побудувати початковий опорний план 1) методом північно-західного кута або його ще називають діагональним методом; 2) методом мінімальної вартості.

Задачу спочатку слід перевірити на збалансованість. Задача не є збалансованою, оскільки 120<150. Маємо $\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j$ то слід ввести фіктивного постачальника A_4, запаси якого дорівнюють $a_4 = \sum_{j=1}^{n} b_j - \sum_{i=1}^{m} a_i = 150 - 120 = 30$, тарифи при цьому $c_{ф} = 0$. Тепер маємо збалансовану задачу:

Постачальники	Споживачі				Запаси
	B_1	B_2	B_3	B_4	
A_1	4	3	2	7	46
A_2	1	1	6	4	34
A_3	3	5	9	4	40
A_4	0	0	0	0	30
Потреби	40	35	30	45	150 150

1) у методі північно-західного кута (діагональному методі) заповнювати таблицю розпочинають з клітинки першого стовпця першого рядка. Для цього порівнюють $a_1 = 46$ з $b_1 = 40$, $b_1 < a_1$ і менший з обсягів записуємо в клітинку, що стає заповненою, при цьому потреби першого споживача задоволені повністю (решту клітинок в першому стовпці обнуляємо), але запаси першого постачальника повністю не вичерпані. Тому потреби другого споживача обсягом на 6 одиниць можна задовольнити за рахунок першого постачальника (перехід до другого стовпця) та решту потреб (36-6=29) обсягом 29 одиниць вантажу привезти з другої бази (перехід на другий рядок). Процес продовжуємо до тих пір, доки не задовольнимо потреби всіх споживачів за рахунок запасів постачальників:

Постачальники	Споживачі				Запаси
	B_1	B_2	B_3	B_4	
A_1	4 40	3 6	2 0	7 0	46/6/0
A_2	1 0	1 29	6 5	4 0	34/5/0
A_3	3 0	5 0	9 25	4 15	40/15/0
A_4	0 0	0 0	0 0	0 30	30
Потреби	40	35/29/0	30/25/0	45	150 150

Побудований план є опорним і невиродженим, оскільки кількість зайнятих клітинок у таблиці дорівнює 7, що дорівнює $m + n - 1 = 4 + 4 - 1 = 7$. Загальна вартість складеного плану перевезення дорівнює сумі добутків обсягів перевезень на відповідні вартості, тобто

$F = 40 \cdot 4 + 6 \cdot 3 + 29 \cdot 1 + 5 \cdot 6 + 25 \cdot 9 + 15 \cdot 4 = 522$ грн;

2) у методі мінімальної вартості із всієї таблиці вартостей перевезення вибирають найменшу вартість і в клітинку, що їй відповідає, розміщують найменше з чисел a_i або b_j. Далі з розгляду виключають або рядок, що відповідає постачальнику, запаси якого повністю використані, або стовпець, що відповідає споживачу, потреби якого повністю задоволені, або і рядок, і стовпець, якщо використані і запаси постачальника і задоволені потреби споживача. Й далі з решти частини таблиці вартостей вибирають найменшу вартість, і процес розподілу запасів продовжують доки всі вони повністю не будуть розподілені, а потреби – задоволені.

Постачальники	Споживачі				Запаси
	B_1	B_2	B_3	B_4	
A_1	4 0	3 16	2 30	7 0	46/16/0
A_2	1 34	1 0	6 0	4 0	34/0
A_3	3 6	5 0	9 0	4 34	40/34/0
A_4	0 0	0 19	0 0	0 11	30
Потреби	40/6/0	35/19/0	30/0	45	150 150

Отже, побудову початкового опорного плану розпочинаємо з клітинки другого рядка і першого стовпця, оскільки $c_{21} = 1$ (хоча можна було б розпочати з $c_{22} = 1$). Отже в клітинку з індексами 2 1 заносимо число 34, тому що воно є мінімальним з чисел $a_2 = 34$ та $b_1 = 40$, а другий рядок виключаємо з подальшого розгляду, тому від другого постачальника повністю вивезений вантаж. З решти таблиці знову здійснюємо вибір клітинки з мінімальною вартістю і вилучаємо з таблиці відповідний рядок або стовпець залежно від того, що вичерпано. Стовпець або рядок, які відповідають фіктивному споживачу або фіктивному постачальнику, розглядаються в останню чергу.

Побудований план є опорним і невиродженим, оскільки кількість зайнятих клітинок у таблиці дорівнює 7, що дорівнює $m + n - 1 = 4 + 4 - 1 = 7$. Загальна вартість складеного плану перевезення дорівнює:

$$F = 3 \cdot 16 + 2 \cdot 30 + 1 \cdot 34 + 6 \cdot 3 + 4 \cdot 34 = 296 \text{ грн.}$$

Маємо, що загальна вартість перевезення вантажу, початково зпланованого методом мінімальної вартості, набагато менша, ніж методом північно-західного кута.

8.3. Критерій оптимальності. Метод потенціалів

Знайдений висхідний опорний план перевіряється на оптимальність методом потенціалів за критерієм, який стверджує наступна теорема.

Теорема. Якщо план $X^*\left(x_{ij}^*\right)$ транспортної задачі є оптимальним, то йому відповідає система з $m+n$ чисел u_i^* і v_j^*, які задовольняють вимоги:

$$u_i^* + v_j^* = c_{ij} \quad \text{для } x_{ij}^* > 0,$$
$$u_i^* + v_j^* \le c_{ij} \quad \text{для } x_{ij}^* = 0.$$

Числа u_i^* і v_j^* називаються потенціалами, відповідно, постачальників і споживачів.

Доведення. Маємо транспортну задачу, в якій мінімізується лінійна функція $F = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij} x_{ij} \to \min$ за обмежень:

$$\begin{cases} x_{i1} + x_{i2} + \ldots + x_{in} = a_i,\ i = \overline{1,m}, \\ x_{1j} + x_{2j} + \ldots + x_{mj} = b_j,\ j = \overline{1,n}, \end{cases}$$
$$x_{ij} \ge 0.$$

Розглянемо її як двоїсту деякій висхідній задачі лінійного програмування, умови якої отримують відомими правилами складання двоїстих задач. Кожному обмеженню виду $x_{i1} + x_{i2} + \ldots + x_{in} = a_i$ висхідної задачі відповідає змінна u_i $\left(i = \overline{1,m}\right)$, а кожному обмеженню виду $x_{1j} + x_{2j} + \ldots + x_{mj} = b_j$ – змінна v_j $\left(j = \overline{1,n}\right)$, при цьому слід максимізувати функцію $Z = \sum_{i=1}^{m} a_i u_i + \sum_{j=1}^{n} b_j v_j \to \max$ при обмеженнях $u_i + v_j \le c_{ij}$. План X^* – оптимальний план двоїстої задачі, тому план $Y^* = \left(u_i^*, v_j^*\right)$ є планом висхідної задачі і на основі теореми двоїстості $F_{\min} = Z_{\max}$ або:

$$\sum_{i=1}^{m} a_i u_i^* + \sum_{j=1}^{n} b_j v_j^* = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij} x_{ij}^*, \quad x_{ij}^* \ge 0.$$

Отже, на основі теорем двоїстості отримуємо, що обмеження висхідної задачі, які відповідають додатним компонентам оптимального плану двоїстої задачі, виконуються як строгі рівності, а відповідні компоненти, дорівнюють нулю – як нерівності, тобто:

$$u_i + v_j = c_{ij} \text{ для } x_{ij}^* > 0,$$

$$u_i + v_j \le c_{ij} \text{ для } x_{ij}^* = 0. \ \blacksquare$$

Отже, з доведеної теореми випливає, що потенціали u_i і v_j знаходять з рівняння $u_i + v_j = c_{ij}$, справедливого для зайнятих клітинок. Одному з потенціалів дається довільне значення, наприклад $u_1 = 0$, тоді інші потенціали визначаються однозначно. Для того, щоб первісний опорний план був оптимальний, необхідне виконання наступних умов:

1) для кожної зайнятої клітини сума потенціалів повинна дорівнювати вартості одиниці перевезення, що стоїть в цій клітинці:

$$u_i + v_j = c_{ij};$$

2) для кожної незайнятої клітини сума потенціалів повинна бути менше або дорівнювати вартості одиниці перевезення, що стоїть в цій клітинці:

$$u_i + v_j \le c_{ij}.$$

Якщо хоча б одна незайнята клітина не задовольняє умову, то опорний план є неоптимальним і його можна поліпшити, вводячи в базис вектор, відповідний клітинці, для якої порушується умова оптимальності.

Таким чином, для перевірки плану на оптимальність необхідно спочатку побудувати систему потенціалів.

Часто обчислюють $\Delta_{ij} = (u_i + v_j) - c_{ij}$. Цю оцінку називають оцінкою вільних клітинок. Якщо $\Delta_{ij} \le 0$, то опорний план є оптимальним. Якщо хоча б одна з оцінок $\Delta_{ij} > 0$, то опорний план не є оптимальним і його можна поліпшити, перейшовши від одного опорного плану до іншого.

Перевіримо опорний, невироджений план, отриманий методом мінімальної вартості, на оптимальність. Для цього побудуємо таблицю потенціалів.

u \ v	$v_1 = 2$	$v_2 = 3$	$v_3 = 2$	$v_4 = 3$
$u_1 = 0$	2 / 4	**3**	**2**	3 / 7
$u_2 = -1$	**1**	2 / 1	1 / 6	2 / 4
$u_3 = 1$	**3**	4 / 5	3 / 9	**4**
$u_4 = -3$	-1 / 0	**0**	-1 / 0	**0**

Для обчислення невідомих u_i і v_j використовуємо рівняння:

$$
\begin{aligned}
&u_1 + v_2 = 3, \quad u_1 = 0,\\
&u_1 + v_3 = 2,\\
&u_2 + v_1 = 1,\\
&u_3 + v_1 = 3,\\
&u_3 + v_4 = 4,\\
&u_4 + v_2 = 0,\\
&u_4 + v_4 = 0.
\end{aligned}
$$

Обчислюємо оцінки Δ_{ij}: $\Delta_{11} = 2 - 4 = -2 < 0$, $\Delta_{14} = 3 - 7 = -4 < 0$, $\Delta_{22} = (-1 + 3) - 1 > 0$, $\Delta_{23} = 1 - 6 = -5 < 0$, $\Delta_{24} = 2 - 4 = -2 < 0$, $\Delta_{32} = 4 - 5 = -1 < 0$, $\Delta_{33} = 3 - 9 = -6 < 0$, $\Delta_{41} = -1 - 0 = -1 < 0$, $\Delta_{43} = -1 - 0 = -1 < 0$. В одному випадку маємо, що $\Delta_{22} = (-1 + 3) - 1 > 0$. Отже, опорний план не є оптимальним і його слід поліпшити.

Перехід до іншого базисного плану

Наявність позитивної оцінки вільної клітини ($\Delta_{ij} > 0$) у ході перевірки опорного розв’язку на оптимальність свідчить про те, що отриманий розв’язок не є оптимальним і для зменшення значення цільової функції треба перейти до іншого опорного рішення. При цьому необхідно перерозподілити вантажі, переміщаючи їх із зайнятих клітин у вільні. Вільна клітина стає зайнятою, а одна з раніше зайнятих клітин – вільною.

Для вільної клітини із $\Delta_{ij} > 0$ будується цикл (ланцюг), всі вершини якого крім однієї знаходяться в зайнятих клітинах; кути прямі, число вершин парне. Біля вільної клітини циклу із $\Delta_{ij} > 0$ ставиться знак «+», потім по черзі проставляються знаки «–» і «+».У вершин зі знаком «–» обирають мінімальний вантаж (ρ), його додають до вантажів, що стоять біля вершин зі знаком «+», і віднімають від вантажів у вершин зі знаком «–». У результаті перерозподілу вантажу отримаємо новий опорний план. Цей план перевіряємо на оптимальність, і так далі до тих пір, поки не отримаємо оптимальний розв’язок.

Оптимальний план повинен розшукуватись тільки серед базисних планів з одним і тим же числом відкритих маршрутів (число ненульових базисних невідомих незмінне і дорівнює рангу матриці системи обмежень задачі). Якщо відкривається новий маршрут (у базис вводитися невідома x_{pq}), один зі старих маршрутів має бути закритий.

При призначенні нових обсягів перевезень x_{ij} треба стежити за дотриманням балансів по рядках і стовпцях таблиці (всі запаси повинні бути вивезені, всі потреби – задоволені).

Продовжимо розвязування прикладу. Оскільки опорний план не є оптимальним і його слід поліпшити, то слід відкрити новий маршрут від A_2 до B_2 з тарифом $c_{22} = 1$ (оскільки $\Delta_{22} > 0$). Необхідно ввести в базис невідому x_{22}.

Постачальники (План 2)	Споживачі				Запаси
	B_1	B_2	B_3	B_4	
A_1	4 0	3 16	2 30	7 0	46
A_2	1 34 −	1 0 $+\rho$	6 0	4 0	34
A_3	3 6 +	5 0	9 0	4 34 −	40
A_4	0 0	0 19 −	0 0	0 11 +	30
Потреби	40	35	30	45	150

Напрямок обходу ланцюга не має значення. Важливо, щоб всі вузли ланцюга, крім початкового, розташовувалися в зайнятих базисних клітинах.

Загальне число маршрутів не повинно збільшитися (в іншому випадку новий план не буде базисним). Вибираємо серед клітин зі знаком «–» вантаж мінімальної величини: $\rho = min\{34, 19, 34\} = 19$. До всіх клітин зі знаком «+» або «–» розносимо вантаж. Маршрут x_{42} закривається.

Новий розподіл вантажу між постачальниками та споживачами такий:

Постачальники	Споживачі				Запаси
	B_1	B_2	B_3	B_4	
A_1	4 0	3 16	2 30	7 0	46
A_2	1 15	1 19	6 0	4 0	34
A_3	3 25	5 0	9 0	4 15	40
A_4	0 0	0 0	0 0	0 30	30
Потреби	40	35	30	45	150

Побудований план є опорним і невиродженим, оскільки кількість зайнятих клітинок у таблиці дорівнює 7. Загальна вартість даного опорного плану перевезення дорівнює:

$$F = 3 \cdot 16 + 2 \cdot 30 + 1 \cdot 15 + 19 \cdot 1 + 3 \cdot 25 + 15 \cdot 4 = 277 \text{ грн.}$$

Методом потенціалів перевіримо опорний план на оптимальність в таблиці:

v u	**v_1 = 3**	**v_2 = 3**	**v_3 = 2**	**v_4 = 4**
u_1 = 0	3 4	**3**	**2**	4 7
u_2 = -2	**1**	**1**	0 6	2 4
u_3 = 0	**3**	3 5	2 9	**4**
u_4 = -4	-1 0	-1 0	-1 0	**0**

За змістом таблиці маємо висновок, що план оптимальний, оскільки всі $\Delta_{ij} = (u_i + v_j) - c_{ij} < 0$.

Економічна оцінка $\Delta_{ij} = c_{ij} - (u_i + v_j) < 0$ показує, на скільки грошових одиниць зменшаться транспортні витрати від завантаження даної клітини одиницею вантажу. Ефективність плану від завантаження потенційної клітини вантажем ρ одиниць становить $\Delta F = \Delta_{ij} \cdot \rho$ грошових одиниць.

8.4. Виродження плану транспортної задачі

У ході розв’язання транспортної задачі може виявитися, що число зайнятих клітин менше, ніж $r = m + n - 1$. У цьому випадку завдання має вироджений розв'язок. Для можливого його виключення доцільно поміняти місцями постачальників і споживачів або ввести у вільну клітку з найменшою вартістю нульову поставку. Нуль поміщають в таку клітину, щоб у кожному рядку і кожному стовпці було не менше однієї зайнятої клітини. Дуже часто план вироджується у процесі його поліпшення, коли при відкритті нового маршруту закривається відразу кілька старих маршрутів.

Продовжимо розгляд прикладу. Якщо ввести в базис x_{24}, то закриваються відразу два маршрути x_{21} і x_{34}. У цьому випадку слід закрити тільки один маршрут з найбільшою вартістю, інші ж залишити в плані, але з нульовими обсягами перевезень, так, щоб загальне число маршрутів формально не змінилося (в плані є базисний нуль $x_{21} = 0$).

Постачальники	Споживачі				Запаси
	B_1	B_2	B_3	B_4	
A_1	4 0	3 16	2 30	7 0	46
A_2	1 15 $-$	1 19	6 0	4 0 $+\rho$	34
A_3	3 25 $+$	5 0	9 0	4 15 $-$	40
A_4	0 0	0	0 0	0 30 $+$	30
Потреби	40	35	30	45	150

Новий план буде виродженим:

Постачальники	Споживачі				Запаси
	B_1	B_2	B_3	B_4	
A_1	4 0	3 16	2 30	7 0	46
A_2	1 0	1 19	6 0	4 15	34
A_3	3 40	5 0	9 0	4 0	40
A_4	0 0	0	0 0	0 30 +	30
Потреби	40	35	30	45	150

Виродження може зустрітися також у процесі складанні початкового плану.

У прикладі, що розглядається, були часткові баланси між групами постачальників і групами споживачів. Досить переставити місцями деяких постачальників або деяких споживачів (що цілком припустимо), і початкові плани будуть виродженими. Нижче наведені умови цієї дещо зміненої задачі, де переставлені місцями споживачі і постачальники:

Постачальники	Споживачі				Запаси
	B_2	B_4	B_1	B_3	
A_2	1	4	1	6	34
A_1	3	7	4	2	46
A_3	5	4	3	9	40
A_4	0	0	0	0	30
Потреби	35	45	40	30	

Складемо початковий план діагональним методом.

Постачальники	Споживачі				Запаси
	B_2	B_4	B_1	B_3	
A_2	1 34	4 0	1 0	6 0	34
A_2	3 1	7 45	4 0	2 0	46
A_3	5 0	4 0	3 40	9 0	40
A_4	0 0	0 0	0 0	0 30	30
Потреби	35	45	40	30	150

Тут через часткові баланси загальне число заповнених кліток виявилося рівним 5, а не $(m + n - 1) = 7$. Треба ввести в план ще два маршрути з нульовим обсягом перевезень. Відомо, що в невиродженому плані, що складений діагональним методом, ненульові компоненти розташовуються на діагоналі таблиці сходами, так що перехід до сусіднього компоненту відбувається або по вертикалі, або по горизонталі. Тому поставимо базисні нулі там, де сходи обриваються, і перехід до сусіднього компоненту плану відбувається по діагоналі. Якщо є кілька варіантів розміщення базисних нулів, вибираємо будь-який. Зараз (план 6) прийнято $x_{34} = x_{33} = 0$. Тепер число базисних невідомих дорівнює $(m + n - 1) = 7$.

Складемо тепер початковий план методом мінімальної вартості.

Постачальники	Споживачі				Запаси
	B_2	B_4	B_1	B_3	
A_2	1 34	4 0	1 0	6 0	34
A_2	3 1	7 15	4 0	2 30	46
A_3	5 0	4 0	3 40	9 0	40
A_4	0	0	0	0	

	0	30	0	0	30
Потреби	35	45	40	30	150

Цей план також виявився виродженим, число ненульових компонентів тут дорівнює 6, що менше $(m + n - 1) = 7$. Однак, на відміну від попередніх випадків, не відомо, куди можна поставити додатковий базисний нуль – адже потрібно гарантувати, що після цього для будь-якої вільної клітинки можна буде побудувати замкнутий ланцюжок, всі інші вузли якого будуть розташовані у заповнених клітинках. Тому у випадку виродження вихідного плану, складеного методом мінімального елемента, застосовується спеціальний прийом, що називається методом «ε–зсуву». Ідея цього методу полягає в усуненні самої причини виродження, для чого до всіх запасів, крім останнього, додається мала величина ε, а з останнього запасу віднімаються всі додані величини (щоб не змінити величини загального запасу).

Складений після ε–зсуву план буде невиродженим. Величину ε заміняємо базисним нулем, що у даній таблиці попадає в клітинку A_3B_4. Далі в процесі поліпшення плану зберігаємо базисні нулі аж до одержання оптимального рішення.

Постачальники	Споживачі				Запаси
	B_2	B_4	B_1	B_3	
A_2	1 $34+\varepsilon$	4 0	1 0	6 0	$34+\varepsilon$
A_2	3 $1-\varepsilon$	7 $15+2\varepsilon$	4 0	2 30	$46+\varepsilon$
A_3	5 0	4 $+\varepsilon$	3 40	9 0	$40+\varepsilon$
A_4	0 0	0 $30-3\varepsilon$	0 0	0 0	$30-3\varepsilon$
Потреби	35	45	40	30	150

Приклад. Фірма здійснює поставку пляшок на три заводи, що займаються виробництвом безалкогольних напоїв. Вона має три склади, причому на складі 1 знаходиться 6 000 пляшок, на складі 2 – 3 000 пляшок і на складі 3 – 4 000 пляшок. Першому заводу потрібно 4000 пляшок, другому заводу – 5000 пляшок, третьому – 1 000 пляшок, четвертому – 3 000. Матрицею

$$\begin{pmatrix} 6 & 4 & 9 & 8 \\ 5 & 3 & 2 & 8 \\ 2 & 3 & 6 & 8 \end{pmatrix}$$ задана вартість перевезення однієї пляшки від кожного складу до кожного заводу. Необхідно так організувати доставку пляшок на заводи, щоб вартість перевезення була мінімальною.

Початковий опорний план отримаємо за методом мінімальної вартості.

Склади	Заводи				запаси
	1	2	3	4	
1	6 0	4 3 000	9 0	8 3 000	6 000
2	5 0	3 2 000	2 1 000	8 0	3 000
3	2 4 000	3 0	6 0	8 0	4 000
потреби	4 000	5 000	1 000	3 000	

Число заповнених клітин 5, $r = m + n - 1 = 6$. Отже задача є виродженою.

Для виключення виродженості необхідно в якусь клітинку ввести нульову поставку. Така клітина стає умовно зайнятою, її доцільно визначити під час обчислення потенціалів зайнятих клітин, вона повинна мати найменшу вартість порівняно з іншими клітинками, які можуть бути умовно зайнятими. Для знаходження потенціалів необхідно помістити нульову поставку в клітинку (3, 2).

u \ v	$v_1 =$ **3**	$v_2 =$ **4**	$v_3 =$ **3**	$v_4 =$ **8**
$u_1 =$ **0**	3 / 6	**4**	3 / 9	**8**
$u_2 =$ **-1**	2 / 5	**3**	**2**	7 / 8
$u_3 =$ **-1**	**2**	**3**	2 / 6	7 / 8

План оптимальний. Вартість транспортних витрат буде мінімальною і становитиме:

$$F = 4 \cdot 3\,000 + 8 \cdot 3\,000 + 3 \cdot 2\,000 + 2 \cdot 1\,000 + 2 \cdot 4\,000 = 52\,000 \text{ грош.од.}$$

Запитання для самоперевірки

1. Опишіть економіко-математичну модель транспортної задачі.

2. Запишіть транспортну задачу в матричній формі і назвіть її властивості.

3. Назвіть методи розв'язування транспортної задачі.

4. Який критерій оптимальності для транспортної задачі в матричній формі?

5. Дайте економічну інтерпретацію методу потенціалів розв'язування транспортної задачі.

6. Як здійснити перехід до іншого базисного плану?

7. Як розв'язати транспортну задачу у випадку виродження?

9. Транспортні задачі з додатковими умовами та задачі економічного змісту, що зводяться до транспортної задачі

9.1. Застосування транспортної задачі до вирішення деяких економічних задач. Транспортні задачі з додатковими умовами.

9.2. Задачі транспортного типу.

9.1. Застосування транспортної задачі до вирішення деяких економічних задач. Транспортні задачі з додатковими умовами

Алгоритми і методи вирішення транспортної задачі можуть бути використані у ході вирішення деяких економічних задач, що не мають нічого спільного з транспортуванням вантажу. У цьому випадку величини вартості c_{ij} мають різний зміст залежно від конкретної економічної задачі. До таких відносяться такі задачі:

1. Оптимальне закріплення за верстатами операцій з обробки деталей. У них c_{ij} є таким економічним показником, як продуктивність. Задача дозволяє визначити, скільки часу і на якій операції потрібно використовувати кожен з верстатів, щоб обробити максимальну кількість деталей. Так як транспортна задача вимагає знаходження мінімуму, то значення c_{ij} беруться з від'ємним знаком.

2. Оптимальні призначення або проблема вибору. Мається m механізмів, які можуть виконувати m різних робіт з продуктивністю c_{ij}. Задача дозволяє

визначити, який механізм і на яку роботу треба призначити, щоб добитися максимальної продуктивності.

3. Задача про скорочення виробництва з урахуванням сумарних витрат на виготовлення і транспортування продукції.

4. Збільшення продуктивності автомобільного транспорту за рахунок мінімізації порожнього пробігу. Зменшення порожнього пробігу скоротить кількість автомобілів для перевезень, збільшивши їх продуктивність.

5. Рішення задач за допомогою методу заборони перевезень. Використовується в тому випадку, якщо вантаж від деякого постачальника з якихось причин не може бути направлений одному із споживачів. Дане обмеження можна врахувати, присвоївши відповідній клітинці достатньо велике значення вартості, тим самим в цю клітинку не будуть робитися перевезення.

Транспортні задачі з додатковими умовами.
Транспортна задача з обмеженими пропускними спроможностями

У транспортних задачах з обмеженими пропускними спроможностями крім тарифів для кожного маршруту додатково задані пропускні спроможності кожного маршруту $x_{ij} \le d_{ij}$. В умовах завдання пропускні спроможності будемо записувати в нижній частині клітинок, а тарифи – у верхній.

Приклад. Маємо три пункти відправлення і три пункти призначення. Відомі запаси a_i продукції у кожного постачальника A_i, потреби b_j кожного споживача B_j і тарифи c_{ij} (вартість перевезення одиниці товару від пункту A_i в пункт B_j, записані в верхній частині клітинки), а також пропускні спроможності кожного маршруту d_{ij} (записані в нижній частині клітинки). Потрібно скласти оптимальний план перевезення, тобто визначити, яка кількість вантажу x_{ij} має бути відправлена з кожного пункту відправлення в кожний пункт призначення, щоб загальна вартість всіх перевезень була найменшою.

Постачальники	Споживачі			Запаси
	B_1	B_2	B_3	
A_1	5 30	2 25	3 25	45

A_2	7 40	9 15	6 20	65
A_3	4 25	7 20	5 15	40
Потреби	70	30	50	

Через двосторонні обмеження $0 \le x_{ij} \le d_{ij}$ у цій задачі з'являється два види вільних невідомих – одні з них в оптимальному плані прирівнюються нулю $x_{ij} = 0$, інші – максимально можливим значенням $x_{ij} = d_{ij}$ (вільні невідомі задаються, базисні невідомі – обчислюються). Значення базисних невідомих, які визначаються системою обмежень-рівностей через вільні невідомі, перебувають в інтервалі $0 < x_{ij} < d_{ij}$ (при виродженні може бути $0 \le x_{ij} \le d_{ij}$).

Функцію цілі також можна виразити через вільні невідомі:

$$f = \sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij} = f_0 + \sum_{(i,j)\notin Б} c'_{ij}x_{ij}\,.$$

Якщо коефіцієнт у перетвореній функції цілі перед вільною невідомою додатний $c'_{ij} > 0$, то збільшення x_{ij} приводить до збільшення витрат f; таку невідому треба зробити якнайменшою, тобто прирівняти її нулю $x_{ij} = 0$.

Якщо коефіцієнт перед вільною невідомою від'ємний $c'_{ij} < 0$, то збільшення x_{ij} приводить до зменшення витрат f; таку невідому треба зробити якнайбільшею, наскільки допускає пропускна спроможність, тобто $x_{ij} = d_{ij}$.

У процесі викладення методу потенціалів було доведено, що різниця між дійсним і непрямим тарифами показує, як змінюється функція цілі у разі збільшення x_{ij} на одиницю.

Для вільної (не базисної) невідомої x_{ij} відповідна їй балансова невідома двоїстої задачі $w_{ij} = c_{ij} – (u_i + v_j)$ буде базисною і саме її значення буде записане у вигляді коефіцієнта в цільовій функції при вільній невідомій x_{ij}. Звідси випливає такий критерій оптимальності плану транспортної задачі із обмеженими пропускними спроможностями:

$$\begin{cases} u_i + v_j \le c_{ij}, \text{ якщо } x_{ij} = 0; \\ u_i + v_j \ge c_{ij}, \text{ якщо } x_{ij} = d_{ij}; \\ u_i + v_j = c_{ij}, \text{ якщо } 0 < x_{ij} < d_{ij}. \end{cases}$$

Ця задача не завжди має розв'язок, наприклад, коли попит не може бути задоволений через обмеженість пропускних спроможностей маршрутів.

Послідовність розрахунків покажемо на прикладі.

У процесі складання першого опорного плану методом мінімальної вартості послідовні невідомі приймаються рівними меншому значенню трьох (а не двох, як звичайно) чисел – залишку запасу, залишку попиту і пропускній спроможності даного маршруту: $x_{ij} = \min\{a_i, b_j, d_{ij}\}$.

Постачальники	Споживачі B_1	B_2	B_3	Запаси
A_1	5 30 0	2 25 *25*	3 25 *20*	45/20/0
A_2	7 40 *40*	9 15 **5**	6 20 **15**	65/50/10/5
A_3	4 25 *25*	7 20 0	5 15 *15*	40/15/0
Потреби	70/45/5	30/5/0	50/30/0	

Базисні невідомі відзначені напівжирним шрифтом. Сірими кольорами і курсивом відзначені вільні невідомі, що дорівнюють пропускним спроможностям маршруту.

У 1-му стовпці і 2-му рядку залишилися нерозподілені залишки в 5 одиниць. Подібно відкритій транспортній задачі уводимо фіктивного постачальника A_4 із запасом $a_4 = 5$ і фіктивного споживача B_4 з потребами $b_4 = 5$. Тариф $c_{44} = 0$ приймаємо рівним нулю, інші тарифи $c_{4j} = c_{i4} = M$, де M – дуже велике число (щоб заблокувати ці маршрути).

Таким чином, одержуємо перший (явно неприпустимий) план:

Постачальники (План 1)	Споживачі B_1	B_2	B_3	B_4	Запаси
A_1	5 $+\rho$ 0	2 *25*	3 – **20**	M	45
A_2	7 *40*	9 5	6 + **15**	M **5**	65
A_3	4 *25*	7 0	5 *15*	M	40
A_4	M – **5**	M	M	0 + **0**	

Потреби	70	30	50		

опорний план не є оптимальним і його слід поліпшити.

Вартість цього плану дуже велика, план явно не оптимальний, оскільки використані заблоковані маршрути x_{41} і x_{24}. Спробуємо поліпшити план, звільнивши неприпустимі маршрути. План вироджений, тому ставимо базисний нуль у клітинку $x_{44} = 0$ і вважаємо $x_{33} = 15$ – базисною змінною (це допустимо, якщо в подальшому обсяг перевезень на цьому маршруті не буде збільшений).

Методом потенціалів знаходимо найбільш перспективну вільну невідому, яку варто ввести в базис.

u \ v		$v_1 =$ 2M-6		$v_2 =$ 3		$v_3 =$ 0		$v_4 =$ M-6	
$u_1 =$	**3**	2M-3	5	6	2	**3**		M-3	M
$u_2 =$	**6**	2M	7	**9**		**6**		**M**	
$u_3 =$	**5**	2M-1	8	8	7	**5**		M-1	M
$u_4 =$	**-M+6**	**M**		-M+9	M	-M+6	M	**0**	

Потенціали визначаємо тільки за базисними клітинками (напівжирний шрифт). У вільних клітинках, виділених сірими кольорами, умова оптимальності вимагає, щоб непрямі тарифи (у верхній частині клітинки) були більше дійсних тарифів (у нижній частині), тобто $u_i + v_j \geq c_{ij}$; в інших вільних клітинках умова оптимальності звичайна $u_i + v_j \leq c_{ij}$ (непрямі тарифи повинні бути менше дійсних). Найбільша розбіжність вийшла для клітинки x_{11}. Уводимо цю невідому в базис, для чого в таблиці плану будуємо замкнутий ланцюг $x_{11} \to x_{41} \to x_{44} \to x_{24} \to x_{23} \to x_{13} \to x_{11}$. Переміщаємо по цьому ланцюжку нерозподілені залишки. При збільшенні x_{23} треба стежити, щоб $x_{23} \leq d_{23} = 20$. Обидва заблокованих маршрути звільнилися, фіктивних постачальника A_4 і споживача B_4 тепер можна відкинути. Це перший реальний план перевезень (всі запаси вивезені, весь попит задоволений). Невідома x_{23} більш не є базисною (тому що $x_{23} = d_{23}$); знайдений план є виродженим.

Постачальники (План 2)	Споживачі				Запаси
	B_1	B_2	B_3	B_4	
A_1	5 **5**	2 *25*	3 **15**	*M* 0	45
A_2	7 *40*	9 **5**	6 20	*M* 0	65
A_3	4	7	5	*M*	40

	25	0	***15***		
A_4	*M* 0	*M*	*M* 0	0 **5**	
Потреби	70	30	50		

Для перевірки плану на оптимальність вважаємо $x_{33} = 15$ – базисною змінною і ставимо базисний нуль у клітинку x_{32} (більше нікуди). Заповнюємо таблицю потенціалів.

u \ v	$v_1 = 2$	$v_2 = 2$	$v_3 = 0$
$u_1 = 3$	**5**	*5* / *2*	**3**
$u_2 = 7$	*9* / *7*	**9**	*7* / *6*
$u_3 = 5$	*7* / *4*	**7**	**5**

Умови оптимальності виконані – у всіх вільних клітинках, де обсяги перевезень дорівнюють пропускним спроможностям, непрямі тарифи більше дійсних. Отриманий опорний план – оптимальний. Його вартість:

$$f = 5\cdot5 + 2\cdot25 + 3\cdot15 + 7\cdot40 + 9\cdot5 + 6\cdot20 + 4\cdot25 + 5\cdot15 = 740.$$

Двоступінчаста транспортна задача

Дуже часто між виробниками A_i і споживачами B_j немає прямих зв'язків, вся продукція спочатку направляється на оптові бази (пункти розподілу) D_k і тільки далі поступає в роздрібну торговельну мережу. У цьому випадку пункти розподілу включаємо як в число споживачів, так і в число постачальників. Прямі перевезення від виробників до споживачів забороняємо, назначаючи дуже великі тарифи на цих маршрутах; також забороняємо перевезення з бази на базу. Особливості цієї задачі розглянемо на такому прикладі.

Тарифи	D_1	D_2	B_1	B_2	B_3	B_4	B_5	Запаси
D_1	0		5	7	2	3	1	3 700
D_2		0	3	1	4	5	2	3 700
A_1	10	27						1 000
A_2	12	14						1 500
A_3	13	8						1 200
Попит	3 700	3 700	800	500	750	1000	650	

Маємо три підприємства з обсягами виробництва 1 000, 1 500, 1 200 і п'ять споживачів з попитом 800, 500, 750, 1 000, 650. Задача закрита, оскільки загальний запас продукції дорівнює загальному попиту 3 700. Спочатку вироблена продукція перевозиться в пункти розподілу D_1 і D_2, які в таблиці умов завдання включені і в число постачальників, і в число споживачів. Товарообіг баз приймаємо рівному загальному попиту (можливо, буде досить однієї бази). Сірим фоном відмічені заборонені маршрути (із тарифами *M*).

Складаємо перший план методом мінімальної вартості:

План 1	D_1	D_2	B_1	B_2	B_3	B_4	B_5
D_1	1200	100 -	$+\rho$		750	1000	650
D_2		2400 +	800 -	500			
A_1	1000						
A_2	1500						
A_3		1200					

Цей план – недопустимий, оскільки зайнята заборонена клітинка $D_1 D_2$. Методом потенціалів знаходимо найбільш перспективний маршрут, який треба ввести в план перевезень. Знайдено відразу два таких еквівалентних маршрути: $D_1 - B_1$, або $A_2 - D_2$. Приймаємо перший варіант і складаємо замкнутий ланцюжок, що починається з клітинки D_1B_1. Переміщаємо по цьому ланцюжку 100 одиниць товару і тим самим розвантажуємо заборонену клітинку.

План 2	D_1	D_2	B_1	B_2	B_3	B_4	B_5
D_1	1 00		100		750	1 000	650
D_2		2 500	700	500			
A_1	1 000						
A_2	1 500						
A_3		1 200					

Методом потенціалів визначаємо, що цей план – оптимальний. Товарообіг першої бази можна знизити до:

$$3\ 700 - 1\ 200 = 25\ 00,$$

а другої $$3\ 700 - 2\ 500 = 1\ 200.$$

Існує також еквівалентний план з тим самим значенням цільової функції:

План 3	D_1	D_2	B_1	B_2	B_3	B_4	B_5

D_1	1 300				750	1 000	650
D_2		2 400	800	500			
A_1	1 000						
A_2	1 400	100					
A_3		1 200					

Тепер найменший товарообіг першої бази дорівнює 3 700 – 1 300 = 2 400, а другої 3 700 – 24 00 = 1 300. Можна комбінувати ці два еквівалентні плани.

На рис. 9.1 ці плани зображені у вигляді орієнтованих графів.

План 2 План 3

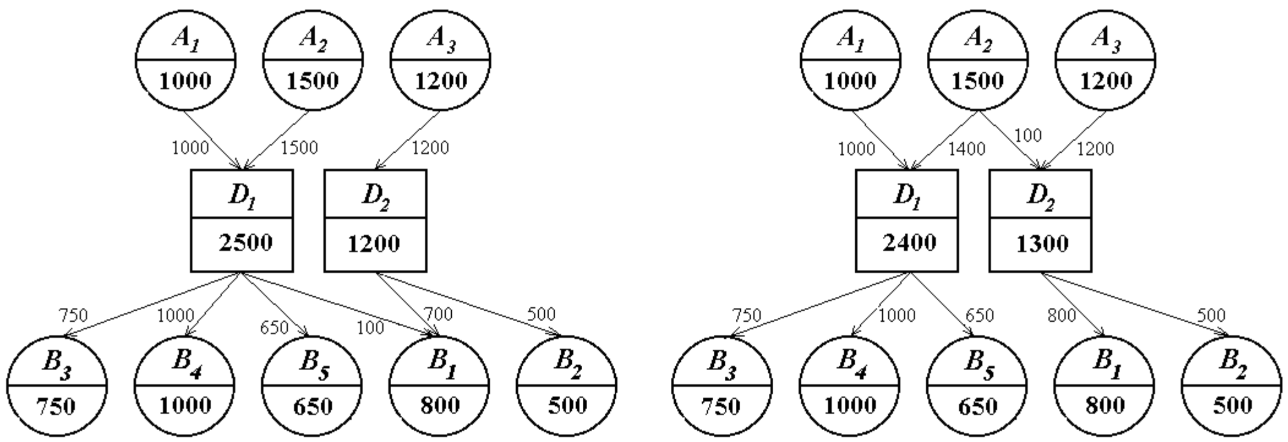


Рис. 9.1. **Оптимальні плани двоступінчатої транспортної задачі**

Транспортна задача за критерієм часу

При перевезеннях швидкопсувної продукції з невеликим терміном зберігання треба якнайшвидше доставити товар споживачам, тому змінюється функція цілі – бажано зменшити найбільший час в плані перевезень:

$$f = max\{t_{ij}\} \to min .$$

Ця функція цілі не є лінійною, тому оптимальний план транспортної задачі за критерієм часу не обов'язково повинен бути базисним. Задачу розв’язуємо у такий спосіб.

На першому етапі складаємо план методом мінімального елементу (мінімальної вартості). Визначаємо для знайденого плану максимальний час перевезень $t_{pq} = max\{t_{ij}\}$ і викреслюємо (забороняємо) всі вільні клітинки таблиці планів з часом перевезень, що не менше t_{pq}. На другому етапі намагаємось розвантажити клітинку з максимальним часом перевезень (ставимо в цю клітинку знак $+\rho$ і будуємо замкнений ланцюжок для перерахування обсягів перевезень). Проте тут є одна особливість. Необхідно було стежити, щоб у разі переходу до нового плану число базисних невідомих не змінювалось, тобто коли закривався один маршрут, відкривався тільки один інший, для чого всі вузли замкненого ланцюжка, крім одного, розташовували у базисних клітинках. Але зараз число нових маршрутів може бути більшим, тобто в базисних клітинках повинні бути тільки ті вузли ланцюжка, що відмічені позначкою «-», а вузли із позначкою «+» можуть бути де завгодно. Таким чином наприкінці одержуємо план з мінімальним часом найтривалішого перевезення, який вже не можна зменшити. Але (це ще одна особливість) пошук оптимального плану на цьому не завершується. Викреслюємо (третій етап) в плані найтриваліший маршрут перевезень і переходимо до скороченої транспортної задачі без одного споживача, поставки до якого вже не можна зменшити. Намагаємось поліпшити план перевезень до інших споживачів, щоб вони одержали свою продукцію якнайшвидше і мали деякий резерв терміну зберігання.

Приклад. Маємо трьох постачальників і п’ять споживачів. Нижче наведені умови закритої задачі (час перевезень на кожному маршруті, запаси 3-х постачальників, попит 5-и споживачів). Перевозиться швидкопсувна продукція з невеликим терміном зберігання, треба якнайшвидше доставити товар споживачам.

Постачальники	Споживачі					Запаси
	B_1	B_2	B_3	B_4	B_5	
A_1	2	6	3	4	8	30
A_2	1	5	6	9	7	35
A_3	3	4	1	6	10	40
Потреби	20	34	16	10	25	

Висхідний план отримаємо методом мінімальної вартості.

	Споживачі	Запаси

Постачальники (План 1)	B_1	B_2	B_3	B_4	B_5	
A_1	2 0+ ρ	6 0	3 0	4 10	8 20 -	30/20/0
A_2	1 20 -	5 10	6 0	9 0	7 5 +	35/15/5/0
A_3	3 0	4 24	1 16	6 0	10 0	40/24/0
Потреби	20/0	34/10/0	16/0	10	25/20/0	

Найтриваліший маршрут A_1–B_5 в цьому плані вимагає 8 годин. Викреслюємо в таблиці планів всі вільні клітинки з таким і більшим часом перевезень (дві клітинки A_2–B_4 і A_3–B_5). Розвантажуємо клітинку A_1–B_5, відкриваючи маршрут A_1–B_1.

Постачальники (План 2)	Споживачі					Запаси
	B_1	B_2	B_3	B_4	B_5	
A_1	2 20 -	6 0	3 0+ ρ	4 10	8 0	30
A_2	1 0+ ρ	5 10 -	6 0	9 0	7 25	35
A_3	3 0	4 24 +	1 16 -	6 0	10 0	40
Потреби	20	34	16	10	25	

Найтриваліший маршрут A_2–B_5 в новому плані 2 вимагає 7 годин і ця величина не може бути зменшена, оскільки поставка продукції споживачеві B_5 другими маршрутами вимагає ще більше часу. План 2 – вже оптимальний, але не найкращий. Розглянемо тепер скорочену задачу без споживача B_5.

У скороченій задачі найтриваліший маршрут A_2–B_2 вимагає 5 годин. Викреслюємо в таблиці плану 2 всі клітинки з таким і більшим часом перевезень і робимо спробу розвантажити клітинку A_2–B_2. Зверніть увагу, що для цього виявилось потрібним відкрити відразу два нових маршрути A_2–B_1 і A_1–B_3.

	Споживачі	Запаси

Постачальники (План 2)	B_1	B_2	B_3	B_4	B_5	
A_1	2 10	6 0	3 10	4 10	8 0	30
A_2	1 10	5 0	6 0	9 0	7 25	35
A_3	3 0	4 34	1 6	6 0	10 0	40
Потреби	20	34	16	10	25	

У другому оптимальному плані (плані 3) найтриваліші маршрути (без споживача B_5) A_3–B_2 і A_1–B_4 вимагають 4 години і ці клітинки неможливо розвантажити навіть частково. Отже одержано найкращий план перевезень, що забезпечує поставку 25 одиниць продукції за 7 годин (це найменший термін поставки споживачу B_5), 44 одиниць продукції за 4 години, (найменший термін поставок споживачам B_2 і B_4), 10 одиниць продукції за 3 години, 10 одиниць продукції за 2 години і 16 одиниць продукції за 1 годину.

Багатопродуктова транспортна задача

Під час перевезення неоднорідного вантажу (вугілля різніх сортів, цементу різних марок, сталевого прокату) є можливість заміни частки попиту на взаємозамінний вид продукту. В багатопродуктовій транспортній задачі потрібно оптимізувати загальний план перевезень, якщо відомі коефіцієнти взаємозамінності одного виду продукту на інший. Особливості вирішення цієї задачі розглянемо на прикладі перевезень 2-х видів вугілля (буре вугілля і антрацит) від 2-х постачальників A_1, A_2 до 2-х споживачів B_1, B_2. Умови задачі (тарифи, запаси і попит на кожний вид продукту) наведені в таблиці:

	B_1	B_2	Запас(1)	Запас(2)
A_1	30	21	300	100
A_2	15	24	400	140
Попит(1)	100	200		
Попит(2)	60	40		
Попит(3)	200(1)	90(1)		

Тут позначено: Запас(1) – це запаси бурого вугілля, Запас(2) – антрациту. Аналогічно, Попит(1) і Попит(2) – це попит на буре вугілля і антрацит (т); ці кількості не можна замінювати на інший вид товару. Попит(3) – це кількості взаємозамінного товару (виражені в одиницях бурого вугілля).

Відомо, що за теплотворною здатністю 3 одиниці бурого вугілля еквівалентні 2-м одиницям антрациту. Отже можна перерахувати всі запаси і попит в одиниці одного виду вугілля, наприклад, в одиниці бурого вугілля. Коефіцієнт $k = {}^{3}/_{2}$ назвемо коефіцієнтом взаємозамінності.

Розщеплюємо кожний пункт постачання на два підпункти $A_1(1)$, $A_1(2)$ і $A_2(1)$, $A_2(2)$ по числу видів товару. Кожний пункт споживання розщеплюємо на три підпункти $B_1(1)$, $B_1(2)$, $B_1(3)$ і $B_2(1)$, $B_2(2)$, $B_2(3)$. Запаси і попит на антрацит виражаємо в одиницях бурого вугілля, для чого помножимо їх на $k = {}^{3}/_{2}$. Одержимо: $a_1(1) = 300$, $a_1(2) = 100\cdot {}^{3}/_{2} = 150$, $a_2(1) = 100$, $a_2(2) = 140\cdot {}^{3}/_{2} = 210$, $b_1(1) = 100$, $b_1(2) = 60\cdot {}^{3}/_{2} = 90$, $b_1(3) = 220$, $b_2(1) = 200$, $b_2(2) = 40\cdot {}^{3}/_{2} = 60$, $b_2(3) = 90$. Нагадаємо, що попит на взаємозамінний товар був відразу виражений в одиницях бурого вугілля, тому його не потрібно було перераховувати.

Тепер перераховуємо тарифи c_{ij}. Затрати на перевезення залежать від маси вантажу, а не від його споживчої властивості (теплотворної здатності). Тому умовно збільшуючи запаси 2-го виду вантажу в k разів, слід відповідно зменшити в k разів тарифи на перевезення цього умовно збільшеного вантажу (в межах $A_1(2)$ і $A_2(2)$ розширеної таблиці тарифів):

	$B_1(1)$	$B_1(2)$	$B_1(3)$	$B_2(1)$	$B_2(2)$	$B_2(3)$	Запаси
$A_1(1)$	30	1000	30	21	1000	21	300
$A_1(2)$	1000	$30\cdot{}^{2}/_{3}=20$	$30\cdot{}^{2}/_{3}=20$	1000	$21\cdot{}^{2}/_{3}=14$	$21\cdot{}^{2}/_{3}=14$	$100\cdot{}^{3}/_{2}=150$
$A_2(1)$	15	1000	15	24	1000	24	100
$A_2(2)$	1000	$15\cdot{}^{2}/_{3}=10$	$15\cdot{}^{2}/_{3}=10$	1000	$24\cdot{}^{2}/_{3}=16$	$24\cdot{}^{2}/_{3}=16$	$140\cdot{}^{3}/_{2}=210$
Попит	100	$60\cdot{}^{3}/_{2}=90$	220	200	$40\cdot{}^{3}/_{2}=60$	90	760 760

Оскільки попит на певний вид товару не може бути задоволеним іншими видами товарів, деякі маршрути забороняємо дуже великими тарифами на перевезення (в таблиці ці клітинки виділені сірим кольором).

Тепер багатопродуктова задача приведена до звичайної транспортної задачі. Наша задача – збалансована, оскільки сумарний запас дорівнює сумарному попиту (760). Як відомо, звичайна збалансована транспортна задача завжди має розв'язок. Але для багатопродуктової задачі треба ще перевірити умови, що запасів досить для задоволення попиту на певні види товарів (що не можуть бути замінені іншими видами):

– сумарний запас бурого вугілля дорівнює $a_1(1) + a_2(1) = 300 + 100 = 400$, що перевищує сумарний попит на незамінну частку на цей вид вугілля $b_1(1) + b_2(1) = 100 + 200 = 300$;

– сумарний запас антрациту (в умовних одиницях) дорівнює $a_1(2) + a_2(2) = 150 + 210 = 360$, що перевершує сумарний попит на незамінну частку на цей вид вугілля $b_1(2) + b_2(2) = 90 + 60 = 150$.

Отже, приведена задача має розв'язок. Оптимальний план в одиницях бурого вугілля (План 1) з мінімальними витратами $f_{min} = 8\ 850$ наведений у таблиці:

План 1	$B_1(1)$	$B_1(2)$	$B_1(3)$	$B_2(1)$	$B_2(2)$	$B_2(3)$	Запаси
$A_1(1)$	10			200		90	300
$A_1(2)$		90			60		150
$A_2(1)$	90		10				100
$A_2(2)$			210				210
Попит	100	90	220	200	60	90	760 / 760

У таблиці План 2 обсяги перевезень антрациту виражені в тоннах.

План 2	$B_1(1)$	$B_1(2)$	$B_1(3)$	$B_2(1)$	$B_2(2)$	$B_2(3)$	Запаси
$A_1(1)$	10			200		90	300
$A_1(2)$		$90 \cdot {}^2/_3 = 60$			$60 \cdot {}^2/_3 = 40$		$150 \cdot {}^2/_3 = 100$
$A_2(1)$	90		10				100
$A_2(2)$			$210 \cdot {}^2/_3 = 140$				$210 \cdot {}^2/_3 = 140$
Попит	100	$90 \cdot {}^2/_3 = 60$	220(1)	200	$60 \cdot {}^2/_3 = 40$	90	

9.2. Задачі транспортного типу

Серед всіх задач математичного програмування транспортні задачі і задачі транспортного типу становлять близько 70 % і ця частка продовжує зростати з опануванням нових класів завдань. Зведення проблеми до задачі транспортного типу гарантує просте її вирішення.

Задача про призначення

Зміст задачі про призначення можна сформулювати як пошук найкра щого виконавця для кожної роботи. Маємо кілька кандидатів на виконання деяких робіт і таблицю продуктивності q_{ij} кожного кандидата на кожній роботі. Слід найкращим чином розподілити виконавців за робочими місцями, щоб загальна продуктивність була найбільшою. Задача легко зводиться до транспортної, де всі запаси і попит дорівнюють одиниці $a_i = b_j = 1$; замість пошуку максимуму загальної продуктивності можна знаходити мінімум витрат, матрицю c_{ij} яких одержують як $c_{ij} = q_{max} - q_{ij}$. Всі змінні x_{ij} тут приймають тільки два значення 0 або 1 ($x_{ij} = 1$ означає, що робітник A_i призначений на роботу B_j).

Приклад. Отже розглянемо задачу призначення чотирьох робітників на виконання чотирьох робіт, якщо задана матриця витрат (скільки буде коштувати виконання кожної роботи кожним виконавцем).

Прочерк у матриці означає, що цей робітник не може виконувати цю роботу. Крім того врахуємо, що з'явився новий робітник (A_5), який може виконувати ці роботи з витратами 60, 45, 30, 80. Чи не буде економічно вигідним призначити на роботу нового виконавця ?

	B_1	B_2	B_3	B_4
A_1	50	50	–	20
A_2	70	40	20	30
A_3	90	30	50	–
A_4	70	20	60	70

Зазвичай для відкритої транспортної задачі вводимо фіктивне робоче місце $B_ф$ з нульовими витратами і складаємо план методом мінімального елементу.

План 1 вироджений, тому методом ε-зсуву знайдені позиції для чотирьох базисних нулів. Робітник A_3 не одержує роботи, якщо прийняти нового працівника A_5. Цей план фактично є оптимальним, але за методом потенціалів потрібно зробити кілька зайвих ітерацій, після яких базисні нулі займуть інші позиції.

План1	B_1	B_2	B_3	B_4	B_5
A_1	**0**			**1**	
A_2	**0**		**1**		
A_3	**0**				**1**
A_4	**0**	**1**			
A_5	**1**				

План2	B_1	B_2	B_3	B_4
A_1				**1**
A_2			**1**	
A_3		**1**		
A_4	**1**			

Якщо призначити на робочі місця штатний персонал (без A_5), то вартість плану 2 збільшиться на 10 цінових одиниць. Робітник A_4 при цьому одержує найбільш невигідну для нього роботу B_1.

Розподільна задача з пропорційними продуктивностями

Якщо матриця продуктивностей має вигляд $q_{ij} = \alpha_i \beta_j$, то загальну розподільну задачу (задачу про оптимальне завантаження обладнання) можна звести до транспортної і одержати її розв'язок набагато простіше. Розглянемо приклад складання оптимальної програми обробки чотирьох видів виробів B_1, B_2, B_3, B_4 на трьох взаємозамінних верстатах A_1, A_2, A_3, якщо відомі продуктивності верстатів q_{ij} для кожного виробу (шт./год.), собівартості одиниці продукції c_{ij} (грн/шт.), фонд завантаження обладнання a_i (год.) і мінімальне планове замовлення b_j (шт.). Умови задачі наведені в таблиці:

Вер-стати	Собівартість, грн./шт.				Фонд часу	Вер-стати	Продуктивність, шт./год.				α_i
	B_1	B_2	B_3	B_4			B_1	B_2	B_3	B_4	
A_1	20	10	5	12	240	A_1	30	50	30	20	5
A_2	8	12	9	8	150	A_2	60	100	60	40	10
A_3	5	10	6	9	150	A_3	18	30	18	12	3
Замовлення	3000	15000	4500	1500		β_j	6	10	6	4	

Потрібно визначити 3×4 = 12 чисел x_{ij} − кількості виробів кожного типу (шт.), що виготовляються на верстаті A_i. Складаємо обмеження задачі.

Об'єм виробленої продукції має бути не меншим, ніж планове замовлення:

$$\begin{cases} x_{11} + x_{21} + x_{31} \geq 3000, \\ x_{12} + x_{22} + x_{32} \geq 15000, \\ x_{13} + x_{23} + x_{33} \geq 4500, \\ x_{14} + x_{24} + x_{34} \geq 1500. \end{cases}$$

Завантаження верстатів не може перевершувати фонду часу:

$$\begin{cases} \dfrac{x_{11}}{30}+\dfrac{x_{12}}{50}+\dfrac{x_{13}}{30}+\dfrac{x_{14}}{20}\le 240, \\ \dfrac{x_{21}}{60}+\dfrac{x_{22}}{100}+\dfrac{x_{23}}{60}+\dfrac{x_{24}}{40}\le 150, \\ \dfrac{x_{31}}{18}+\dfrac{x_{32}}{30}+\dfrac{x_{33}}{18}+\dfrac{x_{34}}{12}\le 150. \end{cases}$$

Всі невідомі − невід'ємні:

$$x_{ij} \ge 0$$

Цільова функція виражає загальні витрати:

$$f = (20x_{11}+8x_{21}+5x_{31})+(10x_{12}+12x_{22}+10x_{32})+(5x_{13}+9x_{23}+6x_{33})+$$
$$+(12x_{11}+8x_{21}+9x_{31}) \to min.$$

Помічаємо, що в цієї задачі продуктивності – пропорційні: $q_{ij} = \alpha_i\,\beta_j$; співмножники α_i, β_j наведені в таблиці. Переходимо до нових змінних $x_{ij} = \beta_j\, y_{ij}$:

$$x_{11} = 6y_{11}, \quad x_{12} = 10y_{12}, \quad x_{13} = 6y_{13}, \quad x_{14} = 4y_{14},$$
$$x_{21} = 6y_{11}, \quad x_{22} = 10y_{22}, \quad x_{23} = 6y_{23}, \quad x_{24} = 4y_{24},$$
$$x_{31} = 6y_{31}, \quad x_{32} = 10y_{32}, \quad x_{33} = 6y_{33}, \quad x_{34} = 4y_{34},$$

Підставляємо ці вирази в обмеження задачі і функцію цілі. Після скорочення загальних співмножників одержуємо:

$$\begin{cases} y_{11}+y_{21}+y_{31} \ge \dfrac{3000}{6} = 500, \\ y_{12}+y_{22}+y_{32} \ge \dfrac{15000}{10} = 1500, \\ y_{13}+y_{23}+y_{33} \ge \dfrac{4500}{6} = 750, \\ y_{14}+y_{24}+y_{34} \ge \dfrac{1500}{4} = 375. \end{cases}$$

$$\begin{cases} y_{11}+y_{12}+y_{13}+y_{14} \le 5\cdot 240 = 1200, \\ y_{21}+y_{22}+y_{23}+y_{24} \le 10\cdot 150 = 1500, \\ y_{31}+y_{32}+y_{33}+y_{34} \le 3\cdot 150 = 450. \end{cases}$$

$$f = 6(20y_{11}+8y_{21}+5y_{31})+10(10y_{12}+12y_{22}+10y_{32})+6(5y_{13}+9y_{23}+6y_{33})+4(12y_{14}+8y_{24}+9y_{34}),$$

$$f = (120y_{11}+100y_{12}+30y_{13}+48y_{14}) +$$
$$+ (48y_{21}+120y_{22}+54y_{23}+32y_{24}) +$$
$$+ (30y_{31}+100y_{32}+36y_{33}+36y_{34}).$$

Таким чином, задача зведена до транспортної задачі з новими тарифами $c'_{ij} = \beta_j c_{ij}$, запасами $a'_i = \alpha_i a_i$, попитом $b'_j = \frac{b_j}{\beta_j}$ відносно невідомих $y_{ij} = \frac{x_{ij}}{\beta_j}$. Умови цієї транспортній задачі і її оптимальний розв'язок наведено нижче.

Тарифи, запаси, попит						Оптимальний план *Yij*						Оптимальний план *Xij*					
	B_1	B_2	B_3	B_4	Запас		B_1	B_2	B_3	B_4	Суми		B_1	B_2	B_3	B_4	Час
A_1	120	100	30	48	1 200	A_1		450	750		1200	A_1		4500	4500		240
A_2	48	120	54	32	1500	A_2	500	600		375	1475	A_2	3000	6000		1500	147,5
A_3	30	100	36	36	450	A_3		450			450	A_3		4500			150
Попит	500	1500	750	375		Суми	500	1500	750	375		Штук	3000	15000	4500	1500	

Транспортна задача виявилася відкритою, тому був введений фіктивний споживач $B_ф$ із попитом 25 одиниць. Оптимальний план відносно y_{ij} знайдений методом потенціалів. Оптимальний план відносно x_{ij} визначений за формулами $x_{ij} = \beta_j y_{ij}$. Наприкінці обчислено реальний час завантаження обладнання.

Більш простий і поширений випадок пропорційності продуктивностей, коли вони є однаковими для всіх споживачів $q_{ij} = \alpha_i$. Сформулюємо умови двох типових задач з такими продуктивностями.

Приклад. Нехай на чотирьох ткацьких верстатах з фондом завантаження 200, 300, 250, 400 годин виробляється тканина трьох артикулів *A, B, C* з продуктивністю 260, 200, 340, 500 м/год. Сумарний попит на тканину кожного артикула − 200, 100, 150 тис. м. Відома матриця прибутку c_{ij}, грн/м. Необхідно найкращим чином розподілити замовлення між верстатами. Умови завдання зведемо в стандартну таблицю:

Верстати	Тканина			Фонд часу	Продуктивність м/год	Запас тис. м
	A	*B*	*C*			
I	2,5	1,6	0,8	200	260	52
II	2,2	1,0	1,0	300	200	60
III	2,0	1,9	0,6	250	340	85
IV	2,8	1,2	0,9	100	500	200
Попит	200	100	150			

В останньому стовпчику таблиці (запас, тис. м) обчислені максимальні об'єми виробництва для кожного верстату (фонд часу × продуктивність), після чого задача безпосередньо зводиться до транспортної задачі. Отже для задач даного типу «продуктивності» – це лише коефіцієнти, що зв'язують розмірності попиту з розмірностями запасів.

Приклад. Запаси вугілля трьох сортів дорівнюють 300, 800,400 т з теплотворною здатністю 1 800, 2 500, 3 000 кал/кг. Вугілля перевозиться до чотирьох споживачів, попит яких виражений у млн кал: 750, 920, 1 100, 800.

Відомі сумарні витрати на видобування і доставку вугілля (матриця c_{ij}, грн/т). Знайти оптимальний план поставок вугілля. Умови задачі записуємо в таблицю:

Вугілля	Витрати, грн/т				Запас т	Теплотворна здатність, кал/кг	Запас млн. кал
	B_1	B_2	B_3	B_4			
I	27	36	18	18	300	1 800	540
II	30	25	15	20	800	2 500	2 000
III	35	30	24	21	400	3 000	1 200
Попит	750	920	1100	800			

Знання теплотворної здатності дозволило перерахувати запаси вугілля кожного сорту в теплові одиниці (млн кал). Якщо визначати план x_{ij} в цих же одиницях, задача буде звичайною транспортною задачею.

Задача про заміну устаткування

Після вироблення ресурсу обладнання можна замінити і купувати нове, або зробити ремонтні роботи різного ступеня складності. Необхідно знайти найбільш вигідну стратегію оновлення устаткування. Звичайно такі задачі намагаються розв’язувати методами динамічного програмування. Проте про оптимальну стратегію заміни устаткування можна звести до задачі транспортного типу.

Приклад. Майстерня використовує щотижня дискові пили у таких кількостях (за днями тижня):

	Пн	Вт	Ср	Чт	Пт	Сб	Нд	*Усього*
Попит	24	12	14	20	18	14	22	124

Можна кожний день купувати нові пили за ціною 12 грошових одиниць, чи віддавати їх на заточення, що коштує 3 грошові одиниці, але ця (ремонтна) робота виконується 2 доби. Є ще можливість віддавати пили на термінове нічне заточення, що буде коштувати 6 грошових одиниць. Яка буде найвигідніша стратегія щотижневого оновлення інструменту ?

Відразу відкинемо стратегію закупівлі всіх 124 пил, що буде коштувати $124 \times 12 = 1\ 488$ грошових одиниць. Можна закупити тільки 24 пили (максимальний щоденний попит) і далі віддавати їх на заточення. Навіть якщо це буде виключно термінові нічні заточення, вартість такої стратегії буде $24 \times 12 + 100 \times 6 = 888$ грошових одиниць, тобто значно менше, ніж 1 488. Чи можна ще знизити вартість оновлення інструменту використовуючи звичайне дводобове заточення ?

Сформулюємо цю задачу як транспортну, де роль споживачів відіграють дні тижня з відомим попитом. Роль постачальників будуть відігравати різні стратегії (способи) постачання інструменту. Перший спосіб – купівля нових пил за ціною 12 грошових одиниць – може забезпечити поставку інструменту в будь-який день тижня, отже запас цього «постачальника» дорівнює 124 пили. Наприкінці кожного робого дня з'являється можливість віддавати на заточення затуплені пили (у кількості, що дорівнює попиту на цей день). Нагадаємо, що звичайне заточення потребує 2 доби, тобто оновлення інструменту на наступні два дні можливе за допомогою термінового нічного заточення і тільки у разі поставок на третій і наступні дні можна розраховувати на більш дешеве звичайне заточення. Отже, маємо ще 7 «постачальників» за днями тижня з запасами, що дорівнюють пилам, використаним у цей день. Затрати на наступні два дні дорівнюють 6, в останні дні 3 грошовим одиницям. До речі, остання 8-ма стратегія наприкінці тижня – нічого не робити, зношений за неділю інструмент списується. Таким чином, маємо 8 стратегій оновлення інструменту і 8 споживачів цього інструменту (8-й споживач – фіктивний, оскільки ця задача відкрита – один перший «постачальник» спроможний забезпечити весь попит).

Нижче в таблиці наведені умови задачі і складено перший план.

Вартість оновлення (зверху курсивом) і план (знизу напівжирне)

Стратегії	Пн	Вт	Ср	Чт	Пт	Сб	Нд	*Залишки*	*Запаси*
Нове	**24** *12*	*12*	*12*	*12*	*12*	*12*	*12*	**100**	124
Пн		**12** *6*	**2** *6*	**10** *3*	*3*	*3*	*3*		24
Вт			**12** *6*	*6*	*3*	*3*	*3*		12
Ср				**10** *6*	*6*	**4** *3*	*3*		14
Чт					**18** *6*	*6*	**2** *3*		20
Пт						**10** *6*	**6** *6*	**2**	18
Сб							**14** *6*		14
Нд								**22**	22
Попит	24	12	14	20	18	14	22	124	

Далі буде пояснено, як був складений цей перший план. Спочатку задовольняється попит першого споживача, що можна зробити тільки одним способом – закупити всі 24 пили. Це – найбільший денний попит, тому більше за цією стратегією інструмент діставати не будемо, залишок запасу першого постачальника переносимо в графу *Залишки*. Задача скорочується на один рядок і один стовпець.

Попит 12 пил у вівторок (якщо перша стратегія вже виключена) можна задовольнити тільки терміновим нічним заточенням інструмента, що відпрацював у понеділок. Запас 24 пили скорочується на 12, тобто є ще 12 пил, що будуть використані у наступні дні (і нам потрібно вирішити, яким способом вони будуть заточені). Задача скорочується ще на один стовпець.

Попит 14 пил у середу можна задовольнити тільки терміновим заточенням, але різними замовленнями. До середи ми маємо знову 24 пили (12 після понеділка і 12 після вівторка). Вирішуємо здати на термінове заточення 12 пил, що маємо після вівторка, ще 2 пили маємо після понеділка, тоді 10 пил можна заточити звичайним більш дешевим способом. Отже до четверга маємо 10 пил після звичайного заточення і ще 10 пил одержуємо після термінового заточення. Задача знову скорочується на два рядки і ще на один стовпець.

Ось таким чином був складений цей план оновлення інструменту. Перевірка методом потенціалів показала, що цей план – оптимальний і є ще багато еквівалентних планів. Вартість оптимального плану складає $f_{min} = 840$ грошових одиниць. Наведемо інструкцію послідовності постачання інструменту за днями тижня:

1) до понеділка слід купувати 24 пили і використати їх у понеділок;

2) наприкінці понеділка здати 10 пил на звичайне дводобове заточення і 14 пил на термінове нічне заточення;

3) у вівторок використати 12 пил (що одержані після нічного заточення) і знову здати їх на термінове нічне заточення;

4) щосереди маємо 14 пил (що одержані після нічного заточення), використовуємо їх, далі 10 пил віддаємо на нічне і 4 – на звичайне дводобове заточення;

5) щочетверга маємо 10 пил після термінового і 10 пил після звичайного заточення, використовуємо їх, далі віддаємо 18 пил на термінове і 2 пили на звичайне дводобове заточення;

6) у п'ятницю використовуємо 18 пил (що одержані після нічного заточення), з них 16 віддаємо на термінове заточення, а 2 пили вже не потрібні (списуємо);

7) щосуботи маємо 4 пили після звичайного заточення і беремо ще 10 після нічного заточення, наприкінці віддаємо всі 14 пил на останнє нічне заточення;

8) щонеділі маємо 22 пили (з них 2 після звичайного і 20 після термінового заточення), використовуємо їх і списуємо;

9) на новий тиждень знову купуємо 24 нові пили. Далі все повторюється.

Запитання для самоперевірки

1. Наведіть приклади економічних задач, для яких доцільно використовувати алгоритми та методи вирішення транспортної задачі.

2. У чому особливість розв'язання задачі вибору оптимального варіанта використання виробничого обладнання?

10. Задачі дробово-лінійного програмування. Основні методи їх розв'язання та аналізу

10.1. Економічна і математична постановка задачі дробово-лінійного програмування.

10.2. Геометрична інтерпретація задачі дробово-лінійного програмування.

10.3. Розв’язування дрібно-лінійної задачі зведенням до задачі лінійного програмування.

10.1. Економічна і математична постановка задачі дробово-лінійного програмування

Часто в реальних оптимізаційних задачах в економіці підприємства критерієм оптимальності використовують рівень рентабельності, продуктивності праці і т. д. Ці фінансові показники є відносними величинами та математично виражаються дробово-лінійними функціями. У цьому випадку загальна економіко-математична модель така:

Максимізувати рентабельність виробництва:

$$Z = \frac{\sum_{j=1}^{n} c_j x_j}{\sum_{j=1}^{n} d_j x_j} \to \max, \qquad (10.1)$$

де c_j – прибуток від реалізації одиниці j-го виду продукції, при цьому загальний прибуток $\sum_{j=1}^{n} c_j x_j$; d_j – витрати на виробництво одиниці j-го виду продукції, при цьому $\sum_{j=1}^{n} d_j x_j$ – загальні витрати на виробництво;

за умови виконання обмежень щодо використання ресурсів підприємств:

$$\sum_{j=1}^{n} a_{ij} x_j \{\le, =, \ge\} b_i$$
$$\left(i = \overline{1,m}\right), \quad x_j \ge 0 \quad \left(j = \overline{1,n}\right).$$

Знаменник цільової функції в області допустимих розв'язків системи обмежень не дорівнює нулю.

Оскільки задача (10.1) відрізняється від звичайної задачі лінійного програмування лише цільовою функцією, тому для її вирішення використовуються модифіковані відомі методи розв'язання задач лінійного програмування.

10.2. Геометрична інтерпретація задачі дробово-лінійного програмування

Нехай дана задача дробово-лінійного програмування для випадку двох змінних:

$$Z = \frac{c_1 x_1 + c_2 x_2}{d_1 x_1 + d_2 x_2} \to \max$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 \le b_1, \\ a_{21}x_1 + a_{22}x_2 \le b_2, \\ \dots \quad \dots \quad \dots \quad \dots \quad \dots \\ a_{m1}x_1 + a_{m2}x_2 \le b_m, \\ x_1 \ge 0, \quad x_2 \ge 0. \end{cases}$$

Як і в лінійному програмуванні будуємо область допустимих розв'язків відповідно до системи обмежень. Для побудови цільової функції її вираз $\dfrac{c_1x_1 + c_2x_2}{d_1x_1 + d_2x_2} = Z$ слід спростити, припустивши $d_1x_1 + d_2x_2 > 0$. Маємо:

$$(c_1 - Zd_1)x_1 + (c_2 - Zd_2)x_2 = 0,$$

звідки виразимо

$$x_2 = -\frac{(c_1 - Zd_1)}{(c_2 - Zd_2)}x_1,$$

$$x_2 = k\ x_1,$$

де $k(Z) = \dfrac{(Zd_1 - c_1)}{(c_2 - Zd_2)}$ – кутовий коефіцієнт, який є функцією від Z.

Рівняння $x_2 = k\ x_1$ представляє пряму з кутовим коефіцієнтом, що проходить через початок координат. Якщо Z приймає постійне значення ($Z = const$), то деяке постійне значення отримає і коефіцієнт k, а пряма займе на площині в прямокутній системі координат певне положення (рис. 10. 1). За іншого постійного значення Z пряма $x_2 = k\ x_1$ займе інше положення. Воно як би повернеться навколо початку координат.

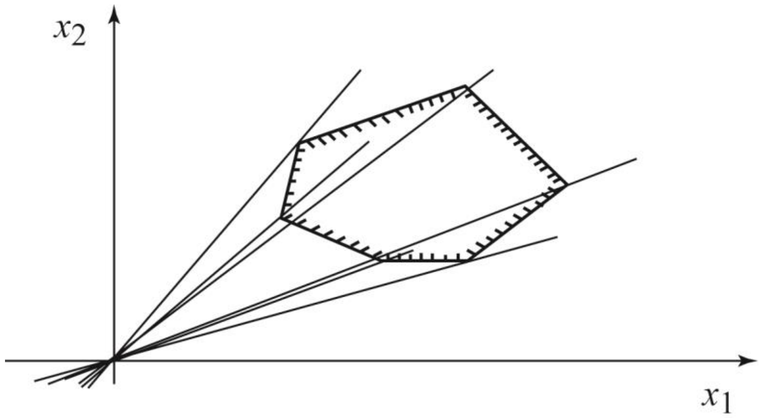


Рис. 10.1.

Необхідно з'ясувати знак похідної кутового коефіцієнта k стосовно Z:

$$k'(Z) = \frac{(c_2d_1 - c_1d_2)}{(c_2 - Zd_2)^2}.$$

Звідси видно, що знаменник дробу завжди додатний, а чисельник не залежить від цільової функції Z. Отже, похідна k' має постійний знак.

Правило 1. Якщо $k'(Z)=\dfrac{(c_2d_1-c_1d_2)}{(c_2-Zd_2)^2}>0$, то оскільки знаменник завжди додатний, слідує, що $(c_2d_1-c_1d_2)>0$, то для відшукання точки максимуму необхідно повертати пряму, яка описує цільову функцію біля початку координат у напрямку проти годинникової стрілки.

Правило 2. Якщо $k'(Z)=\dfrac{(c_2d_1-c_1d_2)}{(c_2-Zd_2)^2}<0$, то $(c_2d_1-c_1d_2)<0$ і для відшукання точки максимуму необхідно повертати пряму, яка описує цільову функцію біля початку координат у напрямку за годинниковою стрілкою.

Зауваження. Багатокутник розв'язків задачі дробово-лінійного програмування може бути: обмежений і максимальне та мінімальне значення досягаються у його кутових точках; необмежений, проте існують кутові точки, в яких досягаються максимальне та мінімальне значення цільової функції; необмежений і досягається лише один із екстремумів; необмежений і точки екстремумів визначити неможливо.

Приклад. На плодоконсервному заводі з трьох видів фруктів (яблука, груші, сливи) виготовляють компот двох видів. Кількість фруктів, необхідна для приготування 1 л компоту, запас фруктів і витрати подані в табл. 10.1. Відомо також, що яблук можна витрачати не більше 500 кг, груш і слив – не менше 400 і 300 кг відповідно. Потрібно скласти план виробництва компоту двох видів і отримати максимальну та мінімальну собівартість 1 л компоту.

Таблиця 10.1

Умова задачі

Фрукти	Запаси, кг	Витрата (кг) на компот виду	
		I	II
Яблука	500	0,3	0,2
Груші	400	0,2	0,3
Сливи	300	0,3	0,2
Витрати на 1 л компоту, грош. од.		0,5	0,6

Позначимо: x_1 – кількість літрів компоту I виду, x_2 – кількість літрів компоту II виду. Тоді система обмежень за витратою фруктів прийме вигляд:

$$\begin{cases} 0{,}3x_1 + 0{,}2x_2 \le 500 \\ 0{,}2x_1 + 0{,}3x_2 \ge 400 \\ 0{,}3x_1 + 0{,}2x_2 \ge 300 \end{cases}$$
$$x_1 \ge 0,\ x_2 \ge 0$$

Витрати на виробництво компоту: $z_1 = 0{,}5x_1 + 0{,}6x_2$, загальний обсяг виробництва компоту представляється функцією $z_2 = x_1 + x_2$, відношення загальних витрат до загального обсягу виробництва дає собівартість 1 л компоту:

$$Z = \frac{0{,}5x_1 + 0{,}6x_2}{x_1 + x_2} \to \min,$$

$$\begin{cases} 0{,}3x_1 + 0{,}2x_2 \le 500, \\ 0{,}2x_1 + 0{,}3x_2 \ge 400, \\ 0{,}3x_1 + 0{,}2x_2 \ge 300, \end{cases}$$
$$x_1 \ge 0,\ x_2 \ge 0.$$

Необхідно створити область допустимих розв'язків (ОДР). Щоб визначити, в якій точці ОДР цільова функція буде мати екстремальне значення, виразимо x_2:

$$x_2 = \frac{0{,}5 - Z}{Z - 0{,}6}x_1,$$

його похідна

$$k' = \frac{0{,}1}{(Z - 0{,}6)^2}.$$

Оскільки похідна за будь-якого значення Z додатна, функція $k = \dfrac{0{,}5 - Z}{Z - 0{,}6}$ зростаюча; із збільшенням Z кутовий коефіцієнт збільшується. Це відповідає обертанню прямої проти ходу годинникової стрілки (рис. 10.2).

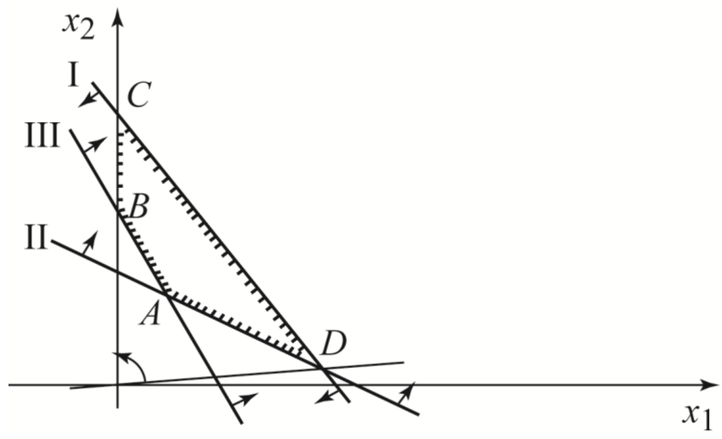


Рис. 10.2

Таким чином, найбільше значення цільова функція буде досягати в вершинах *B* і *C* області припустимих рішень, тобто максимальна собівартість визначається у всіх точках ребра *BC*. Координати точок *B* і *C* рівні відповідно (0; 1 500) і (0; 2 500); цільова функція в цих точках:

$$Z_{\max} = Z(B) = Z(C) = 0{,}6,$$

тобто максимальна собівартість 1 л компоту становить 60 грош. од.

Мінімальна собівартість досягається в точці *D* області припустимих розв’язків; цільова функція в точці *D*:

$$Z_{\min} = Z(D) \approx 0{,}52.$$

Отже, мінімальна собівартість 1 л компоту 52 грош. од, максимальна – 60 грош. од.

На рис. 10.2 видно, що собівартість 1 л компоту залишається постійною, рівною 60 грош. од. у процесі виробництва компоту II виду від 1 500 до 2 500 л., тобто $1\,500 \le x_2 \le 2\,500$.

Зауваження. Розв'язання задач графічним методом може бути простішим. Досить знайти координати крайніх точок області допустимих розв'язків, а потім обчислити значення цільової функції в цих точках і вибрати з них найбільше і найменше.

10.3. Розв’язування дрібно-лінійної задачі зведенням до задачі лінійного програмування

Задачу дробово-лінійного програмування можна звести до задачі лінійного програмування і вирішити симплексним методом.

Позначимо $y_0 = \dfrac{1}{\sum\limits_{j=1}^{n} d_j x_j}$ за умови $\sum\limits_{j=1}^{n} d_j x_j \neq 0$. Цільова функція функція буде мати вигляд:

$$Z = \frac{\sum\limits_{j=1}^{n} c_j x_j}{\sum\limits_{j=1}^{n} d_j x_j} = \sum_{j=1}^{n} c_j x_j \frac{1}{\sum\limits_{j=1}^{n} d_j x_j} = \sum_{j=1}^{n} c_j x_j y_0 = \sum_{j=1}^{n} c_j y_j ,$$

де нова змінна $y_j = y_0 x_j$, звідки $x_j = \dfrac{y_j}{y_0}$. Тоді система обмежен $\sum\limits_{j=1}^{n} a_{ij} x_j = b_i$ буде $\sum\limits_{j=1}^{n} a_{ij} \dfrac{y_j}{y_0} = b_i$ або $\dfrac{\sum\limits_{j=1}^{n} a_{ij} y_j}{y_0} = b_i$. Маємо $\sum\limits_{j=1}^{n} a_{ij} y_j - b_i y_0 = 0.$ Та з урахуванням $y_0 = \dfrac{1}{\sum\limits_{j=1}^{n} d_j x_j}$ або $\dfrac{1}{y_0} = \sum\limits_{j=1}^{n} d_j x_j$ та $\sum\limits_{j=1}^{n} d_j x_j y_0 = 1$, а отже:

$$\sum_{j=1}^{n} d_j y_j = 1,$$
$$\left(i = \overline{1,m}\right), \quad y_j \geq 0 \quad \left(j = \overline{1,n}\right), \quad y_0 > 0.$$

Отже отримаємо таку модель задачі:

$$Z = \sum_{j=1}^{n} c_j y_j \to max,$$

за умови
$$\begin{cases} \sum\limits_{j=1}^{n} a_{ij} y_j - b_i y_0 = 0, \\ \sum\limits_{j=1}^{n} d_j y_j = 1, \\ \left(i = \overline{1,m}\right), \quad y_j \geq 0 \quad \left(j = \overline{1,n}\right), \quad y_0 > 0. \end{cases}$$

Маємо звичайну задачу лінійного програмування, яка розв’язується симплексним методом. Після знаходження оптимального розв'язку отриманої задачі, використовуючи вказані вище співвідношення, знаходять оптимальний розв'язок вихідної задачі дробово-лінійного програмування.

Приклад. Дано задачу дробово-лінійного програмування:

$$F = \frac{2x_1 - x_2}{x_1 + 2x_2 + 1} \to \max \text{ за обмежень:}$$

$$\begin{cases} x_1 - 2x_2 + x_3 = 2, \\ 2x_1 + x_2 + x_4 = 6, \\ x_j \ge 0,\ j = \overline{1,4}. \end{cases}$$

Позначимо $x_1 + 2x_2 + 1 = \dfrac{1}{y_0}$, $y_0 > 0$, тоді $F = 2x_1 y_0 - x_2 y_0$.

Позначимо: $x_1 y_0 = y_1, x_2 y_0 = y_2, x_3 y_0 = y_3, x_4 y_0 = y_4$.

Перетворимо систему обмежень, помноживши обидві частини всіх обмежень на y_0, і перейдемо до змінних y_0, y_1, y_2, y_3, y_4. Задача набуде вигляду $F = 2y_1 - y_2 \to \max$ за обмежень:

$$\begin{cases} -2y_0 + y_1 - 2y_2 + y_3 = 0, \\ -6y_0 + 2y_1 + y_2 + y_4 = 0, \\ y_0 + y_1 + 2y_2 = 1, \\ y_j \ge 0,\ j = \overline{1,4},\ y_0 > 0. \end{cases}$$

№	Базис	$c_{баз}$	c_j	0	2	-1	0	0	Примітки
			A_0	y_0	A_1	A_2	A_3	A_4	
1	A_3	0	0	-2	1	-2	1	0	
2	A_4	0	0	-6	2	1	0	1	
3			1	1	1	2	0	0	
4	A_3	0	2	0	3	2	1	0	
5	A_4	0	6	0	8	13	0	1	
6	y_0	0	1	1	1	2	0	0	
7	Δ_1		0	0	-2	1	0	0	
8	A_1	2	2/3	0	1	2/3	1/3	0	

9	A_4	0	2/3	0	0	2/3	-8/3	1	
10	y_0	0	1/3	1	0	4/3	-1/3	0	
11	Δ_2		4/3	0	0	7/3	2/3	0	

Маємо $Y^* = (1/3, 2/3, 0, 0, 2/3)$.

Тоді $x_1 = \dfrac{y_1}{y_0} = 2, \quad x_2 = \dfrac{y_2}{y_0} = 0, \quad x_3 = \dfrac{y_3}{y_0} = 0, \quad x_4 = \dfrac{y_4}{y_0} = 2,$

$X^* = (2, 0, 0, 2), \quad F_{\max} = 4/3.$

Запитання для самоперевірки

1. Наведіть економічну і математичну постановку задачі дробово-лінійного програмування.

2. Які методи застосовуються для розв'язання задачі дробово-лінійного програмування?

3. Яка геометрична інтерпретація задачі дробово-лінійного програмування?

4. Як розв'язання дробово-лінійної задачі зводиться до задачі лінійного програмування?

11. Цілочислові задачі лінійного програмування. Основні методи їх розв'язання та аналізу

11.1. Економічна постановка задачі цілочислового програмування та її математична модель.

11.2. Графічний метод розв'язання цілочислових задач.

11.3. Основні аналітичні методи розв'язання цілочислових задач. Метод Гоморі.

11.4. Метод гілок і меж.

11.1. Економічна постановка задачі цілочислового програмування та її математична модель

Значна частина задач комерційної діяльності вимагає цілочислового розв'язку. До них належать задачі, у яких змінні величини означають кількість одиниць неділимої продукції, наприклад, розподілу товарів між підприємствами, розкрій матеріалів, число верстатів під час завантаження обладнання, розподіл транспортних засобів за рейсами, продаж автомобілів, розподіл літаків за авіалініями, кількість персональних комп'ютерів у керуючому комплексі і т. д. Лінійні задачі, розв'язок яких має бути отримано в цілих числах, називають задачами цілочислового програмування. У випадку, коли цілочислових значень повинні набувати не всі, а тільки одна або декілька змінних, задача називається частково цілочисловою. До цілочислових задач належать і задачі, в яких змінні набувають значення 0 або 1. Слід відмітити, що умова цілочисловості є нелінійна і може зустрічатись в задачах, які містять як лінійні, так і нелінійні функції.

У загальному вигляді математична модель цілочислового програмування має вигляд:

$$Z = \sum_{j=1}^{n} c_j x_j \to \max(\min), \tag{11.1}$$

$$\sum_{j=1}^{n} a_{ij} x_j = b_i, \tag{11.2}$$

$$\begin{aligned} & i = \overline{1,m},\ j = \overline{1,n}, \\ & x_j \ge 0, \quad x_j - \text{цілі числа}. \end{aligned} \tag{11.3}$$

11.2. Графічний метод розв'язання цілочислових задач

За наявності в задачі лінійного програмування двох змінних, а в системі обмежень – нерівностей, вона може бути розв'язана графічним методом.

У системі координат X_1OX_2 знаходять область допустимих розв'язків, будують вектор $\overline{C}$ і лінію рівня. Переміщуючи лінію рівня по напрямку $\overline{C}$ для задач на максимум, знаходимо найбільш віддалену від початку координат точку і її координати.

У тому випадку, коли координати цієї точки цілочислові, в області допустимих розв'язків будують цілочислову решітку і знаходять на ній такі цілі числа, які задовольняють систему обмежень і за яких значення цільової функції найближче до екстремального не цілочислового розв'язку. Координати такої вершини і є цілочисловим розв'язком. Аналогічно розв'язується задача на мінімум.

Отже, особливість геометричної інтерпретації цілочислової задачі, порівняно зі звичайною задачею лінійного програмування, полягає у визначенні множини допустимих розв’язків. Областю допустимих розв’язків загальної задачі лінійного програмування є опуклий багатогранник. З урахуванням вимоги цілочисловості з даної області вибирають множину дискретних розв’язків, яка складається з окремих точок. Якщо в умові задачі наявні дві змінні, розв’язок задачі можна відшукати графічно, використавши цілочислову сітку, а якщо це не можливо, то слід застосувати спеціальні методи.

Приклад. Для покращення фінансового становища фірма прийняла рішення щодо збільшення випуску конкурентоспроможної продукції, для чого прийнято рішення про встановлення в одному з цехів додаткового обладнання, що займає 19/3 $м^2$ площі. На придбання додаткового обладнання фірма виділила 10 ум. од., при цьому вона може купити обладнання двох видів. Придбання першого комплекту обладнання 1-го виду коштує 1,0 ум. од., другого виду – 3 ум. од. Придбання одного комплекту обладнання 1-го виду дозволяє збільшити випуск продукції в зміну на 2 шт., а одного комплекту обладнання 2-го виду – на 4 шт. Знаючи, що для встановлення одного комплекту обладнання 1-го виду потрібно 2 $м^2$ площі, а для обладнання 2-го виду – 1 $м^2$ площі, визначити такий набір додаткового обладнання, який дає можливість максимально збільшити випуск продукції.

Складемо математичну модель задачі.

$$Z = 2x_1 + 4x_2 \to max$$

$$\begin{cases} 2x_1 + x_2 \le 19/3, \\ x_1 + 3x_2 \le 10, \\ x_1, x_2 - \text{цілі числа.} \end{cases}$$

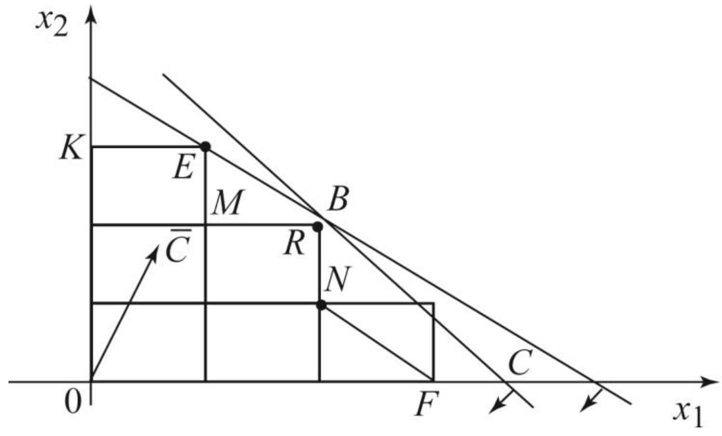


Рис. 11.1.

Без умови цілочисловості чотирикутник *OABC* – область допустимих розв'язків. Оптимальний розв'язок у точці $B\left(\frac{9}{5}, \frac{41}{15}\right)$, $Z_{\max} = \frac{218}{15}$ ум. од.

Умову цілочисловості задовольняють 12 точок. Замінюємо багатокутник *OABC* на *OKEMRNF*, що містить всі допустимі точки з цілочисловими координатами. За градієнтом оптимум у точці $E(1, 3)$, $Z_{\max}(X_{\text{цел}}) = 2 \cdot 1 + 4 \cdot 3 = 14$ ум. од.

Висновок. Фірмі слід придбати один комплект обладнання першого виду і три комплекти обладнання другого виду, що забезпечить їй за наявних обмежень на виробничі площі та кошти максимальне збільшення випуску продукції, що дорівнює 14 ум. од за зміну.

11.3. Основні аналітичні методи розв'язання цілочислових задач. Метод Гоморі

Методи цілочислової оптимізації можна розділити на три основні групи:

а) метод відсікання; б) комбіновані методи; в) наближені методи.

Метод відсікання полягає в поступовому звуженні області допустимих розв’язків. Спочатку задача розв’язується без урахування вимог цілочисловості змінних, продовжується з введенням даної вимоги, при цьому область допустимих розв’язків зменшується до тих пір, доки в оптимальному розв’язку не будуть змінні, значення яких цілі числа.

Розглянемо один з методів відсікання – метод Гоморі.

Алгоритм Гоморі:

1) розв'язується задача без урахування умов цілочисловості;

2) отриманий оптимальний розв'язок (якщо він існує) перевіряється на цілочисловість. Якщо умова цілочисловості виконується за всіма змінними, то оптимальний розв'язок знайдено. Якщо ця умова не виконується, то необхідно перейти до наступного етапу або робити висновок, що задача виявилась нерозв'язаною (випадок: базисна змінна дробова, а всі коефіцієнти в рядку цілі числа);

3) будується додаткове обмеження, що відтинає частину області, в якій міститься оптимальний розв'язок задачі (11.1), (11.2) і не міститься жодного допустимого розв'язку задачі (11.1) – (11.3);

4) останній етап передбачає повернення до задачі лінійного програмування з відкинутою умовою цілочисловості, але з розширеною системою обмежень, в яку включено додаткове обмеження, яке отримане на 3-му кроці. До розширеної системи обмежень знову застосовується симплексна процедура. Якщо знайдений таким чином розв'язок буде знову не цілочисловим, то формується нове додаткове обмеження і процес обчислень повторюється.

Алгоритм Гоморі дозволяє за кінцеве число кроків дійти до оптимального цілочислового розв'язку, якщо він існує. Геометрична інтерпретація розв'язків цілочислової задачі методом Гоморі на площині представлена на рис. 11.2.

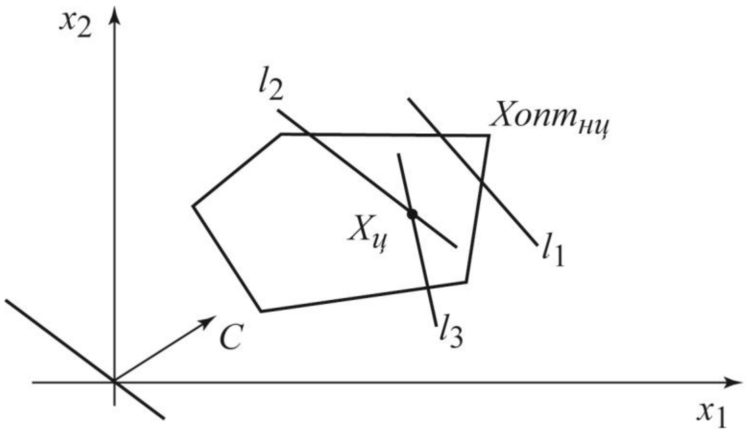


Рис. 11.2.

На рис. 11.2 $X_{опт\ нц}$ – оптимальний розв'язок без урахування цілочисловості; три прямі l_1, l_2, l_3 відповідають трьом додатковим лінійним обмеженням.

Необхідно скласти правильне додаткове обмеження, яке відсікає частину області допустимих розв'язків. Правило відсікання повинно задовольняти наступні умови: 1) бути лінійним; 2) відсікає знайдений оптимальний нецілочисловий розв'язок задачі; 3) не відсікає жодної з цілочислових точок задачі (11.1) – (11.3).

Процедура формування правильного відсікання

Після кожної ітерації система обмежень має вигляд:

$$x_i = \beta_i - \sum_{x_j \in \{\text{ВЗ}\}} \alpha_{ij} x_j \,, x_{\text{j}} \in \{\text{ВЗ}\} - \text{множина вільних змінних},$$

$$x_i \in \{\text{БЗ}\} - \text{множина базисних змінних}.$$

Якщо виконується умова оптимальності задачі, то потрібно знайти оптимальний розв'язок. Якщо всі компоненти оптимального плану є цілочисловими, то задача розв'язана.

Нехай деякі β_0 – нецілі, нехай компонента i_0 – неціла. Розглянемо рівність, для якої виконується умова оптимальності:

$$x_{i0} = \beta_{i0} - \sum_{x_j \in \{\text{ВЗ}\}} \alpha_{i0} x_j \,.$$

Число представляється як: $\alpha = [\alpha] + \{\alpha\}, \quad 0 \le \{\alpha\} < 1$, тоді:

$$\beta_{i0} = [\beta_{i0}] + \{\beta_{i0}\}, \quad \alpha_{i0j} = [\alpha_{i0j}] + \{\alpha_{i0j}\}, \quad x_j \in \{\text{ВЗ}\}.$$

Нехай β_{i0} неціле, то $\{\beta_{i0}\}$>0 і $\{\alpha_{i0j}\} \ge 0$, тоді

$$x_i = \left([\beta_{i0}] - \sum_{x_j \in \{\text{ВЗ}\}} [\alpha_{i0j}] x_j \right) + \left(\{\beta_{i0}\} - \sum_{x_j \in \{\text{ВЗ}\}} \{\alpha_{i0j}\} x_j \right). \tag{11.4}$$

Оскільки перший доданок рівності (11.4) є ціле число, то для того щоб x_{i0} було цілим, необхідно, щоб другий доданок також був цілим, тобто величина $L_{i0} = \{\beta_{i0}\} - \sum_{x_j \in \{\text{ВЗ}\}} \{\alpha_{i0j}\} x_j$ повинна бути цілим числом. Необхідно показати, що $L_{i0} \le 0$ – ціле число. Величина $\sum_{x_j \in \{\text{ВЗ}\}} \{\alpha_{i0j}\} x_j$ не може бути від'ємною; з умови $0 \le \{\beta_{i0}\} < 1$. І враховуючи, що L_{i0} – ціле, то якщо $L_{i0} > 0$, випливає, що $\{\beta_{i0}\} > 1$ – це суперечить визначенню дробової частини числа.

Таким чином, доведено, що будь-який допустимий розв'язок задачі (11.1) – (11.3) має задовольняти нерівності $\{\beta_i\} - \sum\limits_{x_j \in \{ВЗ\}} \{\alpha_{ij}\} x_j \le 0$.

Теорема. Нерівність $\{\beta_i\} - \sum\limits_{x_j \in \{ВЗ\}} \{\alpha_{ij}\} x_j \le 0$ визначає правильне відсікання Гоморі, тобто: 1) є лінійним; 2) відсікає знайдений оптимальний не цілочисловий розв'язок задачі; 3) не відсікає жодного цілочислового плану задачі.

Продовження прикладу. Розв'язати задачу цілочислового програмування методом Гоморі.

$$Z = 2x_1 + 4x_2 \to \max,$$

$$\begin{cases} 2x_1 + x_2 \le 19/3, \\ x_1 + 3x_2 \le 10, \\ x_1, x_2 - \text{цілі числа.} \end{cases} \to \begin{cases} 2x_1 + x_2 + x_3 = 19/3, \\ x_1 + 3x_2 + x_4 = 10, \\ x_1, x_2 - \text{цілі числа.} \end{cases}$$

Розв’язавши задачу без врахування умови цілочисловості, в таблиці отримали нецілочисловий оптимальний розв'язок :

$$X^*\left(\frac{9}{5}, \frac{41}{15}\right),\ Z_{\max} = \frac{218}{15} \text{ ум. од.}$$

Варто формувати додаткове обмеження:

знайти дробові частини чисел 9/5 і 41/15

$\{a\} = a - [a]$**;**

$\left\{\frac{9}{5}\right\} = \frac{9}{5} - 1 = \frac{4}{5}$, $\left\{\frac{41}{15}\right\} = \frac{41}{15} - 2 = \frac{11}{15}$, $\frac{4}{5} = \frac{12}{15} > \frac{11}{15}$ отже, скласти додаткове обмеження для 5-го рядка. Знайти дробові частини $\left\{\frac{3}{5}\right\} = \frac{3}{5} - 0 = \frac{3}{5}$, $\left\{-\frac{1}{5}\right\} = -\frac{1}{5} - (-1) = -\frac{1}{5} + 1 = \frac{4}{5}$, додаткове обмеження буде мати вигляд:

$$\frac{3}{5}x_3 + \frac{4}{5}x_4 \ge \frac{4}{5} \text{ або } \frac{3}{5}x_3 + \frac{4}{5}x_4 - x_5 = \frac{4}{5}$$

і додати в симплекс-таблицю (рядок 7).

№	Базис	$c_{баз}$	c_j	2	4	0	0	0	Примітки
			A_0	A_1	A_2	A_3	A_4	A_5	
1	A_3	0	19/3	2	1	1	0		

2	A_4	0	10	1	3	0	1		
		Δ_1	0	-2	-4	0	0		
3	A_3	0	3	5/3	0	1	-1/3		
4	A_2	4	10/3	1/3	1	0	1/3		
	Δ_2		40/3	-2/3	0	0	4/3		
5	A_1	2	9/5	1	0	3/5	-1/5	0	
6	A_2	4	41/15	0	1	-1/5	2/5	0	
		Δ_3	218/15	0	0	2/5	6/5		
7	A_5	0	-4/5	0	0	-3/5	-4/5	1	
8	A_1	2	1	1	0	0	-1	1	
9	A_2	4	3	0	1	0	2/3	-1/3	
10	A_3	0	4/3	0	0	1	4/3	-5/3	
11		Δ_3	14	0	0	0	2/3	2/3	

Цілочисловий оптимальний розв'язок:

$X_{ц}^{*}(1,3),\ Z_{\max} = 14$ ум. од.

Комбіновані методи цілочислової оптимізації передбачають перебір всіх допустимих цілочислових розв’язків, але при цьому перебір є цілеспрямований у невеликій області розв’язків. Найпоширенішим у цій групі методів є метод гілок і меж.

Наближені методи розв’язування також досить часто використовують. У реальних задачах часто досить наближеного розв’язку. До наближених методів відносять метод локальної оптимізації (метод вектора спаду), модифікації точних методів, методи випадкового пошуку та ін. Багато наближених методів використовують алгоритми точних методів, як наприклад, метод гілок і меж.

11.4. Метод гілок і меж

Основи цього методу доцільно розглянути на чисельному прикладі.

Приклад. Знайти

$$Z = 5x_1 + 4x_2 \to max$$

$$\begin{cases} x_1 + x_2 \le 5, \\ 10x_1 + 6x_2 \le 45, \\ x_1, x_2 \ge 0, \text{цілі числа.} \end{cases}$$

Початкову задачу лінійного програмування (ЛП0) (тобто без урахування цілочисловості) розв'яжемо графічно (рис. 11.3). Її оптимальним розв'язком буде:

$x_1 = 3{,}75;\ x_2 = 1{,}25;\ z = 23{,}75.$

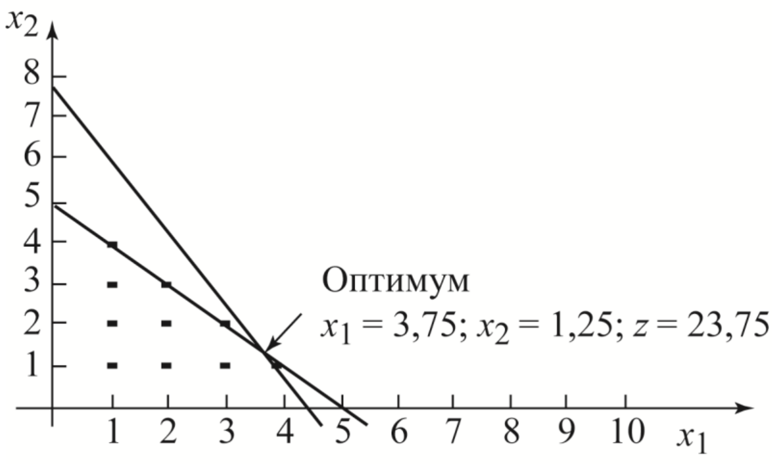


Рис. 11.3.

Оскільки оптимальний розв'язок не цілочисловий, варто застосувати метод гілок і меж, який змінює простір розв'язків задачі лінійного програмування. Спочатку обирається одна з змінних, яка має бути цілочисловою і значення якої в оптимальному розв'язку нецілочислове. Оберемо $x_1 = 3{,}75$. Область $3 < x_1 < 4$ простору допустимих розв'язків задачі ЛП0 не містить цілочислових значень змінної x_1 і може бути виключена з розгляду. Це еквівалентно заміні вихідної задачі ЗЛ0 двома новими задачами лінійного програмування ЛП1 і ЛП2, які визначаються так:

1) простір допустимих розв'язків ЛП1 – простір допустимих розв'язків ЛП0 + ($x_1 \le 3$);

2) простір допустимих розв'язків ЛП2 – простір допустимих розв'язків ЛП0 + ($x_1 \ge 4$).

На рис. 11.4 зображені простори допустимих розв'язків задач ЛП1 і ЛП2.

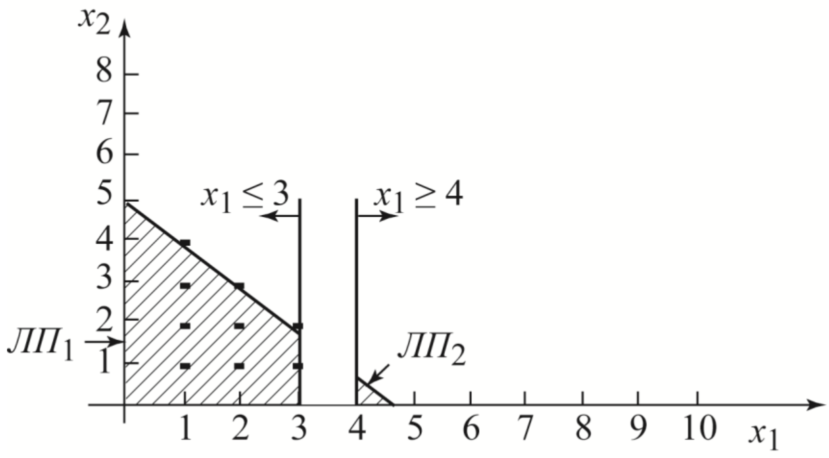


Рис. 11.4.

Далі будемо розв'язувати задачу цілочислового програмування шляхом розв'язання послідовності безперервних задач лінійного програмування.

Нові обмеження $x_1 \le 3$ і $x_1 \ge 4$ взаємовиключні, оскільки ЛП1 і ЛП2 необхідно розглядати як незалежні задачі лінійного програмування (рис. 11.5).

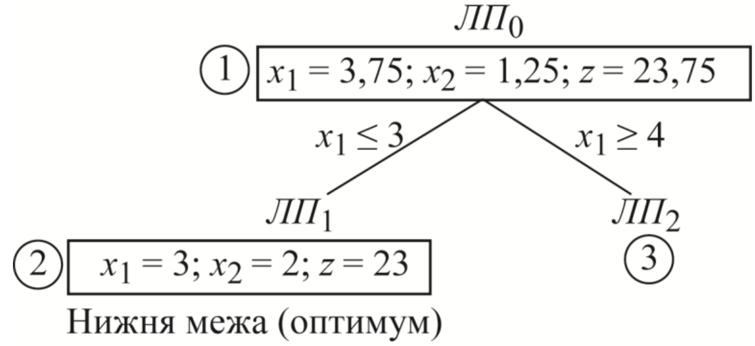


Рис. 11.5.

Діхотомізація задач ЛП – основа концепції розгалуження в методі гілок і меж. У цьому випадку x_1 називається змінною розгалуження.

Оптимальний розв'язок задачі ЦЛП знаходиться в просторі допустимих розв'язків або задачі ЛП1, або ЛП2. Отже, обидві підзадачі повинні бути розв'язані. Потрібно вибирати спочатку задачу ЛП1 (вибір довільний), що має додаткове обмеження $x_1 \le 3$, тобто:

$$Z = 5x_1 + 4x_2 \to max$$
$$\begin{cases} x_1 + x_2 \le 5, \\ 10x_1 + 6x_2 \le 45, \\ x_1 \le 3, \\ x_1, x_2 \ge 0, \text{цілі числа.} \end{cases}$$

Оптимальним розв'язком задачі ЛП1 є $x_1 = 3, x_2 = 2, z = 23$. Оптимальний розв'язок задачі ЛП1 задовольняє вимогу цілочисловості змінних, у цьому випадку говорять, що задача ЛП1 прозондована і можна сказати, що значення $z = 23$ є нижньою межею оптимального (максимального) значення цільової функції вихідної задачі ЦЛП. При значенні нижньої межі $z = 23$ варто дослідити задачу ЛП2. Так як у задачі ЛП0 оптимальне значення цільової функції дорівнює 23,75 і всі її коефіцієнти є цілими числами, то неможливо отримати цілочисловий розв'язок задачі ЛП2 (простір розв'язків якої більш вузький, ніж у задачі ЛП0), який буде краще наявного. У результаті потрібно відкинути підзадачу ЛП2 і вважати її прозондованою.

Необхідно розглянути два моменти: 1) під час вибору підзадачі можна для зондування розв'язати спочатку задачу ЛП2 замість ЛП1; 2) в задачі ЛП0 спочатку вибрати змінну x_2 змінну розгалуження замість x_1.

Ситуація, коли першою розв'язується задача ЛП2, ілюструється схемою обчислень (рис. 11.6). Оскільки значення змінної $x_2 = 0{,}83$ не є цілим числом, то задача ЛП2 досліджується далі. Варто розглянути задачі ЛП3 і ЛП4, використовуючи гілки $x_2 \le 0$ і $x_2 \ge 0$ відповідно. Це означає, що:

1) простір розв'язків ЛП3 = простір розв'язків ЛП2 + ($x_2 \le 0$) = простір розв'язків ЛП0 + ($x_1 \ge 4$) + ($x_2 \le 0$);

2) простір розв'язків ЛП4 = простір розв'язків ЛП2 + ($x_2 \ge 1$) = простір розв'язків ЛП0 + ($x_1 \ge 4$) + ($x_2 \ge 1$).

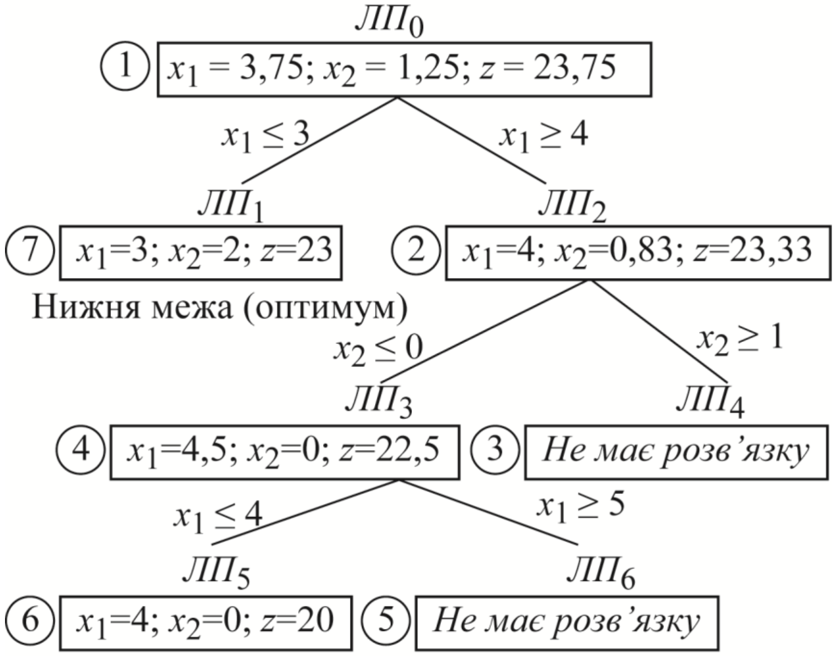


Рис. 11.6.

Є ще три нерозглянуті задачі, які повинні бути розв'язані – ЛП1, ЛП3 і ЛП4. Нехай довільно обрана першою задача ЛП4. Ця задача не має розв'язку, і отже, є прозондованою. Наступною задачею обрана підзадача ЛП3. Її оптимальним розв'язком є $x_1 = 4{,}5, x_2 = 0, z = 22{,}5$. Нецілочислове значення змінної $x_1 = 4{,}5$ породжує дві гілки розв'язку при $x_2 \le 4$ і $x_2 \ge 5$ і відповідні їм підзадачі ЛП5 і ЛП6. При цьому:

1) простір розв'язків

ЛП5 = простір розв'язків ЛП0 + ($x_1 \ge 4$) + ($x_2 \le 0$) +($x_1 \ge 4$).

2) простір розв'язків

ЛП6 = простір розв'язків ЛП0 + ($x_1 \ge 4$) + ($x_2 \le 0$) ($x_2 \ge 5$).

Тепер не розглянуті лише підзадачі ЛП1, ЛП5 і ЛП6. Підзадача ЛП6 прозондована, так як не має допустимих розв'язків. Підзадача ЛП5 має цілочисловий розв'язок $x_1 = 4, x_2 = 0, z = 20$ і отже, породжує нижню межу $z = 20$ оптимального значення цільової функції задачі ЦЛП. Розв'язок підзадачі ЛП1 цілочисловий $x_1 = 3, x_2 = 2, z = 23$. Отже нижню межу значень цільової функції вважаємо рівною 23. Так як усі підзадачі прозондовані, оптимальним розв'язком задачі ЦЛП є розв'язок, що відповідає останній нижній межі, тобто $x_1 = 3, x_2 = 2, z = 23$.

Послідовність розв'язання підзадачі, показана на рис. 11.6 (ЛП0, ЛП2, ЛП4, ЛП3, ЛП6, ЛП5, ЛП1) є найгіршою, тим не менш, вона зустрічається на практиці. Даний приклад вказує на основну слабкість методу гілок і меж: як обрати наступну підзадачу для дослідження і як вибрати для неї змінну розгалуження? Існують евристичні міркування, що дозволяють «вгадати», яка з гілок може призвести до погіршення розв'язання задачі ЦЛП, при цьому не існує строгої теорії, яка завжди забезпечувала б надійні результати.

Розглянута процедура застосована для розв'язання задач максимізації. Для розв'язання задач мінімізації в алгоритмі необхідно замінити нижню межу на верхньою (початкове значення якої дорівнює $z = +\infty$). Алгоритм методу гілок і меж безпосередньо поширюється на задачі частково-цілочислового ЛП, в яких лише деякі із змінних повинні приймати цілочисловий розв'язок.

Запитання для самоперевірки

1. Які задачі в економіці потребують цілочислового розв'язку?
2. Запишіть в загальному вигляді математичну модель цілочислового програмування.
3. Як розв'язати задачу цілочислового програмування графічним методом?
4. Яка геометрична інтерпретація розв'язків цілочислової задачі на площині?
5. Наведіть алгоритм методу Гоморі.
6. У чому суть процедури формування правильного відсікання?
7. Розкажіть загальну ідею реалізації методу гілок і меж.

12. Методи нелінійного програмування

12.1. Загальна постановка задачі опуклого програмування.

12.2. Графічний метод.

12.3. Метод множників Лагранжа.

12.4. Теорема Куна-Таккера.

12.1. Загальна постановка задачі опуклого програмування

Нелінійне програмування застосовується під чса оптимізації промислового виробництва, управління товарними ресурсами, планування обслуговування та ремонту обладнання і т. д.

Для задачі нелінійного програмування на відміну від лінійних задач немає єдиного методу розв'язання. Залежно від виду цільової функції і системи обмежень розроблені спеціальні методи розв'язання, до яких відносяться методи множників Лагранжа, квадратичне і опукле програмування, градієнтні методи, наближені методи розв'язання, графічний метод.

Задача математичного програмування:

$$\begin{aligned}
&\max(\min)\, Z = f(x),\\
&\varphi_i(x)\{\le, =, \ge\}\, b_i \quad \left(i=\overline{1,m}\right),\\
&x \ge 0, \quad x=(x_1, \ldots, x_n),
\end{aligned}$$

у якій або цільова функція, або обмеження, або і те й інше нелінійні, називається нелінійною.

Нелінійні задачі складають широкий клас настільки складних задач, що досі не розроблені досконалі загальні методи, такі як симплекс-метод у лінійному програмуванні, які дозволяли б розв'язувати будь-які нелінійні задачі. Але, незважаючи на відсутність універсальних методів, розроблені способи розв'язання окремих спеціальних класів задач, і насамперед задач з опуклими (увігнутими) функціями $f(x)$ і $\varphi_i(x)$.

Визначення. Функція $f(x)$, визначена на опуклій множині $\boldsymbol{X}$, називається опуклою, якщо для будь-яких точок x_1 і x_2 з цієї множини і будь-якого $0 \le \lambda \le 1$ справедлива нерівність:

$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda) f(x_2).$$

Якщо в даному співвідношенні при $0 < \lambda < 1$ і будь-яких x_1, $x_2 \in X$ $(x_1 \ne x_2)$ має місце строга нерівність, то $f(x)$ називається строго опуклою.

Визначення. Функція $f(x)$, визначена на опуклій множині $\boldsymbol{X}$, називається увігнутою, якщо для будь-яких точок x_1 і x_2 з цієї множини і будь-якого $0 \le \lambda \le 1$ справедлива нерівність:

$$f(\lambda x_1 + (1-\lambda)x_2) \ge \lambda f(x_1) + (1-\lambda) f(x_2).$$

Якщо в даному співвідношенні при $0<\lambda<1$ і будь-яких x_1, $x_2 \in X$ $(x_1 \neq x_2)$ має місце строга нерівність, то $f(x)$ називається строго увігнутою.

Для розв'язання задач нелінійного програмування важливе значення мають такі теореми.

Теорема 1. Якщо $\varphi(x)$ – опукла функція при всіх $\boldsymbol{x} \geq 0$, то буде опуклою і множина розв'язків системи $\varphi(x) \leq b,\ \boldsymbol{x} \geq 0$.

Аналогічна теорема доводиться і для увігнутих функцій. Доводиться, що опукла функція $f(x)$, визначена на опуклій множині $\boldsymbol{X}$, неперервна в будь-якій внутрішній точці цієї множини.

Теорема 2. Опукла функція $f(x)$, визначена на опуклій множині $\boldsymbol{X}$, досягає свого глобального мінімуму в кожній точці $\boldsymbol{x}$, в якій градієнт функції обертається в нуль.

Теорема 3. Локальний мінімум опуклої функції $f(x)$, визначеної на опуклій множині $\boldsymbol{X}$, збігається з її глобальним мінімумом на цій множині.

У теорії опуклого програмування як основна зазвичай розглядається задача мінімізації опуклої функції n змінних $Z = f(x)$ за обмежень $\varphi_i(x) \leq 0$ $\left(i=\overline{1,m}\right), x \geq 0$, де функції $\varphi_i(x)$ передбачаються опуклими. Якщо $f(x)$ і $\varphi_i(x)$ є увігнутими функціями, то маємо задачу максимізації $f(x)$ за обмежень $\varphi_i(x) \geq 0$ $\left(i=\overline{1,m}\right), \boldsymbol{x} \geq 0$.

12.2. Графічний метод

В задачах лінійного програмування точки екстремума є вершинами многокутників розв’язків, а в задачах нелінійного програмування вони можуть лежати в середині області, на ребрі, грані або у вершині многокутника.

Задачі нелінійного програмування з двома змінними, в яких їх цільові функції і системи обмежень можуть бути задані в лінійному і нелінійному вигляді, можуть бути розв'язані графічно.

Задача з лінійною цільовою функцією та нелінійною системою обмежень

Приклад. Знайти глобальні екстремуми функції $F = 2x_1 + x_2$ за обмежень:

$$\begin{cases} x_1^2 + x_2^2 \le 16, \\ x_1, x_2 \ge 0. \end{cases}$$

Область допустимих розв'язків – частина кола з радіусом 4, яка розташована в першій чверті. Лініями рівня цільової функції є паралельні прямі з кутовим коефіцієнтом, рівним – 2.

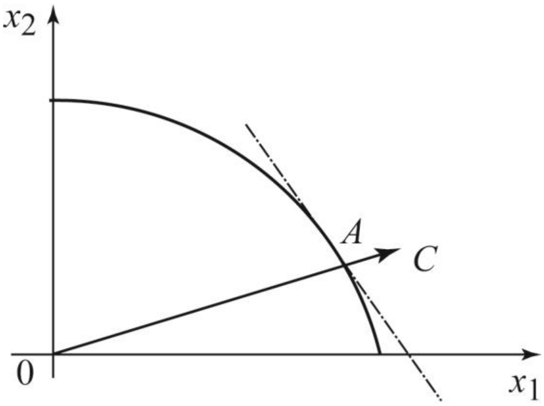


Рис. 12.1.

Необхідно розв’язати систему:

$$\begin{cases} x_1^2 + x_2^2 = 16 \\ x_2 = \frac{1}{2} x_1 \end{cases} \Rightarrow x_1 = 8\sqrt{5}/5, x_2 = 4\sqrt{5}/5, F = 16\sqrt{5}/5 + 4\sqrt{5}/5 = 4\sqrt{5}.$$

Глобальний мінімум, рівний нулю, досягається в точці $O(0, 0)$ (рис. 12.1), глобальний максимум, рівний $4\sqrt{5}$ – в точці $A\left(8\sqrt{5}/5, 4\sqrt{5}/5\right)$.

Задача з нелінійною цільовою функцією та лінійною системою обмежень

Приклад. Знайти глобальні екстремуми функції $F = (x_1 - 2)^2 + (x_2 - 3)^2$ за обмежень:

$$\begin{cases} x_1 + 2x_2 \le 12, \\ x_1 + x_2 \le 9, \\ x_1, x_2 \ge 0. \end{cases}$$

Область допустимих розв'язків – *OABD* (рис. 12.2). Лініями рівня будуть кола з центром у точці O_1. Максимальне значення цільової функції в точці $D(9, 0)$, мінімальне – в точці $O_1(2, 3)$. Тому $F(D) = (9-2)^2 + (0-3)^2 = 58$.

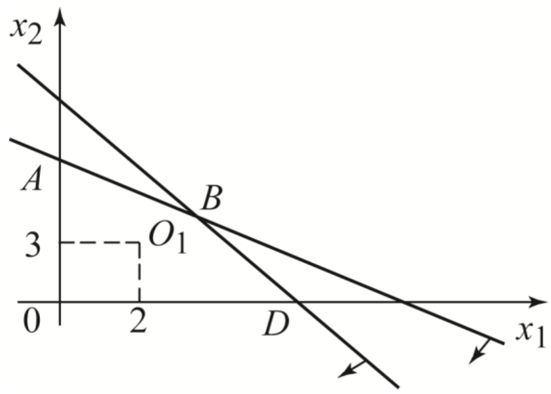


Рис. 12.2.

Отже, глобальний максимум дорівнює 58 і досягається в точці $D(9,0)$, глобальний мінімум дорівнює нулю і досягається в точці $O_1(2,3)$.

12.3. Метод множників Лагранжа

Вважається, що метод множників Лагранжа є класичним методом розв'язування задач математичного програмування, але в практичному застосуванні його можуть виникнути значні обчислювальні труднощі.

Нехай дана задача нелінійного програмування $F = f(x_1, x_2, ..., x_n) \to \max(\min)$ за обмежень: $\varphi_i(x_1, x_2, ..., x_n) = 0, i = \overline{1,m}$.

Припустимо, що функції $f(x_1, x_2, ..., x_n)$ і $\varphi_i(x_1, x_2, ..., x_n)$ неперервні разом зі своїми першими частинними похідними. Обмеження задані у вигляді рівнянь, тому для розв'язання задачі використовується метод відшукання умовного екстремуму функції кількох змінних.

Для розв'язання задачі складають функцію Лагранжа

$$L(x_1, x_2, ..., x_n, \lambda_1, \lambda_2, ..., \lambda_m) = f(x_1, x_2, ..., x_n) + \sum_{i=1}^{m} \lambda_i \varphi_i(x_1, x_2, ..., x_n),$$

де λ_i – множники Лагранжа.

Потім визначаються часткові похідні:

$$\frac{\partial L}{\partial x_j}, \frac{\partial L}{\partial \lambda_i}, i = \overline{1,m}, \ j = \overline{1,n}.$$

Прирівнюючи до нуля частинні похідні, одержимо систему:

$$\frac{\partial L}{\partial x_j} = \frac{\partial f}{\partial x_j} + \sum_{i=1}^{m} \lambda_i \frac{\partial \varphi_i}{\partial x_j} = 0,$$

$$\frac{\partial L}{\partial \lambda_i} = g_i(x_1, x_2, ..., x_n).$$

Розв'язуючи систему, виходить безліч точок, в яких цільова функція L може мати екстремальні значення. Слід зазначити, що умови розглянутої системи є необхідними, але недостатніми. Тому не будь-який отриманий розв'язок визначає точку екстремуму цільової функції. Застосування методу буває виправданим, коли заздалегідь передбачається існування глобального екстремуму, що збігається з єдиним локальним максимумом або мінімумом цільової функції.

Отже, логіка розв'язування задач методом множників Лагранжа така.

1. Скласти функцію Лагранжа.
2. Знайти частинні похідні функції Лагранжа за всіма змінними $x_1, x_2, ..., x_n, \lambda_1, \lambda_2, ..., \lambda_m$ та прирівняти до нуля. Отримаємо систему з $n+m$ рівнянь. Розв'язати отриману систему та знайти всі стаціонарні точки функції Лагранжа.
3. Зі стаціонарних точок, що взяті без $\lambda_1, \lambda_2, ..., \lambda_m$, вибрати точки, в яких функція $f(x_1, x_2, ..., x_n)$ має умовні локальні екстремуми за наявністю обмежень $\varphi_i(x_1, x_2, ..., x_n) = 0, i = \overline{1,m}$. Цей вибір здійснюється з використанням достатніх умов локального екстремума.

Якщо функція $f(x_1, x_2, ..., x_n)$ інтерпретується як дохід або вартість, а вільні члени в обмеженнях − як обсяги деяких ресурсів, то множники Лагранжа $\lambda_1, \lambda_2, ..., \lambda_m$ показують, як зміниться максимальний дохід або мінімальна вартість, коли обсяг ресурсу i-го виду збільшиться на одиницю.

Приклад. Борошномельний комбінат реалізує борошно двома способами: в роздріб через магазин і оптом через торгових агентів. При продажу x_1 кг борошна через магазин витрати на реалізацію складають x_1^2 грош. од., а при продажу x_2 кг борошна за допомогою торгових агентів – x_2^2 грош. од. Визначити, скільки кілограмів борошна слід продавати кожним способом, щоб витрати на реалізацію були мінімальними, якщо за добу виділяється для продажу 5 000 кг борошна.

Необхідно скласти математичну модель задачі.

Слід знайти мінімум сумарних витрат $L = x_1^2 + x_2^2$ за обмежень $x_1 + x_2 = 5\,000$, $x_{1,2} \geq 0$. Для розрахунку моделі використовувати метод множників Лагранжа. Потрібно скласти функцію Лагранжа $F(x_1, x_2, \lambda) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 5\,000)$.

Варто знайти частинні похідні функції F по x_1, x_2, λ, прирівняти їх до нуля, вийде система рівнянь:

$$\begin{cases} 2x_1 + \lambda = 0, \\ 2x_2 + \lambda = 0, \\ x_1 + x_2 - 5\,000 = 0, \end{cases}$$

Звідки $\lambda = -5\,000$, $x_1 = 2\,500, x_2 = 2\,500, L = 12\,500\,000$ грош. од. Надаючи x_1 значення більше і менше 2 500, знаходимо L і з визначення екстремуму функції отримуємо, що L при $x_1 = x_2 = 2\,500$ досягає мінімуму. Таким чином, для одержання мінімальних витрат необхідно витрачати на добу через магазин і торгових агентів по 2 500 кг борошна, при цьому витрати на реалізацію складуть 12 500 000 грош. од.

12.4. Теорема Куна-Таккера

У теорії нелінійного програмування центральне місце займає теорема Куна-Таккера, що узагальнює класичний метод множників Лагранжа на випадок, коли в нелінійній задачі, крім обмежень-рівностей, містяться також і нерівності. Зокрема, для задачі опуклого програмування: мінімізувати $Z = f(x)$ за обмежень $\varphi_i(x) \leq 0$ $\left(i = \overline{1,m}\right), \boldsymbol{x} \geq 0$, де усі функції $f(x)$ і $\varphi_i(x)$ опуклі, теорема Куна-Таккера встановлює зв'язок між оптимальним розв'язком задачі і сідловою точкою функції Лагранжа для цієї задачі:

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

Точка $\left(x^*, \overline{\lambda}^*\right)$ називається сідловою точкою функції $L(x, \overline{\lambda}) = f(x) + \sum_{i=1}^{m} \lambda_i \varphi_i(x)$, якщо *n*- вимірна точка x^* є точкою мінімуму функції $L\left(x, \overline{\lambda}^*\right)$, а *m*- вимірна $\overline{\lambda}^*$

– точкою максимуму функції $L\left(x^*,\overline{\lambda}\right)$, так що для всіх x і $\overline{\lambda}$ виконується нерівність $L\left(x^*,\overline{\lambda}\right)\le L\left(x^*,\overline{\lambda}^*\right)\le L\left(x,\overline{\lambda}^*\right)$.

Теорема Куна-Таккера. Припустимо, що існує вектор $x\ge 0$, такий що $\varphi_i(x)<0$ $\left(i=\overline{1,m}\right)$. Тоді необхідною і достатньою умовою оптимальності вектора x^*, що належить допустимій області, є існування такого вектора $\overline{\lambda}^*$, що для всіх $x\ge 0$ і $\overline{\lambda}\ge 0$ має місце нерівність:

$$L\left(x^*,\overline{\lambda}\right)\le L\left(x^*,\overline{\lambda}^*\right)\le L\left(x,\overline{\lambda}^*\right).$$

Спочатку ця теорема доведена тільки для випадку диференційованих функцій $f(x)$ і $\varphi_i(x)$. Існують узагальнення на випадок опуклих функцій.

Якщо функції $f(x)$ і $\varphi_i(x)$ є диференційованими, то нерівність $L\left(x^*,\overline{\lambda}\right)\le L\left(x^*,\overline{\lambda}^*\right)\le L\left(x,\overline{\lambda}^*\right)$, де $x^*\ge 0$, $\overline{\lambda}^*\ge 0$, еквівалентна наступним «локальним» умовам Куна-Таккера:

$$\frac{\partial L\left(x^*,\overline{\lambda}^*\right)}{\partial x}\ge 0,$$
$$\frac{x^*\partial L\left(x^*,\overline{\lambda}^*\right)}{\partial x}=0,$$
$$x^*\ge 0,$$
$$\frac{\partial L\left(x^*,\overline{\lambda}^*\right)}{\partial \overline{\lambda}}\le 0,$$
$$\frac{\overline{\lambda}^*\partial L\left(x^*,\overline{\lambda}^*\right)}{\partial \overline{\lambda}}=0$$
$$\overline{\lambda}^*\ge 0.$$

При фіксованому $\overline{\lambda}^*$ функція Лагранжа $L\left(x,\overline{\lambda}\right)=f(x)+\sum_{i=1}^{m}\lambda_i\varphi_i(x)$ є опуклою по x, тому відповідно до $f(x_2)-f(x_1)\ge f'(x_1)(x_2-x_1)$ має місце нерівність

$L\left(x,\overline{\lambda}^*\right) - L\left(x^*,\overline{\lambda}^*\right) \geq \frac{\partial L\left(x^*,\overline{\lambda}^*\right)}{\partial x}\left(x - x^*\right)$ та враховуючи $\frac{x^* \partial L\left(x^*,\overline{\lambda}^*\right)}{\partial x} = 0$, отримаємо $L\left(x,\overline{\lambda}^*\right) \geq L\left(x^*,\overline{\lambda}^*\right) + x \cdot \frac{\partial L\left(x^*,\overline{\lambda}^*\right)}{\partial x}$. Оскільки $x \geq 0$, то опускаючи невід'ємний другий доданок у правій частині, отримаємо нерівність $L\left(x,\overline{\lambda}^*\right) \geq L\left(x^*,\overline{\lambda}^*\right)$. Отже, доведено виконання однієї частини нерівності $L\left(x^*,\overline{\lambda}\right) \leq L\left(x^*,\overline{\lambda}^*\right) \leq L\left(x,\overline{\lambda}^*\right)$. Аналогічно доводиться виконання для $\overline{\lambda} \geq 0$ другої частини вказаної нерівності.

Запитання для самоперевірки

1. У чому полягає відмінність задач нелінійного програмування від лінійного?
2. Наведіть загальну постановку задачі опуклого програмування.
3. Як задача нелінійного програмування розв'язується графічно?
4. У чому полягає ідея методу множників Лагранжа?
5. У чому суть теореми Куна-Таккера?

13. Квадратичне програмування

13.1. Основні поняття квадратичного програмування.

13.2. Алгоритм розв'язання задачі квадратичного програмування.

13.1. Основні поняття квадратичного програмування

До задач квадратичного програмування відносять спеціальний клас задач нелінійного програмування, у яких цільова функція $f(x)$ – увігнута (опукла), а всі обмеження лінійні. Прикладом задачі квадратичного програмування є задача максимізації обсягу випуску продукції у випадку квадратичної виробничої функції.

Як основна задач в квадратичному програмуванні приймається задача:

$$f(x)=\sum_{j=1}^{n}c_j x_j+\sum_{i=1}^{n}\sum_{j=1}^{n}d_{ij}x_i x_j\to\max(\min), \qquad (13.1)$$

за обмежень

$$\begin{cases}\sum_{j=1}^{n}a_{ij}x_j\le b_i\,,\\ (i=\overline{1,n}),\\ x_j\ge 0,\\ (j=\overline{1,n}).\end{cases} \qquad (13.2)$$

Для розв'язання задачі необхідно привести визначення.

Квадратичною формою щодо змінних $x_1,\ldots,x_n$ називається числова функція від цих змінних, що має вигляд:

$$F=d_{11}x_1x_1+d_{12}x_1x_2+d_{13}x_1x_3+\ldots+d_{21}x_2x_1+d_{22}x_2x_2+\ldots=\sum_{i=1}^{n}\sum_{j=1}^{n}d_{ij}x_ix_j\ .$$

Матриця $D=(d_{ij})$ є від'ємно визначеною в задачі максимізації і додатньо визначеною – в задачі мінімізації. Це означає, що функція $f(x)$ є строго опуклою по змінним x у разі задачі мінімізації та строго увігнутою – в задачі максимізації. Обмеження в цій задачі передбачаються лінійними, що гарантує опуклість області допустимих розв'язків.

Квадратична форма F називається додатньо (від'ємно)-напіввизначеною, якщо $(F(X)\ge 0)$ $(F(X)\le 0)$ для будь-якого набору значень змінних $X=(x_1,\ldots,x_n)$ і існує такий набір змінних $X'=(x_1',\ldots,x_n')$, де не всі значення змінних одночасно дорівнюють нулю $F(X')=0$.

Теорема 1. Квадратична форма є опуклою функцією, якщо вона додатньо-напіввизначена, а також увігнутою функцією, якщо вона від'ємно-напіввизначена.

У разі, якщо цільова функція не мінімізується, а максимізується або якщо в деяких обмеженнях є знак $\ge$, то такі задачі завжди можна привести до основної форми (13.1), (13.2). Відмінністю задачі квадратичного програмування від задачі лінійного програмування є те, що в останній оптимальний розв'язок знаходиться в вершині багатогранника, а в задачі (13.1), (13.2) оптимальний розв'язок може бути як на межі, так і всередині багатогранника розв'язків.

Для розв'язання задачі (13.1), (13.2) потрібно скласти локальні умови Куна-Таккера, які є необхідними і достатніми умовами оптимальності розв'язання:

$$\frac{\partial L\left(x^*, \overline{\lambda}^{\,*}\right)}{\partial x} \geq 0,$$

$$\frac{x^* \partial L\left(x^*, \overline{\lambda}^{\,*}\right)}{\partial x} = 0,$$

$$x^* \geq 0,$$

$$\frac{\partial L\left(x^*, \overline{\lambda}^{\,*}\right)}{\partial \overline{\lambda}} \leq 0,$$

$$\frac{\overline{\lambda}^{\,*} \partial L\left(x^*, \overline{\lambda}^{\,*}\right)}{\partial \overline{\lambda}} = 0$$

$$\overline{\lambda}^{\,*} \geq 0.$$

Функція Лагранжа в даному випадку має вигляд:

$$L = \sum_{j=1}^{n} c_j x_j + \sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij} x_i x_j + \sum_{i=1}^{m} \lambda_i \left(\sum_{j=1}^{n} a_{ij} x_j - b_i \right).$$

Необхідно знайти частинні похідні даної функції:

$$\frac{\partial L}{\partial x_j} = c_j + 2\sum_{i=1}^{n} d_{ij} x_i + \sum_{i=1}^{m} \lambda_i a_{ij} \quad (j = \overline{1, n}),$$

$$\frac{\partial L}{\partial \lambda_i} = \sum_{j=1}^{n} a_{ij} x_j - b_i \ \left(i = \overline{1, m}\right).$$

Позначимо

$$c_j + 2\sum_{i=1}^{n} d_{ij} x_i + \sum_{i=1}^{m} \lambda_i a_{ij} = v_j,$$

$$\sum_{j=1}^{n} a_{ij} x_j - b_i = -y_i \quad .$$

З урахуванням цих позначень локальні умови Куна-Таккера наберуть вигляд:

$$c_j + 2\sum_{i=1}^{n} d_{ij} x_i + \sum_{i=1}^{m} \lambda_i a_{ij} = v_j \geq 0,$$

$$x_j v_j = 0, \quad x_j \geq 0,$$

$$\sum_{j=1}^{n} a_{ij} x_j - b_i = -y_i \le 0,$$

$$\lambda_i(-y_i) = 0, \quad \lambda_i \ge 0.$$

Об'єднуючи ці співвідношення, отримуємо:

$$\sum_{j=1}^{n} a_{ij} x_j + y_i = b_i \ , \tag{13.3}$$

$$2\sum_{i=1}^{n} d_{ij} x_i - v_j + \sum_{i=1}^{m} \lambda_i a_{ij} = -c_j , \tag{13.4}$$

$$x_j \ge 0 ,\ v_j \ge 0 ,\ y_i \ge 0 ,\ \lambda_i \ge 0 ,$$

$$\sum_{j=1}^{n} x_j v_j + \sum_{i=1}^{m} y_i \lambda_i = 0 \tag{13.5}$$

Отримуємо систему $n+m$ лінійних рівнянь з $2(n+m)$ невідомими $x_1, \ldots, x_n, v_1, \ldots, v_n, y_1, \ldots, y_m, \lambda_1, \ldots, \lambda_m$.

Оскільки функція $f(x)$ строго увігнута, а область допустимих розв'язків становить опуклу множину, допустимий розв'язок, який задовольняє всі ці умови, має бути єдиним і оптимальним.

Отже, в квадратичному програмуванні за допомогою теореми Куна-Таккера нелінійну задачу можна зводити до розв'язання системи лінійних алгебраїчних рівнянь при дотриманні додаткової умови $\sum_{j=1}^{n} x_j v_j + \sum_{i=1}^{m} y_i \lambda_i = 0$, згідно з якою: з кожних двох обмежених за знаком змінних x_j і v_j (відповідно y_i і λ_i) хоча б одна дорівнювала нулю. Іншими словами, не всі розв'язки $X = (x_1, \ldots, x_n, v_1, \ldots, v_n, y_1, \ldots, y_m, \lambda_1, \ldots, \lambda_m)$ системи (13.3), (13.4) задовольняють умову (13.5), а тільки ті, в яких принаймні $n+m$ компонент дорівнюють нулю, тобто стільки, скільки рівнянь в системі і шукати розв'язок слід з базисних розв'язків. Оскільки для знаходження допустимого базисного розв'язку можна застосувати симплекс-метод, то його можна застосовувати також для розв’язування задачі квадратичного програмування.

Використавши метод штучного базису, сформулюємо задачу:

$$Z = M\sum_{j=1}^{n} \psi_j + M\sum_{i=1}^{m} \varphi_i \to \min ,$$

$$2Dx + \lambda A - v + \psi = -c ,$$

$$Ax + y + \varphi = b,$$
$$x \geq 0,\ \lambda \geq 0,\ v \geq 0,\ y \geq 0,\ \psi \geq 0,\ \varphi \geq 0.$$

Якщо в результаті її розв'язування вдалось вивести з базису всі штучні змінні й виконати умови доповняльної нежорсткості, то знайдений базисний розв'язок буде розв'язком задачі квадратичного програмування.

Розв'язання завершується прирівнянням до нуля всіх штучних змінних тільки в тому випадку, коли вихідна задача має непорожню множину допустимих розв'язків.

Крім загальних методів опуклого програмування, спеціально для задачі лінійного квадратичного програмування розроблено велику кількість різних чисельних методів, у тому числі і кінцевих. Найбільш поширеними кінцевими методами є симплексний метод Біла, метод сполучених градієнтів, симплексний метод Вульфа. Проте ефективність того чи іншого методу досить істотно залежить від специфіки розв'язуваної задачі і наявних у розпорядженні обчислювальних засобів.

Розглянемо метод Франка-Вульфа. Переформулюємо задачу: серед довільних базисних розв'язків системи умов (13.3) і (13.4) знайти такий, що перетворює в нуль $v'x + y'\lambda$. Розглянемо вектори:

$$Y = (x_1, ..., x_n, \lambda_1, ..., \lambda_m, v_1, ..., v_n, y_1, ..., y_m),$$
$$\overline{Y} = (v_1, ..., v_n, y_1, ..., y_m, x_1, ..., x_n, \lambda_1, ..., \lambda_m).$$

Тоді:

$$v'x + y'\lambda = \frac{1}{2} Y'\overline{Y} \to \min,$$
$$BY = d,$$
$$Y \geq 0,$$

де $B = \begin{vmatrix} D & A' - E & 0 \\ A & 0 \quad 0 & E \end{vmatrix}$, $d = \begin{vmatrix} -c \\ b \end{vmatrix}$.

Виконавши ряд ітерацій так, що опукла функція $T(Y) = Y'\overline{Y}$ мінімізувалась за таким рекурентним правилом: в результаті k-го кроку маємо довільний розв'язок Y_k, що відповідає системі обмежень задачі, а також довільний розв'язок y_k, який задовольняє тим самим обмеженням, але не обов'язково є базисним. Йому відповідає додатне значення $T(Y)$ (на першому кроці $y_0 = Y_0$).

Починаючи з базисного розв’язку Y_k, мінімізуємо лінійну відносно Y форму $Y'\bar{y}_k$. Маємо послідовність базисних розв’язків:

$$Y^0 = Y_k, Y_1, Y_2, \ldots,$$

причому $(Y_0)'\bar{y}_k > (Y_1)'\bar{y}_k$, … Обчислювання закінчується , якщо $(Y_h)'\overline{Y}_h = 0$; $2(Y_h)'\bar{y}_k \le (y_k)'\bar{y}_k$. Якщо $(Y_h)'\overline{Y}_h = 0$, то Y_h є розв’язком задачі квадратичного програмування, у випадку $2(Y_h)'\bar{y}_k \le (y_k)'\bar{y}_k$ знаходимо $Y_{k+1} = Y_k$, $y_{k+1} = y_k + \mu(Y_{k+1} - y_k)$, де $\mu = \min\left\{1, \dfrac{(y_k - Y_{k+1})'\bar{y}_k}{(Y_{k+1} - y_k)'(Y_{k+1} - \bar{y}_k)}\right\}$.

На кожному кроці маємо або $(Y_h)'\overline{Y}_h = 0$, або $2(Y_h)'\bar{y}_k \le (y_k)'\bar{y}_k$, але через певну кількість кроків тільки $(Y_h)'\overline{Y}_h = 0$.

Існують й інші методи розв’язування задачі (13.1) – (13.2), але вони також мають недоліки.

13.2. Алгоритм розв'язання задачі квадратичного програмування

Алгоритм розв'язання задачі квадратичного програмування складається з таких етапів:

1) складають функцію Лагранжа;

2) записують у вигляді виразів (13.3) – (13.5) необхідні і достатні умови існування сідлової точки для функції Лагранжа;

3) використовуючи метод штучного базису, або встановлюють відсутність сідлової точки для функції Лагранжа, або знаходять її координати;

4) записують оптимальний розв'язок вихідної задачі і знаходять значення цільової функції.

Приклад. Знайти максимальне значення функції $f = 2x_1 + 4x_2 - x_1^2 - 2x_2^2$ за умов:

$$\begin{cases} x_1 + 2x_2 \le 8, \\ 2x_1 - x_2 \le 12, \\ x_1, x_2 \ge 0. \end{cases}$$

Функція f є увігнутою, оскільки становить суму лінійної функції $f_1 = 2x_1 + 4x_2$, яку можна розглядати як увігнуту, і квадратичної форми

$f_2 = -x_1^2 - 2x_2^2$, яка є від'ємно-визначеною і, отже, також увігнутою. Система обмежень задачі включає тільки лінійні нерівності. Отже, можна скористатися теоремою Куна-Таккера.

Складемо функцію Лагранжа:

$$L = 2x_1 + 4x_2 - x_1^2 - 2x_2^2 + \lambda_1(8 - x_1 - 2x_2) + \lambda_2(12 - 2x_1 + x_2).$$

Запишемо необхідні і достатні умови існування сідлової точки функції:

$$\frac{\partial L}{\partial x_1} = 2 - 2x_1 - \lambda_1 - 2\lambda_2 \le 0,$$

$$\frac{\partial L}{\partial x_2} = 4 - 4x_2 - 2\lambda_1 + \lambda_2 \le 0,$$

$$\frac{\partial L}{\partial \lambda_1} = 8 - x_1 - 2x_2 \ge 0,$$

$$\frac{\partial L}{\partial \lambda_2} = 12 - 2x_1 + x_2 \ge 0,$$

$$x_1 \frac{\partial L}{\partial x_1} = x_1(2 - 2x_1 - \lambda_1 - 2\lambda_2) = 0,$$

$$x_2 \frac{\partial L}{\partial x_2} = x_2(4 - 4x_2 - 2\lambda_1 + \lambda_2) = 0,$$

$$\lambda_1 \frac{\partial L}{\partial \lambda_1} = \lambda_1(8 - x_1 - 2x_2) = 0,$$

$$\lambda_2 \frac{\partial L}{\partial \lambda_2} = \lambda_2(12 - 2x_1 + x_2) = 0,$$

$$x_1, x_2, \lambda_1, \lambda_2 \ge 0.$$

Таким чином, маємо систему нерівностей:

$$2x_1 + \lambda_1 + 2\lambda_2 \ge 2,$$

$$4x_2 + 2\lambda_1 - \lambda_2 \ge 4,$$

$$x_1 + 2x_2 \le 8,$$

$$2x_1 - x_2 \le 12.$$

Необхідно перетворити систему нерівностей в рівності, ввівши балансові змінні:

$$2x_1 + \lambda_1 + 2\lambda_2 - v_1 = 2,$$

$$4x_2 + 2\lambda_1 - \lambda_2 - v_2 = 4, \qquad (13.6)$$

$$x_1 + 2x_2 + w_1 = 8,$$
$$2x_1 - x_2 + w_2 = 12.$$

Враховуючи рівності (13.5), варто записати:

$$v_1 x_1 = 0,\ v_2 x_2 = 0,\ w_1 \lambda_1 = 0,\ w_2 \lambda_2 = 0.$$

Для знаходження базисного розв'язку системи лінійних рівнянь (13.6) слід застосувати метод штучного базису. У перше і в друге рівняння системи (13.6) додамо штучні невід'ємні змінні z_1, z_2. Отримаємо розширену задачу:

$$F = -M z_1 - M z_2$$

за умов:

$$\begin{cases} 2x_1 + \lambda_1 + 2\lambda_2 - v_1 + z_1 = 2, \\ 4x_2 + 2\lambda_1 - \lambda_2 - v_2 + z_2 = 4, \\ x_1 + 2x_2 + w_1 = 8, \\ 2x_1 - x_2 + w_2 = 12, \\ x_1, x_2, \lambda_1, \lambda_2, v_1, v_2, w_1, w_2, z_1, z_2 \ge 0. \end{cases}$$

Розв'яжемо задачу симплекс-методом.

				0	0	0	0	0	0	0	0	$-M$	$-M$
N	Базис	$C_б$	A_0	A_{x_1}	A_{x_2}	A_{λ_1}	A_{λ_2}	A_{v_1}	A_{v_2}	A_{w_1}	A_{w_2}	A_{z_1}	A_{z_2}
1	A_{z_1}	$-M$	2	2	0	1	2	-1	0	0	0	1	0
2	A_{z_2}	$-M$	4	0	**4**	2	-1	0	-1	0	0	0	1
3	A_{w_1}	0	8	1	2	0	0	0	0	1	0	0	0
4	A_{w_2}	0	12	2	-1	0	0	0	0	0	1	0	0
		Δ_1	$-6M$	$-2M$	$-4M$	$-3M$	$-M$	M	M	0	0	0	0
5	A_{z_1}	$-M$	2	**2**	0	1	2	-1	0	0	0	1	0
6	A_{x_2}	0	1	0	**1**	$\frac{1}{2}$	$-\frac{1}{4}$	0	$-\frac{1}{4}$	0	0	0	$\frac{1}{4}$
7	A_{w_1}	0	6	1	0	-1	$\frac{1}{2}$	0	$\frac{1}{2}$	1	0	0	$-\frac{1}{2}$
8	A_{w_2}	0	13	2	0	$\frac{1}{2}$	$-\frac{1}{4}$	0	$-\frac{1}{4}$	0	1	0	$\frac{1}{4}$
		Δ_2	$-2M$	$-2M$	0	$-M$	$-2M$	M	0	0	0	0	M

9	A_{x_1}	0	1	**1**	0	$\frac{1}{2}$	1	$-\frac{1}{2}$	0	0	0	$\frac{1}{2}$	0
10	A_{x_2}	0	1	0	**1**	$\frac{1}{2}$	$-\frac{1}{4}$	0	$-\frac{1}{4}$	0	0	0	$\frac{1}{4}$
11	A_{w_1}	0	5	0	0	$-\frac{3}{2}$	$-\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	0	$-\frac{1}{2}$	$-\frac{1}{2}$
12	A_{w_2}	0	11	0	0	$-\frac{1}{2}$	$-\frac{9}{4}$	1	$-\frac{1}{4}$	0	1	-1	$\frac{1}{4}$
		Δ_3	0	0	0	0	0	0	0	0	0	M	M

Допустимий базисний розв'язок:

$$x_1 = 1, x_2 = 1, \lambda_1 = 0, \lambda_2 = 0, v_1 = 0, v_2 = 0, w_1 = 5, w_2 = 11, z_1 = 0, z_2 = 0.$$

Оскільки виконуються умови $v_1 x_1 = 0, v_2 x_2 = 0, w_1 \lambda_1 = 0, w_2 \lambda_2 = 0$, то $X^* = (1, 1)$ є сідловою точкою функції Лагранжа для вихідної задачі. Отже, $X^* = (1, 1)$ – оптимальний план вихідної задачі, $f(X^*) = 3$.

Запитання для самоперевірки

1. Наведіть математичну модель основної задачі в квадратичному програмуванні.

2. У чому відмінність задачі квадратичного програмування від задачі лінійного програмування?

3. У яких цілях в квадратичному програмуванні використовується теорема Куна-Таккера?

4. Назвіть етапи алгоритму розв'язання задачі квадратичного програмування.

14. Теорія ігор. Основні методи їх розв'язання та аналізу

14.1. Основні поняття теорії ігор.

14.2. Розв'язання матричної гри в чистих стратегіях.

14.3. Розв'язання матричної гри в змішаних стратегіях.

14.4. Розв'язання матричної гри графічним методом.

14.5. Розв’язання матричної гри зведенням до задачі лінійного програмування

14.1. Основні поняття теорії ігор

У практичній діяльності часто необхідно узгодити дії фірм, підприємств, об'єднань, міністерств та інших учасників проектів у випадках, коли їх інтереси не збігаються. У цих ситуаціях теорія ігор дозволяє знайти кращий розв'язок у поведінці всіх учасників, зобов'язаних погоджувати дії у разі зіткнення інтересів. Теорія ігор широко проникла в практику економічних рішень і досліджень. Так, можна визначити науково обгрунтовані рівні зниження роздрібних цін і оптимальний рівень товарних запасів, розв'язати задачу планування порядку організації джерел корисних ресурсів у країні і т. д.

Теорію ігор визначають як розділ математики, що вивчає конфліктні ситуації. Це означає, що можна сформулювати оптимальні правила поведінки кожної сторони, яка бере участь у вирішенні конфліктної ситуації. В економіці виявилося недостатньо апарату математичного аналізу, пов'язаного з визначенням екстремумів функції. Теорію ігор слід розглядати як окремий розділ оптимізаційного підходу, що дозволяє розв'язувати нові задачі у прийнятті рішення.

Гра – це спрощена формалізована модель реальної конфліктної ситуації, відмінна від реального конфлікту тим, що ведеться за певними правилами. Іншими словами, гра – це сукупність правил, що визначають можливі дії (чисті стратегії) учасників гри.

Суть гри полягає в тому, що кожен з учасників приймає рішення в конфліктній ситуації, що розвивається, які як він розуміє, можуть забезпечити йому найкращий результат. Результат гри – це значення деякої функції, так званої функції виграшу (платіжної функції). Ця функція задається або таблицею, або аналітичним виразом. Якщо сума виграшів гравців дорівнює нулю, то гру називають грою з нульовою сумою. Якщо в грі беруть участь два гравці, то її називають парною.

Величина виграшу залежить від стратегії, яку використовує гравець. Стратегія – це сукупність правил, що однозначно визначає послідовність дій гравця в кожній конкретній ситуації, що реалізується в процесі гри. Оптимальною називається стратегія, яка під час багаторазового повторення гри забезпечує даному гравцю максимально можливий середній виграш. Будь-яка гра складається з партій. Партією називають кожен варіант гри певним чином. У свою чергу, в партії гравці здійснюють конкретні ходи. Хід – це вибір і реалізація гравцем одного з допустимих варіантів поведінки.

Залежно від кількості стратегій ігри поділяються на скінченні і нескінченні. Залежно від взаємовідносин учасників розрізнюють ігри безкоаліційні (учасники не мають права укладати угоди) або некооперативні, і коаліційні або кооперативні. За характером виграшів ігри поділяють на ігри з нульовою сумою і ігри з ненульовою сумою. У перших − загальний капітал гравців не міняється, а лише перерозподіляється в ході гри, у зв'язку з чим сума виграшів дорівнює нулю (програш приймається як від'ємний виграш). У грі з нульовою сумою сума виграшів відмінна від нуля.

За видом функцій виграшів ігри поділяються на матричні, біматричні, неперервні, опуклі, сепарабельні. У матричній грі (за участю двох учасників) виграші першого учасника задаються матрицею, в біматричних - виграші кожного гравця задаються своєю матрицею. Інші типи такої гри розрізнюються виглядом аналітичного виразу платіжної функції. За кількістю ходів ігри діляться на одноходові (виграші розподіляються після одного ходу кожного гравця) і багатоходові (виграш розподіляється після декількох ходів). Багатоходові ігри у свою чергу діляться на позиційні, стохастичні, диференціальні та ін. Залежно від обсягу наявної інформації розрізнюють гру з повною і неповною інформацією. Надалі автор буде розглядати тільки парну матричну гру з нульовою сумою.

Гра, в якій учасники прагнуть досягти для себе найкращого результату, свідомо вибираючи допустимими правилами гри способи дій, називають стратегічною. Однак в економіці часто доводиться моделювати (формалізувати) ситуації, в якій один з учасників байдужий до результату гри. Така гра називається грою з природою, розуміючи під природою всю сукупність зовнішніх обставин, в яких свідомому гравцеві доводиться приймати рішення. Наприклад, визначення обсягу виробництва випуску сезонної продукції в надії найбільш вигідного дня її реалізації для рівня попиту; формування пакета цінних паперів у розрахунку на високі дивіденди і т. д. Тут в ролі гравця виступає в першому випадку рівень попиту, в другому – розмір очікуваного прибутку. У грі з природою міра невизначеності свідомого гравця зростає.

Розглянемо гру, в якій у кожного з двох гравців A і B є кінцеве число можливих дій, тобто чистих стратегій. Припустимо, що гравець A має у своєму розпорядженні m чистих стратегій, які позначимо $A_1, \ldots, A_m$, а гравець B має n чистих стратегій $B_1, \ldots, B_n$. Для того, щоб гра була повністю визначена, необхідно вказати правило, за яким кожній парі чистих стратегій $\left(A_i; B_j\right)$ ставиться у відповідність число a_{ij} – виграш гравця A за рахунок гравця B або програш гравця B. Розглядається гра з нульовою сумою. При $a_{ij} < 0$ гравець A виплачує гравцю B суму $|a_{ij}|$. Якщо гра складається тільки з точних ходів, то вибір пари чистих стратегій $(A_i; B_j)$ єдиним способом визначає результат гри. Якщо ж протягом гри використовуються випадкові ходи, то результат гри визначається середнім значенням виграшу (його математичним очікуванням).

Нехай відомі значення a_{ij} для кожної пари $(A_i; B_j)$ чистих стратегій, отже, можна скласти матрицю гри, тобто платіжну матрицю, яка є табличним записом функції виграшу (табл. 14.1).

Таблиця 14.1

A_i	B_j				α_i
	B_1	B_2	$\dots$	B_n	
A_1 $\dots$ A_m	a_{11} $\dots$ a_{m1}	a_{12} $\dots$ a_{m2}	$\dots$ $\dots$ $\dots$	a_{1n} $\dots$ a_{mn}	α_1 $\dots$ α_m
β_j	β_1	β_2	$\dots$	β_n	

У теорії матричних ігор завжди передбачається, що в платіжній матриці записані виграші гравця A. Вони можуть виражатися і від'ємними числами, що фактично означає виграш гравця B. Описані ігри називають матричними, або прямокутними. Окрема партія в цих іграх реалізовується таким чином. Гравець A вибирає один із рядків платіжної матриці (одну зі своїх чистих стратегій). Не знаючи результату його вибору, гравець B вибирає один із стовпців (свою чисту стратегію). Елемент матриці, що стоїть на перетині вибраних рядка і стовпця, визначає виграш гравця A (програш гравця B).

14.2. Розв'язання матричної гри в чистих стратегіях

Метою учасників всякої матричної гри є вибір найбільш вигідних стратегій, що забезпечують гравцеві A максимальний виграш, а гравцеві B – мінімальний програш. Стратегію гравця A називають оптимальною, якщо у процесі її застосування виграш гравця A не меншає, якими б стратегіями не користувався б гравець B. Оптимальною для гравця B називають стратегію, у процесі використання якої програш цього гравця не збільшується, які б стратегії не використав гравець A.

Під час пошуку оптимальних стратегій гравці спираються на основоположний принцип теорії ігор – принцип обережності, згідно з яким кожен гравець, вважаючи противника по грі дуже розумним, обирає свої дії в припущенні, що противник ні в якому разі не пропустить жодної можливості використовувати будь-яку його помилку, будь-який прорахунок у своїх інтересах. Тому гравці повинні бути дуже уважні під час вибору кожної своєї чистої стратегії.

Нехай гравцеві A слід зробити свій вибір. Аналізуючи платіжну матрицю (табл. 14.1), він для кожної чистої стратегії A_i $\left(i = \overline{1, m}\right)$ знайде мінімальне значення α_i можливого виграшу: $\alpha_i = \min\limits_j \{a_{ij}\}$ $(i = \overline{1, m})$, а потім з усіх α_i обере найбільше $\alpha = \max\limits_i \{a_i\}$, і прийме відповідну йому чисту стратегію A_i^0. Це і буде переважаюча (яка гарантує) у даних умовах стратегія гравця A. Її називають максимінною, оскільки вона відповідає величині:

$$\alpha = \max_i \min_j \{a_{ij}\}\ . \tag{14.1}$$

Число α називається нижньою чистою ціною гри (максимінною). Воно показує, який мінімальний виграш може отримати гравець A при будь-яких діях гравця B, якщо правильно використовує свої чисті стратегії. У свою чергу гравець B, прагнучи мінімізувати програш, при виборі переважаючої стратегії використовує принцип обережності таким чином: спочатку для кожної чистої стратегії B_j $\left(j = \overline{1, n}\right)$ він визначає максимально можливий програш $\beta_j = \max\limits_i \{a_{ij}\}$, де $j = \overline{1, n}$ (табл. 14.1), а потім серед β_j обере мінімальне значення $\beta = \min\limits_j \{\beta_j\}$, якому і буде відповідати бажана чиста стратегія B_j^0. Її називають мінімаксною, оскільки вона відповідає величині:

$$\beta = \min_j \max_i \{a_{ij}\}\ . \tag{14.2}$$

Число β називається верхньою чистою ціною гри (мінімаксною). Воно показує, який максимальний програш може бути у B гравця під час правильного вибору ним своїх чистих стратегій не залежно від дій гравця A.

Отже, правильно використовуючи чисті стратегії, гравець A забезпечує собі виграш не менше α, а гравець B, правильно використовуючи свої чисті стратеги, не дозволить гравцеві A виграти більше, ніж β.

Теорема. У матричній грі нижня чиста ціна гри не перевищує верхньої чистої ціни гри, тобто $\alpha \le \beta$.

Теорема без доведення.

Якщо в матричній грі нижня і верхня чисті ціни збігаються, тобто $\alpha = \beta$, то гра має сідлову точку в чистих стратегіях і чисту ціну гри $v = \alpha = \beta$.

Позначимо через i_* та j_* номери чистих стратегій, при яких має місце рівність $\alpha = \beta$. Пара чистих стратегій $\left(A_{i*}; B_{j*}\right)$ гравців A і B, при яких досягається ця рівність, називається сідловою точкою матричної гри, а елемент $a_{i_* j_*}$ матриці, який стоїть на перетині i_*-го рядка і j_*-го стовпця, є сідловим елементом платіжної матриці:

$$\max_i \min_j \{a_{ij}\} = \min_j \max_i \{a_{ij}\} = a_{i_* j_*}.$$

Сідловій точці відповідають оптимальні стратегії гравців. Їх сукупність − це розв’язок гри, який має наступну властивість: якщо один із гравців дотримується своєї оптимальної стратегії, то для другого відхилення від його оптимальної стратегії не може бути вигідним.

Приклад. Є три можливих стратегії виробництва продукції:
$S_1 = 986\,200{,}50$ тис. грн., $S_2 = 1\,488\,973$ тис. грн., $S_3 = 1\,967\,962{,}1$ тис. грн.

Залежно від змін ринкової кон'юнктури у зв'язку з наявними можливостями реалізації розраховані варіанти середньорічного прибутку (табл. 14.2).

Таблиця 14.2

Матриця платоспроможного попиту

Обсяг виробництва (оптові закупівлі)	**Розмір прибутку (a_{ij}) залежно від можливих коливань попиту**				$\alpha_i = \min \{a_{ij}\}$
	489 876,17	986 200,50	1 488 973	1 967 962,1	
S_1=986 200,50	49 310,03	197 240,1	197 240,1	197 240,1	49 310,03
S_2=1 488 973	-60	148 897,3	297 794,6	297 794,6	-60
S_3=1 967 962,1	-1 140	98 398,11	196 796,24	393 592,42	-1 140

$\beta_j = \max \{a_{ij}\}$	49 310,03	197 240,1	297 794,6	393 592,42	

Аналіз гри почнемо з позиції максиміну:

$$\alpha_1 = \min_j \{a_{ij}\} = \min\{49\,310{,}01;\ 197\,240{,}10; 197\,240{,}10\} =$$

$$= 49\ 310{,}33 \quad \text{тис. грн.}$$

$$\max_i \{\alpha_i\} = \max_i \min_j \{a_{ij}\} = \{49\,310{,}01;\ -60; -1\,140\} = 49\ 310{,}33 \quad \text{тис. грн.}$$

Стратегія S_1 називається максиминною – це і є нижня ціна гри.

Аналогічно необхідно міркувати з боку «природи»:

$$\beta_1 = \max_i \{49310{,}03; -60; -1140\} = 49310{,}03 \text{ тис. грн}.$$

$$\beta = \min_j \{\beta_j\} = \min_j \{49310{,}03;\ 197240{,}1;\ 297794{,}6;\ 393592{,}42\} =$$

$$= 49310{,}03 \text{ тис. грн.}$$

Ця величина називається верхньою ціною гри. Сідлова точка дорівнює $v = \alpha_1 = \beta_1 = 49310{,}03$ (табл. 14.2). Таким чином, необхідно вибрати першу стратегію з обсягами виробництва 986 200,50 тис. грн.

14.3. Розв'язання матричної гри в змішаних стратегіях

Якщо гра не має сідлової точки, то її розв’язання ускладнюється. Гравцям потрібно так вибирати свої чисті стратегії в черговій партії, щоб партнер не здогадався про них. Це можливо тільки тоді, коли сам не знаєш, яку стратегію використовуватимеш при черговому ході. Тут на допомогу приходить механізм випадкового вибору чистих стратегій. Аналіз гри без сідлової точки показує, що гравець A може виграти більше за максимін α, який він отримав би при максимінній стратегії, якщо в ході гри використовуватиме не одну, а кілька чистих стратегій, випадковим чином змішуючи їх, тобто застосовуючи змішану стратегію. Гравець B теж може програти менше за мінімакс β, який він сплачує гравцеві A при мінімаксній стратегії, якщо використає свою змішану стратегію.

Звернемося до загального випадку матричної гри (табл. 14.3). Позначимо через $p_1, \ldots, p_m$ ймовірності, з якими гравець A використовує у ході гри свої чисті стратегії $A_1, \ldots, A_m$. Оскільки в табл. 14.3 наведено повний набір чистих стратегій гравця A, то для ймовірностей p_i виконуються умови:

$$p_i \geq 0, \left(i = \overline{1, m}\right), \quad \sum_{i=1}^{m} p_i = 1. \tag{14.3}$$

Упорядкована множина $\vec{p} = \left(p_1; ...; p_m\right)$, елементи якої задовольняють умовам (14.3), повністю визначає характер гри гравця A і називається його змішаною стратегією. Таким чином, змішаною стратегією гравця A є повний набір ймовірностей застосування його чистих стратегій.

Таблиця 14.3

Матриця статистичної парної гри

A_i	B_j				p_i
	B_1	B_2	…	B_n	
A_1 … A_m	a_{11} … a_{m1}	a_{12} … a_{m2}	… … …	a_{1n} … a_{mn}	p_1 … p_m
q_j	q_1	q_2	…	q_n	

Випадковість вибору чистих стратегій, якою користується гравець A, забезпечує йому нескінченну множину змішаних стратегій. Будь-яка його чиста стратегія A_i може розглядатися як окремий випадок змішаної стратегії, i-ий компонент якої є 1, а всі інші – 0, тобто $\overline{p} = (0; \ ...; \ 1; \ ...; \ 0)$.

Аналогічно впорядкована множина $\overline{q} = (q_1; \ ...; q_n)$, елементи якої задовольняють співвідношенням:

$$q_j \geq 0, \left(j = \overline{1, n}\right), \quad \sum_{j=1}^{n} q_j = 1,$$

є змішаною стратегією гравця B. Як і гравець A, гравець B має у своєму розпорядженні нескінченну множину змішаних стратегій.

Нехай гравці A і B застосовують змішані стратегії $\overline{p}$ і $\overline{q}$. Це означає, що гравець A використовує стратегію A_i з ймовірністю p_i, а гравець B – стратегію B_j з ймовірністю q_j. Оскільки вони обирають свої чисті стратегії випадково і незалежно один від одного, то ймовірність вибору пари $\left(A_i; B_j\right)$ дорівнює добутку ймовірностей p_i і q_j, тобто $p_i \cdot q_j$. У процесі використання змішаних стратегій гра набуває випадкового характеру, випадковою стає і величина виграшу гравця A (програшу гравця B). У зв'язку з цим можна вести мову лише про середню величину (математичне сподівання) виграшу (програшу). Зрозуміло, що ця величина є функцією від змішаних стратегій $\overline{p}$ і $\overline{q}$. Вона визначається за формулою:

$$f\left(\overline{p},\overline{q}\right)=\sum_{i=1}^{m}\sum_{j=1}^{n}a_{ij}p_iq_j\ . \tag{14.4}$$

Функція $f\left(\overline{p},\overline{q}\right)$ називається ***платіжною функцією*** гри за матрицею, що задана табл. 14.3.

Аналогічно з раніше введеними поняттями нижньої чистої ціни і верхньої чистої ціни вводяться поняття нижньої і верхньої ціни стосовно змішаних стратегій, зберігаючи для них ті ж позначення α і β. Замість виграшу a_{ij} тепер мають на увазі середній виграш $f\left(\overline{p},\overline{q}\right)$, а замість чистих стратегій гравців з номерами i та j – їх змішані стратегії $\overline{p}$ та $\overline{q}$:

$$\alpha=\max_{\overline{p}}\min_{\overline{q}} f\left(\overline{p},\overline{q}\right);\quad \beta=\min_{\overline{q}}\max_{\overline{p}} f\left(\overline{p},\overline{q}\right).$$

Як і в грі, що має сідлову точку в чистих стратегіях, ***оптимальними змішаними стратегіями*** називають такі стратегії $\overline{p}^*$ і $\overline{q}^*$ гравців A і B, за яких має місце рівність:

$$\max_{\overline{p}}\min_{\overline{q}} f\left(\overline{p}^*,\overline{q}^*\right)=\min_{\overline{q}}\max_{\overline{p}} f\left(\overline{p}^*,\overline{q}^*\right)=f\left(\overline{p}^*,\overline{q}^*\right). \tag{14.5}$$

Величина $f(\overline{p}^*,\overline{q}^*)$, що відповідає оптимальним стратегіям $\overline{p}^*$ і $\overline{q}^*$ і отримана за формулою (14.5), називається ціною гри:

$$\nu=f(\overline{p}^*,\overline{q}^*)\ .$$

Існує еквівалентне даному означення оптимальних змішаних стратегій. Стратегії $\overline{p}^*$ і $\overline{q}^*$ називають оптимальними змішаними стратегіями відповідно гравців A і B, якщо ці стратегії утворюють сідлову точку для платіжної функції $f\left(\overline{p},\overline{q}\right)$, тобто задовольняють нерівність:

$$f\left(\overline{p}, \overline{q}^*\right) \le f\left(\overline{p}^*, \overline{q}^*\right) \le f\left(\overline{p}^*, \overline{q}\right). \qquad (14.6)$$

Якщо використати змішані стратегії, то для будь-якої матричної гри можна знайти оптимальні стратегії і ціну гри. В цьому полягає значення основної теореми в теорії ігор ─ теореми Неймана.

Теорема. В змішаних стратегіях будь-яка скінчена матрична гра має сідлову точку. Теорема без доведення.

Критеріями розв'язання матричної гри є наступна теорема.

Теорема. Для того, щоб змішані стратегії $\overline{p}^* = \left(p_1^*; \ldots; p_m^*\right)$ та $\overline{q}^* = \left(q_1^*; \ldots; q_n^*\right)$ були оптимальними для гравців A і B у грі за матрицею $\left\lfloor a_{ij} \right\rfloor_{m\times n}$ з ціною v, необхідно і достатньо виконання нерівностей:

$$\sum_{i=1}^{m} a_{ij} p_i^* \ge v, \quad i = \overline{1,m}, \qquad (14.7)$$

$$\sum_{j=1}^{n} a_{ij} q_i^* \le v, \quad j = \overline{1,n}. \qquad (14.8)$$

Доведення. Нехай $\overline{p}^*$ і $\overline{q}^*$ − оптимальні змішані стратегії. Доведемо, що тоді виконуються нерівності (14.7) і (14.8). Нерівність (14.7) виходить із справедливого для оптимальних стратегій співвідношення (14.6), яке записане в розгорненій формі:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} p_i q_j^* \le v \le \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} p_i^* q_j, \qquad (14.9)$$

де в праву частину (14.9) замість довільного вектора $\overline{q} = \left(q_1; \ldots; q_n\right)$ підстановлено одиничний вектор $(0; \ldots; 1; \ldots; 0)$, у якому i-ий компонент дорівнює 1, а всі інші − 0. Тим самим доведена умова необхідності. Доведемо тепер достатність. Нехай нерівності (14.7) і (14.8) виконуються. Покажемо, що тоді $\overline{p}^*$ і $\overline{q}^*$ є оптимальними змішаними стратегіями, тобто має місце співвідношення (14.9). Покажемо спочатку, що з (14.7) витікає права частина нерівності (14.9). Нехай $\overline{q} = \left(q_1; \ldots; q_n\right)$─ довільний вектор. Перетворимо суму в правій частині з урахуванням (14.7):

$$\sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij} p_i^* q_j = \sum_{j=1}^{n} q_j \sum_{i=1}^{m} a_{ij} p_i^* \ge \sum_{j=1}^{n} q_j v = v \cdot \sum_{j=1}^{n} q_j = v \cdot 1 = v.$$

Отже,

$$\sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij} p_i^* q_j \ge v,$$

а це і є права частина (14.9). Аналогічно можна довести, що з (14.8) витікає ліва частина нерівності (14.9). Тобто доведено, що виконання нерівностей (14.7) і (14.8) обумовлює виконання (14.9), а це і свідчить про оптимальність змішаних стратегій $\overline{p}^*$ і $\overline{q}^*$. ■

З теореми маємо такий наслідок: щоб розв'язати матричну гру, необхідно знайти невід'ємні розв'язки $(p_1;...;p_m;q_1;...;q_n;v)$ системи $n+m$ лінійних нерівностей (14.7) і (14.8) і лінійних рівнянь (14.9).

Чисті стратегії гравця, що входять до його оптимальної змішаної стратегії з ймовірностями, відмінними від нуля, називаються ***активними стратегіями*** гравця. Число активних стратегій не перевищує $\min\{m;n\}$.

Теорема. Якщо один з гравців дотримується своєї оптимальної змішаної стратегії, то його виграш залишається незмінним і рівним ціні гри незалежно від того, яку стратегію застосовує другий гравець, якщо тільки той не вийде за межі своїх активних стратегій. Теорема без доведення.

Розв'язання гри можна істотно спростити, якщо своєчасно виявити домінування одних стратегій, які присутні в платіжній матриці, над іншими, оскільки це дозволить скоротити вимірність матриці. Якщо в платіжній матриці всі елементи k- ого рядка не менші за відповідні елементи s- ого рядка, тобто $a_{kj} \ge a_{sj}$ $\left(j=\overline{1,n}\right)$, то виграш гравця A за стратегією A_k не буде меншим, ніж за стратегією A_s, якою б стратегією не користувався гравець B. Отже, стратегія A_k домінує над стратегією A_s, відповідно стратегію A_k називають ***домінуючою***, а стратегію A_s – ***домінуємою.***

Аналогічно, якщо всі елементи l- ого стовпця не перевершують відповідних елементів r- ого стовпця, тобто $a_{il} \ge a_{ir}$ $\left(i=\overline{1,m}\right)$, то гравцеві B за будь-яких умов невигідно застосовувати стратегію B_r, оскільки в цьому випадку він програватиме більше (не менше), ніж під час використання стратегії B_l. Тому кажуть, що стратегія B_l домінує над стратегією B_r, і називають їх відповідно домінуючою і домінуємою. Окремим випадком домінування є дублювання стратегій, коли $a_{kj} = a_{sj}$ $\left(j = \overline{1, n}\right)$, або $a_{il} = a_{ir}$ $\left(i = \overline{1, m}\right)$.

Спрощення платіжних матриць шляхом виключення явно невигідних гравцям чистих стратегій можливо завдяки наступній теоремі.

Теорема. Нехай I – гра, в матриці якої k - а стратегія гравця A домінує над s - ою, а I' – гра, матриця якої отримана з матриці гри I шляхом виключення s - го рядка. Тоді: а) ціна гри I' дорівнює ціні гри I; б) для гравця B оптимальна змішана стратегія $\overline{q}^* = \left(q_1^*; \ldots; q_n^*\right)$ в грі I' є також його оптимальною змішаною стратегією в грі I; в) якщо вектор $\overline{p}'^* = \left(p_1^*; \ldots; p_{s-1}^*; p_{s+1}^*; \ldots; p_m^*\right)$є оптимальною змішаною стратегією гравця A в грі I', то в грі I для нього оптимальною є змішана стратегія $\overline{p}^* = \left(p_1^*; \ldots; p_{s-1}^*; 0; p_{s+1}^*; \ldots; p_m^*\right)$. Теорема без доведення.

Отже, якщо стратегія A_k домінує над стратегією A_s, то ймовірність застосування останньої в оптимальній змішаній стратегії $\overline{p}^*$ гравця A дорівнює нулю, тому s - й рядок з платіжної матриці можна виключити.

Аналогічна теорема існує у випадку домінування стратегій гравця B.

Приклад. Виконати можливі спрощення платіжної матриці:

$$\begin{pmatrix} 7 & 4 & 8 & 6 & 5 \\ 5 & 4 & 6 & 1 & 5 \\ 7 & 4 & 8 & 6 & 5 \\ 9 & 6 & 3 & 2 & 7 \end{pmatrix}.$$

Відповідні елементи першого і третього рядка є рівними, тому один з них (наприклад третій) можна виключити. Елементи другого рядка не перевищують відповідних елементів першого, тому другий рядок теж виключаємо і приходимо до матриці:

$$\begin{pmatrix} 7 & 4 & 8 & 6 & 5 \\ 9 & 6 & 3 & 2 & 7 \end{pmatrix}.$$

Елементи першого стовпця даної матриці перевищують відповідні елементи другого стовпця, елементи третього ─ елементи четвертого, а елементи п'ятого ─ елементи другого. Домінуємі перший, третій і п'ятий стовпці опускаємо. У результаті отримаємо матрицю:

$$\begin{pmatrix} 4 & 6 \\ 6 & 2 \end{pmatrix}.$$

Якщо цю матрицю проаналізувати з позицій гравця A, то ніяких подальших спрощень виконати неможливо. Отже, замість розв’язання гри з матрицею вимірності 4×5 досить вирішити гру вимірності 2×2: ймовірності активних стратегій гравців в обох іграх будуть одними і тими ж, що випливає з теореми.

14.4. Розв'язання матричної гри графічним методом

Найбільш проста матрична гра – це гра, в якій кожен з гравців має дві стратегії. Матриця $\mathbf{A}$ має вигляд:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

Якщо сідлової точки немає, то розв'язок гри знаходиться в змішаних стратегіях, а саме $\mathbf{X} = \left(x_1, x_2 \right)$, $\mathbf{Y} = \left(y_1, y_2 \right)$. Це є ймовірності, з якими застосовують гравці свої стратегії.

Необхідно записати систему рівнянь

$$\begin{cases} a_{11}x_1 + a_{21}x_2 = v \\ a_{12}x_1 + a_{22}x_2 = v \end{cases}.$$

З урахуванням $x_1 + x_2 = 1$, знаходимо

$$x_1 = \frac{a_{22} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}}, \quad x_2 = \frac{a_{11} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}}. \tag{14.10}$$

Ціна гри:

$$v = \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}}. \tag{14.11}$$

Складаючи аналогічну систему рівнянь, знаходимо оптимальну стратегію для гравця B:

$$y_1 = \frac{a_{22} - a_{12}}{a_{11} + a_{22} - a_{12} - a_{21}}, \quad y_2 = \frac{a_{11} - a_{21}}{a_{11} + a_{22} - a_{12} - a_{21}}. \tag{14.12}$$

Приклад. Розв'язати гру, задану платіжною матрицею:

$$\mathbf{A}=\begin{pmatrix}1 & 3\\ 2 & 1\end{pmatrix}.$$

Маємо, що $\alpha = 1$, $\beta = 2$, матриця не має сідлової точки. За формулами знаходимо оптимальні стратегії та ціну гри:

$$\mathbf{X}=\left(\frac{1}{3};\frac{2}{3}\right),\ \mathbf{Y}=\left(\frac{2}{3};\frac{1}{3}\right),\ \nu=\frac{5}{3}.$$

Гру з матрицею 2×2 можна розв'язати графічно за допомогою таких побудов (рис. 14.1 − 14.2).

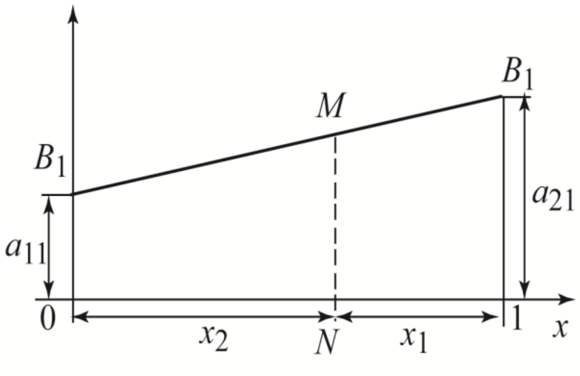


Рис. 14.1.

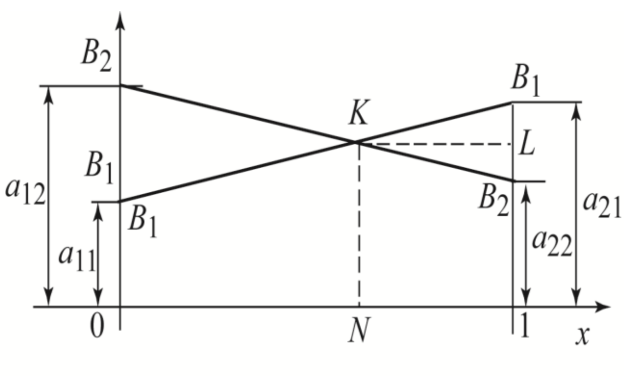


Рис. 14.2.

За віссю абсцис відкладаємо відрізок, довжина якого дорівнює одиниці. Лівий кінець відрізка (точка $x = 0$) відповідає стратегії A_1, правий – стратегії A_2. Проміжні точки x відповідають змішаним стратегіям (x_1, x_2), де $x_1 = 1 - x$, $x_2 = x$. На кінцях цього відрізка перпендикулярно до осі абсцис проводять прямі і відкладають виграш при відповідних чистих стратегіях. Якщо гравець B використовує стратегію B_1, то виграш під час використання чистих стратегій A_1 і A_2 складе відповідно a_{11} і a_{21}. Відкладемо ці точки і з'єднаємо їх прямою B_1B_1. Якщо гравець A використовує змішану стратегію, то його виграшу відповідає деяка точка M, яка лежить на цій прямій. Аналогічно можна побудувати пряму B_2B_2, відповідну стратегії B_2 гравця *B*. Ламана B_1KB_2 – нижня межа виграшу, який отримає гравець *A*. Точка K, в якій виграш максимальний, визначає ціну гри і її розв'язок. Для знаходження оптимальної стратегії гравця B використовуються формули:

$$y_1 = \frac{LB_2}{LB_2 + LB_1}, \quad y_2 = \frac{LB_1}{LB_2 + LB_1}.$$

Справедливість цих співвідношень доводиться, якщо у формули, що виражають y_1 і y_2, підставити замість LB_2 і LB_1 їх значення, маємо:

$$LB_2 = v - a_{22};\ LB_1 = a_{21} - v.$$

Тут же можна розглянути задачу мінімізації верхньої межі виграшу для гравця *B*, помінявши місцями у розв'язанні гравців A і B.

Використовуючи геометричну інтерпретацію, можна знайти розв'язання ігор, заданих матрицею $2 \times n$. Кожній із n стратегій гравця B відповідає пряма. Побудувавши ці прямі, знаходять нижню межу виграшу. Точка K, що на нижній межі і для якої величина виграшу найбільша, визначає ціну гри і її розв'язок. При цьому визначаються активні стратегії гравця B (відповідні їм прямі перетинаються в точці K): з геометричних міркувань можна знайти значення y_j, що відповідають активним стратегіям гравця B.

Аналогічно може бути розв'язана гра з матрицею $m \times 2$, тільки при цьому будують верхню межу виграшу і на ній визначають мінімум.

Геометричні побудови доцільно використовувати для визначення активних стратегій гравців. Потім розв'язання знаходять за аналітичними формулами (14.10) − (14.12), або відповідні значення ***X***, ***Y*** і v знаходять з геометричних міркувань.

Приклад. Знайти розв'язок гри, заданої матрицею:

$$\mathbf{A} = \begin{pmatrix} 2 & 3 & 1 & 4 \\ 4 & 2 & 3 & 1 \end{pmatrix}.$$

Прямі на рис. 14.3 відповідають стратегіям гравця B. Ламана B_3KB_4 відповідає нижній межі виграшу.

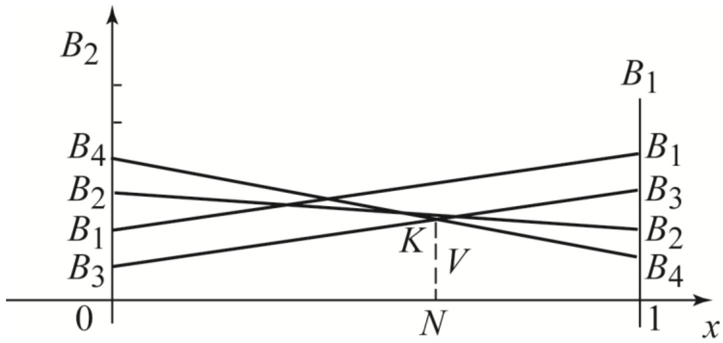


Рис. 14.3. **Графічний розв'язок гри, матриця якої має розміри 2×4**

Оптимальні стратегії гравця B – третя і четверта. За формулами (1) – (3) можна знайти розв'язання гри: $\mathbf{X}=(0{,}4;\,0{,}6)$; $\mathbf{Y}=(0;\,0;\,0{,}6;\,0{,}4)$, $v=2{,}2$. Отже, гравець A використовує стратегію A_1 з імовірністю 0,4, а стратегію A_2 – з імовірністю 0,6. При цьому його виграш в середньому складе 2,2 одиниць.

14.5. Розв'язання матричної гри зведенням до задачі лінійного програмування

Розглянемо випадок, коли в платіжній матриці $\lfloor a_{ij} \rfloor_{m\times n}$, $\alpha \neq \beta$ і всі елементи $a_{ij} \geq 0$. Ясно, що тоді і ціна гри $v>0$. Знайдемо спочатку оптимальну змішану стратегію $\bar{q}=(q_1;\,...;q_n)$ гравця *B*. Використовуючи її, він програє не більше v при будь-якій чистій стратегії A_i гравця A, тобто:

$$\sum_{j=1}^{n} a_{ij} q_j \leq v \qquad \left(i=\overline{1,m}\right).$$

Розділимо обидві частини нерівності на v:

$$\sum_{j=1}^{n} a_{ij} \frac{q_j}{v} \leq 1 \qquad \left(i=\overline{1,m}\right).$$

Позначивши

$$\frac{q_j}{v}=y_j \qquad \left(j=\overline{1,n}\right), \tag{14.13}$$

отримаємо систему обмежень для нових змінних:

$$\sum_{j=1}^{n} a_{ij} y_j \leq 1 \quad \left(i=\overline{1,m}\right) \tag{14.14}$$

$$y_j \geq 0 \quad \left(j=\overline{1,n}\right). \tag{14.15}$$

Крім того, y_j задовольняє умову:

$$\sum_{j=1}^{n} y_j = \sum_{j=1}^{n} \frac{q_j}{v} = \frac{1}{v} \sum_{j=1}^{n} q_j = \frac{1}{v} \cdot 1 = \frac{1}{v}.$$

Гравець B прагне зробити свій гарантований програш v можливо меншим, а значить, можливо більшою величину:

$$\frac{1}{v} = \sum_{j=1}^{n} y_j = \varphi. \qquad (14.16)$$

Враховуючи співвідношення (14.13)–(14.15), приходимо до наступної математичної моделі задачі:

$$\varphi = \sum_{j=1}^{n} y_j \quad \to \max \qquad (14.17)$$

за лінійних обмежень основної системи:

$$\sum_{j=1}^{n} a_{ij} y_j \le 1 \quad \left(i = \overline{1, m}\right) \qquad (14.18)$$

та обмежень на знак:

$$y_j \ge 0 \quad \left(j = \overline{1, n}\right). \qquad (14.19)$$

Це типова задача лінійного програмування, що записана в симетричній формі. Вирішивши її, знайдемо оптимальний вектор $\mathbf{Y}^* = \left(y_1^*; \ldots; y_n^*\right)$ і відповідне йому максимальне значення цільової функції $\varphi^* = \varphi_{\max}$, а потім, використовуючи вирази (14.13) і (14.15), визначимо ціну гри і компоненти оптимальної змішаної стратегії:

$$v = \frac{1}{\varphi_{\max}}, \quad q_j^* = v y_j^* \quad \left(j = \overline{1, n}\right). \qquad (14.20)$$

Міркуючи аналогічно, приходимо до задачі лінійного програмування:

$$f = \sum_{i=1}^{m} x_i \quad \to \min, \qquad (14.21)$$

$$\sum_{i=1}^{m} a_{ij} x_i \ge 1, \quad j = \overline{1, n}, \qquad (14.22)$$

$$x_i \ge 0, \qquad i = \overline{1, m}, \qquad (14.23)$$

розв’язуючи яку, знайдемо оптимальний вектор $\mathbf{X}^* = \left(x_1^*; \ldots; x_m^*\right)$ і $f = f_{min}$, а потім визначимо:

$$v = \frac{1}{f_{\min}}, \quad p_i^* = v x_i^* \quad \left(i = \overline{1, m}\right), \qquad (14.24)$$

тобто оптимальну змішану стратегію $\overline{p}^* = \left(p_1^*; \ldots; p_m^*\right)$ гравця A.

Задачі (14.17) − (14.19) і (14.21) − (14.23) утворюють симетричну пару двоїстих спряжених задач лінійного програмування, а тому немає необхідності вирішувати обидві задачі. Знайшовши розв'язок однієї з них, скористаємося відповідністю між двоїстими змінними, що встановлюється стосовно канонічних форм задач

$$\begin{array}{cc} \overbrace{x_1 \; \dots \; x_m}^{\text{Вільні}} & \overbrace{x_{m+1} \; \dots \; x_{m+n}}^{\text{Базисні}} \\ \underbrace{y_{n+1} \; \cdots \; y_{n+m}}_{\text{Базисні}} & \underbrace{y_1 \; \cdots \; y_n}_{\text{Вільні}} \end{array} .$$

Потім з рядка цільової функції останньої симплекс-таблиці, що містить компоненти оптимального вектора двоїстої задачі, випишемо їх.

Приклад. Торгівельна фірма розробила декілька варіантів плану продажу товарів на майбутній ярмарці з урахуванням мінливої кон'юнктури ринка та попиту покупців. Рівні доходу від можливих сполучень представлені в табл. 14.4. Визначити оптимальний план продажу товарів.

Позначимо ймовірність застосування торгівельною фірмою стратегії $П_1 - x_1$, стратегії $П_2 - x_2$, $П_3 - x_3$, ймовірність використання стратегії $K_1 - y_1$, стратегії $K_2 - y_2$, $K_3 - y_3$.

Таблиця 14.4

План продажу	Величина доходу, грош. од.		
	K_1	K_2	K_3
$П_1$	8	4	2
$П_2$	2	8	4
$П_3$	1	2	8

Для першого гравця (торгівельної фірми) математична модель задачі має вигляд:

$$f = X_1 + X_2 + X_3 \to \min,$$

за обмежень:

$$\begin{aligned} &8X_1 + 2X_2 + X_3 \geq 1, \\ &4X_1 + 8X_2 + 2X_3 \geq 1, \\ &2X_1 + 4X_2 + 8X_3 \geq 1, \\ &X_i \geq 0, \quad i = \overline{1,3}, \end{aligned}$$

де $x_i = X_i v$.

Для другого гравця (кон'юнктури ринку та попиту покупців) математична модель задачі має вигляд:

$$z = Y_1 + Y_2 + Y_3 \to \max,$$

за обмежень:

$$\begin{aligned} &8Y_1 + 4Y_2 + 2Y_3 \le 1, \\ &2Y_1 + 8Y_2 + 4Y_3 \le 1, \\ &Y_1 + 2Y_2 + 8Y_3 \le 1, \\ &Y_j \ge 0, \quad j = \overline{1,3}. \end{aligned}$$

Знайдемо оптимальний розв'язок задачі для другого гравця симплексним методом. Остання таблиця має вигляд (табл. 14.5).

Таблиця 14.5

Базис	$C_б$	C	1	1	1	0	0	0
		A_0	A_1	A_2	A_3	A_4	A_5	A_6
$У_1$	1	$\frac{1}{14}$	1	0	0	$\frac{1}{7}$	$-\frac{1}{14}$	0
$У_2$	1	$\frac{11}{196}$	0	1	0	$-\frac{3}{98}$	$\frac{31}{1996}$	$-\frac{1}{14}$
$У_3$	1	$\frac{5}{49}$	0	0	1	$-\frac{1}{98}$	$-\frac{3}{98}$	$\frac{1}{7}$
	Δ	$\frac{45}{196}$	0	0	0	$\frac{5}{49}$	$\frac{11}{196}$	$\frac{1}{14}$

З табл. 14.5 слідує, що $Y^* = \left(\frac{1}{14}, \frac{11}{196}, \frac{5}{49}\right)$, $z\left(Y^*\right) = \frac{45}{196}$.

Ціна гри $v = \frac{1}{z\left(Y^*\right)} = \frac{196}{45}$. Оскільки $y_i = Y_i v$, то $y_1 = \frac{14}{45}$, $y_2 = \frac{11}{45}$, $y_3 = \frac{20}{45}$.

Оптимальна стратегія другого гравця: $y^* = \left(\frac{14}{45}, \frac{11}{45}, \frac{20}{45}\right)$.

Стратегії першого гравця знайдемо з останньої симплекс-таблиці, використовуючи метод відповідності змінних висхідної і двоїстої задач. Отримаємо:

$$x^* = \left(\frac{20}{45}, \frac{11}{45}, \frac{14}{45}\right).$$

Таким чином, торгівельна фірма на ярмарці повинна дотримуватись стратегії

$x^* = \left(\frac{20}{45}, \frac{11}{45}, \frac{14}{45}\right)$, при цьому вона одержить доход не менше $v = \frac{196}{45}$ грош. од.

Запитання для самоперевірки

1. Які задачі в економіці можна розв'язувати методами теорії ігор?
2. Наведіть основні поняття теорії ігор.
3. Як розв'язати матричну гру в чистих стратегіях?
4. Як розв'язати матричну гру в змішаних стратегіях?
5. Наведіть основну теорему теорії ігор.
6. Як розв'язати матричну гру графічним методом?

15. Аналіз та управління ризиком в економіці на базі концепції теорії ігор

15.1. Суть ризиків та управління ними в економіці.

15.2. Аналітичні методи оцінки ризиків в економіці.

15.3. Ігри з природою.

15.4. Критерії для прийняття рішень.

15.1. Суть ризиків та управління ними в економіці

Згідно економіко-математичним енциклопедичним словником ризик (економічний, господарський) визначається як небезпека виникнення непередбачуваних збитків, утрат прибутку, що очікується, доходу або майна, коштів у зв’язку з випадковою зміною умов економічної діяльності, несприятливими обставинами [36, с. 457- 458]. Виникає ризик, коли комерційна, виробнича діяльність здійснюється в умовах невизначеності через нестачу інформації й через це досягнення позитивного результату не гарантується. Ризик вважається невід’ємною компонентою ринкової економіки та вимірюється частотою, ймовірністю виникнення збитків і їх рівнем. Ухвалюючи те чи інше рішення в економіці управлінці завжди оцінюють ймовірність збитків та вибирають надійний варіант з альтернативних на основі допустимого рівня ризику. Кількісна оцінка рівня господарського ризику є обов’язковою складовою техніко-економічного обґрунтування будь-якого проекту або пропозиції.

Управління ризиком передбачає: 1) визначення найбільш важливих наслідків критичних ситуацій; 2) розробку схеми поведінки суб’єкту господарювання в різних варіантах розвитку критичної ситуації для зменшення збитків від непередбачених факторів; 3) формування умов подолання негативних наслідків ухваленого управлінського рішення або проекту для продовження виробничо-господарської діяльності фірми чи підприємства. Управляти ризиком означає не тільки впливати на результат з метою зменшення негативних наслідків ризикових ситуацій, але й активно впливати на кожну функцію управління. Відомо, що найпоширенішим методом управління є диферсифікація ризику, що передбачає розділення його. Прикладом цього є збалансований портрет інвесторів. Іншим відомими методами є вивчення ринку та його тенденцій, клієнтів та партнерів, маркетингове дослідження, страхування ризику, зазвичай створюється резервний фонд для покриття збитків.

В економіці виокремлюють розумні ризики, за яких рівень матеріальних збитків не призведе до кризової ситуації, та катастрофічні ризики. У свою чергу розумні ризики мають зони допустимого та критичного ризику. Катастрофічний ризик призводить до руйнування всієї діяльності та до її кінця. В зоні допустимого ризику діяльність суб’єкта господарювання зберігає свою доцільність й очікувані збитки менше очікуваного прибутку й існують загрози недоотримання наміченого прибутку. Зона критичного ризику характеризується випадковими збитками, розміри яких перевищують величину очікуваного прибутку, величину всіх засобів, вкладених у справу, при цьому не тільки недоотримується прибуток, але наявні збитки в сумі, що дорівнює всім витратам. У зоні катастрофічного ризику випадкові збитки досягають величини, що дорівнює всьому капіталу підприємця, що призводить до банкрутства, закриття підприємства, припинення діяльності.

Фахівці з проблем ризику в економіці класифікують ризики за такими ознаками: місцем виникнення (зовнішні і внутрішні), ступенем розповсюдженості (загальні інформаційні, особливі банківські та виробничі), рівнем прийняття рішення (макроекономічні та мікроекономічні), тривалістю дії (короткострокові та постійні), рівнем збитків (мінімальні, середні, оптимальні, максимальні, допустимі, критичні, катастрофічні), ступенем правомірності (оправдані та неоправдані), можливістю страхування (ті, що можна застрахувати та ті, що не можна застрахувати), рівнем сталості збитків (статичні та данамічні), відносно аспектів ризику (психологічні, соціальні, економічні, юридичні, політичні, медико-біологічні, комбіновані), об'єктивністю (з об'єктивною ймовірністю, зі суб'єктивною ймовірністю та суб'єктивно-об'єктивною ймовірністю), часом ухвалення рішення (випереджувані, своєчасні, запізнілі), типами (обґрунтовані, необґрунтовані, азартні), кількістю осіб, що ухвалюють рішення (індивідуальні, групові), видами (пов'язані з природними лихами та соціально-економічні).

15.2. Аналітичні методи оцінки ризиків в економіці

До складу аналітичних методів оцінки ризиків в економіці входять: статистичні методи, метод з використанням інструментів теорії ігор, метод експертних оцінок, метод побудови дерева рішень, метод аналогій та ін. Ризики оцінюють ймовірністю настання несприятливих негативних наслідків: $W = p$ або $W = px$, де $x-$ величина негативних наслідків, величина збитків. Ризики обчислюють і як $W = M(x) = m_x-$ математичне сподівання несприятливих наслідків та дисперсію $W = D(x) = \sigma_x^2$. Саме дисперсію найчастіше використовують як показник ризику, велике значення якого говорить про великі коливання значень показників, що досліджуються. Ризики оцінюються за допомогою коефіцієнта варіації: $W = V(x) = \dfrac{\sigma(x)}{\bar{x}}$.

15.3. Ігри з природою

З метою зменшення несприятливих наслідків у кожному конкретному випадку управлінського рішення слід враховувати ступінь ризику і наявну інформацію. Тут особа, яка приймає рішення (ОПР), вступає в ігрові відносини з деякою абстрактною особою, яку умовно називають «природою». При цьому маються деякі імовірнісні характеристики стану природи. Таку ситуацію прийнято називати іграми з природою.

Будь-яку господарську діяльність людини можна розглядати як гру з природою. У широкому сенсі під «природою» розуміють сукупність невизначених чинників, що впливають на ефективність прийнятих рішень.

Задачею особи, що приймає рішення (ОПР), є прийняття найкращого управлінського рішення в кожній конкретній ситуації. Якість прийнятого рішення залежить від інформованості ОПР про ситуацію, в якій приймається рішення. Байдужість природи до результату гри (виграшу) і можливість отримання ОПР додаткової інформації про її стан відрізняють гру з природою від звичайної матричної гри.

Ігри з природою становлять основну модель теорії прийняття рішень в умовах часткової невизначеності.

Велику кількість станів природи необхідно позначити через Π, окремий стан – $\Pi_j, \Pi_j \in \Pi \left(j = \overline{1, n}\right)$. Безліч рішень (стратегій) ОПР слід позначити через $A_i, A_i \in A \left(i = \overline{1, m}\right)$.

У взаємовідносинах з природою ОПР може використовувати будь-які з стратегій $A_1, ..., A_m$ залежно від станів Π_j природи. ОПР відшукує оптимальну поведінку, яка і буде її оптимальною стратегією. При цьому вона може користуватися як чистими, так і змішаними стратегіями.

На підставі платіжної матриці гри з природою:

$$\begin{pmatrix} a_{11}\ a_{12} \dots a_{1n} \\ a_{21}\ a_{22} \dots a_{2n} \\ \dots\dots\dots\dots\dots \\ a_{m1}\ a_{m2} \dots a_{mn} \end{pmatrix},$$

де a_{ij} – виграш ОПР, якщо вона використовує стратегію A_i під час стану природи Π_j.

Відмінність гри з природою від звичайної матричної гри полягає насамперед у спрощенні гри. Виявлення дублюючих і домінуючих стратегій проводиться тільки для стратегій ОПР.

Іноді у разі розв'язання гри з природою використовується матриця ризиків. Елементи r_{ij} матриці ризиків рівні різниці між максимально можливим виграшем і тим виграшем, який статистик отримає в тих же умовах Π_j, застосовуючи стратегію A_i, тобто $r_{ij} = \beta_j - a_{ij}$, де $\beta_j = \max_i a_{ij}$.

15.4. Критерії для прийняття рішень

Оптимальну стратегію ОПР можна визначити, використовуючи ряд критеріїв. Розглянемо дві групи критеріїв, що використовують і не використовують апріорні ймовірності q_j станів природи. До першої групи належать критерії Байєса і Лапласа. У цьому випадку користуються середнім значенням $\overline{a_i}$ виграшу.

Так, при відомому розподілі ймовірностей різних станів $П_j$ природи використовується критерій Байєса. Показником у цьому критерії служить або величина середнього виграшу, або величина середнього ризику.

Платіжна матриця $(a_{ij})_{m \times n}$ представлена у вигляді табл. 15.1.

Таблиця 15.1

Платіжна матриця $(a_{ij})_{m \times n}$

Стратегія ОПР A_i	Стан природи $П_j$				Середній виграш $\overline{a_i}$
	$П_1$	$П_2$	…	$П_n$	
A_1	a_{11}	a_{12}	…	a_{1n}	$\overline{a}_1$
…	…	…	…	…	…
A_m	a_{m1}	a_{m2}	…	a_{mn}	$\overline{a_m}$
q_j	q_1	q_2	…	q_n	

За критерієм Байєса за оптимальну приймається та чиста стратегія A_i, за якої максимізується середній виграш $\overline{a}_i$ ОПР, тобто позначається $\overline{a} = \max_i \overline{a}_i$, де $\overline{a}_i = \sum_{j=1}^{n} a_{ij} q_j$.

Матриця ризиків представлена у вигляді табл. 15.2.

Таблиця 15.2

Матриця ризиків (r_{ij})

Стратегія ОПР A_i	Стан природи Π_j				Середній ризик $\overline{r_i}$
	Π_1	Π_2	$\dots$	Π_n	
A_1	r_{11}	r_{12}	$\dots$	r_{1n}	$\overline{r}_1$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
A_m	r_{m1}	r_{m2}	$\dots$	r_{mn}	$\overline{r_m}$
q_j	q_1	q_2	$\dots$	q_n	

За оптимальну стратегію ОПР приймається чиста стратегія A_i, за якої мінімізується середній ризик, тобто забезпечується $\overline{r} = \min\limits_i \overline{r}_i$, де $\overline{r_i} = \sum\limits_{j=1}^{n} r_{ij} q_j$.

У разі, коли ймовірності станів природи правдоподібні, для їх оцінки використовується принцип Лапласа, згідно з яким всі стани природи вважаються рівноймовірними, тобто $q_1 = \dots = q_n = \frac{1}{n}$. Оптимальною вважається стратегія, що забезпечує максимум середнього виграшу.

До другої групи критеріїв вибору оптимальної стратегії ОПР, що використовуються при невідомих апріорних ймовірностях станів природи, відносяться критерії Вальда, Севіджа і Гурвіца.

Максимальний критерій Вальда збігається з критерієм вибору максимальної стратегії, що дозволяє отримувати нижню чисту ціну α в парній грі з нульовою сумою. При критерії Вальда за оптимальну приймається чиста стратегія, яка в найгірших умовах гарантує максимальний виграш, тобто $\alpha = \max\limits_i \min\limits_j \{a_{ij}\}$.

Критерій мінімального ризику Севіджа рекомендує вибирати оптимальну стратегію ту, за якої величина максимального ризику мінімізується в найгірших умовах, тобто забезпечується $\min\limits_i \max\limits_j \{r_{ij}\}$. Суть критерію Севіджа полягає у виборі такої стратегії, щоб не допустити надмірно великих збитків, до яких вона може призвести. Спочатку знаходять матрицю ризиків, елементи якої показують, який збиток матиме підприємство, якщо для кожного стану природи не буде вибрана найкраща стратегія. Елементи матриці ризиків знаходять за формулою $r_{ij} = \max a_{ij} - a_{ij}$, де $\max a_{ij}$ – максимальний елемент у стовпці висхідної матриці. Далі оптимальна стратегія за даним критерієм знаходиться за формулою: $\min\left(\max\left(\max a_{ij} - a_{ij}\right)\right)$.

Критерії Вальда і Севіджа орієнтують ОПР на найбільш несприятливі стани природи, тобто ці критерії висловлюють песимістичну оцінку ситуації.

Критерій Гурвіца є критерієм песимізму – оптимізму. За оптимальну приймається та стратегія, для якої виконується співвідношення:

$$\max_i \left\{ \lambda \min_j \{a_{ij}\} + (1-\lambda) \max_j \{a_{ij}\} \right\}, \quad \text{де} \quad 0 \le \lambda \le 1 .$$

При $\lambda = 0$ маємо критерій крайнього оптимізму, а при $\lambda = 1$ – критерій песимізму Вальда. Якщо $0 < \lambda < 1$, то маємо середнє. При бажанні підстрахуватися в даній ситуації λ приймають близьким до одиниці. У загальному випадку число λ вибирають виходячи з досвіду чи суб'єктивних міркувань.

Приклад. Продовжимо розв'язання прикладу теми 14 в табл. 15.3.

Таблиця 15.3

Матриця платоспроможного попиту

Обсяг виробництва (оптові закупівлі)	Розмір прибутку (a_{ij}) залежно від можливих коливань попиту				$\alpha_i = \min \{a_{ij}\}$
	489 876,17	986 200,50	1 488 973	1 967 962,1	
S_1=986 200,50	49 310,03	197 240,1	197 240,1	197 240,1	49 310,03
S_2=1 488 973	-60	148 897,3	297 794,6	297 794,6	-60
S_3=1 967 962,1	-1 140	98 398,11	196 796,24	393 592,42	-1 140
$\beta_j = \max \{a_{ij}\}$	49 310,03	197 240,1	297 794,6	393 92,42	

Побудуємо матрицю ризиків (табл. 15.4).

Таблиця 15.4

Матриця ризиків

r_{ij}	489 876,17	986 200,50	1 488 973	1 967 962,1	$\max \{r_i\}$	$S_{\text{опт}}$
S_1	0	0	100 554,5	196 352,32	196 352,32	
S_2	49 370,03	48 342,80	0	95 797,82	95 797,82	95 797,82
S_3	50 450,03	98 841,99	100 998,36	0	100 998,36	

Значення критерію Севіджа полягає в прагненні уникнути більшого ризику під час вибору рішень. Отже, відповідно до критерію, необхідно виробляти продукцію в обсязі $S_2 = 1488973$ тис. грн.

Нехай дотримується песимістична оцінка (критерій Гурвіца) і вважається, що $\lambda = 0{,}8$, тоді для кожної стратегії відповідно маємо:

$$y_1 = 0{,}8 \cdot 49\,310{,}03 + (1 - 0{,}8) \cdot 197\,240{,}1 = 78\,896{,}04 \text{ тис. грн},$$

$$y_2 = 0{,}8 \cdot (-60) + (1 - 0{,}8) \cdot 297\,794{,}6 = 59\,510{,}92 \text{ тис. грн},$$

$$y_3 = 0{,}8 \cdot (-1140) + (1 - 0{,}8) \cdot 393\,592{,}42 = 77\,806{,}48 \text{ тис. грн}.$$

За критерієм Гурвіца раціональним є такий варіант обсягу виробництва:

$$Y = \max_i \{y_i\} = \max\{78\,896{,}04; 59\,510{,}92; 77\,806{,}48\} = 78\,896{,}04 \text{ тис. грн} \rightarrow S_1.$$

Приклад. Фірма «Фармацевт» виробляє ліки та біомедичну продукцію. Пік попиту на деякі лікарські препарати (сердечно-судинні, анальгетики) приходиться на літній період, інші (антиінфекційні, від кашлю) − на осінній та весняний періоди. Затрати на 1 умовну одиницю продукції за вересень – жовтень склали: в першій групі лікарських препаратів – 20 грош. од.; у другій групі – 15 грош. од. За даними спостережень протягом декількох останніх років відділом маркетингу фірми встановлено, що вона може реалізувати протягом вересня – жовтня в умовах теплової погоди 3 050 умовних одиниць першої групи препаратів і 1 100 умовних одиниць другої групи препаратів; в умовах прохолодної погоди – 1 525 умовних одиниць першої групи препаратів і 3 690 умовних одиниць другої групи препаратів. У зв’язку з можливими змінами погоди виникає задача – визначити стратегію фірми з випуску продукції, яка забезпечує максимальний дохід від реалізації за ціною продажу 40 грош. од. за 1 умовну одиницю продукції першої групи і 30 грош. од. за 1 умовну одиницю продукції другої групи.

Визначимо виробничу програму підприємства в умовах ризику з використанням матричних ігор. Відомо, що фірма володіє двома стратегіями: A_1 − якщо в цьому році буде тепла погода, і A_2 − якщо в цьому році буде холодна погода. Якщо фірма прийме стратегію A_1 і в дійсності буде тепла погода (стратегія природи B_1), то випущена продукція (3 050 умовних одиниць першої групи препаратів і 1 100 умовних одиниць другої групи препаратів) буде повністю реалізована і дохід складе $3\,050(40 - 20) + 1100(30 - 15) = 77\,500$ грош. од. У випадку прохолодної погоди (стратегія природи B_2) препарати другої групи будуть продані повністю, а першої групи тільки в кількості 1 525 умовних одиниць і частина препаратів залишиться нереалізованою. Дохід складе $1525(40-20)+1100(30-15)-20(3\,050-1\,525)=16\,500$ грош. од. Аналогічно, якщо фірма прийме стратегію A_2 і в дійсності буде холодна погода, то дохід складе $1\,525(40-20)+3\,690(30-15)=85\,850$ грош. од. За теплої погоди дохід складе: $1\,525(40-20)+1100(30-15)-(3\,690-1100)15=8\,150$ грош. од.

Розглядаючи фірму і погоду як два гравця, отримаємо платіжну матрицю:

$$\begin{array}{ccc} & B_1 & B_2 \\ A_1 & 77\,500 & 16\,500 \\ A_2 & 8\,150 & 85\,850 \end{array}$$

$$\alpha=\max(16\,500,8\,159)=16\,500 \text{ грош. од.},$$

$$\beta=\min(77\,500,85\,850)=77\,500 \text{ грош. од.}$$

Ціна гри лежить в діапазоні $16\,500 \le v \le 77\,500$ грош. од.

З платіжної матриці видно, що за всіма умовами дохід фірми буде не менший 16 500 грш. одн., але якщо погодні умови співпадуть з обраною стратегією, то дохід фірми може скласти 77 500 грош. од.

Позначимо ймовірність застосування фірмою стратегії A_1 через x_1, а стратегії A_2– через x_2, причому $x_1=1-x_2$. Розв'язавши гру графічним методом, отримаємо $x_{опт}=(0{,}56;\,0{,}44)$ при цьому ціна гри $v=46\,986$ грош. од. Оптимальний план виробництва лікарських препаратів складе:

$$0{,}56(3\,050;1100)+0{,}44(1\,525;3\,690)=(2\,379;\,2\,239{,}6).$$

Таким чином, фірмі доцільно випускати протягом вересня і жовтня 2 379 грош. од. препаратів першої групи і 2 239,6 грош. од. препаратів другої групи, тоді за будь-якої погоди вона отримає дохід не менше 46 986 грош. од.

В умовах невизначеності, якщо немає можливості фірмі використовувати змішані стратегії, застосовують критерії природи.

1. *Критерій Вальда*:

$$\max(\min a_{ij})=\max(16\,500,8\,150)=16\,500 \text{ грош. од.}$$

фірмі доцільно використовувати стратегію A_1.

2. *Критерій Гурвіца*: для визначеності приймемо $\lambda=0{,}4$, тоді для стратегії фірми A_1

$$\max_i \left\{ \lambda \min_j \{a_{ij}\} + (1-\lambda) \max_j \{a_{ij}\} \right\}, \quad \text{де} \quad 0 \le \lambda \le 1,$$

$$0{,}4 \cdot 16\,500 + (1 - 0{,}4) \cdot 77\,500 = 53\,100 \text{ грош. од.}$$

для стратегії A_2

$$0{,}4 \cdot 8\,150 + (1 - 0{,}4) \cdot 85\,850 = 54\,770 \text{ грош. од.,}$$

Отже $\max(53\,100,\, 54\,770) = 54\,770$ грош. од. і фірмі доцільно використовувати стратегію A_2.

3. *Критерій Севіджа*. Максимальний елемент в першому стовпці 77 500, а в другому – 85 850. Елементи матриці ризиків будуть мати вигляд:

$$r_{ij} = \begin{pmatrix} 0 & 69\,350 \\ 69\,350 & 0 \end{pmatrix},$$

$$\min\left(\max\left(\max a_{ij} - a_{ij}\right)\right) = \min(69\,350,\, 69\,350) = 69\,350 \text{ грош. од.}$$

За даним критерієм фірмі доцільно використовувати стратегію A_1 або A_2. Слід відмітити, що кожний з даних критеріїв не може бути визнаним задовільним для кінцевого вибору рішення, проте їх сумісний аналіз дозволяє наочно представити наслідки прийняття тих чи інших управлінських рішень.

За умови відомого розподілу ймовірностей різних станів природи критерієм рішення є максимум максимального сподівання. Якщо відомо, що ймовірності теплої й холодної погоди рівні і складають 0,5, тоді оптимальна стратегія фірми визначається так:

$$\max\left((0{,}5 \cdot 77\,500 + 0{,}5 \cdot 16\,500); (0{,}5 \cdot 8\,150 + 0{,}5 \cdot 85\,850)\right) = \max(47\,000; 47\,000) =$$
$$= 47\,000.$$

Фірмі доцільно використати стратегію A_1 або A_2.

Запитання для самоперевірки

1. Для яких задач в економіці застосовуються ігри з природою?

2. Наведіть основні критерії для прийняття рішень в задачах ігор з природою.

3. У чому суть критерію Байєса?

4. У чому суть критерію Вальда, Севіджа і Гурвіца?

16. Динамічне програмування

16.1. Загальні теоретичні викладки.

16.2. Розв'язання задачі динамічного програмування в аналітичній формі.

16.3. Розв'язання задачі про заміну обладнання.

16.1. Загальні теоретичні викладки

Основним методом динамічного програмування є розроблений американським математиком Р. Беллманом метод рекурентних співвідношень, в основі якого лежить принцип оптимальності: якщо управління процесу оптимальне то воно буде оптимальним і для процесу, що залишається після виконання першого кроку. Цей принцип дозволяє встановити співвідношення між екстремальними значеннями цільової функції в задачах з різною тривалістю процесу з різними початковими станами [1, с. 125].

Математичне формулювання принципу оптимальності частіше наводиться для задач з адитивним критерієм оптимальності (сепарабальна функція мети).

Нехай необхідно знайти максимум функції:

$$Z = z_1(x_1) + z_2(x_2) + z_3(x_3)$$

за обмежень

$$\begin{cases} u_1(x_1) + u_2(x_2) + u_3(x_3) \le a_0, \\ v_1(x_1) + v_2(x_2) + v_3(x_3) \le b_0 \end{cases}$$

і будь-якими двосторонніми обмеженнями на невідомі.

Зручно інтерпретувати x_j як обсяг виробництва певного товару, $z_j(x_j)$ – прибуток від реалізації цієї продукції, $u_j(x_j)$, $v_j(x_j)$ – витрати двох видів ресурсів на виробництво, a_0, b_0 – наявні запаси ресурсів.

Необхідно звести цю багатовимірну задачу до послідовності одновимірних задач, у кожній з яких враховується тільки одна невідома. Слід позначити через $f_1(a_1, b_1)$, $f_2(a_2, b_2)$, $f_3(a_3, b_3)$ – максимальний прибуток, отриманий від випуску тільки одного виду товару, тільки перших двох видів і всіх трьох найменувань при заданих ресурсах в обсягах $a_j \le a_0, b_j \le b_0$. За таких позначень оптимальний розв’язок можна записати як $Z_{\max} = f_3(a_0, b_0)$, де

a_0, b_0 – конкретні числові значення обсягів ресурсів, а a_j, b_j – ще невідомі параметри.

Перший (початковий) етап. Нехай задані всі невідомі, крім першої, і необхідно знайти оптимальний план випуску продукції одного першого найменування при наявних залишках ресурсів a_1, b_1. Іншими словами, необхідно знайти $f_1(a_1, b_1) = \max \{z_1(x_1)\}$ за обмежень $u_1(x_1) \le a_1$, $v_1(x_1) \le b_1$ і двосторонніх обмежень на x_1.

Другий етап. Визначимо оптимальний план виробництва двох видів продукції при наявних залишках ресурсів a_2, b_2.

Необхідно знайти $f_2(a_2, b_2) = \max \{z_1(x_1), z_2(x_2)\}$

за обмежень

$$\begin{cases} u_1(x_1) + u_2(x_2) \le a_2, \\ v_1(x_1) + v_2(x_2) \le b_2 \end{cases}$$

і двосторонніх обмежень на невідомі.

Зведемо цю двовимірну задачу до одновимірної. Нехай невідома x_2 – задана, тоді запаси ресурсів на виробництво видів продукції, що залишилися, скорочуються на $u_2(x_2)$, $v_2(x_2)$ і максимум прибутку (при заданій x_2) буде дорівнювати:

$$z_2(x_2) + f_1(a_2 - u_2(x_2), b_2 - v_2(x_2)),$$

де $f_1(a_2 - u_2(x_2), b_2 - v_2(x_2)) = f_1(a_1, b_1)$ знайдений на попередньому етапі найбільший прибуток від випуску першого виду товару при залишках ресурсів, які зараз залежать від заданого значення x_2.

Загальний прибуток від випуску двох видів товару виявився вираженим через одну невідому x_2. Перебираючи її можливі значення, знаходимо максимум прибутку у процесі виробництва двох видів продукції в такій формі:

$$f_2(a_2, b_2) = \max_{x_2} \{z_2(x_2) + f_1(a_2 - u_2(x_2), b_2 - v_2(x_2))\} .$$

Отримали відоме функціональне (рекурентне) рівняння Беллмана. Невідома x_2 – задається, далі додається оптимальний розв'язок попереднього етапу $f_1(a_1, b_1)$ для залишків ресурсів:

$$a_1 = a_2 - u_2(x_2), b_1 = b_2 - v_2(x_2),\ b_1 = b_2 - v(x_2).$$

Шукається максимум отриманої суми за всіма можливими x_2.

Третій етап. Повторюючи міркування попереднього етапу, отримуємо аналогічне функціональне рівняння:

$$f_3(a_3, b_3) = \max_{x_3} \{ z_3(x_3) + f_2(a_3 - u_3(x_3), b_3 - v_3(x_3)) \} .$$

Оскільки третій етап – заключний, то можна не розв'язувати задачу із змінними запасами ресурсів, а відразу покласти їх конкретним числовим значенням заданих запасів ресурсів $a_3 = a_0, b_3 = b_0$. На заключному етапі визначається остання компонента оптимального плану x_3^0.

Зворотним ходом процедури визначаємо інші компоненти оптимального плану. Оскільки вже відомо значення x_3^0, знаходимо залишки ресурсів $a_2 = a_3 - u(x_3^0), b_2 = b_3 - v(x_3^0)$ та із загального розв'язку другого етапу знаходимо значення другої компоненти оптимального плану x_2^0. Далі обчислюємо $a_1 = a_2 - u(x_2^0), b_1 = b_2 - v(x_2^0)$ і із загального розв'язку першого етапу знаходимо першу компоненту x_1^0.

16.2. Розв'язання задачі динамічного програмування в аналітичній формі

Знайти максимум сепарабельної функції цілі:

$$z = z(x_1) + z(x_2) = x_1 + (2x_2 - x_2^2)$$

за обмежень

$$\begin{cases} 3x_1 + 2x_2 \le 6, \\ x_1 + 2x_2 \le 4, \\ x_1 \ge 0,\ x_2 \ge 0. \end{cases}$$

Перший (початковий) етап. На 1-му етапі необхідно знайти максимум одновимірної функції:

$$f_1(a_1, b_1) = \max z_1(x_1) = \max x_1$$

за обмежень

$$\begin{cases} 0 \le 3x_1 \le a_1, \\ 0 \le x_1 \le b_1. \end{cases}$$

Задача проста і її легко розв'язати аналітично для всіх a_1, b_1. Змінна x_1 приймає максимальне значення на правій межі інтервалу варіювання в точці $x_1 = \min\left\{\frac{a_1}{3}, b_1\right\}$. Значення $f_1(a_1, b_1)$ дорівнює цьому ж виразу.

Другий (заключний) етап. Необхідно виписати рекурентне співвідношення:

$$f_2(a_2, b_2) = \max_{x_2} \{z_2(x_2) + f_1(a_2 - 2x_2, b_2 - 2x_2)\} =$$
$$= \max_{x_2} \left\{\left(2x_2 - x_2^2\right) + f_1(6 - 2x_2, 4 - 2x_2)\right\}.$$

Оскільки цей етап – заключний, то слід підставити відразу відомі значення параметрів $a_2 = 6, b_2 = 4$.

Знаходимо $x_1 = \min\left\{\frac{a_1}{3}, b_1\right\} = \min\left\{\left(\frac{6 - 2x_2}{3}, 4 - 2x_2\right)\right\}$ для $0 \le x_2 \le 2$.

На інтервалі $0 \le x_2 \le 1{,}5$ справедлива нерівність $\frac{6 - 2x_2}{3} \le (4 - 2x_2)$, звідки $x_1 = \frac{6 - 2x_2}{3}$.

На інтервалі $1{,}5 \le x_2 \le 2$ справедлива протилежна нерівність $\frac{6 - 2x_2}{3} \ge (4 - 2x_2)$, звідки $x_1 = (4 - 2x)$, $f_1(6 - 2x_2, 4 - 2x_2) = \max x_1$:

$$f_1(6 - 2x_2, 4 - 2x_2) = \begin{cases} 2 - \frac{2}{3}x_2, & (0 \le x_2 \le 1{,}5) \\ 4 - 2x_2, & (1{,}5 \le x_2 \le 2) \end{cases}.$$

З урахуванням виразу для f_1 отримуємо:

$$z_{\max} = f_2(6,4) = \max\begin{cases}(2x_2 - x_2^2) + (2 - \frac{2}{3}x_2), & (0 \le x_2 \le 1{,}5)\\ (2x_2 - x_2^2) + (4 - 2x_2), & (1{,}5 \le x_2 \le 2)\end{cases} =$$

$$= \max\begin{cases}(2 + \frac{4}{3}x_2 - x_2^2), & (0 \le x_2 \le 1{,}5)\\ (4 - x_2^2), & (1{,}5 \le x_2 \le 2)\end{cases} =$$

$$= \max\begin{cases}2{,}444, & (x_2 = 0{,}667)\\ 1{,}750, & (x_2 = 1{,}5)\end{cases} =$$

$$= 2{,}444 \quad (x_2 = 0{,}667)$$

При $x_2 \le 1{,}5$: $z_{\max} = 2 + \dfrac{4}{3}x_2 - x_2^2$. Максимум квадратичного тричлена досягається при $x_2 = \dfrac{2}{3} = 0{,}667$ і дорівнює $z_{\max} = \dfrac{22}{9} = 2{,}444$. При $x_2 \le 1{,}5$: $z = 4 - x_2^2$. Максимум цього виразу досягається при $x_2 = 1{,}5$ і дорівнює $z = \dfrac{7}{4} = 1{,}75$. Більше з отриманих значень дорівнює 2,444, при цьому x_2^0 = = $^2/_3$ = 0,667. Зворотним ходом отримуємо $x_1^0 = 2 - \dfrac{2}{3}x_2^0 = \dfrac{14}{9}$.

16.3. Розв'язання задачі про заміну обладнання

Однією з основних задач, які вирішуються на підприємстві за допомогою оптимізаційних методів, є задача про заміну обладнання. Старіння обладнання включає фізичний і моральний знос, у результаті чого збільшуються виробничі витрати, витрати на ремонт і обслуговування, знижується продуктивність праці і ліквідна вартість обладнання. Необхідно розробити оптимальну стратегію використання і своєчасної заміни застарілого обладнання з метою знизити до мінімуму сумарні витрати на придбання обладнання і його експлуатацію, що еквівалентно, щоб отримати максимальний прибуток. Залежності, що описують зростання експлуатаційних витрат, зниження продуктивності та ліквідної вартості обладнання, можуть бути задані аналітично або в табличній формі.

У цій задачі послідовність етапів не може бути довільною. Час є визначальним чинником, тому допустимі тільки дві можливі послідовності етапів - від початку до кінця (алгоритм «прямої прогонки»), або від кінця до початку (алгоритм «зворотної прогонки»). Розв'язок при цьому підрозділяється на ряд етапів, кожен з яких відповідає черговому року експлуатації.

Нехай обладнання експлуатується n = 5 років, після чого продається. На початку кожного року необхідно прийняти рішення – зберегти старе обладнання, або ж замінити його новим (можливі також й інші рішення, наприклад, провести ремонтні роботи різного ступеня складності). Вартість нового обладнання p, ліквідна вартість знижується з кожним пропрацьованим роком удвічі і після t років експлуатації становить $s(t) = p \cdot 2^{-t}$. Витрати на експлуатацію протягом року залежать від віку устаткування і збільшуються з кожним пропрацьованим роком на одну і ту ж величину. Знайдемо оптимальну політику своєчасної заміни обладнання за таких умов завдання.

Алгоритм прямої прогонки

У цьому методі потрібно знайти величину експлуатаційних витрат за весь термін служби обладнання, для чого доведеться знаходити суму членів арифметичної прогресії:

$$v(t) = u(1) + u(2) + \ldots + u(t) = 0{,}15\,p(1 + 2 + \ldots + t) = 0{,}075\,pt(t+1).$$

Перевіримо вираз: за перший рік роботи нового обладнання ці витрати складуть $0{,}75\,p \cdot 1 \cdot 2 = 0{,}15\,p$ – так і було задано за умовою задачі. Таким чином, сумарні експлуатаційні витрати тут є квадратичною функцією віку обладнання (t – кількість відпрацьованих років).

Тепер введемо позначення і виведемо основне рекурентне співвідношення. Позначимо через f_k мінімальні витрати на експлуатацію та оновлення обладнання протягом строку k років, після чого обладнання продається. У цих позначеннях метою є визначення $f_n = f_5$.

У принципі обладнання можна експлуатувати необмежено довго, так що до кінця $k-$го року вік устаткування може бути будь-яким у межах $1 \le t \le k$. Якщо припустити, що відомий цей вік, то легко визначаються загальні витрати за весь час роботи обладнання $w(t) = p - s(t) + v(t)$, а витрати за попередні роки виражаються функцією f_{k-t}. Оскільки вік t нам заздалегідь невідомий, перебираємо всі його можливі значення з інтервалу $1 \le t \le k$ і знаходимо мінімум наступного виразу:

$$f_k = \min\{w(t) + f_{k-t}\}.$$

Це є шукане рекурентне співвідношення (функціональне рівняння Беллмана). Для початку обчислювального процесу вважаємо $f_0 = 0$.

Значення витрат (з точністю до множника p) беремо з табл. 16.1 залежно від віку $1 \le t \le 5$:

Таблиця 16.1

Дані

t	$s(t)$	$v(t)$	$w(t)$
1	$\frac{1}{2}$ = 0,500	0,075·1·2 = 0,150	1 – 0,500 + 0,150 = 0,650
2	$\frac{1}{4}$ = 0,250	0,075·2·3 = 0,450	1 – 0,250 + 0,450 = 1,200
3	$\frac{1}{8}$ = 0,125	0,075·3·4 = 0,900	1 – 0,125 + 0,900 = 1,775
4	$\frac{1}{16}$ = 0,062	0,075·4·5 = 1,500	1 – 0,062 + 1,500 = 2,438
5	$\frac{1}{32}$ = 0,031	0,075·5·6 = 2,250	1 – 0,031 + 2,250 = 3,218

Послідовно обчислюємо значення $f_k = \min\{w(t) + f_{k-t}\}$:

$$f_1 = \min\{w(1) + f_0\} = 0{,}650 + 0 = 0{,}650; \qquad t = 1.$$

$$f_2 = \min\begin{Bmatrix} w(1) + f_1 \\ w(2) + f_0 \end{Bmatrix} = \min\begin{Bmatrix} 0{,}650 + 0{,}650 \\ 1{,}200 + 0 \end{Bmatrix} = \min\begin{Bmatrix} 1{,}300 \\ 1{,}200 \end{Bmatrix} = 1{,}200; \quad t = 2.$$

$$f_3 = \min\begin{Bmatrix} w(1) + f_2 \\ w(2) + f_1 \\ w(3) + f_0 \end{Bmatrix} = \min\begin{Bmatrix} 0{,}650 + 1{,}200 \\ 1{,}200 + 0{,}650 \\ 1{,}775 + 0 \end{Bmatrix} = \min\begin{Bmatrix} 1{,}850 \\ 1{,}850 \\ 1{,}775 \end{Bmatrix} = 1{,}775; \quad t = 3.$$

$$f_5 = \min\begin{Bmatrix} w(1) + f_4 \\ w(2) + f_3 \\ w(3) + f_2 \\ w(4) + f_1 \\ w(5) + f_0 \end{Bmatrix} = \min\begin{Bmatrix} 0{,}650 + 2{,}400 \\ 1{,}200 + 1{,}775 \\ 1{,}775 + 1{,}200 \\ 2{,}438 + 0{,}650 \\ 3{,}218 + 0 \end{Bmatrix} = \min\begin{Bmatrix} 3{,}050 \\ 2{,}975 \\ 2{,}975 \\ 3{,}050 \\ 3{,}218 \end{Bmatrix} = 2{,}975; \quad \begin{matrix} t = 2, \\ t = 3. \end{matrix}$$

Задача розв'язана. Оптимальна стратегія при такому співвідношенні витрат полягає в заміні устаткування після кожних двох років роботи. Оскільки нормативний термін $n = 5$ років виявився непарним, запропоновано два рівноцінних розв'язки – перший раз обладнання працює три роки, нове обладнання працює два роки, або навпаки, два роки і три роки після оновлення. Мінімальні витрати на експлуатацію та оновлення за такої стратегії становлять $2{,}975\,p$.

Алгоритм зворотної прогонки

Необхідно позначити через $\varphi_k(\tau)$ – мінімальні витрати за останні k-років за умови, що на початку цього періоду є обладнання віку τ. Тоді $\varphi_1(\tau)$ – мінімальні витрати за останній рік, де $1 \le \tau \le n-1$; $\varphi_2(\tau)$ – мінімальні витрати за останні два роки, де $1 \le \tau \le n-2$; і так далі до $\varphi_n(0)$ – шукані мінімальні витрати за всі n років (на самому початку обладнання нове). На кожному етапі треба розглядати дві альтернативи – зберегти старе обладнання або замінити його.

Нехай $k = 1$. Якщо обладнання віку τ зберегти, річні експлуатаційні витрати будуть дорівнювати $u(\tau+1)$. Устаткування після цього зноситься на рік і його ліквідна вартість буде дорівнювати $s(\tau+1)$. Останній рік роботи дещо відрізняється від інших, оскільки за умовами задачі після $n=5$ років обладнання продається за залишковою вартістю незалежно від його віку, тому при збереженні обладнання в останній рік роботи загальні витрати складуть $u(\tau+1)-s(\tau+1)$. Якщо ж в останній рік обладнання замінити, загальні витрати будуть дорівнювати $p-s(\tau)+u(1)-s(1)$ – вартість нового обладнання за вирахуванням ліквідної вартості старого, плюс експлуатаційні витрати за 1 рік роботи нового обладнання, нарешті, повернення грошей від продажу щойно придбаного устаткування. Остаточно отримуємо:

$$\varphi_1(\tau)=\min\begin{cases} u(\tau+1)-s(\tau+1) & \text{– зберегти,}\\ p-s(\tau)+u(1)-s(1) & \text{– замінити.}\end{cases}$$

Усі обчислення зведені в табл. 16.2, де для кожного значення віку обладнання τ обчислені ліквідна вартість обладнання $s(\tau)$, експлуатаційні витрати за попередній рік роботи $u(\tau)$, загальні витрати, якщо це обладнання зберегти, загальні витрати, якщо обладнання замінити, і мінімальні витрати в цих двох альтернативних варіантах.

Таблиця 16.2

Обчислення

τ	$s(\tau)$	$u(\tau)$	$u(\tau+1)-s(\tau+1)$	$(1+u(1)-s(1))-s(\tau)$	$\varphi_1(\tau)$
1	0,500	0,15	0,30–0,250=0,050	0,65–0,500=0,150	0,050 (зберегти)
2	0,250	0,30	0,45–0,125=0,325	0,65–0,250=0,400	0,325 (зберегти)
3	0,125	0,45	0,60–0,062=0,538	0,65–0,125=0,525	0,525 (замінити)

4	0,062	0,60	0,75–0,031=0,719	0,65–0,062=0,588	0,588 (замінити)
5	0,031	0,75			

До останнього 5-го року роботи вік обладнання може бути будь-яким − від 1-го до 4-х років. У табл. 16.2 є 5-ий напівпорожній рядок для обчислення $u(\tau+1)$ і $s(\tau+1)$ при $\tau=4$. Обчислені мінімальні витрати у процесі вибору найкращого варіанта заміни обладнання, звідки випливає, що навіть на початку останнього року роботи обладнання доцільно замінити, якщо його вік більше 2-х років.

Нехай тепер $k=2$, тобто розглядаються останні два роки роботи. Цей етап найважливіший для розуміння, інші етапи будуть однотипними. Отже, зазвичай, на початку цього періоду обладнання можна зберегти або замінити. Якщо обладнання віку τ зберегти, то експлуатаційні витрати в 1-й рік роботи будуть дорівнювати $u(\tau+1)$, далі обладнання постарішає на рік і подальші витрати в останній рік (при оптимальній політиці заміни зношеного обладнання) складуть $\varphi_1(\tau+1)$. Якщо ж обладнання на початку періоду замінити, то загальні витрати будуть дорівнювати $p-s(\tau)+u(1)+\varphi_1(1)$ – вартість нового обладнання за вирахуванням ліквідної вартості старого плюс експлуатаційні витрати за 1 рік роботи нового обладнання, нарешті, мінімальні витрати за наступний рік. Остаточно отримуємо:

$$\varphi_2(\tau)=\min\begin{cases} u(\tau+1)+\varphi_1(\tau+1) & \text{– зберегти,} \\ p-s(\tau)+u(1)+\varphi_1(1) & \text{– замінити.} \end{cases}$$

З точністю до множника p ці обчислення зведені в табл. 16.3.

Таблиця 16.3.

Обчислення

τ	$s(\tau)$	$u(\tau)$	$u(\tau+1)-\varphi_1(\tau+1)$	$(1+u(1)+\varphi_1(1))-s(\tau)$	$\varphi_2(\tau)$
1	0,500	0,15	0,30+0,325=0,625	1,20–0,500=0,700	0,625 (зберегти)
2	0,250	0,30	0,45+0,525=0,975	1,20–0,250=0,950	0,950 (замінити)
3	0,125	0,45	0,60+0,588=1,188	1,20–0,125=1,075	1,075 (замінити)
4	0,062	0,60			

Аналогічно при $k=3$ отримуємо рекурентне співвідношення:

$$\varphi_3(\tau) = \min \begin{cases} u(\tau+1) + \varphi_2(\tau+1) & - \text{ зберегти,} \\ p - s(\tau) + u(1) + \varphi_2(1) & - \text{ замінити.} \end{cases}$$

У табл. 16.4 обчислені значення цих витрат (без множника p).

Таблиця 16.4

Обчислення витрат без множника

τ	$s(\tau)$	$u(\tau)$	$u(\tau+1) - \varphi_2(\tau+1)$	$(1 + u(1) + \varphi_2(1)) - s(\tau)$	$\varphi_3(\tau)$
1	0,500	0,15	0,30+0,950=1,250	1,775–0,500=1,275	1,250 (зберегти)
2	0,250	0,30	0,45+1,075=1,525	1,775–0,250=1,525	1,525 ($^{\text{зберег}}/_{\text{замін}}$)
3	0,125	0,45			

При k = 4 маємо рекурентне співвідношення:

$$\varphi_4(\tau) = \min \begin{cases} u(\tau+1) + \varphi_3(\tau+1) & - \text{ зберегти,} \\ p - s(\tau) + u(1) + \varphi_3(1) & - \text{ замінити,} \end{cases}$$

Обчислюємо $\varphi_4(\tau)$ – мінімальні витрати за чотири останні роки роботи (табл. 16.5)

Таблиця 16.5

Витрати

τ	$s(\tau)$	$u(\tau)$	$u(\tau+1) + \varphi_3(\tau+1)$	$(1 + u(1) + \varphi_3(1)) - s(\tau)$	$\varphi_4(\tau)$
1	0,500	0,15	0,30+1,525=1,825	2,400–0,500=1,900	1,825 (зберегти)
2	0,250	0,30			

Обчислюємо мінімальні витрати за 5 років роботи – це тільки один варіант:

$$\varphi_5(0) = p \cdot (1 + u(1) + \varphi_4(1)) = (1 + 0{,}15 + 1.825)p = 2{,}975\,p\,.$$

Устаткування слід замінювати після кожних двох років роботи.

Мінімальні витрати на експлуатацію і заміну обладнання дорівнюють $\varphi_5(0) = 2{,}975\,p$.

Алгоритми прямої і зворотної прогонки настільки відрізняються один від одного, що свідчить, що методика динамічного програмування ще далека від формального завершення, в ній багато що залежить від мистецтва, від уміння побачити найпростіший шлях вирішення.

Запитання для самоперевірки

1. Які загальні теоретичні викладки методів динамічного програмування?
2. Наведіть етапи реалізації методів динамічного програмування.
3. Наведіть сутність етапів розв'язання задачі про заміну обладнання.

4. У чому ідея алгоритму прямої прогонки?
5. У чому ідея алгоритму зворотної прогонки?

Розділ 2. Економетрика

17. Особливості економетричних моделей та принципи їх побудови

17.1. Особливості економетричних моделей. Роль і місце економетричних моделей в аналізі соціально-економічних систем.

17.2. Формування сукупності спостережень. Поняття однорідності спостережень. Точність вихідних даних.

17.3. Основні етапи побудови економетричних моделей. Особливості обгрунтування форми економетричної моделі.

17.1. Особливості економетричних моделей. Роль і місце економетричних моделей в аналізі соціально-економічних систем

В економіко-математичному енциклопедичному словнику економетрика визначена як навчальна дисципліна, що дозволяє на базі положень економічної теорії і результатів економічних вимірювань надавати конкретні кількісні вирази загальним (кількісним) закономірностям, обумовленим економічною теорією [1, с. 597]. Економетрика грунтується на математичній статистиці, насамперед, багатовимірному статистичному аналізі, економічній теорії та економічній статистиці. Так, математична статистика забезпечує прикладним математичним інструментарієм, економічна теорія націлює на виявлення об'єктивно існуючих економічних законів між економічними показниками, вказує напрямки формалізації, методи специфікації та ідентифікації відповідних економетричних моделей, економічна статистика забезпечує інформацією економетричне моделювання.

Предметом економетрики є економічні та соціально-економічні системи, а саме модельний опис конкретних кількісних взаємозв'язків, що існують між показниками, які аналізуються [1, с. 597].

Результатом застосування інструментів економетрики є побудова економетричної моделі. Економетрична модель містить набір регресійних рівнянь, що описують стохастичні зв'язки між досліджуваними економічними показниками, а також окремі тотожності, що характеризують співвідношення між економічними показниками.

Вважається, що найбільш поширений математичний вид досліджуваних зв'язків – лінійний (щодо аналізованих змінних) і адитивні форми. При цьому можливі ситуації, коли одні й ті ж показники в одних рівняннях моделі є такі, що пояснюються, а в інших – що пояснюють (такі моделі прийнято називати системами одночасних рівнянь) [33, с. 416].

Економетричне моделювання націлене на отримання двох типів кінцевих прикладних результатів:

1) опис різних станів соціально-економічних систем (СЕС), імітація різних можливих сценаріїв їх розвитку;

2) прогнозування економічних показників, що характеризують стан і розвиток СЕС.

Економетричні моделі використовуються для аналізу соціально-економічних систем на всіх рівнях їх управління, а саме на макрорівні – країни світу, світовій економіці, на мезорівні – регіони всередині країни, адміністративні райони і на мікрорівні – підприємства, різні структури бізнесу, домашнє господарство. За допомогою економетричних моделей проводиться аналіз усіх видів діяльності суб'єктів господарювання.

У економетричному моделюванні прийнято розрізняти основні типи змінних: ендогенні, екзогенні, лагові. Під екзогенними змінними розуміються ознаки або показники, значення яких задаються ззовні моделі і вони є керованими або запланованими. Ендогенні змінні формуються в процесі функціонування соціально-економічної системи і залежать від екзогенних змінних або від їх взаємодії. Лагові змінні є ендогенні змінні, значення яких виміряні в минулі моменти часу і є вже заданими.

17.2. Формування сукупності спостережень. Поняття однорідності спостережень. Точність вихідних даних

Основою економетричних моделей є статистичні дані, які формує економічна статистика. Статистична база для економетричної моделі може складатися з часових та просторових (структурних) рядів даних. Часовим рядом називається ряд значень статистичного показника, упорядкованого за часом. Розрізняють короткострокові (до 1 року), середньострокові (1 – 3 роки) і довгострокові (понад 5 років) часові рядки.

У економетричному дослідженні використовують поняття відносної однорідності. У процесі формування сукупності спостережень слід дотримуватися порівнянності даних у просторі і в часі. Це означає, що вихідні дані повинні мати:

1) однаковий ступінь агрегування;

2) однорідну структуру одиниць сукупності;

3) одні й ті ж методи розрахунку показників у часі;

4) однакову періодичність обліку окремих змінних;

5) порівняльні ціни й однакові інші зовнішні економічні умови.

Критерії математичної статистики дозволяють кількісно перевірити однорідність. Це критерій Фішера (розглядався в теорії ймовірностей і математичній статистиці) і критерій Бартлетта.

Критерій Бартлетта є найбільш загальним критерієм у вирішенні проблем однорідності сукупності спостережень:

$$K_Б = \frac{A}{B},\ A = k\ n\ ln\sigma^2_{заг} - n\sum_i ln\sigma_i^2;\ \ B = 1 + \frac{k+1}{3nk},$$

де k – кількість груп, n – загальна кількість спостережень; $\sigma^2_{заг}$ – загальна дисперсія, σ_i^2 – дисперсія i-і групи спостережень. Якщо $\frac{A}{B}$ не перевищує χ^2 за обраних рівні значимості α і $k-1$ ступенів свободи, то можна стверджувати, що дисперсії взяті з однорідної сукупності спостережень.

Точність вимірювання – це його адекватність. Критерій точності кожного виду вимірювання визначається згідно з цілями цього виміру. Похибки вимірювання не зводяться до арифметичних похибок.

17.3. Основні етапи побудови економетричних моделей. Особливості обґрунтування форми економетричної моделі

При побудові економетричної моделі слід дотримуватися загальних методологічних принципів економіко-математичного моделювання. Відповідно до цього процедура побудови економетричної моделі повинна містити шість основних етапів:

1) постановочний етап, який передбачає визначення кінцевих цілей моделювання, набору ознак або показників;

2) апріорний етап, де проводиться теоретико-логічний аналіз соціально-економічної системи та її частин, генезис вихідних даних;

3) параметризація – коли виконується власне моделювання, тобто вибір загального виду моделі, в тому числі складу і форми зв'язків, що до неї входять;

4) інформаційний етап, що передбачає збір необхідної статистичної інформації;

5) ідентифікація моделі, згідно з якою виконується статистичний аналіз моделі, перш за все, статистичне оцінювання невідомих параметрів моделі;

6) верифікація моделі передбачає зіставлення реальних і модельних даних, перевірку адекватності моделі, оцінку точності модельних даних.

Мета побудови економетричної моделі – кількісний опис взаємозв'язків між змінними, тому вона насамперед будується за допомогою методів регресії і кореляції. При цьому проста регресія має вигляд: $\hat{y}_x = f(x)$, множинна: $\hat{y} = f(x_1, x_2, ..., x_n)$. Формування специфікації моделі – це формулювання виду моделі, виходячи з відповідної економічної теорії зв'язку між змінними. Розрізняють пояснюючі змінні – фактори. Фактори впливають на результат, тому виділяють результативну ознаку. З кола факторів, що впливають на результативну ознаку, необхідно виділити фактори, що найбільш суттєво впливають. Припустимо, що висувається гіпотеза: величина попиту *y* на товар знаходиться в зворотній залежності від ціни *x*, тобто $\hat{y}_x = a - bx$, при цьому необхідно знати, які інші фактори передбачаються незмінними, можливо, надалі їх доведеться врахувати в моделі і від простої регресії перейти до множинної.

Практично в кожному окремому випадку величина результативної ознаки складається з двох доданків: $y_j = \hat{y}_j + \varepsilon_j$, де y_j – фактичне значення результативної ознаки; $\hat{y}_j$ – теоретичне значення результату, знайдене з рівняння регресії; ε_j – випадкова величина, відхилення реального значення від теоретичного. Якщо правильно обрана специфікація моделі, то величина випадкових помилок менше.

У парній регресії вибір математичної функції $\hat{y}_x = f(x)$ може бути здійснений трьома методами: графічним, аналітичним, експериментальним. Графічний метод заснований на полі кореляції. Аналітичний метод заснований на вивченні матеріальної природи зв'язку досліджуваних ознак. У процесі обробки інформації на комп'ютері вибір виду рівняння регресії зазвичай здійснюється експериментальним методом, тобто шляхом порівняння величини залишкової дисперсії $D_{зал} = \frac{1}{n}\sum\left(y-\hat{y}_x\right)^2$, розрахованої при різних моделях. Чим менше величина залишкової дисперсії, тим меншою мірою спостерігається вплив не врахованих у рівнянні регресії факторів і краще рівняння регресії підходить до вихідних даних. Під час обробки статистичних даних на комп'ютері перебираються різні математичні функції в автоматичному режимі і з них вибирається та, для якої залишкова дисперсія є найменшою.

Методи верифікації засновані на процедурах статистичної перевірки гіпотез і на статистичному аналізі характеристик точності різних статистичних прийомів оцінювання параметрів. Найбільш поширеним і ефективним підходом до верифікації економетричної моделі визнаний принцип так званих ретроспективних розрахунків, які залежать від кінцевих цілей моделювання.

Зазвичай економетрика включає чотири розділи: виклад методів побудови класичної регресійної моделі та застосування класичного методу найменших квадратів; побудова узагальненої лінійної регресії і застосування узагальненого методу найменших квадратів; статистичний аналіз часових рядів; аналіз систем одночасних рівнянь.

Запитання для самоперевірки

1. Як визначається економетрика? На якій науковій основі вона будується?

2. У чому відмінність економетричної моделі від інших моделей?

3. Які основні цілі розробки економетричних моделей?

4. Які типи змінних розрізняють в економетричному моделюванні?

5. У чому суть поняття однорідності спостережень? Які вимоги пред'являються до вихідних даних при дотриманні однорідності?

6. Які основні етапи побудови економетричних моделей?

7. Розкрийте суть процесу формування специфікації економетричної моделі.

8. З яких розділів складається економетрика і, відповідно, які типи економетричних моделей розрізняють?

Тема 18. Парна регресія і кореляція в економетричних дослідженнях

18.1. Загальні поняття регресійного аналізу. Типи зв'язків.

18.2. Лінійна регресія і кореляція: зміст і оцінка параметрів.

Оцінювання параметрів лінійної моделі парної регресії за допомогою методу найменших квадратів (МНК).

18.3. Нелінійна регресія. Приклад обчислення оцінок параметрів квадратичної моделі методом найменших квадратів.

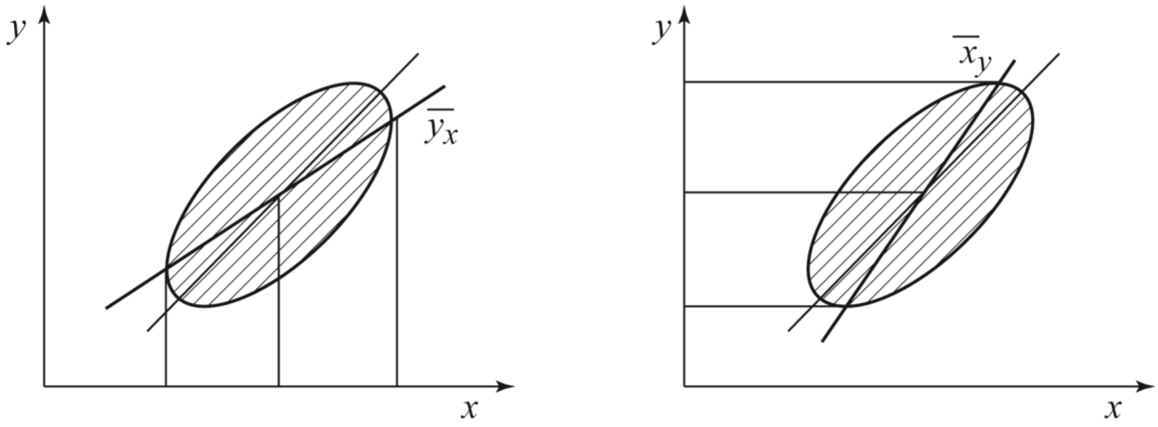
18.1. Загальні поняття регресійного аналізу. Типи зв'язків

Кореляційно-регресійний аналіз є одним з основних методів побудови економетричної моделі. Мета кореляційно-регресійного аналізу полягає у встановленні факту наявності або відсутності залежностей між кількома показниками і опису цих зв'язків досить простими виразами. Серед всіх заданих показників (ознак, змінних) один показник вважається результативною ознакою y (відгуком) і на цей показник впливають інші пояснюючі змінні x_j (фактори). При цьому результативна ознака має бути виміряна в метричній неперервній шкалі.

За визначенням *функціональних зв'язків*, кожному значенню аргументу (набору значень аргументів) відповідає єдине значення результативної ознаки. У природі часто зустрічається інший тип залежностей, коли одному і тому ж набору аргументів кожен раз відповідає кілька різних значень відгуку. У таких стохастичних (ймовірнісних) залежностях кожному значенню аргументу відповідає свій ряд розподілу результативної ознаки. При зміні аргументу X змінюються всі характеристики розподілу Y – змінюється його центр групування, діапазон, вид розподілу. У кореляційно-регресійному аналізі необхідно стежити за зміною тільки однієї характеристики розподілу залежної змінної – центру групування Y при кожному значенні X (тобто за зміною умовного математичного сподівання $M(y \mid x)$ при зміні аргументу x): $M(y \mid x)=f(x)$. Такий окремий вид стохастичної залежності називається *кореляційною залежністю*, її графік – *лінією регресії*, рівняння – *рівнянням регресії*.

Як для кореляційних, так і для функціональних залежностей має місце однозначна відповідність між значеннями аргументу x і середніми значеннями результативної ознаки (вибіркова оцінка умовного математичного очікування). Однак між цими видами залежностей залишається принципова відмінність – кореляційні залежності незворотні щодо заміни напрямку причинно-наслідкових зв'язків.

Рис. 18.1 демонструє відмінності між спряженими парними лініями регресії $\bar{y}_x = f_1(x)$ і $\bar{x}_y = f_2(y)$.



***а*)** **б)**

Рис. 18.1. **Парні лінії регресії:**

***а*) *x* – аргумент, *y* – функція, б) *y* – аргумент, *x* – функція**

Даний приклад відповідає найбільш поширеному випадку сумісного нормального розподілу двох випадкових величин (*X*, *Y*). При рівні довіри P = 0,95 практично всі точки (95 %) потрапляють в еліптичну область, так як для двовимірного нормального закону хмара розсіювання точок (*X*, *Y*) має форму витягнутого еліпса.

Для функціональної залежності не має значення, щодо якої змінної розв'язане рівняння – графік від цього не зміниться. У регресійному аналізі дослідник повинен сам призначити, яка ознака є результативною, а яка факторною. Якщо X – причина, а Y – наслідок (рис.18.1а), то лінія регресії $M(y \mid x)= f_1(x)$ буде співпадати з діаметром еліпса, що спряжений з сімейством вертикальних хорд, тобто серединами вертикальних хорд. Якщо Y – причина, а X – наслідок, то лінія регресії $M\ (x \mid y)= f_2(y)$ буде співпадати з діаметром еліпса, що спряжений з сімейством горизонтальних хорд, тобто серединами горизонтальних хорд.

Обидві спряжені лінії регресії не збігаються з головною віссю еліпса розсіювання даних – це наслідок вибору тільки однієї з змінних як випадкової, якій відповідно до гіпотез Гаусса-Маркова приписують всі погрішності – як випадкові, так і невипадкові. У цих теоретичних залежностях передбачається, що одна змінна визначається іншою. Проте буває, що обидві змінні (x, y) є різними наслідками однієї загальної причини (наприклад, внаслідок тимчасового тренда), що й обумовлює зв'язок між цими змінними. У таких задачах обидві змінні рівноправні і випадкові з спільним законом розподілу системи випадкових величин. Для опису таких залежностей розроблено спеціальний математичний інструмент – так звана діагональна регресія Фріша. Графік діагональної регресії збігається з головною віссю розсіювання даних. Слід зазначити, що діагональна регресія не є регресією за визначенням, цей тип зв'язку відрізняється від кореляційного, оскільки точки діагоналі еліпса не є середніми значеннями однієї із змінних при заданому значенні іншої змінної.

Розрізняють лінійні і нелінійні регресії. Лінійна регресія знайшла широке застосування в економетриці у вигляді чіткої економічної інтерпретації її параметрів. Реальні значення залежної змінної не завжди збігаються з її умовними математичними очікуваннями і можуть бути різними за одним і тим же значенням змінної що пояснює, тому залежність повинна бути доповнена доданком, що є випадковою величиною і вказує на стохастичну суть залежності. Присутність в моделі випадкової величини обумовлено трьома причинами: специфікацією моделі, вибірковим характером вихідних даних, особливостями виміру змінних.

Лінійна парна регресія зводиться до знаходження рівняння виду: $\hat{y}_x = a + b \cdot x$ або $y = a + b \cdot x + \varepsilon$. Нелінійні регресії діляться на два класи: регресії, нелінійні щодо включених в аналіз змінних, що пояснюють, але лінійні за параметрами, що оцінюють, і регресії, нелінійні за параметрам, що оцінюють.

Залежність декількох змінних, що виражається функцією $M(Y \mid x_1, x_2, ..., x_n) = f(x_1, x_2, ..., x_n)$, називають множинною регресією.

18.2. Лінійна регресія і кореляція: зміст і оцінка параметрів.

Оцінювання параметрів лінійної моделі парної регресії за допомогою методу найменших квадратів (МНК)

Розроблення регресійної моделі передбачає: 1) статистичну оцінку рівняння регресії, яке отримується в результаті точкових оцінок параметрів регресії; 2) перевірку статистичних гіпотез відносно параметрів регресії; 3) перевірку адекватності математичної моделі.

Рівняння виду $\hat{y}_x = a + b \cdot x$ дозволяє за заданим значенням фактора x обчислити теоретичні значення результативної ознаки. На графіку теоретичні значення представляють лінію регресії (рис 18.2) [30, с. 41].

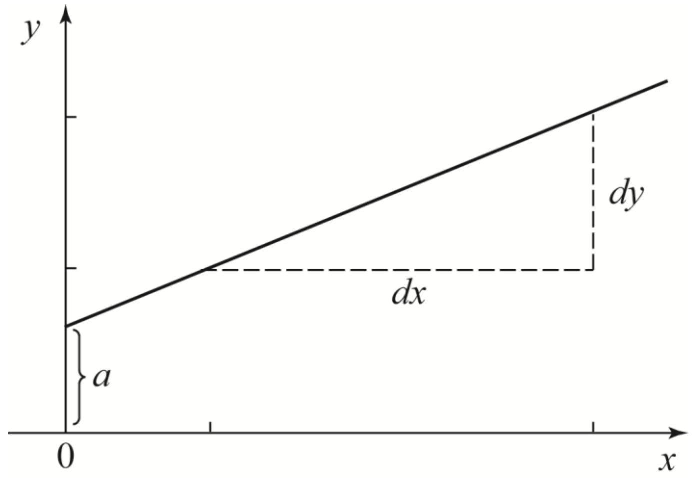


Рис. 18.2. **Графічна оцінка параметрів лінійної регресії**

Побудова лінійної регресії зводиться до оцінки її параметрів a і b. Відомі фахівці з проблем кореляційно-регресійного аналізу рекомендують постійну регресію a розглядати як коефіцієнт за фіктивною змінною, що приймає у всіх спостереженнях значення 1. Стала a визначає точку перетину прямої регресії з віссю ординат. Оскільки відповідно до загальної інтерпретації рівняння регресії a є середнім значенням y у точці $x = 0$, звідси випливає, що економічна інтерпретація a часто дуже складна або взагалі неможлива [29, с. 56 – 57]. Стала a виконує в рівнянні регресії функцію вирівнювання та, завдяки їй, функція регресії непомилкова. Оскільки рівняння регресії інтерпретоване тільки в області скупчення точок, а отже, тільки між найменшим і найбільшим значеннями змінної, які спостерігаються x, то сталу a не обов'язково інтерпретувати. В економічних дослідженнях у параметра a інтерпретується тільки знак: якщо $a > 0$, то відносна зміна результату відбувається повільніше, ніж зміна фактора.

Параметр b називається коефіцієнтом регресії. Він характеризує нахил прямої до осі абсцис або $b = tg\alpha$, де α – кут, який пряма регресія утворює з віссю абсцис. Коефіцієнт регресії є мірою залежності змінної y від змінної x або мірою впливу зміною x змінної на змінну y. Коефіцієнт b вказує середню величину змінення змінної y при зміненні пояснюючої змінної x на одну одиницю, при цьому знак b вказує напрямок цього змінення. Якщо $b > 0$ маємо додатну лінійну регресію, що демонструє поступальний характер зміни залежної змінної при збільшенні значень пояснюючої змінної x. Якщо $b < 0$, то маємо від'ємну лінійну регресію, при якій із збільшенням значень x значення змінної y зменшуються. Стала a має ту ж розмірність, що й змінна y, а розмірність коефіцієнта регресії b становить відношення розмірності залежної змінної до розмірності пояснюючої змінної.

Класичний підхід до оцінювання параметрів лінійної регресії заснований на ***методі найменших квадратів*** (МНК), що дозволяє отримати такі оцінки параметрів, за яких сума квадратів відхилень фактичних значень результативної ознаки y від теоретичних $\hat{y}_x$ мінімальна:

$$\sum_{i=1}^{n}\left(y_i - \hat{y}_{x_i}\right)^2 \to \min, \varepsilon_i = y_i - \hat{y}_{x_i}, \quad \sum_{i=1}^{n}\varepsilon_i^2 \to \min .$$

Позначимо $S = \sum_{i=1}^{n}\left(y_i - \hat{y}_{x_i}\right)^2 = \sum_{i=1}^{n}\left(y_i - a - b \cdot x_i\right)^2$. Щоб знайти мінімум даної функції, треба обчислити частині похідні по кожному з параметрів a і b та прирівняти їх до нуля.

$$\frac{dS}{da} = -2\sum_{i=1}^{n} y + 2n \cdot a + 2 \cdot b\sum_{i=1}^{n} x = 0;$$

$$\frac{dS}{db} = -2\sum_{i=1}^{n} y \cdot x + 2 \cdot a\sum_{i=1}^{n} x + 2 \cdot b\sum_{i=1}^{n} x^2 = 0.$$

Отримаємо таку систему нормальних рівнянь для оцінки параметрів a і b:

$$\begin{cases} n \cdot a + b\sum_{i=1}^{n} x = \sum_{i=1}^{n} y, \\ a\sum_{i=1}^{n} x + b\sum_{i=1}^{n} x^2 = \sum_{i=1}^{n} x \cdot y. \end{cases}$$

Методом виключення знаходимо:

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \; a = \frac{\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n} x_i}{n}.$$

Або

$$a=\bar{y}-b\cdot\bar{x},\ b=\frac{\overline{yx}-\bar{y}\,\bar{x}}{\overline{x^2}-\bar{x}^2},$$

де $\bar{y}=\dfrac{\sum\limits_{i=1}^{n}y_i}{n}$ і $\bar{x}=\dfrac{\sum\limits_{i=1}^{n}x_i}{n}$, $\overline{yx}=\dfrac{\sum\limits_{i=1}^{n}y_i x_i}{n}$.

Тісноту зв'язку досліджуваних явищ оцінює лінійний коефіцієнт парної кореляції:

$$r_{xy}=\frac{n\sum\limits_{i=1}^{n}x_i y_i-\sum\limits_{i=1}^{n}x_i\sum\limits_{i=1}^{n}y_i}{\sqrt{\left(n\sum\limits_{i=1}^{n}x_i^2-\left(\sum\limits_{i=1}^{n}x_i\right)^2\right)\left(n\sum\limits_{i=1}^{n}y_i^2-\left(\sum\limits_{i=1}^{n}y_i\right)^2\right)}}.$$

Відомо, що $-1\le r_{xy}\le 1$. Якщо $r_{xy}=0$, то відсутній лінійний кореляційний взаємозв’язок між результативною ознакою та фактором. Чим ближче значення коефіцієнта кореляції за модулем до 1, тим тісніший лінійний кореляційний взаємозв’язок між результативною ознакою та фактором. На рис. 18.3 представлені можливі ситуації при різних значеннях r_{xy}.

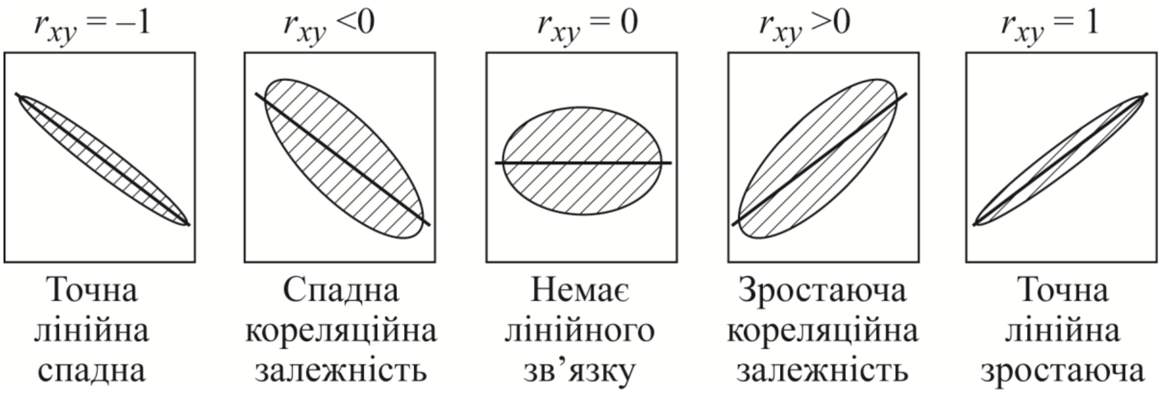


Рис.18.2. **Різні випадки тісноти зв’язку**

Існує зв'язок між коефіцієнтом кореляції і коефіцієнтом регресії:

$$r_{xy}=b\frac{\sigma_x}{\sigma_y}=\frac{\overline{yx}-\bar{y}\,\bar{x}}{\sigma_x\sigma_y}.$$

Рівняння парної лінійної регресії зручно записати в стандартизованих змінних:

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

або при зміні напрямку причинно-наслідкових зв'язків, отримаємо рівняння спряженої регресії:

$$\frac{x-\bar{x}}{\sigma_x} = r_{xy}\frac{y-\bar{y}}{\sigma_y}.$$

Для оцінки якості підбору парної лінійної функції використовують квадрат лінійного коефіцієнта кореляції, який називається коефіцієнтом детермінації, тобто $R^2 = r^2$.

Приклад. Вивчається залежність собівартості одиниці виробу (y, тис. грн) від величини випуску продукції (x, тис. шт.) за групами підприємств за звітний період. Економіст обстежив $n = 5$ підприємств і отримав результати, що є значеннями змінних x та y (табл. 18.1).

Таблиця 18.1

Номер підприємства	1	2	3	4	5
x	2	3	4	5	6
y	1,9	1,7	1,8	1,6	1,4

1. Вважаючи, що між змінними x, y має місце лінійна залежність, визначити вибіркове рівняння лінійної регресії. Необхідні обчислення виконаємо в таблиці (табл. 18.2).

Таблиця 18.2

Номер підприємства	x	y	x^2	xy
1	2	1,9	4	3,8
2	3	1,7	9	5,1
3	4	1,8	16	7,2
4	5	1,6	25	8
5	6	1,4	36	8,4
Сума	20	8,4	90	32,5

За формулами знайдемо параметри рівняння парної лінійної регресії:

$$b=\frac{n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2}=\frac{5\cdot 32{,}5-20\cdot 8{,}4}{5\cdot 90-20^2}=-0{,}11;$$

$$a=\frac{\sum_{i=1}^{n}y_i-b\sum_{i=1}^{n}x_i}{n}=\frac{8{,}4\text{ - }(\text{-}0{,}11)\cdot 20}{5}=2{,}12\,.$$

Отже, лінійна залежність собівартості одиниці виробу (y, тис. грн) від величини випуску продукції (x, тис. шт.) за групами підприємств за звітний період описується рівнянням $y=a+bx=2{,}12+(-0{,}11)x$.

2. Знайти залишки ε, коефіцієнт кореляції Пірсона і коефіцієнт детермінації. Необхідні обчислення виконаємо в таблиці (табл. 18.3).

Таблиця 18.3

Номер	x	y	y^2	$\hat{y}=2{,}12+(-0{,}11)x$	$\varepsilon=y-\hat{y}$
1	2	1,9	3,61	1,90	0,00
2	3	1,7	2,89	1,79	-0,09
3	4	1,8	3,24	1,68	0,12
4	5	1,6	2,56	1,57	0,03
5	6	1,4	1,96	1,46	-0,06
Сума	20	8,4	14,26		

$$r=\frac{n\sum_{i=1}^{n}x_iy_i-\sum_{i=1}^{n}x_i\sum_{i=1}^{n}y_i}{\sqrt{\left(n\sum_{i=1}^{n}x_i^2-\left(\sum_{i=1}^{n}x_i\right)^2\right)\left(n\sum_{i=1}^{n}y_i^2-\left(\sum_{i=1}^{n}y_i\right)^2\right)}}=\frac{5\cdot 32{,}5-20\cdot 8{,}4}{\sqrt{\left(5\cdot 90-20^2\right)\left(5\cdot 14{,}26-8{,}4^2\right)}}=-0{,}904$$

Значення коефіцієнта кореляції близьке до -1, що свідчить про дуже сильний від'ємний зв'язок (зі зростанням значень x значення y зменшуються). Знаки $b=-0{,}11$ і $r=-0{,}904$ збігаються. Коефіцієнт детермінації дорівнює $r^2=(-0{,}904)^2=0{,}817$, тобто 81,7% загальної варіації собівартості y залежить від випуску продукції x. Побудована модель не пояснює 18,3 % варіації собівартості. Ця частина варіації пояснюється факторами, не включеними у модель.

3. Виконати прогноз на основі побудованої парної лінійної моделі.

Обчислимо прогнозне значення собівартості y при випуску продукції 5,5 тис. шт.

$$y = 2{,}12 + (-0{,}11)x;$$

$$y(5{,}5) = 2{,}12 - 0{,}11 \cdot 5{,}5 = 1{,}515 \text{ тис. грн.}$$

18.3. Нелінійна регресія. Приклад обчислення оцінок параметрів квадратичної моделі методом найменших квадратів

Багато залежностей в економіці не є лінійними і тому їх моделювати лінійними рівняннями не доцільно. Наприклад, для дослідження попиту y на деякий товар від ціни x даного товару в деяких випадках можна обмежитись лінійним рівнянням регресії, але якщо необхідно визначити еластичність попиту залежно від ціни, то необхідно обчислити логарифмічну модель. В аналізі витрат y від обсягу випуску x найбільше обґрунтованою є логарифмічна модель. В розгляді виробничих функцій лінійна модель не є реальною, в даному випадку використовують степеневі моделі. Наприклад, широко відома виробнича функція Кобба−Дугласа $y = AK^{\alpha}B^{\beta}$, де y − обсяг випуску, K і B − витрати капітала і праці відповідно, A, α, β − параметри моделі. Достатньо широко використовують в сучасному економічному аналізі й інші моделі, як наприклад, обернені та експоненціальна моделі.

Регресії, нелінійні за пояснюючою змінною, можуть мати різний вигляд:

1) поліноми різних ступенів, наприклад, $y = a + b_1x + b_2x^2 + b_3x^3 + \varepsilon$;

2) рівнобічна гіпербола $y = a + \dfrac{b}{x} + \varepsilon$.

В економічних задачах часто використовують регресії, нелінійні за параметрами, що оцінюються, такого вигляду:

1) степенева $y = a \cdot x^b \cdot \varepsilon$;

2) показникова $y = a \cdot b^x \cdot \varepsilon$;

3) експоненціальна $y = e^{a + b \cdot x} \cdot \varepsilon$.

Відомо, що для успішного використання МНК бажано, щоб модель була лінійною відносно параметрів. Для двопараметричних моделей, які лінійно залежать від параметрів або можуть бути приведені до такої форми перетвореннями, існує графічний спосіб перевірки їх придатності для опису даних (перевірки адекватності моделі).

Нехай $Y = F(x, y)$ і $X = \Phi(x, y)$ – такі функціональні перетворення, після яких форма зв'язку формально набуває лінійний вигляд $Y = a + bX.$ Слід зазначити, що графіком лінійної залежності є пряма, яка виділяється серед багатьох інших кривих. Звідси випливає висновок, що якщо емпіричні точки у перетворених координатах не групуються навколо деякої прямої, то прийнята форма зв'язку повинна бути відхилена. Із застосуванням сучасних програмних продуктів графіки легко побудувати і перетворити, тому з застосуванням комп'ютера такий спосіб ідентифікації є надзвичайно ефективним.

Найчастіше застосовується або логарифмування, або перехід до зворотних величин. На рис. 18.4 наведені додаткові відомості про двопараметричні лінійні залежності, які можуть бути зведені до лінійних зазначеними функціональними перетвореннями.

З рис. 18.4 видно, що серед двопараметричних залежностей відсутні форми зв'язку з екстремумами. Це є певним обмеженням. Зазвичай у стандартних обчисленнях розглядається тільки лінійна модель або одна з розглянутих двопараметричних залежностей, але потім подальший аналіз покаже відсутність оптимумів – і не тому, що їх дійсно немає, а тому, що обрана форма зв'язку не допускає їх наявності в принципі ні за яких значень параметрів.

Якщо ж передбачається наявність оптимумів, то слід використовувати трипараметричні форми зв'язку, наприклад, квадратичну модель. Квадратична модель повинна бути прийнята як базова при аналізі даних з екстремумами. Узагальнена квадратична модель з функціональними перетвореннями здатна адекватно описати дуже широкий клас залежностей з екстремумами.

Побудоване рівняння нелінійної регресії доповнюється показником кореляції, а саме індексом кореляції (R):

$$R = \sqrt{1 - \frac{\sum (y - \hat{y}_x)^2}{\sum (y - \bar{y})^2}}, \ 0 \le R \le 1.$$

Чим ближче величина даного показника, тим тісніше зв'язок між розглянутими ознаками і тим надійніше побудоване рівняння регресії.

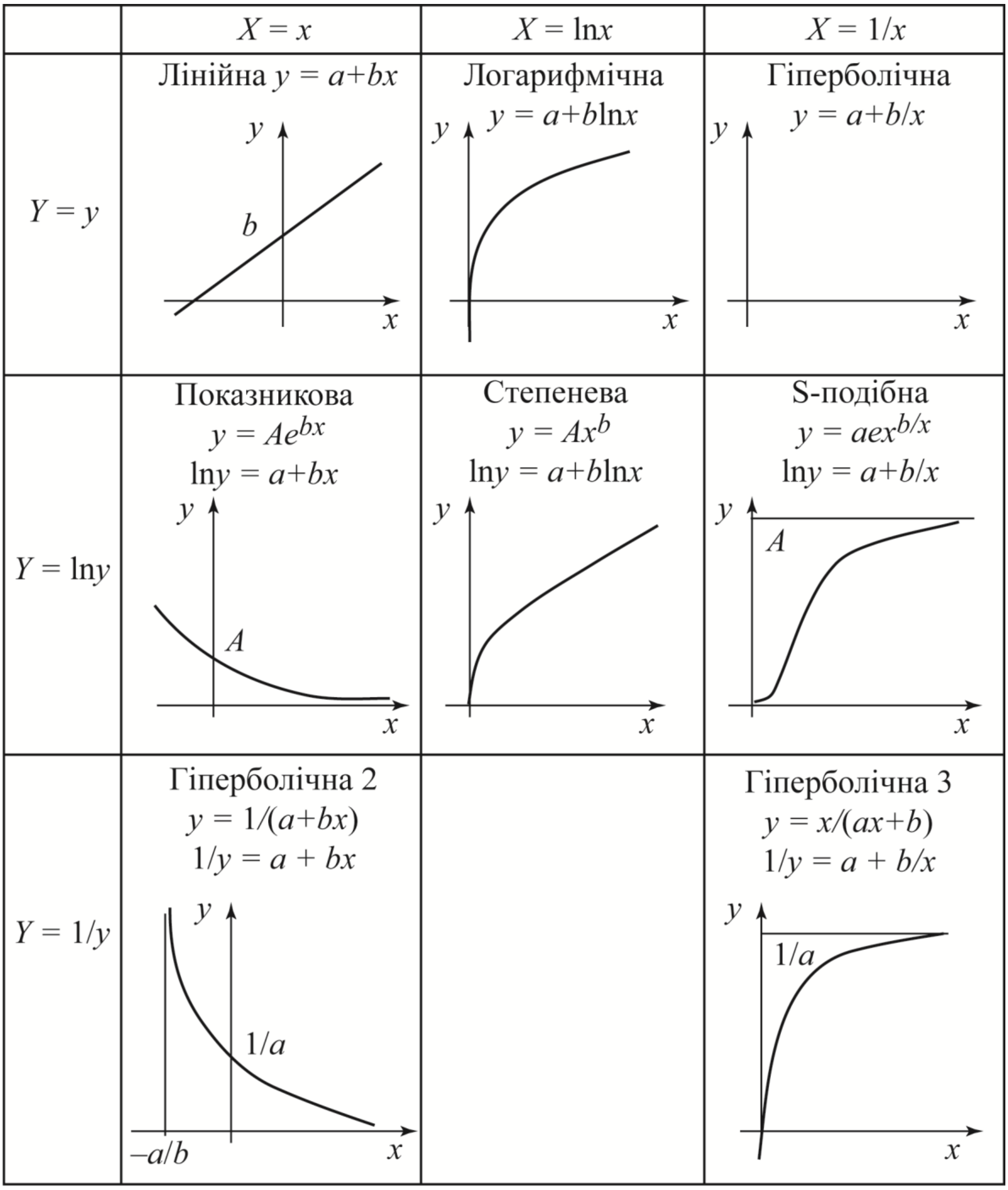


Рис. 18.4. **Двопараметричні залежності** $Y = a + bX$

Приклад обчислення оцінок параметрів квадратичної моделі методом найменших квадратів $y = a + b_1 x + b_2 x^2 + \varepsilon$**.** Дана функція формально зводиться до двофакторної лінійної моделі заміною змінних $x_1 = x, x_2 = x^2$. При цьому вияснясться, що аргументи x_1, x_2 не є незалежними змінними в загальному розумінні, вони можуть бути зв'язані між собою, при цьому визначник системи нормальних рівнянь повинен бути відмінним від нуля. Крім того, виявляється, що одній пояснюючій змінній в нелінійній моделі може відповідати не один, а відразу декілька членів, що необхідні для опису не лінійності. Так дані на рис. 18.5 (емпіричні точки) явно відхиляються від прямої та видно наявність максимума залежності, тому відповідно квадратична модель $y = a + b_1 x + b_2 x^2 + \varepsilon$ буде адекватніше описувати цю нелінійну залежність, ніж лінійна $y = a + b_1 x + \varepsilon$.

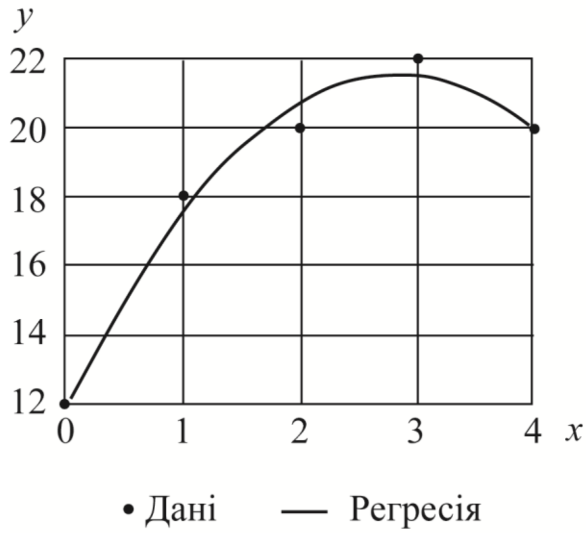


Рис.18.5. Квадратична залежність

Умова ортогональності похибок до кожного члену квадратичної моделі $\sum_{i=1}^{n} \varepsilon_i = 0, \sum_{i=1}^{n} \varepsilon_i x_i = 0, \sum_{i=1}^{n} \varepsilon_i x_i^2 = 0,$ приводить до такої системи нормальних рівнянь:

$$\begin{cases} \sum_{i=1}^{n} y_i = n \cdot a + b_1 \sum_{i=1}^{n} x_i + b_2 \sum_{i=1}^{n} x_i^2, \\ \sum_{i=1}^{n} y_i x_i = a \sum_{i=1}^{n} x_i + b_1 \sum_{i=1}^{n} x_i^2 + b_2 \sum_{i=1}^{n} x_i^3, \\ \sum_{i=1}^{n} y_i x_i^2 = a \sum_{i=1}^{n} x_i^2 + b_1 \sum_{i=1}^{n} x_i^3 + b_2 \sum_{i=1}^{n} x_i^4. \end{cases}$$

Усі необхідні суми обчислені в табл. 18.4.

Таблиця 18.4

	Дані		Обчислення сум						Обчислені значення	
№	x	y	x^2	x^3	x^4	yx	yx^2	y^2	y_p	e
1	0	12	0	0	0	0	0	144	12,114	-0,114
2	1	18	1	1	1	18	18	324	17,547	0,457
3	2	20	4	8	16	40	80	400	20,696	-0,686
4	3	22	9	27	81	66	198	484	21,543	0,457
5	4	20	16	64	256	80	320	400	20,114	-0,114
Суми	10	92	30	100	354	204	616	1 752		0

Обчислені суми підставимо в систему нормальних рівнянь:

$$\begin{cases} 92 = 5 \cdot a + b_1 10 + b_2 30, \\ 204 = 10 \cdot a + b_1 30 + b_2 100, \\ 616 = 30 \cdot a + b_1 100 + b_2 354. \end{cases}$$

Розв'язавши дану систему, маємо $a = 12{,}114, b_1 = 6{,}571, b_2 = -1{,}143$. Обчислені значення $\hat{y} = 12{,}114 + 6{,}571x - 1{,}143x^2$ наведені в тій же таблиці разом з похибками $\varepsilon = y - \hat{y}$. Графік обчисленої квадратичної залежності зображено на рис. 18.5, при цьому спостерігається добре згладжування висхідних даних.

Переконуємось, що сума всіх похибок дорівнює нулю $\sum_{i=1}^{n} \varepsilon_i = 0$. Підрахуємо суму квадратів похибок безпосередньо:

$$\sum_{i=1}^{n} \varepsilon_i^2 = (-0{,}114)^2 + (0{,}457)^2 + (-0{,}686)^2 + (0{,}457)^2 + (-0{,}114)^2 = 0{,}914\,.$$

Для перевірки обчислимо цю ж суму квадратів за формулою:

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} y_i^2 - a\sum_{i=1}^{n} y_i - b_1 \sum_{i=1}^{n} y_i x_i - b_2 \sum_{i=1}^{n} y_i x_i^2 =$$
$$= 1752 - 12{,}114 \cdot 92 - 6{,}571 \cdot 204 + 1{,}143 \cdot 616 = 1{,}116.$$

Розходження в результатах обчислень двома способами пояснюється погрішностями в обчисленні параметрів моделі з трьома десятковими знаками. При обчисленні цих параметрів з чотирма десятковими знаками, то для суми квадратів похибок отримаємо значення $\sum_{i=1}^{n} \varepsilon_i^2 = 0{,}945$, а з п'ятьма – $\sum_{i=1}^{n} \varepsilon_i^2 = 0{,}915$.

Запитання для самоперевірки

1. Яка мета кореляційно-регресійного аналізу?
2. Які типи зв'язків існують і в чому їх сутність?
3. Як розташовуються сполучені лінії регресії?
4. Які види регресії розрізняють?
5. Які основні причини наявності у регресійній моделі випадкового відхилення?
6. Як інтерпретуються параметри a і b в парній лінійної регресії?
7. У чому полягає суть методу найменших квадра тів?
8. Назвіть основні етапи регресійного аналізу.

9. Як пов'язані емпіричні коефіцієнти лінійної регресії з вибірковим коефіцієнтом кореляції між змінними рівняння регресії?

10. Які види парної нелінійної регресії розрізняють?

11. Які види залежностей можна звести функціональним перетворенням до лінійних?

19. Перевірка якості рівняння регресії

19.1. Елементи дисперсійного аналізу. Коефіцієнт детермінації. Перевірка якості побудованої парної лінійної моделі.

19.2. Оцінка статистичної значущості коефіцієнтів регресії і кореляції.

19.3. Обчислення інтервалів прогнозу за лінійною парною регресією.

19.1. Елементи дисперсійного аналізу. Коефіцієнт детермінації. Перевірка якості побудованої парної лінійної моделі

Безпосередньому розрахунку *F*-критерію для перевірки якості побудованої регресійної моделі передує дисперсійний аналіз. Дисперсійний аналіз – математичний апарат для порівняння середніх декількох популяцій (груп, шарів, класів), які визначаються рівнями деяких величин (факторів), покладених в основу класифікації. Міжнародним позначенням дисперсійного аналізу є абревіатура *ANOVA* (*analysis of variance*). У дисперсійному аналізі передбачається, що групи розрізняються тільки середнім рівнем результативної змінної, дані в кожній групі розподілені нормально з однаковою дисперсією.

З математичної статистики відомо розкладання загальної суми квадратів відхилень змінної y від середнього значення $\bar{y}$ на дві частини – «пояснену» і «непояснену»:

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

загальна сума квадратів відхилень = сума квадратів відхилень, пояснена регресією + залишкова сума квадратів відхилень або $SS_y = SS_{y_x} + SS_\varepsilon$. Цю суму квадратів розглядають як суму детермінованої і випадкової компонент.

Такий розклад на складові має число ступенів свободи:

$$df_y = df_{y_x} + df_\varepsilon,$$

де $df_y = n-1$, оскільки на n відхилень $(y_i - \bar{y})$ накладений один зв'язок − сума всіх цих відхилень дорівнює нулю $\sum_{i=1}^{n}(y_i - \bar{y}) = 0$;

$df_\varepsilon = n-1-m$, тут $m-$ число пояснюючих змінних.

Для числа ступенів свободи обчислених значень має виконуватись:

$$df_{y_x} = df_y - df_\varepsilon = (n-1)-(n-1-m) = m.$$

Розділивши кожну суму квадратів на відповідне число ступенів свободи, отримаємо дисперсії на одну ступінь свободи. Р. Фішер запропонував усі викладки дисперсійного аналізу оформляти у вигляді стандартної таблиці ($m-$ число параметрів, що оцінюють у рівнянні регресії при незалежних змінних, n − число спостережень) (табл. 19.1).

Таблиця 19.1

Компоненти дисперсії	Число ступенів свободи df	Дисперсія на одну ступінь свободи	Дисперсійне відношення F	Рівень значимості
Факторна	$df_{y_x} = m$	$S^2{}_{y_x} = \frac{\sum(\hat{y}_x - \bar{y})^2}{m}$	$F_p = \frac{S^2_{y_x}}{S^2_\varepsilon}$	α
Залишкова	$df_\varepsilon = n-m-1$	$S^2_\varepsilon = \frac{\sum(y - \hat{y}_x)^2}{n-m-1}$		
Загальна	$df_y = n-1$	$S^2_y = \frac{\sum(y-\bar{y})^2}{n-1}$		

При $F_p > F_{0,01}(df_{y_x}, df_\varepsilon)$ результативна ознака Y і фактор X не є незалежними, вони пов'язані між собою, між ними є статистична (кореляційна) залежність. При $F_p < F_{0,05}\left(df_{y_x}, df_\varepsilon\right)$ результативна ознака Y і фактор X є незалежними.

Введемо міру тісноти кореляційного зв'язку.

Розглядаючи $SS_y = SS_{y_x} + SS_\varepsilon$ як суму детермінованої і випадкової компонент, у відносних одиницях маємо: $1 = \frac{SS_{y_x}}{SS_y} + \frac{SS_\varepsilon}{SS_y}$. Відносний внесок детермінованої частини називається індексом детермінації:

$$\eta^2 = \frac{SS_{y_x}}{SS_y} = 1 - \frac{SS_\varepsilon}{SS_y}.$$

Індекс детермінації змінюється від 0 до 1 ($0 \le \eta^2 \le 1$). Якщо індекс детермінації дорівнює нулю, то дорівнює нулю сума квадратів SS_{y_x}, отже, рівні нулю всі її члени, звідки для будь-якої групи всі середні групові рівні спільному середньому, між середніми груповими немає відмінностей, результативна змінна Y не залежить від класифікаційного показника X. Якщо індекс детермінації дорівнює одиниці, то дорівнює нулю сума квадратів SS_ε, отже, рівні нулю всі її члени, тобто ніякого випадкового розкиду немає, кожній групі (кожному рівню класифікаційного показника X) відповідає єдине значення результативної змінної $\hat{y}_x$. Чим ближче індекс детермінації до одиниці, тим ближче кореляційна залежність до функціональної. Індекс детермінації η^2 показує, яка частина повної мінливості визначається класифікаційним фактором (відмінностями між групами спостережень). Для сумісності з заходами тісноти зв'язків іншої природи прийнято отримувати корінь квадратний з індексу детермінації $\eta = \sqrt{\eta^2}$. Характеристика η називається кореляційним відношенням.

Для спряжених залежностей індекси детермінації обчислюються за формулами:

$$\eta^2_{y/x} = \frac{SS_{y_x}}{SS_y}, \quad \eta^2_{x/y} = \frac{SS_{x_y}}{SS_x}.$$

Обидва кореляційні відношення перевищують абсолютну величину коефіцієнта кореляції $\eta_{y/x}, \eta_{x/y} > \left|r_{xy}\right|$. Якщо одно з кореляційних відношень перевищує інше, то це є доведенням на користь вибору відповідного напрямку причинно-наслідкових зв'язків.

На відміну від коефіцієнта детермінації при обчисленні індексу детермінації не використовуються ніякі передбачення про форму кореляційного зв'язку. Для оцінки якості підбору лінійної функції розраховується коефіцієнт детермінації ($\eta^2 = R^2$), для парної лінійної він дорівнює $R^2 = r^2$, де $r^2 = \frac{S^2_{y_x}}{S^2_y}$.

Оцінка значущості рівняння регресії в цілому виконується за допомогою *F*-критерію Фішера. При цьому висувається нульова гіпотеза (H_0), що коефіцієнт регресії дорівнює нулю, тобто $H_0 : b = 0$.

Виразимо суму квадратів $SS_{y_x} = R^2 SS_y$ і $SS_\varepsilon = (1 - R^2) SS_y$ через загальну суму квадратів SS_y і коефіцієнт детермінації R^2. Отримаємо вираз дисперсійного відношення Фішера:

$$F_p = \frac{SS_{y_x}}{SS_\varepsilon} \cdot \frac{df_\varepsilon}{df_{y_x}} = \frac{R^2}{1 - R^2} \cdot \frac{n - 1 - m}{m},$$

який потрібно порівнювати з табличними значеннями: $F_{0,01}(df_{y_x}, df_\varepsilon)$ та $F_{0,05}(df_{y_x}, df_\varepsilon)$.

Значення *F*-критерію Фішера для парної лінійної регресії розраховується $F_p = \frac{r^2}{1 - r^2} \cdot (n - 2)$. Якщо $F_p > F_{0,01}(1, n - 2)$, то H_0 відкидається і робиться висновок про суттєвості (значущості) зв'язку, що вивчається. Якщо $F_p < F_{0,05}(1, n - 2)$, то H_0 не може бути відхилена без серйозного ризику зробити неправильний висновок про наявність зв'язку і рівняння регресії вважається статистично незначущим.

Слід відмітити, що для лінійної однофакторної залежності міра тісноти зв'язку r_{xy} і характеристика її значимості F_p однакові для обох спряжених моделей.

19.2. Оцінка статистичної значущості коефіцієнтів регресії і кореляції

Рівняння регресії обчислюється на основі скінченного набору статистичних даних, що є вибіркою. Якщо переходити до іншої вибірки, то коефіцієнти рівняння регресії є випадковими величинами. Тому необхідно порівняти обчислені коефіцієнти регресії a і b з деякими теоретично очікуваними значеннями a^* і b^*. Такий аналіз здійснюється згідно зі статистичною перевіркою гіпотез: $H_0 : b = b^*$, $H_1 : b \neq b^*$, та обчислюється значення $t_p = \frac{b - b^*}{S_b}$, яке порівнюється з табличним значенням критерію Стьюдента при числі ступенів свободи $v = n - 2$ та рівнем значимості α. Якщо $\left|t_p\right| = \left|\frac{b - b^*}{S_b}\right| \geq t_{\alpha, v}$, то нульова гіпотеза, що $b = b^*$ відхиляється.

На початковому етапі розроблення регресійної моделі перевіряється гіпотеза про наявність лінійної залежності між y та x: $H_0 : b = 0$, $H_1 : b \neq 0$. Така гіпотеза ще називається гіпотезою про статистичну значущість коефіцієнта регресії. У разі відхилення нульової гіпотези коефіцієнт регресії вважається статистично значущим, що свідчить про наявність лінійної залежності між y та x.

Для перевірки гіпотези обчислюються t-значення Стьюдента: $t_b = \frac{b}{S_b}$; $t_a = \frac{a}{S_a}$ і порівнюються з табличним за числом ступенів свободи $n - 2$. Якщо $t_b > t_{\alpha = 0{,}05}(n - 2)$, то гіпотезу про неістотність коефіцієнта регресії можна відхилити. При цьому стандартні похибки S_b і S_a обчислюють за формулами:

$$S_b = \sqrt{\frac{\sum (y - \hat{y}_x)^2 / (n - 2)}{\sum (x - \bar{x})^2}} = \sqrt{\frac{S_\varepsilon^2}{\sum (x - \bar{x})^2}};$$

$$S_a = \sqrt{\frac{\sum (y - \hat{y}_x)^2}{n - 2} \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}} = \sqrt{S_\varepsilon^2 \cdot \frac{\sum x^2}{n \sum (x - \bar{x})^2}}.$$

Значимість лінійного коефіцієнта кореляції перевіряється на основі величини похибки коефіцієнта кореляції m_r:

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

Обчислене значення t – критерію Стьюдента знаходиться за формулою:

$$t_r = \frac{r}{\sqrt{1-r^2}} \cdot \sqrt{n-2}.$$

Даний вираз свідчить, що в парній лінійній регресії $t_r^2 = F$, а також $t_b^2 = F$, а значить $t_r^2 = t_b^2$. Тобто перевірка гіпотез про значущість коефіцієнтів регресії і кореляції рівносильна перевірці гіпотези про істотності лінійного рівняння регресії.

Знаючи значення стандартних похибок S_b і S_a, можна обчислити довірчі інтервали для параметрів рівняння регресії:

$$b \pm t_{\alpha,n-2} S_b; \; a \pm t_{\alpha,n-2} S_a.$$

За допомогою дисперсійного аналізу перевіряється значущість найтіснішого кореляційного зв’язку. Якщо в результаті дисперсійного аналізу виявиться, що кореляційний зв'язок не значущий, то немає сенсу проводити регресійний аналіз заданої форми, вона також буде не значимою. Такий підхід має сенс, коли характер вибіркових даних припускає їх групування пояснюючої змінної і можливість підрахунку середніх $\bar{y}_j$ для кожного j – го інтервалу групування.

Для перевірки значущості кореляційного зв’язку y/x скористаємось $SS_{y_x} = \eta^2 SS_y$ і $SS_\varepsilon = (1-\eta^2) SS_y$. Маємо вираз для дисперсійного відношення Фішера:

$$F_\eta = \frac{S_{y_x}^2}{S_\varepsilon^2} \cdot \frac{df_\varepsilon}{df_{y_x}} = \frac{\eta^2}{1-\eta^2} \cdot \frac{n-k}{k-1},$$

де k – число інтервалів групування по x.

Обчислене значення порівнюємо з табличними значеннями $F_{0,01}(df_{y_x}, df_\varepsilon)$ та $F_{0,05}(df_{y_x}, df_\varepsilon)$. Якщо $F_\eta < F_{0,05}(df_{y_x}, df_\varepsilon)$, то робимо висновок про відсутність кореляційного зв’язку.

Для перевірки адекватності моделі візьмемо до уваги, що похибки моделі називаються залишками моделі, оскільки крім випадкових похибок до них відносяться систематичні похибки вибору неправильної форми зв’язку (похибки специфікації моделі). Якщо маємо міру чисто випадкової мінливості, то залишки моделі e можна розкласти на дві компоненти − випадкову ε і систематичну Ω : $e = \Omega + \varepsilon$. Так само розкладається сума квадратів відхилень: $SS_e = SS_\Omega + SS_\varepsilon$ і число ступенів свободи: $df_e = df_\Omega + df_\varepsilon$. Враховуючи, що $SS_\Omega = \left(\eta^2 - R^2\right)SS_y$, $SS_\varepsilon = \left(1-\eta^2\right)SS_y$, $df_\Omega = k - 1 - m, df_\varepsilon = n - k$, маємо

$$F_A = \frac{S_\Omega^2}{S_\varepsilon^2} \cdot \frac{df_\varepsilon}{df_\Omega} = \frac{\eta^2 - R^2}{1-\eta^2} \cdot \frac{n-k}{k-1-m}.$$

Обчислене дисперсійне відношення Фішера F_A необхідно порівняти з табличними значеннями $F_{0,01}\left(df_\Omega, df_\varepsilon\right)$ та $F_{0,05}\left(df_\Omega, df_\varepsilon\right)$. Якщо $F_A < F_{0,05}\left(df_\Omega, df_\varepsilon\right)$, то систематичною похибкою можна знехтувати і вважати модель адекватною. Але якщо виявиться, що $F_A > F_{0,01}\left(df_\Omega, df_\varepsilon\right)$, то систематичною похибкою знехтувати не можна, слід шукати більш підходящу форму зв'язку.

19.3. Обчислення інтервалів прогнозу за лінійною парною регресією

Одним з основних завдань побудови регресійної моделі є використання її для обчислення прогнозу. Тут можливий двоїстий підхід: прогноз умовного математичного сподівання залежної змінної при встановлених значеннях пояснюючої змінної (прогноз середнього значення) або прогноз деякого конкретного значення залежної змінної (прогноз конкретного значення).

Точкове прогнозне значення $\hat{y}_x$ при $x_p = x_k$ отримуємо у ході підстановки в рівняння регресії $\hat{y}_x = a + bx$ відповідного значення x_p. Проте він вважається не реалістичним.

Прогноз результативного показника буде реалістичним, якщо обчислити довірчий інтервал для значення змінної при заданому значенні змінної x:

$$\hat{y}_x - m_{\hat{y}_x} \le y_p \le \hat{y}_x + m_{\hat{y}_x},$$

де $m_{\hat{y}_x}$ – стандартна похибка.

Виведення формули стандартної похибки $\hat{y}_x$ ґрунтується на таких викладках. В рівняння $\hat{y}_x = a + bx$ підставимо замість a вираз $a = \bar{y} - b\bar{x}$, тоді рівняння регресії набуде вигляду $\hat{y}_x = \bar{y} - b\bar{x} + bx = \bar{y} + b\left(x - \bar{x}\right)$. Звідси слідує, що стандартна похибка $m_{\hat{y}_x}$ залежить від похибки $\bar{y}$ та похибки коефіцієнта регресії S_b, тобто $m_{\hat{y}_x}^2 = m_{\bar{y}}^2 + S_b^2(x-\bar{x})^2$. Враховуючи, що $m_{\bar{y}}^2 = \frac{S_\varepsilon^2}{n}$ та $S_b^2 = \frac{S_\varepsilon^2}{\sum(x-\bar{x})^2}$ при $x_p = x_k$ отримаємо формулу обчислення стандартної похибки прогнозованого значення:

$$m_{\hat{y}_x}^2 = \frac{S_\varepsilon^2}{n} + \frac{S_\varepsilon^2}{\sum(x-\bar{x})^2}(x-\bar{x})^2 = S_\varepsilon^2\left(\frac{1}{n} + \frac{(x-\bar{x})^2}{\sum(x-\bar{x})^2}\right) \text{ або}$$

$$m_{\hat{y}_x} = S_\varepsilon\sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum(x-\bar{x})^2}}.$$

Величина стандартної похибки прогнозного середнього значення за заданого значення $x_p = x_k$ характеризує похибку положення лінії регресії. Величина стандартної похибки $m_{\hat{y}_x}$ досягає мінімуму при $\bar{x}$ і зростає в міру віддалення від нього. Обчислення прогнозного значення $\hat{y}_x$ з 95 % довірчим інтервалом виконується за формулою $\hat{y}_{x_p=x_k} \pm t_\alpha \cdot m_{\hat{y}_x}$.

На графіку довірчі межі для $\hat{y}_x$ становлять гіперболи, розташовані по обидві сторони від лінії регресії. Фактичні значення y варіюють біля середнього значення $\hat{y}_x$. Окремі значення y можуть відхилятьсь від $\hat{y}_x$ на величину випадкової похибки ε, дисперсія якої є залишковою дисперсією на одну ступінь свободи, звідси помилка прогнозного індивідуального значення y включає і стандартну похибку $m_{\hat{y}_x}$, і випадкову похибку S_ε. Отже середня похибка прогнозованого індивідуального значення $y(x_k)$ складатиме:

$$m_{y(x_k)} = S\sqrt{1 + \frac{1}{n} + \frac{(x_k-\bar{x})^2}{\sum(x-\bar{x})^2}}.$$

Щоб мати загальне судження про якість моделі з відносних відхилень по кожному спостереженню, визначають середню похибку апроксимації як середню арифметичну просту:

$$A = \frac{1}{n}\sum\left|\frac{(y-\hat{y}_x)}{y}\right| \cdot 100.$$

В аналізі моделей впливу фактора на результат корисним є розрахунок коефіцієнта еластичності: $Э = f'(x)\frac{x}{y}$, де $f'(x)$ – перша похідна, характеризує співвідношення приростів результату і фактора для відповідної форми зв'язку. Оскільки коефіцієнт еластичності для лінійної функції не є величиною постійною, а залежить від відповідного значення *x*, то розраховується середній показник еластичності за формулою: $\overline{Э} = b\frac{\bar{x}}{\bar{y}}$.

Приклад. 1. Перевірити гіпотезу про наявність лінійного зв'язку між змінними: собівартістю одиниці виробу (*y*, тис. грн) і величиною випуску продукції (*x*, тис. шт.) (продовження прикладу теми 2) у генеральній сукупності. Довірча ймовірність $p = 95\,\%$, $n = 5$.

Маємо H_0: $r_г = 0$, тобто між змінними відсутній лінійний зв'язок у генеральній сукупності H_1: $r_г \neq 0$, тобто між змінними є лінійний зв'язок у генеральній сукупності. $\alpha = 0{,}05$, було розраховано, що $r = -0{,}904$, тому

$$t = \sqrt{\frac{r^2(n-2)}{1-r^2}} = \sqrt{\frac{0{,}817(5-2)}{1-0{,}817}} = 3{,}66\,;\; t > t_{\alpha=0{,}05,\,3};\quad 3{,}66 > 3{,}1825\,.$$

Відхиляємо гіпотезу H_0 і приймаємо гіпотезу H_1 на рівні значимості 5 %. Між собівартістю одиниці виробу (*y*, тис. грн) і величиною випуску продукції (*x*, тис. шт.) є лінійний зв'язок у генеральній сукупності.

2. Знайти довірчий інтервал для коефіцієнта регресії залежності собівартості одиниці виробу (*y*, тис. грн.) і величини випуску продукції (*x*, тис. шт.) при $p = 95\,\%$, $n = 5$.

$$b \pm t_{\alpha,\,n-2} m_b\,;$$

Номер	*x*	*y*	$(x-\bar{x})^2$	$(y-\hat{y})^2$
1	2	1,9	4	0,00
2	3	1,7	1	0,0081
3	4	1,8	0	0,0144
4	5	1,6	1	0,0009
5	6	1,4	4	0,0036
Сума	20	8,4	10	0,027

$$m_b = \sqrt{\frac{\sum(y-\hat{y}_x)^2/(n-2)}{\sum(x-\bar{x})^2}} = \sqrt{\frac{S_\varepsilon^2}{\sum(x-\bar{x})^2}} = \sqrt{\frac{0{,}027/3}{10}} = 0{,}03.$$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{20}{5} = 4;\ t_{\alpha=0{,}05,\,3} = 3{,}182$$

$$b \pm t_{\alpha,\,n-2} m_b = -0{,}11 \pm 3{,}182 \cdot 0{,}03 = -0{,}11 \pm 0{,}1,\ \text{тобто}\ -0{,}21 < b < -0{,}01.$$

3. Знайти довірчий інтервал для середнього значення змінної y при заданому значенні $x_0 = 5{,}5$ тис. шт. Довірча ймовірність $p = 95\,\%$, $n = 5$.

Номер	x	x^2	$(y-\hat{y})^2$
1	2	4	0,00
2	3	9	0,0081
3	4	16	0,0144
4	5	25	0,0009
5	6	36	0,0036
Сума	20	90	0,027

$$y = a + bx_0 \pm t_{\alpha,\,n-2} S \sqrt{\frac{1}{n} + \frac{(x_0-\bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}},\ S = \sqrt{\frac{\sum_{i=1}^{n}\varepsilon_i^2}{n-2}};$$

$$\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n} = 90 - \frac{400}{5} = 10;\ S = \sqrt{\frac{0{,}027}{3}} = 0{,}0949;$$

$$(x_0-\bar{x})^2 = (5{,}5-4)^2 = 2{,}25;$$

$$y = 2{,}12 + (-0{,}11)5{,}5 \pm 3{,}182 \cdot 0{,}0949\sqrt{\frac{1}{5} + \frac{2{,}25}{10}} = 1{,}515 \pm 0{,}197.$$

Тобто довірчий інтервал для середнього значення змінної y при заданому значенні $x_0 = 5{,}5$ тис. шт. дорівнює $1{,}318 < y < 1{,}712$, тобто $(1{,}318; 1{,}712)$.

4. Знайти довірчий інтервал для індивідуального значення змінної y при заданому значенні $x_0 = 5{,}5$ тис. шт. Довірча ймовірність $p = 95\,\%$, $n = 5$.

$$a + bx_0 \pm t_{\alpha,\, n-2} S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}}\,,$$

$$2{,}12 + (-0{,}11)5{,}5 \pm 3{,}182 \cdot 0{,}0949\sqrt{1 + \frac{1}{5} + \frac{2{,}25}{10}} = 1{,}515 \pm 0{,}3605\,.$$

Тобто довірчий інтервал для індивідуального значення змінної y при заданому значенні $x_0 = 5{,}5$ тис. шт. дорівнює $1{,}1545 < y < 1{,}8755$, тобто $(1{,}1545;\, 1{,}8755)$. Довірчий інтервал для індивідуальних значень змінної y ширше довірчого інтервалу для середнього значення змінної y (при заданому значенні x_0).

Запитання для самоперевірки

1. Яка основна ідея дисперсійного аналізу?

2. Як перевірити якість побудованої парної лінійної моделі за допомогою критерію Фішера?

3. Як розраховується і що показує індекс детермінації?

4. Як оцінити значимість коефіцієнтів регресії?

5. Як оцінити значимість коефіцієнта кореляції?

6. Як побудувати довірчі інтервали для параметрів рівняння регресії

7. Як побудувати довірчий інтервал для значення змінної при заданому значенні змінної?

8. Чи можна визначити якість моделі з відносних відхилень по кожному спостереженню?

9. Яке призначення коефіцієнта еластичності в регресійному аналізі?

20. Лінійні моделі множинної регресії

20.1. Загальні питання побудови множинної регресійної моделі.

20.2. Матрична форма регресійного аналізу.

20.3. Регресійна модель в стандартизованих змінних.

20.4. Частинні рівняння регресії.

20.5. Множинна кореляція.

20.1. Загальні питання побудови множинної регресійної моделі

У практичній діяльності враховують вплив не одного чинника на результат, а кількох. Так, на рентабельність виробництва впливають і продуктивність праці, і матеріаломісткість, і фондовіддача, і оборотність оборотних коштів, і ритмічність виробництва. Тому найчастіше для аналізу або для прогнозування необхідно враховувати багато факторів, тобто необхідно розробити множинну регресійну модель

Основна мета множинної регресії – побудувати модель з великим числом факторів, визначивши при цьому вплив кожного з них окремо, а також сукупний їх вплив на результативний показник.

Побудова моделі починається з визначення її специфікації. Суть проблеми специфікації моделі передбачає вирішення відбору факторів і вибору виду рівняння регресії. Фактори повинні бути кількісними, не корелювати між собою і не бути функціонально залежними. Відбір виконується на основі теоретико-економічного аналізу.

Оскільки фактори повинні пояснювати варіацію залежної змінної, то при додатковому включенні в регресію $m+1$ фактора коефіцієнт детермінації повинен зростати, а залишкова дисперсія зменшується:

$$R_{m+1}^2 \geq R_{m+1}^2 \text{ і } S_{m+1}^2 \leq S_{m+1}^2.$$

Якщо цього не відбувається і показники практично мало відрізняються один від одного, то включений в аналіз фактор x_{m+1} не покращує модель і практично є зайвим фактором. Перенасиченість моделі зайвими факторами призводить до статистичної не значимості параметрів регресії за $t-$критерієм Стьюдента.

Відбір факторів у модель є одним з найважливіших етапів практичного використання методів регресії. Найпоширенішими методами побудови рівняння множинної регресії є: метод виключення; метод включення; покроковий регресійний аналіз. За методом виключення передбачається відбір фактора з повної їх сукупності, за методом включення − додаткове введення фактора, в покроковому регресійному аналізі − виключення раніше введеного фактору. В процедурі відсіювання факторів часто використовують матрицю частинних коефіцієнтів кореляції. При цьому слід, також, пам’ятати, що число включених у модель факторів має бути в 6 – 7 разів менше обсягу сукупності даних. Якщо це співвідношення не виконується, то число ступенів свободи залишкової варіації дуже мале, що призводить до того, що параметри рівняння регресії виявляються статистично не значимими, а $F-$критерій менший табличного значення.

Виходячи з чіткої інтерпретації параметрів, найбільш широко використовуються в економіці лінійна і степенева залежності. Рівняння лінійної множинної регресії $y = a + b_1 x_1 + b_2 x_2 + ... + b_m x_m + \varepsilon$ (рівняння «чистої» регресії). Коефіцієнти «чистої» регресії b_i характеризують середню зміну результату зі зміною відповідного фактора на одиницю при незмінному значенні інших факторів, закріплених на середньому рівні. У степеневій функції $\hat{y}_x = a \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot ... \cdot x_m^{b_m}$ коефіцієнти b_i (коефіцієнти еластичності) показують на скільки відсотків змінюється в середньому результат із зміною відповідного фактора на 1 % при незмінності дії інших факторів. Цей вид рівняння регресії отримав найбільше поширення у виробничих функціях, у дослідженнях попиту і споживання.

Для моделювання залежностей в економіці використовуються й інші рівняння множинної регресії, що можна лінеаризувати:

експоненціальна – $y = e^{a + b_1 x_1 + b_2 x_2 + ... + b_m x_m + \varepsilon}$;

гіперболічна – $y = \dfrac{1}{a + b_1 x_1 + b_2 x_2 + ... + b_m x_m}$.

Параметри рівняння множинної регресії оцінюються, як і в парній регресії, методом найменших квадратів (МНК). При його застосуванні будується система нормальних рівнянь, розв'язуючи яку отримують оцінки параметрів рівняння. Для рівняння $y = a + b_1 x_1 + b_2 x_2 + ... + b_m x_m + \varepsilon$ система нормальних рівнянь буде:

$$\begin{cases} \sum y = n \cdot a + b_1 \cdot \sum x_1 + b_2 \cdot \sum x_2 + ... + b_m \cdot \sum x_m, \\ \sum y \cdot x_1 = a \cdot \sum x_1 + b_1 \cdot \sum x_1^2 + b_2 \cdot \sum x_1 x_2 + ... + b_m \cdot \sum x_1 x_m, \\ ... \quad ... \quad ... \quad ... \quad ... \quad ... \quad ... \quad ... \\ \sum y \cdot x_m = a \cdot \sum x_m + b_1 \cdot \sum x_1 \cdot x_m + b_2 \cdot \sum x_2 x_m + ... + b_m \cdot \sum x_m^2. \end{cases}$$

Її розв'язок можна знайти за правилом Крамера: $a = \dfrac{\Delta a}{\Delta}, b_1 = \dfrac{\Delta b_1}{\Delta}, ..., b_m = \dfrac{\Delta b_m}{\Delta}$, де Δ – визначник системи, при цьому:

$$\Delta=\begin{vmatrix} n & \sum x_1 & \sum x_2 & \ldots \sum x_m \\ \sum x_1 & \sum x_1^2 & \sum x_2 x_1 & \ldots \sum x_m x_1 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 & \ldots \sum x_m x_2 \\ \ldots\ \ldots & \ldots\ \ldots & \ldots\ \ldots & \ldots\ \ldots \\ \sum x_m & \sum x_1 x_m & \sum x_2 x_m & \ldots \sum x_m^2 \end{vmatrix} \neq 0,$$

$\Delta a, \Delta b_1, \ldots, \Delta b_m$ – частинні визначники, які одержані шляхом заміни відповідного стовпця визначника системи даними лівої частини.

20.2. Матрична форма регресійного аналізу

Необхідно позначити стовпці значень змінних (економічних показників) y, x_i, ε буквами Y, X_i, E. Варто записати рівняння регресії

$$y = a + b_1 x_1 + b_2 x_2 + \ldots + b_m x_m + \varepsilon$$

для всіх спостережень змінних: $a = b_0$

$$\begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ \ldots \\ 1 \end{pmatrix} b_0 + \begin{pmatrix} x_{11} \\ x_{12} \\ \ldots \\ x_{1n} \end{pmatrix} b_1 + \begin{pmatrix} x_{21} \\ x_{22} \\ \ldots \\ x_{2n} \end{pmatrix} b_2 + \ldots + \begin{pmatrix} x_{m1} \\ x_{m2} \\ \ldots \\ x_{mn} \end{pmatrix} b_m + \begin{pmatrix} e_1 \\ e_2 \\ \ldots \\ e_n \end{pmatrix},$$

де слід позначити $X = (X_0\ X_1\ X_2 \ldots X_m) = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{m1} \\ 1 & x_{12} & x_{22} & \ldots & x_{m2} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{mn} \end{pmatrix}$,

$Y' = (y_1, y_2, \ldots, y_n)$, $X_0 = (1, 1, \ldots, 1)$, $B = (b_0, b_1, b_2, \ldots, b_m)'$, $E = (e_1, e_2, \ldots, e_n)'$.

Тоді у векторній формі рівності пишуться як:

$$Y = X_0 b_0 + X_1 b_1 + X_2 b_2 + \ldots + X_m b_m + E,$$

а в матричній формі: $Y = XB + E$. Помножимо цю рівність на матрицю X' і врахуємо, що $X_i' E = 0$, отримаємо в матричній формі систему нормальних рівнянь:

$$X' Y = X' X B.$$

Звідки знаходимо вектор коефіцієнтів регресії у вигляді:

$$B = (X' X)^{-1} X' Y = CX' Y, \ B = CX' Y,$$

де $C = \left(X'X\right)^{-1}$ – матриця, обернена до матриці системи нормальних рівнянь $X'X$, тобто $C\,(X'X) = I$ (одинична матриця).

20.3. Регресійна модель в стандартизованих змінних

Для визначення рейтингу впливу факторів у моделі обчислюють регресійну модель в стандартизованих змінних. Усі формули регресійного аналізу в стандартизованих змінних набувають набагато простішого вигляду. Позначимо $t_y = \dfrac{y - \bar{y}}{s_y}$, $t_{x_i} = \dfrac{x_i - \bar{x}_i}{s_{x_i}}$.

Тоді $y - \bar{y} = b_1\left(x_1 - \bar{x}_1\right) + b_2\left(x_2 - \bar{x}_2\right) + \ldots + b_m\left(x_m - \bar{x}_m\right)$,

$$\frac{y - \bar{y}}{s_y} = \beta_1 \frac{\left(x_1 - \bar{x}_1\right)}{s_{x_1}} + \beta_2 \frac{\left(x_2 - \bar{x}_2\right)}{s_{x_2}} + \ldots + \beta_m \frac{\left(x_m - \bar{x}_m\right)}{s_{x_m}} + \frac{e}{s_y},$$

$$t_y = \beta_1 t_1 + \beta_2 t_2 + \ldots + \beta_m t_m + \varepsilon,\ \varepsilon = \frac{e}{s_y}.$$

Коефіцієнти «чистої» регресії пов'язані зі стандартизованими коефіцієнтами регресії: $\beta_i = b_i \dfrac{s_{x_i}}{s_y}$, $a = \bar{y} - b_1\bar{x}_1 - b_2\bar{x}_2 - \ldots - b_m\bar{x}_m$.

Застосовуючи МНК до рівняння множинної регресії в стандартизованому масштабі, після перетворень отримуємо систему нормальних рівнянь:

$$\begin{cases} r_{yx_1} = \beta_1 + \beta_2 r_{x_1x_2} + \ldots + \beta_m r_{x_1x_m}, \\ r_{yx_2} = \beta_1 r_{x_2x_1} + \beta_2 + \ldots + \beta_m r_{x_2x_m}, \\ \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \quad \ldots \\ r_{yx_m} = \beta_1 r_{x_mx_1} + \beta_2 r_{x_mx_2} + \ldots + \beta_m. \end{cases}$$

Стандартизовані коефіцієнти регресії показують, на скільки сигм (середньоквадратичних відхилень) зміниться в середньому результат, якщо відповідний фактор x_i зміниться на одну сигму при незмінному середньому рівні інших факторів. Оскільки змінні центровані та нормовані, стандартизовані коефіцієнти регресії β_i порівнянні між собою і тому можна ранжувати фактори за силою їх впливу на результат.

20.4. Частинні рівняння регресії

З рівняння лінійної множинної регресії $y = a + b_1 x_1 + b_2 x_2 + \ldots + b_m x_m + \varepsilon$ отримують частинні рівняння регресії результативної ознаки з фактором x при закріпленні інших факторів на середньому рівні:

$$
\begin{aligned}
y_{x_1 \cdot x_2 x_3 \ldots x_m} &= a + b_1 x_1 + b_2 \overline{x_2} + \ldots + b_m \overline{x_m} + \varepsilon, \\
y_{x_2 \cdot x_1 x_3 \ldots x_m} &= a + b_1 \overline{x_1} + b_2 x_2 + \ldots + b_m \overline{x_m} + \varepsilon, \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots, \\
y_{x_m \cdot x_1 x_2 \ldots x_{m-1}} &= a + b_1 \overline{x_1} + b_2 \overline{x_2} + \ldots + b_{m-1} \overline{x_{m-1}} + b_m x_m + \varepsilon,
\end{aligned}
$$

або

$$
\begin{aligned}
y_{x_1 \cdot x_2 x_3 \ldots x_m} &= A_1 + b_1 x_1, \\
y_{x_2 \cdot x_1 x_3 \ldots x_m} &= A_2 + b_2 x_2, \\
&\ldots\ldots\ldots\ldots\ldots, \\
y_{x_m \cdot x_1 x_2 \ldots x_{m-1}} &= A_m + b_m x_m,
\end{aligned}
$$

де

$$
\begin{aligned}
A_1 &= a + b_2 \overline{x_2} + \ldots + b_m \overline{x_m}, \\
A_2 &= a + b_1 \overline{x_1} + \ldots + b_m \overline{x_m}, \\
&\ldots\ldots\ldots\ldots\ldots\ldots, \\
A_m &= a + b_1 \overline{x_1} + b_2 \overline{x_2} + \ldots + b_{m-1} \overline{x_{m-1}}.
\end{aligned}
$$

Відомо, що на відміну від парної регресії частинні рівняння регресії характеризують ізольований вплив фактора на результат при закріпленні інших факторів на незмінному середньому рівні. Оскільки влив інших факторів приєднується до вільного члена рівняння множинної регресії, то частинні коефіцієнти еластичності можна обчислити за формулою:

$$Э_{y_{x_i}} = b_i \frac{x_i}{\hat{y}_{x_i \cdot x_1 x_3 \ldots x_{i-1} x_{i+1} \ldots x_m}}.$$

20.5. Множинна кореляція

Показник множинної кореляції характеризує тісноту зв'язку розглянутого набору факторів з досліджуваною ознакою або оцінює тісноту спільного впливу чинників на результат.

Незалежно від форми зв'язку показник множинної кореляції може бути знайдений як індекс множинної кореляції:

$$R_{yx_1 \ldots x_m} = \sqrt{1 - \frac{S_\varepsilon^2}{S_y^2}},$$

де S_y^2 – загальна дисперсія результативної ознаки; S_ε^2 – залишкова дисперсія. $0 \le R_{yx_1 \dots x_m} \le 1$, чим ближче його значення до 1, тим тісніший зв'язок результативної ознаки з усім набором досліджуваних факторів. При правильному виборі факторів у регресійний аналіз величина індексу множинної кореляції буде істотно відрізнятись від індексу кореляції парної залежності.

У разі лінійної залежності ознак формула індексу кореляції може бути представлена у вигляді:

$$R = \sqrt{\beta_{x_1} \cdot r_{yx_1} + \beta_{x_2} \cdot r_{yx_2} + \dots + \beta_{x_m} \cdot r_{yx_m}} = \sqrt{\sum \beta_{x_i} \cdot r_{yx_i}} .$$

Формула індексу множинної кореляції для лінійної регресії має назву лінійного коефіцієнта множинної кореляції або сукупного коефіцієнта кореляції. У разі лінійної залежності лінійний коефіцієнт множинної кореляції визначається через визначники парних коефіцієнтів кореляції:

$$R_{yx_1 \dots x_m} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}} ,$$

де $$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_m} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{yx_2} & r_{x_2x_1} & 1 & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_m} & r_{x_mx_1} & r_{x_2x_m} & \dots & 1 \end{vmatrix}; \quad \Delta r_{11} = \begin{vmatrix} 1 & r_{x_1x_2} & \dots & r_{x_1x_m} \\ r_{x_2x_1} & 1 & \dots & r_{x_2x_m} \\ \dots & \dots & \dots & \dots \\ r_{x_mx_1} & r_{x_2x_m} & \dots & 1 \end{vmatrix}.$$

Ранжування факторів за силою впливу на результат можна провести через $\beta-$коефіцієнти або ж за допомогою частинних коефіцієнтів кореляції (для лінійних зв'язків) і частинних індексів детермінації (за нелінійних зв'язків).

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

Порядок частинного коефіцієнта кореляції визначається кількістю чинників, вплив яких виключається. Наприклад $r_{yx_1 \cdot x_2}$, – коефіцієнт частинної кореляції першого порядку. Коефіцієнти частинної кореляції більш високих порядків можна визначити через коефіцієнти частинної кореляції більш низьких порядків за рекурентною формулою:

$$r_{yx_i\cdot x_1x_2...x_m} = \frac{r_{yx_i\cdot x_1x_2...x_{m-1}} - r_{yx_m\cdot x_1x_2...x_{m-1}} \cdot r_{x_ix_m\cdot x_1x_2...x_{m-1}}}{\sqrt{\left(1-r^2{}_{yx_m\cdot x_1x_2...x_{m-1}}\right)\cdot\left(1-r^2{}_{x_ix_m\cdot x_1x_2...x_{m-1}}\right)}}.$$

При двох факторах і $i=1$ дана формула матиме вигляд: $r_{yx_1\cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{\left(1-r^2{}_{yx_2}\right)\cdot\left(1-r^2{}_{x_1x_2}\right)}}$. Відповідно при $i=2$ і двох факторах частинний коефіцієнт кореляції y з фактором x_2 можна визначити за формулою $r_{yx_2\cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{\left(1-r^2{}_{yx_1}\right)\cdot\left(1-r^2{}_{x_1x_2}\right)}}$.

Якщо маємо рівняння регресії з трьома факторами $y = a + b_1x_1 + b_2x_2 + b_3x_3 + \varepsilon$, то частинні коефіцієнти кореляції другого порядку визначаються на основі частинних коефіцієнтів кореляції першого порядку. Наприклад, при $i=1$ частинний коефіцієнт кореляції дорівнює:

$$r_{yx_1\cdot x_2x_3} = \frac{r_{yx_1\cdot x_2} - r_{yx_3x_2} \cdot r_{x_1x_2x_3}}{\sqrt{\left(1-r^2{}_{yx_3\cdot x_2}\right)\cdot\left(1-r^2{}_{x_1x_3\cdot x_2}\right)}}.$$

Існує взаємозв'язок між β-коефіцієнтами і коефіцієнтами кореляції:

$$\begin{cases} \beta_{x_1} = \dfrac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1-r^2_{x_1x_2}}, \\ \beta_{x_2} = \dfrac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1-r^2_{x_1x_2}}. \end{cases}$$

Порівнюючи їх з рекурентними формулами обчислення частинних коефіцієнтів кореляції $r_{yx_1\cdot x_2}, r_{yx_2\cdot x_1}$ можна встановити, що:

$$r_{yx_1\cdot x_2} = \beta_{x_1}\sqrt{\frac{1-r^2{}_{x_1x_2}}{1-r^2{}_{yx_2}}},$$

$$r_{yx_2\cdot x_1} = \beta_{x_2}\sqrt{\frac{1-r^2{}_{x_1x_2}}{1-r^2{}_{yx_1}}}.$$

У економетриці частинні коефіцієнти кореляції зазвичай не мають самостійного значення. В основному їх використовують на стадії формування моделі, зокрема в процедурі відсіву факторів.

Знаючи частинні коефіцієнти кореляції можна визначити сукупний коефіцієнт кореляції:

$$R_{yx_1x_2\ldots x_m} = \sqrt{1-\left(1-r^2_{yx_1}\right)\left(1-r^2_{yx_2\cdot x_1}\right)\left(1-r^2_{yx_3\cdot x_2x_1}\right)\ldots\left(1-r^2_{yx_m\cdot x_1x_2\ldots x_{m-1}}\right)}.$$

При повній залежності результативної ознаки від факторів коефіцієнт сукупного їх впливу дорівнює одиниці.

Приклад. У процесі вивчення залежності прибутку (тис. грн) y від вироблення продукції на одного працівника (од.) x_1 та індексу цін на продукцію (%) x_2 отримані дані по 30 підприємствах (табл. 20.1).

Таблиця 20.1

Вихідні дані

Ознака	Середнє значення	Середньоквадратичне відхилення	Парний коефіцієнт кореляції
y	250	38	$r_{yx_1} = 0{,}68$
x_1	47	12	$r_{yx_2} = 0{,}63$
x_2	112	21	$r_{x_1x_2} = 0{,}42$

1. Побудувати рівняння множинної регресії в стандартизованій формі і рівняння «чистої» регресії.

2. Розрахувати лінійні коефіцієнти частинної кореляції і коефіцієнт множинної кореляції, порівняти їх з лінійними коефіцієнтами парної кореляції, пояснити відмінності між ними.

1. $\beta_{x_1} = \dfrac{r_{yx_1} - r_{yx_2}\cdot r_{x_1x_2}}{1-r^2_{x_1x_2}} = \dfrac{0{,}68-0{,}63\cdot 0{,}42}{1-0{,}42^2} = 0{,}5044$;

$$\beta_{x_2} = \frac{r_{yx_2} - r_{yx_1}\cdot r_{x_1x_2}}{1-r^2_{x_1x_2}} = \frac{0{,}63-0{,}68\cdot 0{,}42}{1-0{,}42^2} = 0{,}4182.$$

Отримаємо рівняння $t_y = 0{,}5044t_{x_1} + 0{,}4182t_{x_2}$.

Таким чином, перший фактор – вироблення продукції на одного працівника (од.) x_1 впливає більше на зміну прибутку (тис. грн) y, порівняно з індексом цін на продукцію (%) x_2.

Для побудови рівняння «чистої» регресії розрахуємо b_1 і b_2, використовуючи формули для переходу від β_i до b_i:

$$\beta_i = b_i \frac{s_{x_i}}{s_y}; \; b_i = \beta_i \frac{s_y}{s_{x_i}}.$$

$$b_1 = 0{,}5044 \frac{38}{12} = 1{,}5973; \; b_2 = 0{,}4182 \frac{38}{21} = 0{,}7567.$$

Значення a визначаємо із співвідношення

$$a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 = 250 - 1{,}5973 \cdot 47 - 0{,}7567 \cdot 21 = 159{,}0362;$$

$$\hat{y}_x = 159{,}0362 + 1{,}5973 x_1 + 0{,}7567 x_2.$$

Таким чином, зі зміною вироблення продукції на одного працівника (од.) x_1 на одиницю при незмінному значенні другого чинника, закріпленого на середньому рівні, середня зміна прибутку складе 1,5973 тис. грн, а із зміною індексу цін на продукцію на 1 % (x_2) середня зміна прибутку складе 0,7567 тис. грн.

2. Коефіцієнти частинної кореляції:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{\left(1 - r^2_{yx_2}\right) \cdot \left(1 - r^2_{x_1 x_2}\right)}} = \frac{0{,}68 - 0{,}63 \cdot 0{,}42}{\sqrt{\left(1 - 0{,}63^2\right) \cdot \left(1 - 0{,}42^2\right)}} = 0{,}5894;$$

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{\sqrt{\left(1 - r^2_{yx_1}\right) \cdot \left(1 - r^2_{x_1 x_2}\right)}} = \frac{0{,}63 - 0{,}68 \cdot 0{,}42}{\sqrt{\left(1 - 0{,}68^2\right) \cdot \left(1 - 0{,}42^2\right)}} = 0{,}5176.$$

Зважаючи на наявність міжфакторного зв'язку $r_{x_1 x_2} = 0{,}42$, спостерігаємо розбіжність між значеннями коефіцієнтами парної та частинної кореляції:

$$r_{yx_1} = 0{,}68; \qquad r_{yx_2} = 0{,}63;$$

$$r_{yx_1 \cdot x_2} = 0{,}5894; \qquad r_{yx_2 \cdot x_1} = 0{,}5176.$$

Висновки щодо тісноти і напрямку зв'язку на основі коефіцієнтів парної і частинної кореляції збігаються.

Коефіцієнт множинної кореляції дорівнює:

$$R = \sqrt{\beta_{x_1} \cdot r_{yx_1} + \beta_{x_2} \cdot r_{yx_2}} = \sqrt{0{,}5044 \cdot 0{,}68 + 0{,}4182 \cdot 0{,}63} = \sqrt{0{,}606456} = 0{,}77875.$$

Залежність y від x_1 і x_2 характеризується як тісна, в якій 60 % варіації середнього прибутку (тис. грн) визначаються варіацією виробленням продукції на одного працівника (од.) та індексом цін на продукцію (%). Інші фактори, що не включені в модель, становлять відповідно 40 % від загальної варіації y.

Запитання для самоперевірки

1. У чому суть специфікації множинної регресійної моделі?

2. Як оцінюються параметри множинної лінійної регресійної моделі?

3. Яка мета побудови регресійної моделі в стандартизованих змінних?

4. Який зв'язок між коефіцієнтами множинної лінійної «чистої» регресії і в стандартизованих змінних?

5. Як інтерпретуються β-коефіцієнти?

6. Як обчислюється індекс множинної лінійної кореляції?

7. Яке основне призначення коефіцієнтів приватної кореляції?

21. Оцінка надійності результатів множинної лінійної моделі

21.1. Перевірка значущості рівняння множинної регресії.

21.2. Перевірка статистичної значущості коефіцієнтів рівняння регресії.

21.1. Перевірка значущості рівняння множинної регресії

Значимість рівняння множинної регресії в цілому оцінюється за допомогою F-критерію Фішера. Перевіряється гіпотеза H_0 про статистичну значимість коефіцієнта детермінації ($H_0 : R^2 = 0$). Для перевірки даної гіпотези використовується F-статистика:

$$F_p = \frac{S_{y_x}^2}{S_\varepsilon^2} = \frac{R^2}{1 - R^2} \cdot \frac{n - 1 - m}{m},$$

де $S_{y_x}^2$ – факторна сума квадратів на одну ступінь свободи;

S_ε^2 – залишкова сума квадратів на одну ступінь свободи;

R^2 – коефіцієнт детермінації;

m – число параметрів при змінних (у лінійної регресії збігається з числом включених у модель факторів);

n – число спостережень.

Величина F у разі виконання передумов МНК і при справедливості H_0 має розподіл Фішера, аналогічний розподілу $F-$статистики.

Показники F і R^2 рівні або нерівні нулю одночасно. Якщо F=0, то R^2=0, і лінія регресії $y = \bar{y}$ є найкращою за МНК, і, отже, змінна y лінійно не залежить від $x_1, x_2, ..., x_m$. Для перевірки нульової гіпотези при заданому рівні значущості α за таблицями критичних точок розподілу Фішера знаходиться критичне значення $F_{кр} = F_\alpha(m, n-m-1)$. Нульова гіпотеза відхиляється, якщо $F > F_{кр}$. Це рівнозначно тому, що $R^2 > 0$, тобто R^2 статистично значущий.

21.2. Перевірка статистичної значущості коефіцієнтів рівняння регресії

Зважаючи на кореляції між факторами значущість одного і того ж фактора може бути різною залежно від послідовності його введення в модель. Мірою для оцінки включення фактора в модель служить частинний $F-$критерій Фішера. Він побудований на порівнянні приросту факторної дисперсії, обумовленого впливом додатково включеного фактора, із залишковою дисперсією на одну ступінь свободи за регресійною моделлю в цілому:

$$F_{x_1} = \frac{R^2_{yx_1x_2...x_m} - R^2_{yx_2...x_m}}{1 - R^2_{yx_1x_2...x_m}} \cdot \frac{n-m-1}{1},$$

де $R^2_{yx_1x_2...x_m}$ – коефіцієнт множинної детермінації для моделі з повним набором факторів;

$R^2_{yx_2...x_m}$ – той же показник, але без включення в модель фактора x_1.

У чисельнику $R^2_{yx_1x_2...x_m} - R^2_{yx_2...x_m}$ – приріст частки поясненої варіації *y* за рахунок додаткового включення в модель фактора x_1.

У загальному вигляді для фактора x_i частинний $F-$критерій визначається як:

$$F_{x_i} = \frac{R^2_{yx_ix_1x_2..x_i...x_m} - R^2_{yx_1,..., x_{i-1}, x_{i+1},...,x_m}}{1 - R^2_{yx_1x_2..x_i...x_m}} \cdot \frac{n-m-1}{1}.$$

У чисельнику $R^2_{yx_ix_1x_2..x_i...x_m} - R^2_{yx_1,..., x_{i-1}, x_{i+1},...,x_m}$ – приріст частки поясненої варіації *y* за рахунок додаткового включення в модель фактора x_i.

Коли в моделі два фактори маємо:

$$F_{x_1}=\frac{R^2_{yx_1x_2}-r^2_{yx_2}}{1-R^2_{yx_1x_2}}\cdot\frac{n-m-1}{1};\ F_{x_2}=\frac{R^2_{yx_1x_2}-r^2_{yx_1}}{1-R^2_{yx_1x_2}}\cdot\frac{n-m-1}{1}.$$

Частинний $F-$критерій Фішера оцінює значимість коефіцієнтів чистої регресії. Знаючи величину F_{x_i} можна визначити і $t-$критерій для коефіцієнта регресії при $i-$му факторі: $t_{b_i}=\sqrt{F_{x_i}}$. Якщо розрахункове значення t_{b_i} більше табличного, то підтверджується значимість включеного в модель фактора.

Якщо в множинній лінійній регресії більше ніж два фактори, наприклад, маємо $y=a+b_1x_1+b_2x_2+b_3x_3+\varepsilon$, то визначається послідовно $F-$критерій для рівняння з одним фактором x_1, далі $F-$критерій для додаткового включення в модель фактора x_2, тобто для переходу від однофакторного рівняння регресії до двофакторного, та $F-$критерій для додаткового включення в модель фактора x_3. Таким чином, проводиться оцінка значущості фактора x_3 після включення в модель факторів x_1 та x_2. З $t-$критерієм Стьюдента пов'язаний саме частинний $F-$критерій, послідовний же $F-$критерій корисний на етапі формування моделі.

Оцінка значущості коефіцієнтів чистої регресії за $t-$критерієм Стьюдента може бути проведена без розрахунків частинних $F-$критеріїв:

$$t_{b_i}=\frac{b_i}{S_{b_i}},$$

де b_i- коефіцієнт «чистої» регресії при факторі x_i;

$S_{b_i}-$ середньоквадратична похибка коефіцієнтів регресії b_i.

Для рівняння множинної регресії $\hat{y}=a+b_1x_1+b_2x_2+...+b_mx_m$ середньоквадратична похибка коефіцієнтів регресії b_i визначається за формулою:

$$S_{b_i}=\frac{S_y\sqrt{1-R^2_{yx_1...x_m}}}{S_{x_i}\sqrt{1-R^2_{x_ix_1...x_m}}}\cdot\frac{1}{\sqrt{n-m-1}},$$

де S_y- середньоквадратичне відхилення для результативної ознаки y;

$S_{x_i}-$ середньоквадратичне відхилення для фактора x_i;

$R^2_{yx_1\ldots x_m}-$ коефіцієнт детермінації для рівняння множинної регресії;

$R^2_{x_i x_1\ldots x_m}-$ коефіцієнт детермінації для залежності фактора x_i зі всіма іншими факторами рівняння множинної регресії;

$n-m-1-$ число ступенів свободі для залишкової суми квадратів відхилень.

Враховуючи, що

$$b_i=\frac{S_y\sqrt{R^2_{yx_1\ldots x_m}-R^2_{yx_1\ldots x_{i-1}x_{i+1}\ldots x_m}}}{S_{x_i}\sqrt{1-R^2_{x_i x_1\ldots x_m}}},$$

можна переконатись, що $t_{b_i}=\dfrac{b_i}{S_{b_i}}=\sqrt{F_{x_i}}$. На основі визначень b_i та S_{b_i} отримаємо:

$$t_{b_i}=\frac{S_y\sqrt{R^2_{yx_1\ldots x_m}-R^2_{yx_1\ldots x_{i-1}x_{i+1}\ldots x_m}}}{S_{x_i}\sqrt{1-R^2_{x_i x_1\ldots x_m}}}:\frac{S_y\sqrt{1-R^2_{yx_1\ldots x_m}}}{S_{x_i}\sqrt{1-R^2_{x_i x_1\ldots x_m}}}\cdot\frac{1}{\sqrt{n-m-1}}=$$

$$=\frac{\sqrt{R^2_{yx_1\ldots x_m}-R^2_{yx_1\ldots x_{i-1}x_{i+1}\ldots x_m}}}{\sqrt{1-R^2_{x_i x_1\ldots x_m}}}\cdot\sqrt{n-m-1}=\sqrt{F_{x_i}}.$$

Взаємозв'язок частинного коефіцієнта кореляції, частинного $F-$критерію і $t-$критерію Стьюдента для коефіцієнтів чистої регресії використовується в процедурах відбору чинників. Відбір чинників методом виключення здійснюється не тільки за частинними коефіцієнтами кореляції, виключаючи чинник з найменшим значенням частинного коефіцієнта кореляції, а й за величинами t_{b_i} і F_{x_i}. Частинний $F-$критерій широко використовується і під час побудови моделі методами включення і покрокового відбору змінних.

Приклад. (Продовжимо приклад, що в темі 20) Розрахувати загальний і частинні $F-$критерій Фішера.

$$F_p=\frac{R^2}{1-R^2}\cdot\frac{n-m-1}{1}=\frac{0{,}6065}{1-0{,}6065}\cdot\frac{30-2-1}{1}=41{,}615;\; F_{\alpha=0{,}05}(1,\,27)=4{,}21.$$

$F_p>F_{табл}$.

Отже, рівняння залежності прибутку (тис. грн) y від виробітку продукції на одного працівника (од.) x_1 та індексу цін на продукцію (%) x_2 статистично значимо.

Перевіримо значущість коефіцієнтів регресії в рівнянні множинної регресії. Частинний $F-$критерій для фактора x_1 визначимо за формулою:

$$F_{x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot \frac{n-m-1}{1} = \frac{0{,}6065 - 0{,}63^2}{1 - 0{,}6065} \cdot \frac{30-2-1}{1} = 14{,}3817 .$$

Табличне значення дорівнює $F_{\alpha=0{,}05}(1,27)$=4,21. Коефіцієнт регресії в моделі статистично значущий. Включення в модель фактора x_1 після фактора x_2 статистично виправдане – частка поясненої варіації зросла на $\left(0{,}6065 - 0{,}63^2\right)\cdot 100 = 20{,}96\,\%$.

Частинний $F-$критерій для фактора x_2 визначимо за формулою:

$$F_{x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot \frac{n-m-1}{1} = \frac{0{,}6065 - 0{,}68^2}{1 - 0{,}6065} \cdot \frac{30-2-1}{1} = 9{,}8824 ,$$

що більше табличного. Коефіцієнт регресії в моделі статистично значущий. Включення в модель фактора x_2 після фактора x_1 статистично виправдане – частка поясненої варіації зросла на $\left(0{,}6065 - 0{,}68^2\right)\cdot 100 = 14{,}41\,\%$.

Приклад. Визначити статистичну значущість коефіцієнтів лінійної множинної регресійної моделі залежності обсягу пропозиції товару від ціни товару x_1 і зарплати працівників x_2:

$$\hat{y} = 90{,}74 + 0{,}88x_1 - 7{,}32x_2 ,$$

якщо $n = 10$; $m_a = 24{,}87$; $m_{b_1} = 0{,}36$; $m_{b_2} = 1{,}91$. Довірча ймовірність 95 %.

Позначимо $H_0 : b_i = 0$, тобто змінна, що пояснює x_i, не впливає на результативну ознаку y.

$H_1 : b_i \neq 0$, тобто змінна, що пояснює x_i впливає на результативну ознаку y.

Обчислимо значення $t-$статистики:

$$t_a = \frac{a}{m_a} = \frac{90,74}{24,87} = 3,649;$$

$$t_{b_1} = \frac{b_1}{m_{b_1}} = \frac{0,88}{0,36} = 2,44;$$

$$t_{b_2} = \frac{b_2}{m_{b_2}} = -\frac{7,32}{1,91} = -3,832.$$

Оскільки граничні точки $\pm t_{\alpha,\, n-m-1} = \pm 2,365$, відкидаємо гіпотезу H_0 і приймаємо H_1 на рівні значущості 5%. Всі коефіцієнти статистично значущі, тобто x_1 і x_2 впливають на результативну ознаку y.

Запитання для самоперевірки

1. Як використовувати критерій Фішера для перевірки загальної якості рівняння регресії?

2. Які існують методи перевірки статистичної значущості коефіцієнтів рівняння регресії?

3. Чи існує зв'язок між методами перевірки статистичної значущості коефіцієнтів рівняння регресії?

22. Мультиколінеарність, її наслідки та методи усунення

22.1. Передумови методу найменших квадратів.

22.2. Суть мультиколінеарності. Наслідки мультиколінеарності.

22.3. Визначення мультиколінеарності. Методи усунення мультиколінеарності.

22.1. Передумови методу найменших квадратів

Найпоширенішим методом оцінки параметрів рівняння множинної лінійної регресії $y = a + b_1x_1 + b_2x_2 + ... + b_mx_m + \varepsilon$ є метод найменших квадратів. Властивості оцінок коефіцієнтів регресії, якість побудованої регресії істотно залежать від властивостей випадкової складової. Доведено, що для отримання за МНК найкращих результатів необхідно виконання передумов щодо випадкового відхилення.

Передумови методу найменших квадратів (умови Гаусса−Маркова):

1. Математичне сподівання випадкового відхилення ε_i дорівнює нулю для всіх спостережень $M(\varepsilon_i)=0,\ i=\overline{1,n}$.

Дана умова означає, що випадкове відхилення в середньому не впливає на залежну змінну. В кожному конкретному спостереженні вільний член може бути або додатним, або від'ємним, але він не повинен мати систематичне зміщення. При цьому виконання $M(\varepsilon_i)=0$ спонукає виконання $M(y\,|\,x=x_i)=a^*+b^*x_i$ (для парної регресії).

2. Гомоскедастичність (сталість дисперсії відхилень). Дисперсія випадкових відхилень ε_i стала: $D(\varepsilon_i)=D(\varepsilon_j)=\sigma^2$ для будь-яких спостережень i і j.

Дана умова має на увазі: незважаючи на те, що у процесі кожного конкретного спостереження випадкове відхилення може бути або великим, або меншим, не повинно бути ніякої апріорної причини, що викликатиме велику помилку (відхилення). Виконання даної передумови називається гомоскедастичністю (сталість дисперсії відхилень); невиконання цієї передумови − гетероскедастичністю (несталість дисперсії відхилень). Оскільки $D(\varepsilon_i)=M(\varepsilon_i-M(\varepsilon_i))^2=M(\varepsilon_i^2)$, то можна записати $M(\varepsilon_i^2)=\sigma^2$.

3. Відсутність автокореляції.

Випадкові відхилення ε_i і ε_j є незалежними один від одного для всіх $i\neq j$.

Виконання даної передумови передбачає, що відсутній систематичний зв'язок між будь-якими випадковими відхиленнями або, іншими словами, величина і визначений знак будь-якого випадкового відхилення не повинні бути причинами величини і знака будь-якого іншого відхилення.

Виконання даної передумови спонукає таке співвідношення:

$$\sigma_{\varepsilon_i\varepsilon_j}=\operatorname{cov}(\varepsilon_i,\varepsilon_j)=\begin{cases}0 & \text{если } i\neq j\\ \sigma^2 & \text{если } i=j\end{cases}.$$

Якщо дана умова здійснюється, то говорять про відсутність автокореляції. З урахуванням здійснення передумови 1 дане співвідношення можна записати у вигляді $M(\varepsilon_i\varepsilon_j)=0\ (i\neq j)$.

4. Випадкове відхилення має бути незалежно від змінних, що пояснюють: $\sigma_{\varepsilon_i x_j} = 0$.

Зазвичай ця умова виконується автоматично, якщо змінні, що пояснюють, не є випадковими в даній моделі.

Вираз $\sigma_{\varepsilon_i x_j} = 0$, обґрунтовується як:

$$\sigma_{\varepsilon_i x_j} = cov(\varepsilon_i , x_i) = M((\varepsilon_i - M(\varepsilon_i))(x_i - M(x_i))) =$$
$$= M(\varepsilon_i (x_i - M(x_i))) = M(\varepsilon_i \; x_i) - M(\varepsilon_i) M(x_i) = M(\varepsilon_i \; x_i) = 0.$$

5. Модель є лінійною щодо параметрів.

Для випадку множинної лінійної регресії істотними є ще дві передумови.

6. Відсутність мультиколінеарності.

Між пояснюючими змінними відсутня строга (сильна) лінійна залежність.

7. Помилки $\varepsilon_i , i = \overline{1,n}$ мають нормальний розподіл ($\varepsilon_i \sim N(0,\sigma)$).

Здійсненність даної передумови важлива для перевірки статистичних гіпотез і побудови інтервальних оцінок.

Теорема Гаусса–Маркова. Якщо передумови 1 – 5 виконані, то оцінки отримані за МНК, мають наступні властивості:

1. Оцінки є незміщеними, тобто $M(a \;) = a^*$, $M(b_i) = b_i^*$. Це витікає з того, що $M(\varepsilon_i) = 0$ та свідчить про відсутність систематичної похибки у визначенні положення лінії регресії.

2. Оцінки спроможні, оскільки дисперсія оцінок параметрів при зростанні числа n спостережень прямує до нуля: $D(a) \xrightarrow[n\to\infty]{} 0$, $D(b_i) \xrightarrow[n\to\infty]{} 0$. Іншими словами, у процесі збільшення обсягу вибірки надійність оцінок збільшується.

3. Оцінки ефективні, тобто вони мають найменшу дисперсію порівняно з будь-якими іншими оцінками даних параметрів, лінійних відносно величин y_i.

Таким чином, у ході виконання передумов МНК щодо помилок ε_i оцінки $b_0, b_1, ..., b_m$ параметрів множинної лінійної регресії за МНК є незміщеними, спроможними та ефективними, тобто *BLUE*-оцінками (*Best Linear Unbiased Estimators*) – найкращими лінійними незміщеними оцінками.

Якщо друга і третя передумови порушені, тобто дисперсія відхилень не стала і значення ε_i, ε_j зв'язані між собою, то властивості незміщеності і спроможності зберігаються, а властивість ефективності не зберігається.

22.2. Суть мультиколінеарності. Наслідки мультиколінеарності

Серйозною проблемою у процесі побудови моделей множинної лінійної регресії за МНК є мультиколінеарність – лінійний взаємозв'язок двох або декількох пояснюючих змінних. Мультиколінеарність може бути проблемою лише у разі множинної регресії.

Причиною виникнення мультиколінеарності в економічних дослідженнях є існування співвідношень між пояснюючими змінними. Це стосується регресії, побудованої як на результатах одночасних обстежень, так і за даними, отриманими з часових рядів. Наприклад, у ході дослідження залежності витрат від обсягу виробництва і введених у дію основних фондів необхідно врахувати, що обсяг виробництва залежить також від основних фондів. Тому коефіцієнти регресії не будуть точно відображати залежність від даних двох факторів, тому що основні фонди роблять також вплив на витрати через обсяг виробництва.

Система нормальних рівнянь має розв'язок, якщо визначник матриці $X'X$ відмінний від нуля і матриця $X'X$ не вироджена. Чим сильніша кореляція між факторами, тим менше визначник матриці $X'X$. Це призводить до значного зниження точності оцінки параметрів регресії, спотворення оцінок дисперсії залишків, дисперсії коефіцієнтів регресії і коваріації між ними.

Зазвичай виділяються такі наслідки мультиколінеарності [2, с. 271 – 281]:

1) великі дисперсії (стандартні помилки) оцінок. Це ускладнює знаходження істинних значень визначених величин і розширює інтервальні оцінки, погіршуючи їх точність;

2) зменшуються t-статистики коефіцієнтів, що може призвести до невиправданого висновку про суттєвості впливу відповідної пояснюючої змінної на залежну змінну;

3) оцінки коефіцієнтів з МНК і їх стандартні помилки стають дуже чутливими до найменших змін даних, тобто вони стають нестійкими;

4) ускладнюється визначення внеску кожної зі змінних, що пояснюють дисперсію, що пояснюється рівнянням регресії залежної змінної;

5) можливе отримання неправильного знака у коефіцієнта регресії.

22.3. Визначення мультиколінеарності. Методи усунення мультиколінеарності

Існує кілька ознак, за якими може бути встановлено наявність мультиколінеарності:

1) коефіцієнт детермінації R^2 досить високий, але деякі з коефіцієнтів регресії статистично незначущі, тобто вони мають низькі $t-$статистики;

2) парна кореляція між малозначущими пояснюючими змінними досить висока. За великої їх кількості доцільнішим є використання частинних коефіцієнтів кореляції;

3) високі частинні коефіцієнти кореляції.

Окремі коефіцієнти кореляції визначають силу лінійної залежності між двома змінними без урахування впливу на них інших змінних. Однак під час вивчення багатовимірних зв'язків у ряді випадків парні коефіцієнти кореляції можуть давати зовсім неправильні уявлення про характер зв'язку між двома змінними. Це пояснюється тим, що обидві змінні змінюються в одному напрямку під впливом інших змінних, як врахованих у моделі, так і неврахованих. Тому необхідно вимірювати дійсну силу лінійного зв'язку між двома змінними, очищену від впливу на розглянуту пару змінних інших факторів. Отже, для обгрунтованішого висновку про кореляції між парами пояснюючих змінних необхідно обчислювати частинні коефіцієнти кореляції;

4) сильна допоміжна (додаткова регресія), тобто будь-яка з пояснюючих змінних є лінійною комбінацією інших пояснюючих змінних.

У цьому випадку для аналізу будуються рівняння регресії кожної з пояснюючих змінних від тих, що залишились. Обчислюються відповідні коефіцієнти детермінації R_j^2 і розраховується їх статистична значимість за критерієм Фішера:

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

Якщо коефіцієнт R_j^2 статистично незначущий, то відповідна змінна не є лінійною комбінацією інших змінних і її можна залишити в рівнянні регресії. В іншому випадку є підстави вважати, що ця змінна істотно залежить від інших пояснюючих змінних і має місце мультиколінеарність.

Єдиного методу усунення мультиколінеарності не існує, але є рекомендації, які часто використовуються [30, с. 214 – 222]:

1. Вилучення змінної з моделі. Цей метод полягає в тому, що пояснюючі змінні, які високо корелюють, виключаються з регресії, і вона заново оцінюється. Практикою доведено, що якщо $|r_{ij}| > 0{,}8$, то одну зі змінних слід виключити. Яку саме змінну необхідно виключити визначають на підставі економічного аналізу залежної змінної, оскільки можна допустити помилку специфікації. Наприклад, у ході дослідження попиту на деякий товар пояснюючі змінні можна взяти ціну даного товару та ціну його замінника, які часто корелюються один з одним. Виключивши з моделі ціну замінника, скоріше допустимо помилку специфікації. Внаслідок цього можна отримати зміщені оцінки і висновки будуть необґрунтованими.

2. Отримання додаткових даних або нової вибірки.

Мультиколінеарність прямо залежить від вибірки, тому у процесі переходу до іншої вибірки мультиколінеарность буде відсутня, або вона не буде великою. В такому випадку для зменшення мультиколінеарності достатньо збільшити обсяг вибірки. Але й тут можуть виникнути проблеми, пов'язані з отриманням нової вибірки, або розширенням старої.

3. Зміна специфікації моделі, тобто змінюється або форма моделі, або добавляються пояснюючі змінні, не враховані в первісній моделі, але такі, що істотно впливають на залежну змінну.

4. Використання попередньої інформації про деякі параметри. Іноді на основі раніше побудованих регресійних рівнянь або проведених економічних досліджень сформовано вже уявлення про величину або співвідношення двох або декількох коефіцієнтів регресії. Тут можуть виникнути труднощі, обумовлені способом отримання попередньої інформації та малою ймовірністю того, що виділений коефіцієнт регресії буде одним і тим же у різних моделей.

5. Перетворення змінних. Часто у ході розв'язування економічних задач для усунення проблеми мультиколінеарності використовується метод перетворення змінних. Наприклад, нехай емпіричне рівняння регресії має вигляд: $\hat{y} = b_0 + b_1 x_1 + b_2 x_2$, причому змінні x_1, x_2 корелюють між собою, можна визначити регресійні залежності відносних величин:

$$\frac{\hat{y}}{x_1} = b_0 + b_1 \frac{x_2}{x_1} \text{ або } \frac{\hat{y}}{x_2} = b_0 + b_1 \frac{x_1}{x_2}.$$

Цілком імовірно, що в цих моделях мультиколінеарність буде відсутня. Можливі й інші перетворення, що схожі до наведених.

6. Вирішенням проблеми мультиколінеарності може стати коригування самого математичного методу оцінки параметрів регресійної моделі. Так було запропоновано (Гоерл) перед оберненням матриці $X'X$ додавати до її діагональних елементів малі числа k. Це попереджає виродження матриці, збільшує її визначник і зменшує дисперсію параметрів. Нові оцінки (ридж-оцінки) параметрів обчислюються за формулою:

$$b = \left(X'X + kI\right)^{-1} X'Y.$$

Від такої операції (збільшення діагональних елементів матриці $X'X$ на мале число k) сума квадратів помилок збільшується, але типовий характер залежності залишкової дисперсії від k подібний графіку на рис. 22.1 − до якоїсь межі $k < k^*$ залишкова дисперсія зростає незначно.

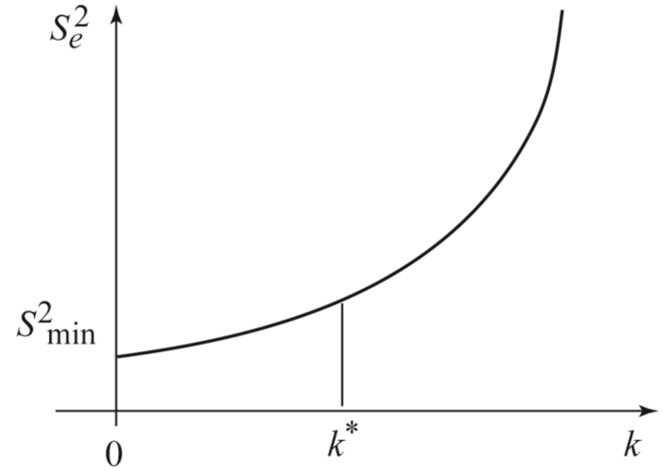


Рис. 22.1. **Зростання залишкової дисперсії s_e^2**
при збільшенні діагональних елементів матриці $X'X$ на k

Стійкість оцінок параметрів характеризується сумою їх дисперсій:

$$L(b) = \sum_{i=1}^{m} s_{b_i}^2.$$

Цю формулу можна перетворити, якщо знайти всі власні числа λ_i і власні вектори U_i матриці $X'X = UD_\lambda U'$. Тут через D_λ позначена діагональна матриця власних чисел, а через U – ортогональна матриця власних векторів.

Обчислюємо $U'b = g$. Тепер сума дисперсій параметрів моделі може бути представлена у вигляді:

$$L(b) = \sum_{i=1}^{m} \frac{\lambda_i s_e^2(k) + k^2 g_i}{(\lambda_i + k)^2}.$$

При мультиколінеарності (виродженні матриці $X'X$) одне або кілька власних чисел λ_i близькі до нуля і при $k = 0$ виходить дуже велике (якщо не нескінченне) значення $L(b)$. Введення навіть дуже малого k відразу виправляє ситуацію.

Існує кілька рекомендацій з вибору величини k; при комп'ютерній реалізації обрано найбільш радикальний шлях – обчислюється ряд оцінок за різних значень параметра $0 < k < 0{,}1$ і будуються графіки «сліду» залежностей коефіцієнтів регресії і залишкової дисперсії від цього параметра. Якщо є мультиколінеарність, то при $k \to 0$ починається різке зростання деяких (якщо не всіх) коефіцієнтів регресії. Необхідно вибрати таке найменше значення параметра k, за якого поведінка всіх коефіцієнтів регресії вже стабілізувалася, але ще не почалося інтенсивне зростання залишкової дисперсії.

Для порівняння різних альтернативних схем регресійного аналізу фахівцями було зроблено численні перевірки на даних з різними особливостями. Перше місце у всіх таких перевірках зайняли ридж-оцінки. МНК-оцінки у всіх перевірках виявлялися гіршими за всіма критеріями якості. У зв'язку з цим є рекомендації: якщо мультиколінеарності немає, то ридж-оцінки близькі до МНК-оцінок, а в разі мультиколінеарності вони значно краще, причому перевага ридж-оцінок збільшується з ростом числа змінних m.

7. Метод головних компонент. Результатом застосування методу головних компонент на вихідній системі пояснюючих змінних є нові змінні, які є лінійною комбінацією вихідних. Висхідну стандартизовану систему пояснюючих змінних X_i перетворюють некорельованою системою компонент F_i, що має властивості: перша компонента F_1 пояснює максимум загальної мінливості даних, друга компонента F_2 – максимум залишку мінливості і т. д. Кілька перших (головних) компонент пояснюють майже всю мінливість даних, а інші компоненти відображають випадкові помилки і можуть бути відкинуті. У моделі з головними компонентами виконують зворотний перехід до вихідних

змінних X_i, тим самим обходять проблему мультиколінеарності. Компоненти складаються як лінійні комбінації:

$$F_i = u_1 X_1 + u_2 X_2 + ... + u_m X_m,$$

коефіцієнти якої є елементами власних векторів $U_i = (u_1, u_2, ..., u_m)'$ кореляційної матриці $R = (X'X)/n$; дисперсії компонент рівні відповідним власним числам λ_i кореляційної матриці.

Приклад. Побудувати модель залежності трудомісткості обробки деталі (y) від кількості операцій (x_1), кількості оброблених поверхонь (x_2), кількості деталей (x_3) за даними, наведеними в табл. 22.1.

Таблиця 22.1

Вихідні і проміжні дані

№	x_1	x_2	x_3	y	y_p	e
1	2	3	4	5	6	7
1	239	449	52	72,4	63,54	8,86
2	240	450	50	72,7	60,33	12,37
3	251	465	53	73,4	63,29	10,11
4	256	470	65	75,5	73,27	2,23
5	258	475	67	78,5	80,48	-1,98
6	263	478	74	88,0	80,18	7,82
7	264	479	80	91,5	86,44	5,06
8	267	482	85	93,0	89,82	3,18
9	270	485	89	95,5	92,02	3,48
10	272	487	97	98,8	99,80	-1,00
11	274	492	99	105,8	107,01	-1,21
12	275	493	108	107,2	116,82	-9,62
13	276	494	111	107,9	119,53	-11,63
14	278	498	117	114,8	129,30	-14,50
15	283	502	128	122,7	135,91	-13,21
16	286	504	134	134,6	138,30	-3,70
17	290	509	137	147,5	140,64	6,86
18	291	512	144	150,3	152,45	-2,15
19	293	520	153	170,7	174,50	-3,80
20	295	525	168	192,1	197,11	-5,01
21	298	528	172	194,6	199,30	-4,70
22	300	532	173	198,8	203,15	-4,35

23	302	537	175	202,7	210,37	-7,67
24	304	539	178	206,1	212,23	-6,13
25	305	541	181	208,8	217,12	-8,32
26	308	546	185	216,6	223,67	-7,07
27	311	549	189	220,7	225,87	-5,17
28	318	554	193	229,0	220,32	8,68

Закінчення табл. 22.1

29	319	557	205	239,0	238,05	0,95
30	322	560	215	254,6	247,35	7,25
31	325	564	224	267,9	257,64	10,26
32	327	569	231	275,0	270,78	4,22
33	328	573	238	275,4	284,77	-9,37
34	332	574	244	290,5	281,94	8,56
35	335	580	248	295,7	290,68	5,02
36	338	584	256	311,2	299,80	11,40
37	340	590	265	322,9	317,48	5,42
38	344	598	274	336,5	333,47	3,03
39	346	605	281	346,8	350,97	-4,17
Середнє	295,46	524,33	157,38	181,68	181,68	0
ДИСПР	872,92	1 803,4	4 644,4	7 163,32	7 110,01	53,31

Стандартним методом найменших квадратів отримано рівняння регресії:

$$\begin{array}{l} \hat{y} = -253{,}61 - 3{,}027x_1 + 2{,}181x_2 + 1{,}184x_3. \\ (t_b) \qquad\qquad (5{,}7) \qquad (5{,}0) \qquad (5{,}9) \end{array}$$

Ця модель значуща в цілому і має тільки значущі фактори (значення $t-$статистик Стьюдента $t_b > t_{0,05} = 2{,}0$). Коефіцієнт множинної кореляції дуже високий $R = 0{,}996$. У табл. 22.1 для кожного варіанта даних наведені розрахункові значення y_p і залишки моделі $\varepsilon = y - y_p$. В останньому рядку табл. 22.1 обчислені дисперсії всіх змінних. Коефіцієнт детермінації можна обчислити двома способами:

$$R^2 = 1 - \frac{s_\varepsilon^2}{s_y^2} = 1 - \frac{53{,}31}{7163{,}32} = 0{,}9926\,; \; RR = \frac{s_p^2}{s_y^2} = \frac{7\,110{,}01}{7\,163{,}32} = 0{,}9926.$$

Можна також перевірити, що немає кореляції між залишками і розрахунковими значеннями: $r_{\varepsilon y_p} = 0$ (згідно з МНК, залишки ортогональні до розрахункових значень). На перший погляд отримана модель повністю доброякісна. Однак, маємо від'ємний коефіцієнт регресії при x_1. Це інтерпретується так: чим більше операцій у процесі виготовлення деталі, тим менше трудомісткість, що є сумнівним. Слід побудувати цю модель залежності в стандартизованих змінних:

$$\hat{Y} = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3,$$

$$\hat{Y} = -1{,}057\,X_1 + 1{,}094\,X_2 + 0{,}953\,X_3,$$

тут β-коефіцієнти виявилися більшими одиниці (за модулем). Однією з причин такої ситуації може бути мультиколінеарність. З табл. 22.2 випливає, що між змінними, що пояснюють, присутні дуже великі кореляції. Фактично маємо не три змінні, що пояснюють, а три різні форми однієї змінної.

Таблиця 22.2

Кореляційна таблиця

	x_1	x_2	x_3	y
x_1	1	0,9967	0,9939	0,9813
x_2	0,9967	1	0,9955	0,9901
x_3	0,9939	0,9955	1	0,9925
y	0,9813	0,9901	0,9925	1

Метод покрокової регресії в даному прикладі не усунув проблему підключення в модель тісно пов'язаних груп показників. Розглянемо покроково процес складання регресійної моделі. Результативна ознака тісніше пов'язана з кількістю деталей x_3. Облік однієї змінної x_3 пояснює $(0{,}9925)^2 = 0{,}9851$ $(98{,}51\,\%)$ загальної мінливості результативної ознаки y. Трьохфакторна модель пояснює 99,26 % загальної мінливості y, тобто додатковий внесок інших факторів x_1, x_2 становить тільки 0,75 % мінливості y. Облік x_3 фактично вичерпує мінливість всіх змінних, від яких залишаються тільки малі випадкові залишки. Проте на наступному кроці в модель підключається x_1 (точніше малі залишки від цієї змінної) і вже виходить абсурдна модель: $\hat{y} = 254{,}03 - 1{,}174x_1 + 1{,}738x_3$. Далі підключається x_2. Отже, метод покрокової регресії не усуває проблему мультиколінеарності і в модель може увійти група тісно пов'язаних показників.

Використовуємо метод гребневої (ридж-) регресії, згідно з якою до діагональних елементів кореляційної матриці додається мале число (ридж-параметр k). Врахування цього малого параметра веде до невеликої систематичної помилки в оцінках коефіцієнтів регресії, однак істотно знижує їх випадкову мінливість. Оцінки параметрів стають дещо зміщеними, але більш ефективними. Відомо, що дисперсії випадкових помилок коефіцієнтів регресії пропорційні діагональним елементам оберненої матриці $C = \left(r_{x_i x_j}\right)^{-1}$. При мультиколінеарності визначник кореляційної матриці може виявитися близьким до нуля (виродження), тоді елементи оберненої матриці будуть дуже великими. Введення навіть дуже малого ридж-параметра перешкоджає надмірному зростанню елементів оберненої матриці. У табл. 22.3 наведені обернені матриці – вихідна (C) і з урахуванням параметра $k = 0{,}05$ (C_k).

Таблиця 22.3

Обернені матриці C і C_k

***k* = 0**			***k* = 0,05**		
157,15398	-127,646	-29,1128	12,279638	-6,29038	-5,659
-127,6463	215,8722	-88,0455	-6,290378	12,64669	-6,03662
-29,1128	-88,0455	117,5862	-5,659001	-6,03662	12,0323

Таким чином, введення малого параметра $k = 0{,}05$ зменшило елементи оберненої матриці більш ніж в 10 разів. При цьому множник пропорційності *(MSE)* змінився, але не істотно. При $k = 0{,}05$ модель має вигляд:

$$\hat{y} = -446{,}8 + 0{,}590x_1 + 0{,}713x_2 + 0{,}515x_3.$$
$$(t_b) \qquad\qquad (2{,}5) \qquad (4{,}8) \qquad (5{,}0)$$

Коефіцієнти детермінації R^2 і RR тепер не рівні:

$$R^2 = 0{,}9812 \text{ і } RR = 0{,}9473.$$

Між залишками і розрахунковими значеннями допущена невелика кореляція:

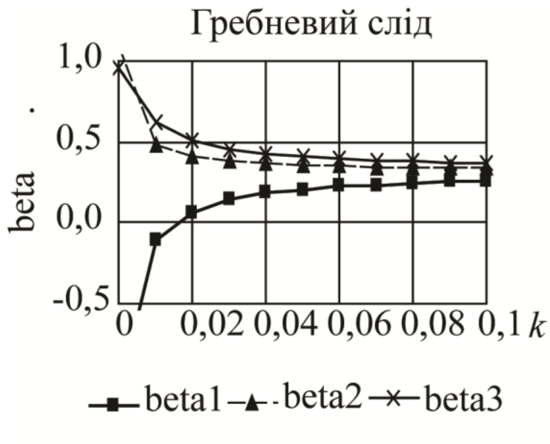
$$r_{\varepsilon y_p} = 0{,}127.$$

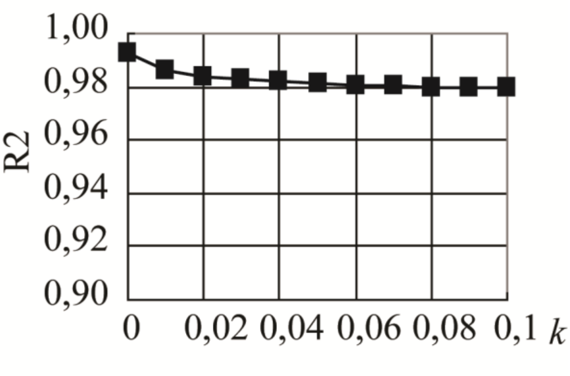
На відміну від висхідної моделі $k = 0$ тут немає непорозумінь. У табл. 22.4 наведені для різних k значення β-коефіцієнтів, а також R^2, RR і $r_{\varepsilon y_p}$. На підставі даних табл. 22.3 побудовані графіки гребневого сліду для кожної з названих характеристик (рис. 22.2).

Таблиця 22.4

Значення коефіцієнтів регресії в стандартизованих змінних

k	β_1	β_2	β_3	R^2	RR	r_{e,y_p}
0	-1,057	1,094	0,953	0,993	0,993	0,000
0,01	-0,105	0,476	0,617	0,987	0,974	0,054
0,02	0,067	0,407	0,510	0,984	0,967	0,069
0,03	0,141	0,381	0,460	0,983	0,960	0,087
0,04	0,181	0,367	0,431	0,982	0,954	0,107
0,05	0,206	0,358	0,411	0,981	0,947	0,127
0,06	0,223	0,351	0,397	0,981	0,941	0,148
0,07	0,235	0,346	0,387	0,980	0,935	0,168
0,08	0,245	0,342	0,378	0,980	0,929	0,188
0,09	0,252	0,339	0,371	0,980	0,923	0,208
0,1	0,257	0,336	0,365	0,979	0,917	0,228





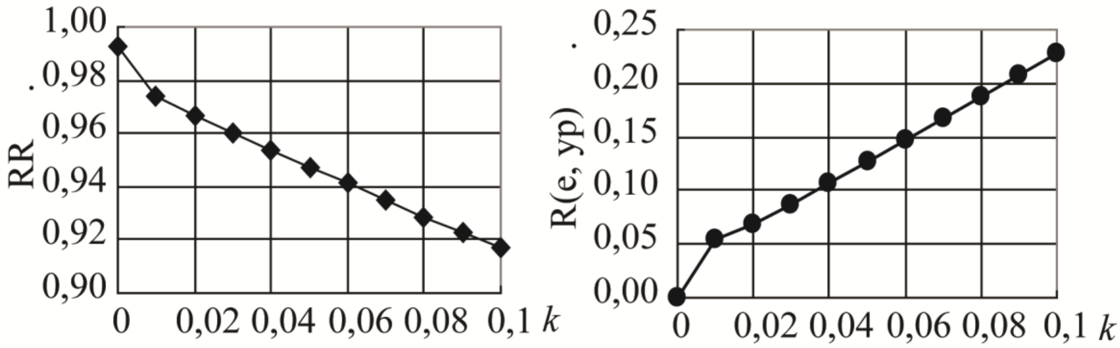


Рис. 22.2. **Графіки залежності від ридж-параметра** k

Можна спостерігати, що оцінка коефіцієнта детермінації за формулою $R^2 = 1 - \frac{s_\varepsilon^2}{s_y^2}$ із зростанням ридж-параметра k зменшується поступово, не так інтенсивно, як оцінка за формулою $RR = \frac{s_p^2}{s_y^2}$. Тому як оцінку рекомендується завжди використовувати $R^2 = 1 - \frac{s_\varepsilon^2}{s_y^2}$. Із зростанням ридж-параметра k збільшується кореляція розрахункових значень із залишками моделі (це ефект внесення систематичної помилки). Графік на рис. 22.2 демонструє, наскільки нестійкими були початкові оцінки і як істотно вони змінилися при зовсім невеликих значеннях ридж-параметра. При $k = 0{,}02$ раніше від'ємна оцінка β_1 стала додатною, а при $k = 0{,}04$ значення всіх оцінок стабілізувалися. При цьому коефіцієнт детермінації R^2 практично не змінився (він знизився від 0,996 до 0,982), між розрахунковими значеннями і залишками з'явилася невелика кореляція $r_{\varepsilon y_p} = 0{,}110$.

При $k > 5$ усі β-коефіцієнти стають рівні між собою – такий результат можна отримати, використовуючи метод головних компонент (табл. 22. 5).

Таблиця 22.5

Власні вектори і власні числа кореляційної матриці

№	U_1	U_2	U_3
1	0,577325	-0,60081	0,552918
2	0,577633	-0,17808	-0,79664

3	0,577092	0,779302	0,24424
λ	2,990734	0,006256	0,00301
%	99,7 %	0,2 %	0,1 %
Σ	99,7 %	99,9 %	100,0 %

З табл. 22.5 випливає, що внесок однієї першої компоненти становить 99,7 %, на другу і третю – припадає 0,3 %. Отже, достатньо врахувати одну першу (головну) компоненту, яка є комбінацією вихідних змінних з майже однаковими коефіцієнтами (елементи першого власного вектора U_1):

$$F_1 = 0{,}5773X_1 + 0{,}5776X_2 + 0{,}5771X_3.$$

Наприкінці одержимо рівняння регресії в стандартизованих змінних:

$$\hat{Y} = 0{,}3303X_1 + 0{,}3305X_2 + 0{,}3302X_3.$$

У однофакторній моделі $\hat{Y} = 0{,}9925X_3$ весь вклад трьох рівноправних факторів присвоєно одній змінній.

У сучасних умовах розвитку програмного забезпечення і комп'ютерів усі перераховані методи усунення мультиколінеарності можна досить легко застосувати. Для підтвердження правильності розв’язку рекомендується використовувати декілька.

Запитання для самоперевірки

1. Перерахуйте передумови МНК. Які наслідки їх можливості виконання або невиконання?

2. Якими властивостями повинні володіти оцінки, отримані за МНК?

3. Поясніть суть проблеми мультиколінеарності при побудові множинної регресійної моделі.

4. Які основні наслідки мультиколінеарності?

5. Як можна виявити мультиколінеарність?

6. Перерахуйте основні методи усунення мультиколінеарності. Поясніть їх відмінність.

23. Гетероскедастичність та методи її визначення.

Узагальнений метод найменших квадратів

23.1. Суть гетероскедастичності.

23.2. Наслідки гетероскедастичності.

23.3. Методи визначення та пом'якшення гетероскедастичності.

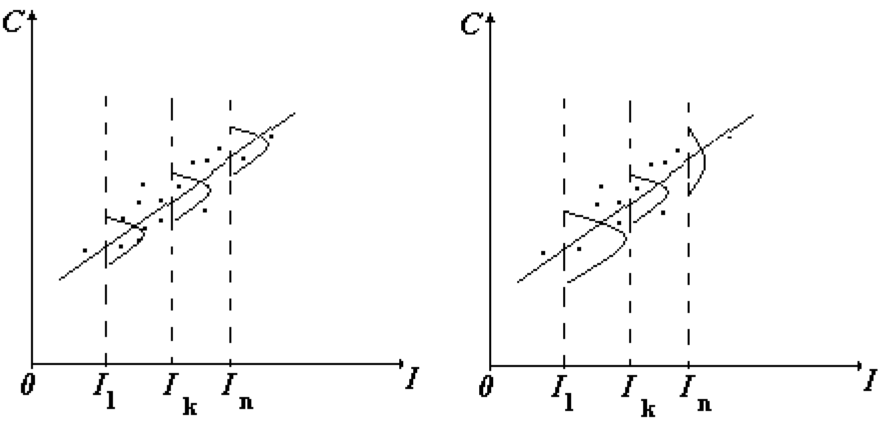
23.1. Суть гетероскедастичності

Однією з ключових передумов МНК є умова сталості дисперсій випадкових відхилень, тобто $D(\varepsilon_i) = D(\varepsilon_j) = \sigma^2$ для будь-яких спостережень i і j. Як було вже сказано, виконання даної передумови називається гомоскедастичністю, невиконання − гетероскедастичністю.

На практиці гетероскедастичність не така вже й рідкісна. Під час розгляду вибіркових даних маємо справу з конкретними реалізаціями залежної змінної y_i і відповідно з визначеними випадковими відхиленнями ε_i. Проте до здійснення вибірки ці показники апріорі могли приймати довільні значення на основі деяких ймовірнісних розподілів. Але при цьому не може бути ніякої апріорної причини, яка б обумовлювала великі відхилення в одних спостереженнях і малі − в інших.

Часто є підстава вважати, що ймовірнісні розподіли випадкових відхилень при різних спостереженнях будуть різними. Це не означає, що випадкові відхилення ε_i обов'язково будуть великими за певних спостережень і малими – при інших, але це означає, що апріорна ймовірність цього велика.

На рис. 23.1 наведено графіки, що ілюструють гомо- та гетеросекдастичність.



а) б)

Рис. 23.1. **а) гомоскедастичність залишків; б) гетероскедастичність**

залишків

В обох випадках із зростанням доходу зростає і середнє значення споживання. Але якщо на рис. 23.1а) дисперсія споживання залишається однією і тією ж для різних рівнів доходу, то на рис. 23.1б) при аналогічній залежності середнього споживання від доходу дисперсія споживання не залишається сталою, а збільшується з ростом доходу. Фактично це означає, що в другому випадку суб'єкти з великим доходом у середньому споживають більше, ніж суб'єкти з меншим доходом, і крім того, розкид в їх споживанні більш істотний для більшого рівня доходу. Люди з великим доходом мають більший простір для його розподілу. Реалістичність данії ситуації не викликає сумнівів. Проблема гетероскедастичності характерна для перехресних даних (об'єкти – час) і досить рідко зустрічається під час розгляду часових рядів.

23.2. Наслідки гетероскедастичності

У ряді невиконання даної передумови, тобто при гетероскедастичності, наслідки застосування МНК будуть такими:

1. Оцінки коефіцієнтів як і раніше залишаться незміщеними і лінійними.

2. Оцінки не будуть ефективними, не будуть навіть асимптотично ефективними. Збільшення дисперсії оцінок знижує ймовірність отримання максимально точних оцінок.

3. Дисперсії оцінок розраховуватимуться зі зміщенням. Зміщеність з'являється внаслідок того, що не пояснена рівнянням регресії дисперсія:

$$S^2 = \frac{\sum \varepsilon_i^2}{n-m-1},$$

яка використовується у процесі обчислення оцінок дисперсій всіх коефіцієнтів, не є більш незміщеною.

4. Унаслідок всі висновки, одержувані на основі відповідних $t-$ і $F-$ статистик, а також інтервальні оцінки будуть ненадійними. Отже, статистичні висновки, що отримуються у процесі стандартних перевірок якості оцінок, можуть бути помилковими і призводити до неправильних висновків щодо побудованої моделі.

23.3. Методи визначення та пом'якшення гетероскедастичності

У ряді випадків, знаючи характер даних, поява проблеми гетероскедастичності можна передбачати і спробувати усунути цей недолік ще на етапі специфікації. Однак значно частіше цю проблему доводиться вирішувати після побудови рівняння регресії. Зараз не існує будь-якого однозначного методу визначення гетероскедастичності й виявити її в кожному випадку є досить складною задачею. До теперішнього часу для такої перевірки розроблено досить велике число тестів і критеріїв для них. Найбільш популярними і наочними є: графічний аналіз відхилень, тест рангової кореляції Спірмена, тест Парку, тест Глейзера, тест Голдфельда−Квандта.

Графічний аналіз залишків

Даний метод є найпростішим і візуальне представлення відхилень дозволяє визначитися з наявністю гетероскедастичності. У цьому випадку по осі абсцис відкладаються значення (x_i) пояснюючої змінної X (або лінійної комбінації пояснюючих змінних $Y = a + b_1x_1 + ... + b_mx_m$), а по осі ординат або відхилення ε_i, або їх квадрати ε_i^2. Приклади таких графіків наведено на рис. 32.2.

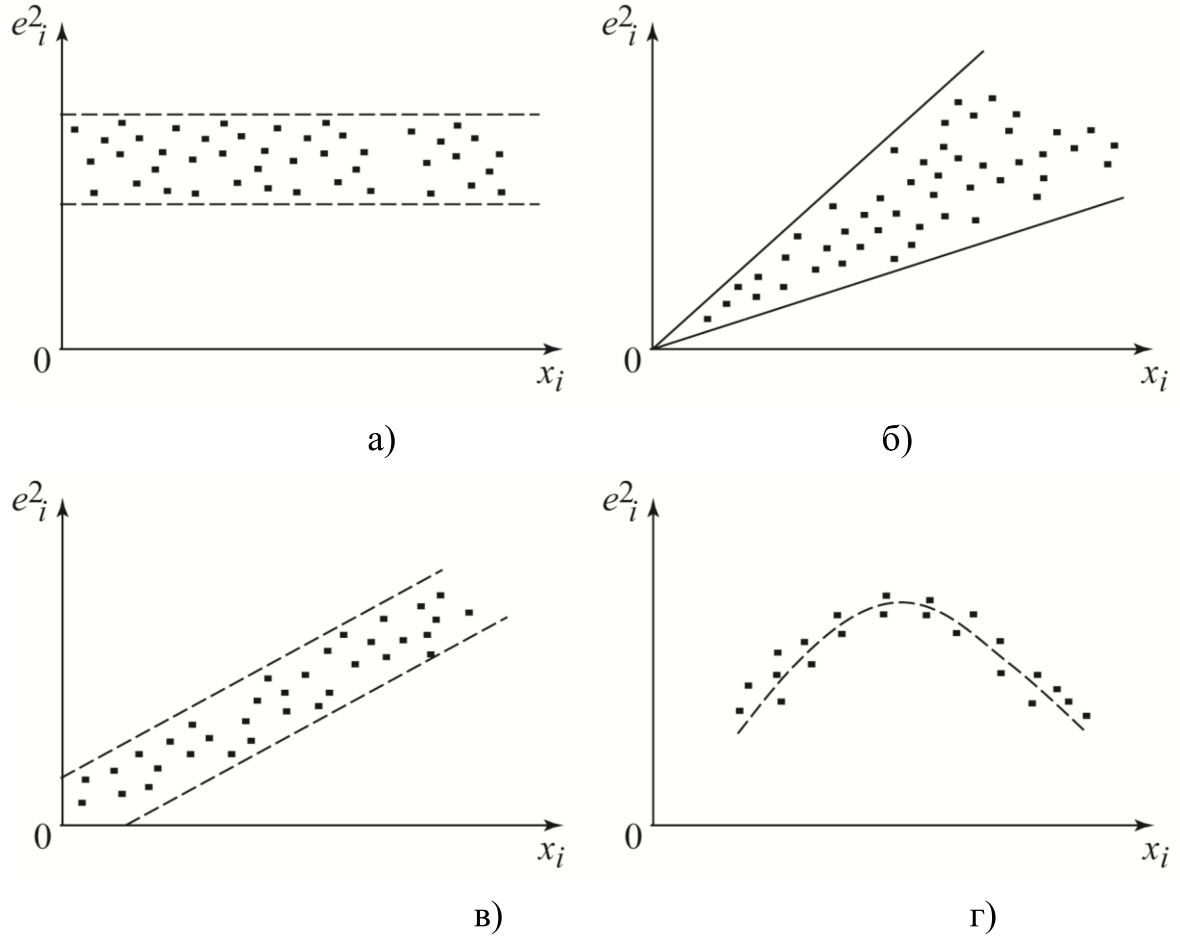


Рис. **23.2. Залежності ε_i^2 від x_i**

На рис. 23.2а) всі відхилення знаходяться всередині смуги постійної ширини, паралельної осі абсцис. Це говорить про незалежність дисперсій ε_i^2 від значень змінної X і їх сталості, тобто в цьому випадку виконуються умови гомоскедастичності. Інші рис. 23.2б), в), г) відображають велику ймовірність наявності гетероскедастичності для розглянутих статистичних даних.

Такий аналіз найбільш доцільний за великої кількості змінних, що пояснюють.

Тест рангової кореляції Спірмена

У ході використання даного тесту передбачається, що дисперсія відхилень буде або збільшуватися, або зменшуватися із збільшенням значень x. Тому для регресії, побудованої за допомогою МНК, абсолютні величини відхилень ε_i і значення x_i випадкової величини X будуть корельовані. Значення x_i і ε_i ранжуються (упорядковуються за величинами). Потім визначається коефіцієнт рангової кореляції: $r_{x\varepsilon} = 1 - 6 \cdot \frac{\sum d_i^2}{n\left(n^2 - 1\right)}$, де d_i – різниця між рангами x_i і ε_i. Доведено, що якщо коефіцієнт кореляції $\rho_{x\varepsilon}$ для генеральної сукупності дорівнює нулю, то статистика $t = \frac{r_{x\varepsilon}\sqrt{n-2}}{\sqrt{1 - r_{x\varepsilon}^2}}$ має розподіл Стьюдента з числом ступенів свободи $\nu = n - 2$. Отже, якщо значення t-статистики перевищує $t_{\text{кр}} = t_{\frac{\alpha}{2}, n-2}$, то необхідно відхилити гіпотезу про рівність нулю коефіцієнта кореляції $\rho_{x,\varepsilon}$, а, отже, і про відсутність гетероскедастичності. В іншому випадку гіпотеза про відсутність гетероскедастичності приймається.

Якщо в моделі регресії більше ніж одна змінна, що пояснює, то перевірка гіпотези може здійснюватися за допомогою t-статистики для кожної з них окремо.

Приклад. *Продовження прикладу лекції 18.* Перевірити гіпотезу про відсутність гетероскедастичності за допомогою тесту рангової Спірмена. Довірча ймовірність $p = 95\,\%$.

x_i	ε_i	$\lvert\varepsilon_i\rvert$	d_1	d_2	$d = d_1 - d_2$	d^2
2	0	0	5	5	0	0
3	-0,09	0,09	4	2	2	4
4	0,12	0,12	3	1	2	4
5	0,03	0,03	2	4	-2	4
6	-0,06	0,06	1	3	-2	4
Сума						16

$n = 5$, $r_{x,\varepsilon} = 1 - 6 \cdot \dfrac{\sum d_i^2}{n(n^2 - 1)} = 1 - 6 \cdot \dfrac{16}{5(5^2 - 1)} = 0{,}2$.

$\alpha = \dfrac{1 - p}{2} = \dfrac{1 - 0{,}95}{2} = 0{,}25$. За t-таблицями знаходимо граничну точку $t_{\text{кр}} = t_{\frac{\alpha}{2}, n-2} = t_{0{,}025;\,5-2} = 3{,}182$.

Статистика $t = \dfrac{r_{x,\varepsilon}\sqrt{n-2}}{\sqrt{1 - r_{x,\varepsilon}^2}} = \dfrac{0{,}2\sqrt{5-2}}{\sqrt{1 - 0{,}2^2}} = 0{,}352 < 3{,}182$. Приймаємо гіпотезу про відсутність гетероскедастичності на рівні значимості 5 %.

Тест Парка

Р. Парк запропонував критерій визначення гетероскедастичності, який доповнює графічний метод деякими формальними залежностями. Передбачається, що дисперсія $\sigma_i^2 = \sigma^2(\varepsilon_i)$ є функцією $i-$го значення x_i пояснюючої змінної.

Парк запропонував таку залежність $\sigma_i^2 = \sigma^2 x_i^{\beta} \varepsilon^{v_i}$. Після логарифмування даного виразу отримаємо $\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln x_i + v_i$. Оскільки дисперсії σ_i^2 зазвичай невідомі, тому їх замінюють оцінками ε_i^2.

Критерій Парка передбачає наступні етапи:

1. Будується рівняння регресії $y_i = a + b_1 x_i + \varepsilon_i$.

2. Для кожного спостереження визначається $\ln \varepsilon_i^2 = \ln(y_i - \hat{y}_i)^2$.

3. Будується регресія $\ln \varepsilon_i^2 = \alpha + \beta \ln x_i + v_i$, де $\alpha = \ln \sigma^2$. Для множинної регресії будується ця залежність для кожної змінної, що пояснює.

4. Перевіряється статистична значимість коефіцієнта β на основі $t-$статистики $t = \dfrac{\beta}{\sigma_\beta}$. Якщо коефіцієнт β статистично значущий, то це означає

наявність зв'язку між $\ln \varepsilon_i^2$ і $\ln x_i$, тобто гетероскедастичності в статистичних даних.

Застосування тесту Парка в деяких випадках може призвести до необґрунтованих висновків, тому він доповнюється іншими тестами.

Тест Глейзера

Вважається, що тест Глейзера за своєю суттю аналогічний тесту Парка та доповнює його аналізом інших залежностей між дисперсіями відхилень σ_i та значеннями змінної x_i. Тут оцінюється регресійна залежність модулів відхилень $|\varepsilon_i|$ від $x|\varepsilon_i|$. При цьому залежність моделюється таким рівнянням:

$$|\varepsilon_i| = \alpha + \beta x_i^k + v_i .$$

Змінюючи значення k можна побудувати різні регресії, при цьому статистична значущість коефіцієнтів β в кожному випадку означає наявність гетероскедастичності. Якщо для декількох регресій коефіцієнт β виявляється значущим, то орієнтуються на найкращу з них. Як і в тесті Парка, в тесті Глейзера для відхилень v_i може порушуватись умова гетероскедастичності, проте в багатьох випадках запропоновані моделі є добрими для визначення гетероскедастичності.

Тест Голдфельда–Квандта

У тесті Голдфельда–Квандта передбачається, що стандартне відхилення σ_ε пропорційно значенню змінної x в цьому спостереженні: $\sigma_{\varepsilon_i}^2 = \sigma^2 x_i^2 \quad \left(i = \overline{1, n}\right)$. Також передбачається, що ε_i має нормальний розподіл і відсутня автокореляція і всі n спостережень упорядковуються по величині x. Ця упорядкована вибірка ділиться на три приблизно рівні частини $k, n-2k, k$ відповідно. Для кожної з вибірок обсягу k оцінюється своє рівняння регресії і знаходяться суми квадратів відхилень $S_1 = \sum_{i=1}^{k} \varepsilon_i^2$ і $S_3 = \sum_{i=n-k+1}^{k} \varepsilon_i^2$ відповідно.

При заданій довірчій ймовірності $p, \alpha = 1 - p$ за F-таблицями знаходимо граничну точку $F_{\alpha,\, k-m-1,\, k-m-1}$, де $m-$число факторів моделі; розраховуємо значення $F = \dfrac{S_3}{S_1}$. Якщо $F < F_{\alpha,\, k-m-1,\, k-m-1}$, то на рівні значущості α приймається гіпотеза про відсутність гетероскедастичності. В іншому випадку гіпотеза про відсутність гетероскедастичності відхиляється. Для множинної регресії тест зазвичай проводиться для того фактора, який максимальною мірою пов'язаний з σ_{ε_i}. При цьому обирають $k > m+1$. Для переконливості даний тест можна розрахувати для кожного фактора.

Приклад. Є регресійна лінійна модель з $m=2$ фактора, $n=30$ спостережень. Для перших і останніх $k=11$ спостережень суми квадратів відхилень $S_1=20$ і $S_3=45$ відповідно. За допомогою тесту Голдфельда–Квандта перевірити гіпотезу про відсутність гетероскедастичності.

Нехай довірча ймовірність $p=95\,\%$. Тоді $\alpha=0{,}05$. За F-таблицями знаходимо граничну точку $F_{0{,}05\,8,\,8}=3{,}44$. Необхідно розрахувати $F=\dfrac{S_3}{S_1}=\dfrac{45}{20}=2{,}25<3{,}44$. На рівні значущості 5 % приймається гіпотеза від відсутності гетероскедастичності.

Коли встановлений факт присутності гетероскедастичності, то виникає необхідність перетворення моделі для усунення цього недоліку. В даному випадку розрізняють підходи, що передбачають відома чи не відома дисперсія σ_i^2 залишків ε_i.

Метод зважених найменших квадратів (МЗНК)

Якщо відомі σ_ε^2, то рекомендується застосовувати метод зважених найменших квадратів (МЗНК). При цьому слід розділити кожне спостереження на відповідне йому значення дисперсії. Наприклад, нехай маємо $y_i=a+b_1x_i+\varepsilon_i$. Поділимо обидві частини рівняння на відоме $\sigma_{\varepsilon_i}=\sqrt{\sigma_{\varepsilon_i}^2}$:

$$\frac{y_i}{\sigma_{\varepsilon_i}}=\frac{a}{\sigma_{\varepsilon_i}}+b_1\frac{x_i}{\sigma_{\varepsilon_i}}+\frac{\varepsilon_i}{\sigma_{\varepsilon_i}}.$$

Нехай $\dfrac{y_i}{\sigma_{\varepsilon_i}}=y_i^*$, $\dfrac{x_i}{\sigma_{\varepsilon_i}}=x_i^*$, $\dfrac{\varepsilon_i}{\sigma_{\varepsilon_i}}=v_i$, $\dfrac{1}{\sigma_{\varepsilon_i}}=z_i$. Отримали рівняння регресії без вільного члена, але з додатковою змінною, що пояснює Z і з перетвореним відхиленням v_i:

$$y_i^*=az_i+b_1x_i^*+v_i\,.$$

При цьому для всіх v_i виконується умова гомоскедастичності.

Отже, метод зважених найменших квадратів містить такі етапи. Перший етап передбачає, що всі спостереження (x_i, y_i) ділять на відому величину σ_{ε_i}. Таким чином, спостереження з найменшими дисперсіями набувають найбільшої ваги, а, отже, вони будуть значущими у процесі оцінювання коефіцієнтів регресії, спостереження з максимальними дисперсіями, навпаки, набувають найменшої ваги та будуть менш значущими. За другим етапом за методом найменших квадратів для перетворених значень $\left(\frac{1}{\sigma_i}, \frac{x_i}{\sigma_i}, \frac{y_i}{\sigma_i}\right)$ будується рівняння регресії без вільного члена з гарантованими якостями оцінок.

Часто в практичних задачах значення дисперсії σ_i^2 залишків ε_i невідомо, але щоб використати метод зважених найменших квадратів слід передбачити, наприклад, що σ_i^2 пропорційні значенням x_i або значенням x_i^2. Розглянемо перший випадок, а саме $\sigma_{\varepsilon_i}^2 = \sigma^2 x_i^2$, де $\sigma^2 -$ коефіцієнт пропорційності. Тоді рівняння $y_i = a + b_1 x_i + \varepsilon_i$ розділимо на $\sqrt{x_i}$:

$$\frac{y_i}{\sqrt{x_i}} = \frac{a}{\sqrt{x_i}} + b_1 \frac{x_i}{\sqrt{x_i}} + \frac{\varepsilon_i}{\sqrt{x_i}},$$

а отже, $\frac{y_i}{\sqrt{x_i}} = a\frac{1}{\sqrt{x_i}} + b_1\sqrt{x_i} + v_i$.

Оскільки $\sigma_{\varepsilon_i}^2 = \sigma^2 x_i^2$, то $\sigma^2(v_i) = \sigma^2\left(\frac{\varepsilon_i}{\sqrt{x_i}}\right) = \frac{1}{x_i}\sigma^2 x_i = \sigma^2 = const$ й далі знайшовши за методом найменших квадратів оцінки a і b повертаємось до початкового рівняння $y_i = a + b_1 x_i + \varepsilon_i$.

Іноді рекомендують використовувати $\hat{Y} = a + b_1 X_1 + b_2 X_2 + ... + b_m X_m$ замість конкретної пояснюючої змінної X_i, тобто отримуємо:

$$\frac{y_i}{\sqrt{\hat{y}_i}} = a' \frac{1}{\sqrt{\hat{y}_i}} + b' \frac{x_{1i}}{\sqrt{\hat{y}_i}} + b_2' \frac{x_{2i}}{\sqrt{\hat{y}_i}} + ... + b_m' \frac{x_{mi}}{\sqrt{\hat{y}_i}} + \varepsilon_i \frac{1}{\sqrt{\hat{y}_i}}.$$

Залежність σ_i^2 від x_i іноді виражають не лінійною функцією, а квадратичною, то перетворенням рівняння буде:

$\frac{y_i}{x_i} = \frac{a}{x_i} + b_1 + \frac{\varepsilon_i}{x_i}$, а отже $\frac{y_i}{x_i} = \frac{a}{x_i} + b_1 + v_i$, де $v_i = \frac{\varepsilon_i}{x_i}$ і для v_i виконується умова гомоскедастичності. Після визначення за методом найменших квадратів оцінок коефіцієнтів регресії повертаються до початкового рівняння.

На практиці рекомендують застосовувати декілька методів визначення гетероскедастичності та способів її коректування.

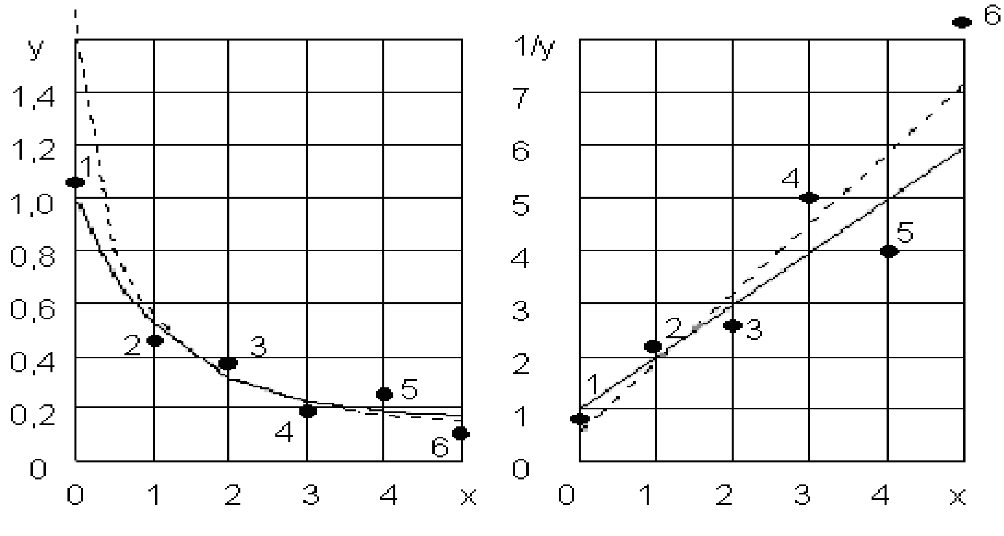
Приклад. Нехай значення y породжуються точною залежністю $\eta = \frac{1}{1+x}$, яким присвоєно похибки $\varepsilon = \pm 0{,}05$, тобто $y = \eta(x) + \varepsilon$; аргумент x набуває послідовні цілі значення 0, 1, 2, 3, 4, 5, 6. Варто припустити, що форма зв'язку відома $y_p = \frac{1}{a + b_1 x}$.

Після функціонального перетворення $Y = \frac{1}{y}$ модель формально стає лінійною $Y_p = a + b_1 x$ і її параметри можуть бути оцінені стандартною процедурою МНК. У табл. 23.1 і на рис. 23.3 наведені початкові дані $(x, \eta, \varepsilon, y)$, перетворені значення залежної змінної $\left(\frac{1}{\eta}, \frac{1}{y}\right)$ і нова система помилок $\left(e = \frac{1}{\eta} - \frac{1}{y}\right)$.

Таблиця 23.1

n	x	η	ε	y	$\frac{1}{\eta}$	$\frac{1}{y}$	e
1	0	1	0,05	1,050	1	0,952	-0,048
2	1	0,5	-0,05	0,450	2	2,222	0,222
3	2	0,333	0,05	0,383	3	2,611	-0,389
4	3	0,25	-0,05	0,200	4	5	1
5	4	0,2	0,05	0,250	5	4	-1
6	5	0,167	-0,05	0,117	6	8,547	2,546

З табл. 23.1 видно, що збільшення розкиду перетворених даних $\frac{1}{y}$ при збільшенні змінної x. На рис. 23.3б це чітко видно, причому остання точка $\left(x, \frac{1}{y}\right)$ вийшла за поле графіка і схожа на викид. Цей викид перетягнув у свій бік всю лінію регресії.



а) *б*)

Рис. 23.3. **Нелінійна залежність у вихідних змінних (а) і після функціональних перетворень (б),**

де • – емпіричні точки; ——— – породжуюча залежність; - - - - - –відновлена за МНК.

Маємо такі викладки для обчислення параметрів моделі $\frac{1}{y} = a + b_1 x + e$.

Система нормальних рівнянь має вигляд:

$$\left[\frac{1}{y}\right] = b_0 + b_1[x];$$

$$\left[\frac{x}{y}\right] = b_0[x] + b_1\left[x^2\right].$$

Ці рівняння реалізують умови ортогональності помилок для кожного члена моделі $[e] = 0,\ [ex] = 0$. Усі необхідні розрахунки виконані в табл. 23.2.

Таблиця 23.2

n	x	x^2	y	$\frac{1}{y}$	$\frac{x}{y}$	$\frac{1}{y_p}$	y_p
1	0	0	1,050	0,952	0	0,624	1,601

2	1	1	0,450	2,222	2,222	1,930	0,518
3	2	4	0,383	2,611	5,222	3,236	0,309
4	3	9	0,200	5,000	15,000	4,541	0,220
5	4	16	0,250	4,000	16,000	5,847	0,171
6	5	25	1,117	8,547	42,735	7,153	0,140
Суми	15	55		23,332	81,179		

Отримаємо таку систему нормальних рівнянь:

$$23{,}332 = a \cdot 6 + b_1 \cdot 15;$$

$$81{,}179 = a \cdot 15 + b_1 \cdot 55.$$

Розв'язавши систему рівнянь, отримаємо такі значення параметрів моделі $a = 0{,}6245, \quad b_1 = 1{,}3057$, вони істотно відрізняються від параметрів залежності $\eta = \frac{1}{1+x}$, де $a = b_1 = 1$. Обчислені значення:

$$\frac{1}{y_p} = 0{,}6245 + 1{,}3057x;$$

$$y_p = \frac{1}{0{,}6245 + 1{,}3057x};$$

наведені в останніх двох стовпцях таблиці, графіки ліній регресії зображені пунктирними лініями на рис. 23.3. При $x = 0$ замість очікуваного значення $\eta = 1$ (з помилкою близькою до $\varepsilon = 0{,}05$) отримаємо $y_p = 1{,}6$, що є неправильним. Вид графіка 23.3б) і його розташування щодо емпіричних точок викликає сумнів щодо правильності вибору самої форми зв'язку.

Отже, будь-які порушення початкових гіпотез призводять до помилок в оцінюванні параметрів моделі, причому помилки можуть бути виправлені шляхом модернізації стандартної процедури оцінювання. Тому для усунення гетероскедастичності в даному прикладі помножимо обидві частини рівняння

$$\frac{1}{y} = a + b_1 x + e$$

на y^2 і отримаємо ($\varepsilon = ey^2$):

$$y = ay^2 + b_1 xy^2 + \varepsilon.$$

Для цієї моделі система нормальних рівнянь має вигляд:

$$\lfloor y^3 \rfloor = a \lfloor y^4 \rfloor + b_1 \lfloor xy^4 \rfloor;$$

$$\lfloor xy^3 \rfloor = a\lfloor xy^4 \rfloor + b_1 \lfloor x^2 y^4 \rfloor;$$

(умови ортогональності помилок до кожного члена моделі $\lfloor \varepsilon y^2 \rfloor = 0, \quad \lfloor \varepsilon x y^2 \rfloor = 0$). Необхідні суми підраховані в табл. 23.3.

Таблиця 23.3

n	x	y	y^3	y^4	xy^3	xy^4	x^2y^4	$\frac{1}{y_p}$	y_p	η
1	0	1,050	1,158	1,216	0	0	0	0,959	1,043	1
2	1	0,450	0,091	0,041	0,091	0,041	0,041	1,901	0,526	0,5
3	2	0,383	0,056	0,026	0,112	0,043	0,086	2,843	0,352	0,333
4	3	0,200	0,008	0,002	0,024	0,480	0,014	3,786	0,264	0,25
5	4	0,250	0,016	0,004	0,062	0,016	0,062	4,728	0,212	0,2
6	5	1,117	0,002	0,000	0,008	0,001	0,005	5,670	0,176	0,167
Суми			1,330	1,284	0,298	0,105	0,209			

Звідки отримаємо таку систему нормальних рівнянь:

$$1{,}330 = a \cdot 1{,}284 + b_1 \cdot 0{,}105;$$

$$0{,}298 = a \cdot 0{,}105 + b_1 \cdot 0{,}209.$$

Розв'язанням нової системи нормальних рівнянь є значення параметрів $a = 0{,}959, \quad b_1 = 0{,}942$, які близькі до очікуваних $a = \beta_1 = 1$ (параметри моделі $\eta = \frac{1}{1+x}$). В останніх колонках таблиці обчислені значення $\frac{1}{y_p}$ і y_p. Для порівняння наведені ординати η. Відповідність між заданою (η) та відтворювальною методом найменших квадратів залежностями дуже велика. Таким чином, заміна стандартної процедури МНК на узагальнену процедуру зважених найменших квадратів забезпечує виконання другої гіпотези Гаусса − Маркова про рівноточність спостережень.

Запитання для самоперевірки

1. Поясніть суть гетероскедастичності.
2. Які відомі наслідки гетероскедастичності?
3. Перерахуйте методи визначення гетероскедастичності.

4. У чому суть тесту рангової кореляції Спірмена?

5. У чому суть тесту Парка?

6. У чому суть тесту Голдфельда − Квандта?

7. Перерахуйте методи пом'якшення гетероскедастичності.

8. У чому суть методу зважених найменших квадратів?

9. Чому за наявності гетероскедастичності ЗНК дозволяє одержати більш ефективні оцінки, ніж звичайний МНК?

24. Автокореляція залишків моделі та методи її усунення

24.1. Суть і причини автокореляції.

24.2. Наслідки автокореляції. Методи визначення автокореляції.

24.3. Методи усунення автокореляції.

24.1. Суть і причини автокореляції

У динамічних рядах важливу роль відіграє чинник часу, який вміщуючи у собі багато інших чинників розвитку, обумовлює напрямок зміни соціально-економічних систем. Члени одного і того ж динамічного ряду пов'язані між собою: попередні члени впливають на наступні. Це явище називається автокореляцією. Тому, перш ніж знаходити кількісну оцінку зв'язку між часовими рядами, необхідно перевірити існування автокореляції. Розрізняють автокореляцію змінних y і x_i і автокореляцію залишків моделі.

Важливою передумовою побудови якісної регресійної моделі за МНК є незалежність значень випадкових відхилень ε_i від значень відхилень у всіх інших спостереженнях. Відсутність залежності гарантує відсутність корельованості між будь-якими відхиленнями $\sigma(\varepsilon_i, \varepsilon_j) = 0$ при $i \neq j$ і, зокрема, між сусідніми відхиленнями $\sigma(\varepsilon_{i-1}, \varepsilon_i) = 0$.

Автокореляція (послідовна кореляція) визначається як кореляція між значеннями показників, що спостерігаються, впорядкованими в часі (часові ряди) або у просторі (перехресні дані). Автокореляція залишків (відхилень) зазвичай зустрічається в регресійному аналізі при використанні даних часових рядів. Найчастіше зустрічається так звана додатна автокореляція, порівняно з від’ємною автокореляцією. В більшості випадків додатна автокореляція обумовлюється спрямованою постійною дією деяких не врахованих у моделі факторів. Наприклад, нехай щомісячно досліджується попит Y на охолоджувальні напої залежно від доходу X. Трендова залежність, яка відображає збільшення попиту зі зростанням доходу, може бути представлена лінійною функцією $Y = a + b \cdot X$, що зображена на рис. 24.1. Але фактичні точки спостереження зазвичай будуть перевищувати трендову лінію в літні періоди і будуть знаходитись нижче її в зимові періоди.

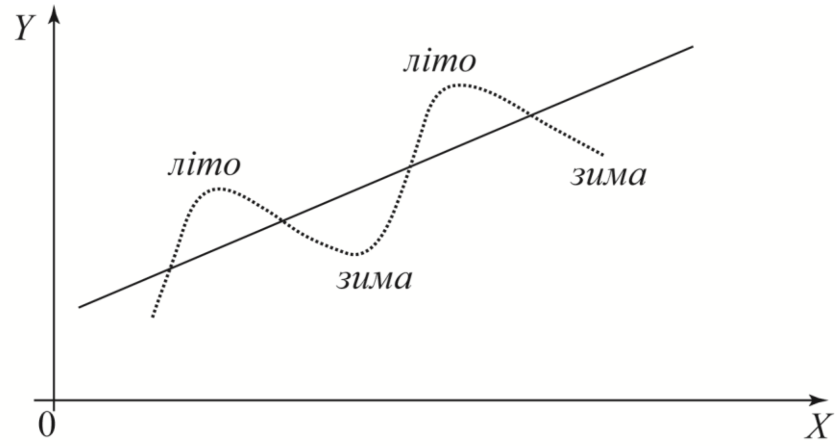


Рис. 24.1.

Аналогічну картину можемо мати в макроекономічному аналізі з урахуванням циклів ділової активності.

Серед основних причин, що викликають появу автокореляції, виділяють помилки специфікації, інерцію у зміні економічних показників, ефект павутини, згладжування даних. Неврахування в моделі важливої змінної або неправильний вибір форми залежності приводить до системних відхилень точок спостережень від ліній регресії, що може зумовити автокореляцію. Багато економічних показників (інфляція, безробіття, ВНП та ін.) володіють певною циклічністю, пов'язаною з хвилеподібністю ділової активності. У багатьох виробничих та інших сферах економічні показники реагують на зміну економічних умов із запізненням (часовим лагом). А також, дані по деякому тривалому часовому періоду отримують усередненням даних за складовими його підінтервалами. Це може призвести до певного згладжування коливань, які були всередині розглянутого періоду, що у свою чергу може послужити причиною автокореляції.

24.2. Наслідки автокореляції. Методи визначення автокореляції

Наслідки автокореляції певною мірою такі ж як і наслідки гетероскедастичності. Зазвичай виокремлюють такі наслідки:

1. Оцінки параметрів, залишаючись лінійними і незміщеними, не є ефективними, а отже, вони перестають мати властивості найкращих лінійних незміщених оцінок (BLUE-оцінок).

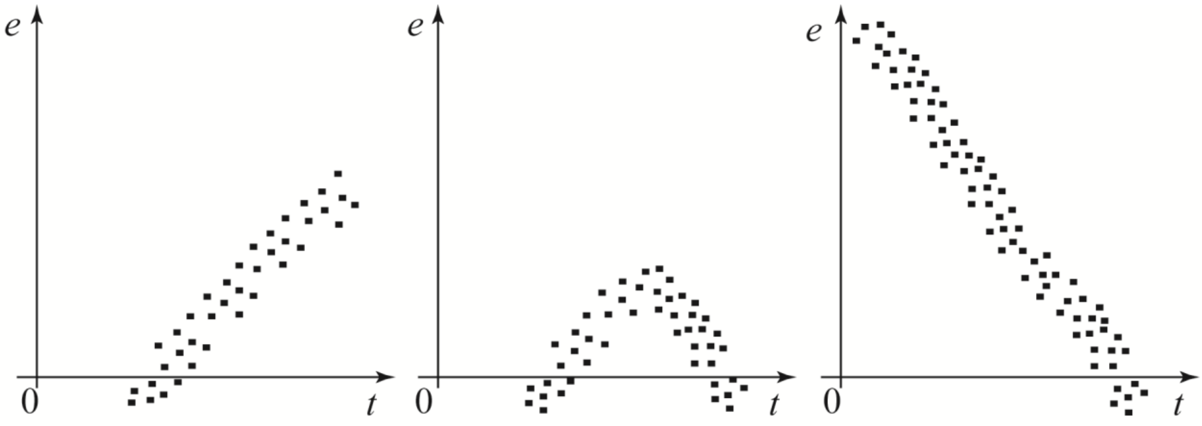
2. Дисперсії оцінок будуть зміщеними. Часто дисперсії, які розраховуються за стандартними формулами, є заниженими. Це призводить до збільшення t-статистик. Що, у свою чергу, тягне за собою визнання значущими змінні, що пояснюють, які насправді незначущі. А також, висновки по $t-$статистиці і $F-$статистиці можуть бути не правильними, що погіршує статистичну якість моделі.

3. Оцінка дисперсії регресії $S^2 = \sum_{t=1}^{T} \frac{\varepsilon_t^2}{T-m-1}$ є незміщеною оцінкою справжнього значення σ^2, при цьому в багатьох випадках занижуючи її. Тут $T = n, t = i$.

4. Підсумовуючи попередні наслідки, маємо, що висновки за $t-$статистикою і $F-$статистикою, які визначають значущість коефіцієнтів регресії і коефіцієнта детермінації, можуть бути не правильними, що погіршує прогнозні якості моделі.

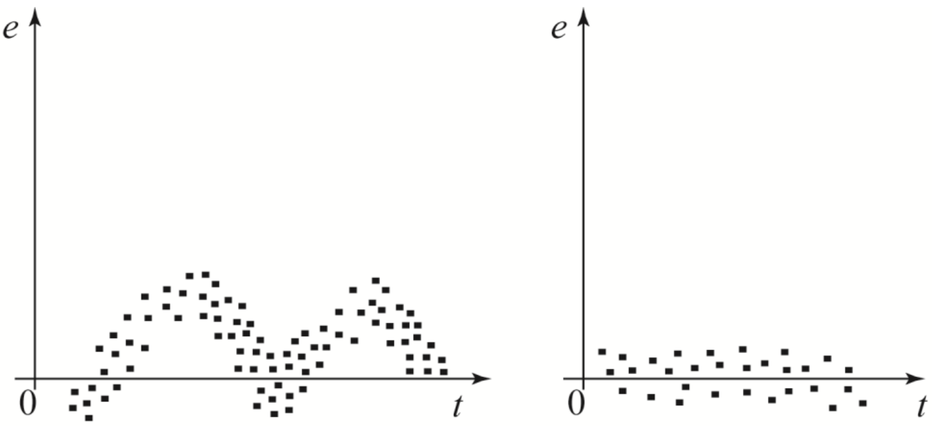
Існує кілька методів виявлення автокореляції.

Графічний метод. Розглянемо графічний метод, що пов'язує відхилення ε_i з моментами t їх отримання (їх порядковими номерами i). Це так звані послідовно-часові графіки (рис. 24.2).



а) б) с)

Рис. 24.2. **Послідовно-часові графіки**



д) е)

Закінчення рис. 24.2

На рис. 24.2а) – д) автокореляція має місце, а рис. 24.2е) свідчить про відсутність автокореляції. У сучасних комп'ютерних прикладних програмах для розв'язання задач з економетрики аналітичний вираз регресії доповнюється графічним представленням результатів. На графік реальних коливань залежної змінної накладається графік коливань змінної за рівнянням регресії. Порівнюючи два графіка, висувають гіпотезу про наявність автокореляції залишків. Якщо ці графіки перетинаються рідко, то можна передбачити існування додатної автокореляції залишків.

Метод рядів. У даному методі послідовно визначають знаки відхилень $\varepsilon_t, t=\overline{1,T}$. Наприклад,

(- - - - -)(+ + + + + + +)(- - -)(+ + + +)(-),

маємо 5 «-», 7 «+», 3 «-», 4 «+», 1 «-» з 20 спостережень.

Ряд є неперервна послідовність однакових знаків, при цьому довжина ряда визначається кількістю знаків у ряді. Якщо рядів замало порівняно з кількістю спостережень n, то ймовірна додатна автокореляція. Якщо ж рядів дуже багато, то ймовірна від’ємна автокореляція.

Нехай $n-$ обсяг вибірки, n_1- загальна кількість знаків «+» в n спостереженнях, n_2- загальна кількість знаків «-» в n спостереженнях, $k-$ кількість рядів. Для $n_1 > 10$ і $n_2 > 10$ та відсутності автокореляції випадкової величини k має місце асимптотично нормальний розподіл з:

$$M(k)=\frac{2n_1n_2}{n_1+n_2}+1;\quad D(k)=\frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)^2(n_1+n_2+1)}+1.$$

Якщо $M(k)-u_{\frac{\alpha}{2}}D(k)<k<M(k)+u_{\frac{\alpha}{2}}D(k)$, то гіпотеза про відсутність автокореляції не відхиляється.

Для невеликої кількості спостережень $n_1<20$ і $n_2<20$ вчені Свед і Ейзенхарт розробили таблиці критичних значень кількості рядів для n спостережень. З таблиць визначають нижнє k_1 та верхнє k_2 значення за рівня значущості $\alpha=0{,}05$. Якщо $k_1<k<k_2$, то говорять про відсутність автокореляції; якщо $k<k_1$, то говорять про додатну автокореляцію; якщо $k>k_2$, то говорять про від'ємну автокореляцію залишків.

Для наведеного приклад: маємо n = 20, n_1 = 11, n_2 = 9, k = 5. З таблиць визначаємо k_1 = 6, k_2 = 16. Оскільки k = 5<6, то приймається передбачення про наявність додатної автокореляції за рівня значущості $\alpha=0{,}05$.

Критерій Дарбіна−Уотсона. Найбільш відомим критерієм виявлення автокореляції першого порядку є критерій Дарбіна−Уотсона. На основі обчисленої статистики DW Дарбіна−Уотсона робиться висновок про автокореляцію:

$$DW=\frac{\sum_{t=2}^{T}(\varepsilon_t-\varepsilon_{t-1})^2}{\sum_{t=1}^{T}\varepsilon_t^2}. \qquad (24.1)$$

Таким чином статистика Дарбіна-Уотсона тісно пов'язана з вибірковим коефіцієнтом кореляції $r_{\varepsilon_t\varepsilon_{t-1}}$: $DW\approx 2(1-r_{\varepsilon_t\varepsilon_{t-1}})$ і якщо $0\le DW\le 4$, то її значення можуть вказати на наявність або відсутність автокореляції. Якщо $r_{\varepsilon_t\varepsilon_{t-1}}$ = 0 (автокореляція відсутня), то $DW\approx 2$. Якщо $r_{\varepsilon_t\varepsilon_{t-1}}$ = 1 (додатна автокореляція), то $DW\approx 0$. Якщо $r_{\varepsilon_t\varepsilon_{t-1}}$ = -1 (від'ємна автокореляція), то $DW\approx 4$. У критерії Дарбіна−Уотсона при заданому рівні значимості α, числі спостережень n і кількості пояснювальних змінних m визначаються два значення: d_1 – нижня межа і d_u – верхня межа. Загальна схема критерію Дарбіна−Уотсона така:

1. Для побудованого емпіричного рівняння регресії $\hat{y} = a + b_1x_1 + b_2x_2 + ... + b_mx_m$ визначають значення відхилень $\varepsilon_t = y_t - \hat{y}_t$.

2. За формулою $DW = \dfrac{\sum_{t=2}^{T}\left(\varepsilon_t - \varepsilon_{t-1}\right)^2}{\sum_{t=1}^{T}\varepsilon_t^2}$ розраховують статистику DW.

3. За таблицею критичних точок Дарбіна−Уотсона визначаються два числа d_l і d_u, потім робляться висновки за правилом:

$0 \le DW \le d_l$ – існує додатна кореляція;

$d_l \le DW \le d_u$ – висновок про наявність автокореляції не визначено;

$d_u \le DW \le 4 - d_u$ – автокореляція відсутня;

$4 - d_u \le DW \le 4 - d_l$ – висновок про наявність автокореляції не визначено;

$4 - d_l \le DW \le 4$ – існує від'ємна автокореляція.

У процесі використання критерія Дарбіна−Уотсона слід враховувати обмеження: 1) критерій Дарбіна−Уотсона застосовується лише для моделей з вільним членом; 2) передбачається, що випадкові відхилення ε_t визначаються за ітераційною схемою $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$, яка називається авторегресійною схемою першого порядку $AR(1)$, де v_t – випадковий член; 3) статистичні дані повинні мати однакову періодичність; 4) критерій Дарбіна−Уотсона не слід застосовувати для регресійних моделей, що містять в складі пояснюючих змінних лагову змінну y_{t-1}, тобто для випадку:

$$y = a + b_1x_{t1} + b_2x_2 + ... + b_mx_{t\,m} + \gamma\, y_{t-1} + \varepsilon_t.$$

Для авторегресійних моделей розроблені спеціальні тести виявлення автокореляції, зокрема h-статистика Дарбіна, яка обчислюється за формулою:

$$h = \hat{\rho}\sqrt{\frac{n}{1 - nD(g)}},$$

де $\hat{\rho}$ – оцінка ρ авторегресії першого порядку, $D(g)$ – вибіркова дисперсія коефіцієнта при лаговій змінній y_{t-1}; n – число спостережень. Для великого n справедлива нульова гіпотеза H_0: $\rho = 0$ статистика h має стандартизований нормальний розподіл ($h \sim N(0,1)$). Для заданого рівня значущості α визначається критична точка $u_{\frac{\alpha}{2}}$ з умови $\Phi\left(u_{\frac{\alpha}{2}}\right) = \dfrac{1-\alpha}{2}$ і порівнюється h з $u_{\frac{\alpha}{2}}$. Якщо $|h| > u_{\frac{\alpha}{2}}$, то нульову гіпотезу про відсутність автокореляції відхиляють, в противному випадку вона не відхиляється.

24.3. Методи усунення автокореляції

Основною причиною наявності випадкового члена в моделі є недосконалі знання про причини і взаємозв'язки, що визначають те чи інше значення залежної змінної. Тому властивості випадкових відхилень, в тому числі й автокореляція, в першу чергу залежить від вибору формули залежності і складу змінних, що пояснюють. Так як автокореляція найчастіше викликається неправильною специфікацією моделі, то необхідно, насамперед, скорегувати саму модель. Також можна спробувати змінити форму залежності (наприклад, лінійну на лог-лінійну або гіперболічну). Можна скористатися авторегресійним перетворенням, а саме авторегресійною схемою першого порядку $AR(1)$.

Розглянемо модель парної лінійної регресії: $y = a + b \cdot x + \varepsilon$, тоді спостереженням t і $(t-1)$ відповідають формулі:

$$y_t = a + b \cdot x_t + \varepsilon_t, \tag{24.2}$$

$$y_{t-1} = a + b \cdot x_{t-1} + \varepsilon_{t-1}. \tag{24.3}$$

Нехай випадкові відхилення мають авторегресію першого порядку:

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t,$$

де v_t – випадкові відхилення, що задовольняють всім передумовам МНК, а коефіцієнт ρ відомий. Віднімемо з виразу (24.2) вираз (24.3), помножений на ρ:

$$y_t - \rho y_{t-1} = a(1-\rho) + b(x_t - \rho x_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1}).$$

Поклавши $y_1^* = y_t - \rho y_{t-1}$, $x_t^* = x_t - \rho x_{t-1}$, $\beta_0^* = \beta_0(1-\rho)$ отримаємо $y_t^* = a^* + b \cdot x_t^* + v_t$. Оскільки коефіцієнт ρ відомий, то y_t^*, x_t^*, v_t обчислюються просто і v_t − випадкові відхилення, що задовольняють всім передумовам МНК, тому оцінки a^*, b мають властивості найкращих лінійних незміщених оцінок.

Слід відмітити, що в даних перетвореннях губиться перше спостереження і число ступенів свободи зменшується на одиницю, що в великих вибірках не так істотно, але в малих вибірках може призвести до втрати ефективності. Ця проблема зазвичай долається за допомогою поправки Прайса−Вінстена: $x_1^* = \sqrt{1-\rho^2} \cdot x_1$, $y_1^* = \sqrt{1-\rho^2} \cdot y_1$. Авторегресійне перетворення може бути узагальнено на довільну кількість пояснюючих факторів. Авторегресійне перетворення першого порядку $AR(1)$ може бути узагальнено на перетворення більш високих порядків $AR(2)$, $AR(3)$. Проте на практиці значення коефіцієнта ρ зазвичай невідомо і його необхідно оцінити. Існує декілька методів оцінювання. Розглянемо деякі з них.

Визначення ρ на основі статистики Дарбіна−Уотсона.

Статистики Дарбіна−Уотсона тісно пов'язана з коефіцієнтом кореляції між сусідніми відхиленнями через співвідношення $DW \approx 2(1 - r_{\varepsilon_t \varepsilon_{t-1}})$. Тоді як оцінка коефіцієнта ρ може бути взятий коефіцієнт $r = r_{\varepsilon_t \varepsilon_{t-1}}$. Тоді маємо $r \approx 1 - \dfrac{DW}{2}$. Цей метод оцінювання непоганий при великій кількості спостережень. У цьому випадку оцінка r параметра ρ буде достатньо точною.

Ітераційний метод Кохрана−Оркатта

Розглянемо метод Кохрана−Оркатта на прикладі парної регресії: $y = a + b \cdot x + \varepsilon$ і авторегресійній схемі першого порядку: $\varepsilon_t = \rho\varepsilon_{t-1} + v_t$. Ітераційний процес складається з таких етапів: 1) оцінюється за МНК регресія $y = a + b \cdot x + \varepsilon$ і для неї визначаються оцінки e_t відхилень ε_t; 2) з використанням схем $AR(1)$ оцінюється регресійна залежність $e_t = \hat{\rho} e_{t-1} + v_t$, де $\hat{\rho}$ − оцінка коефіцієнта ρ; 3) на основі даної оцінки будується рівняння:

$$y_t - \hat{\rho} y_{t-1} = a^*(1-\hat{\rho}) + b^*(x_t - \hat{\rho} x_{t-1}) + (\varepsilon_t - \hat{\rho}\varepsilon_{t-1}),$$

за допомогою якого оцінюються коефіцієнти a^*, b^* (у даному випадку значення $\hat{\rho}$ відомо); 4) значення $a = a^*(1-\hat{\rho}), b = b^*(x_t - \hat{\rho} x_{t-1})$ підставляються в рівняння $y = a + b \cdot x + \varepsilon$ і знову обчислюють оцінки e_t відхилень і процес повертається до етапу 2. Послідовність етапів буде продовжуватись до тих пір, доки не буде досягнута необхідна точність.

Метод Хілдрета–Лу.

За даним методом регресія:

$$y_t - \rho y_{t-1} = a(1-\rho) + b(x_t - \rho x_{t-1}) + (\varepsilon_t - \rho\varepsilon_{t-1})$$

оцінюється для кожного можливого значення ρ з відрізка [-1, 1] з будь-яким кроком, наприклад, 0,001; 0,01 і т.д. Величина $\hat{\rho}$, що забезпечує найменшу стандартну похибку регресії, приймається як оцінка коефіцієнта ρ. Значення a і b оцінюються з наведеного рівняння за даним значенням $\hat{\rho}$. Цей ітераційний метод широко використовується в пакетах прикладних програм.

Отже, якщо за наявності автокореляції після виправлення специфікації моделі вона залишається, то це пов'язують з внутрішніми властивостями ряда відхилень ε_t. В даному випадку пропонуються перетворення, що усувають автокореляцію. Відмінною є авторегресійна схема першого порядку $AR(1)$, для якої необхідно оцінити коефіцієнт кореляції між відхиленнями, що можна зробити на основі статистики Дарбіна–Уотсона, Кохрана–Оркатта, Хілдрета–Лу та ін. У випадку наявності серед пояснюючих змінних лагової залежної змінної наявність автокореляції встановлюється за допомогою $h-$статистики Дарбіна, а для її усунення застосувати метод Хілдрета–Лу.

Приклад. Нехай обсяг пропозиції товару залежить від ціни товару x_1 і зарплати співробітників x_2: $\hat{y} = 90{,}74 + 0{,}88x_1 - 7{,}32x_2$: (табл. 24.1). Визначити наявність автокореляції за допомогою критерію Дарбіна–Уотсона.

Таблиця 24.1

Вихідні дані

y_i	x_1	x_2	$\hat{y}_i$	ε_i	ε_i^2	$\varepsilon_i - \varepsilon_{t-1}$	$(\varepsilon_i - \varepsilon_{t-1})^2$
20	10	12	11,7	8,3	68,89		
35	15	10	30,74	4,26	18,15	-4,04	16,32
30	20	9	42,46	-12,46	155,25	-16,72	279,56
45	25	9	46,86	-1,86	3,46	10,6	112,36
60	40	8	67,38	-7,38	54,46	-5,52	30,47
70	37	8	64,74	5,26	27,67	12,64	159,77
75	43	6	84,66	-9,66	93,32	-14,92	22,61
90	35	4	92,26	-2,26	5,11	7,4	54,76
105	40	4	96,66	8,34	69,56	10,6	112,36
110	55	5	102,54	7,46	55,65	-0,88	0,77
640	**320**	**75**			**551,52**		**988,98**

$$DW = \frac{\sum_{t=2}^{T}(\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{T}\varepsilon_t^2} = \frac{988{,}98}{551{,}52} = 1{,}793 .$$

За таблицею розподілу Дарбіна-Уотсона знаходимо d_l = 0,697 і d_u = 1,641. Тоді 4−d_u = 4−1,641 = 2,359. Оскільки $d_u \le DW \le 4 - d_u$ (1,641<1,793<2,359), то гіпотеза про відсутність автокореляції залишків не відхиляється на рівні значущості 0,05. Це є одним з підтверджень високої якості моделі.

Приклад. У табл. 24.2 представлені відомості протягом 24 років про витрати на житло в США (*C*, млрд дол.). Залежно від прибутку (*W*, млрд дол.) та індексу реальних цін на житло (*P*, для 1972 прийнято *P* = 100) [4].

Таблиця 24.2

Витрати на житло (*C*) залежно від прибутку (*W*) та індексу цін (*P*)

Рік	*C*	*W*	*P*	*Y = lnC*	*Const*	$X_1 = lnW$	$X_2 = lnP$	$Y_p = lnCp$	*e*
1960	64,0	489,7	104,5	4,159	1	6,194	4,649	4,140	0,019
1961	67,0	503,8	105,1	4,205	1	6,222	4,655	4,171	0,033
1962	70,7	524,9	105,0	4,258	1	6,263	4,654	4,220	0,039
1963	74,0	542,3	104,8	4,304	1	6,296	4,652	4,259	0,045
1964	77,4	580,8	104,5	4,349	1	6,364	4,649	4,341	0,008

1965	81,6	616,3	104,0	4,402	1	6,424	4,644	4,412	-0,010

Закінчення таблиці 24.2

1966	85,3	646,8	102,6	4,446	1	6,472	4,631	4,474	-0,027
1967	89,1	673,5	102,2	4,490	1	6,512	4,627	4,523	-0,033
1968	93,5	701,3	100,9	4,538	1	6,553	4,614	4,575	-0,037
1969	98,4	722,5	100,0	4,589	1	6,583	4,605	4,613	-0,024
1970	102,0	751,6	99,6	4,625	1	6,622	4,601	4,661	-0,036
1971	106,4	779,2	100,0	4,667	1	6,658	4,605	4,702	-0,035
1972	112,5	810,3	100,0	4,723	1	6,697	4,605	4,748	-0,025
1973	118,2	865,3	99,1	4,772	1	6,763	4,596	4,828	-0,056
1974	124,2	858,4	95,1	4,822	1	6,755	4,555	4,833	-0,011
1975	128,3	875,8	93,3	4,854	1	6,775	4,536	4,864	-0,009
1976	134,9	906,8	93,7	4,905	1	6,810	4,540	4,903	0,002
1977	141,3	942,9	94,5	4,951	1	6,849	4,549	4,946	0,005
1978	148,5	988,8	94,7	5,001	1	6,896	4,551	5,001	-0,001
1979	154,8	1015,5	93,8	5,042	1	6,923	4,541	5,036	0,006
1980	159,8	1021,6	93,0	5,074	1	6,929	4,533	5,046	0,028
1981	164,8	1049,3	94,2	5,105	1	6,956	4,545	5,073	0,032
1982	167,5	1058,3	96,7	5,121	1	6,964	4,572	5,074	0,047
1983	171,3	1095,4	99,2	5,143	1	6,999	4,597	5,105	0,038

Методом найменших квадратів у логарифмічному масштабі отримано наступне рівняння з коефіцієнтів детермінації $R^2 = 0{,}9908$.

$$\ln C = -1{,}51 + 1{,}18 \ln W - 0{,}35 \ln p + e,$$
$$(tb) \qquad (0{,}8) \quad (21{,}0) \quad (1{,}1)$$

У дужках наведені значення статистик Стьюдента для кожного коефіцієнта регресії. Незначний за критерієм Стьюдента член – $0{,}35 \ln P$ – не слід прибирати з моделі, оскільки при цьому збільшується *MSE* – незміщена оцінка залишкової дисперсії. У логарифмічному масштабі коефіцієнти регресії є еластичностями результативної ознаки за кожним чинником.

Отримане рівняння регресії значимо в цілому і пояснює 99,1 % мінливості $\ln C$, проте оцінка коефіцієнта еластичності за прибутком виявилася більше одиниці b_1 = 1,18. В останніх колонках таблиці обчислені значення lnC_p і залишки моделі $e = lnC – lnC_p$. Графік залишків моделі за роками зображений на рис. 24.3.

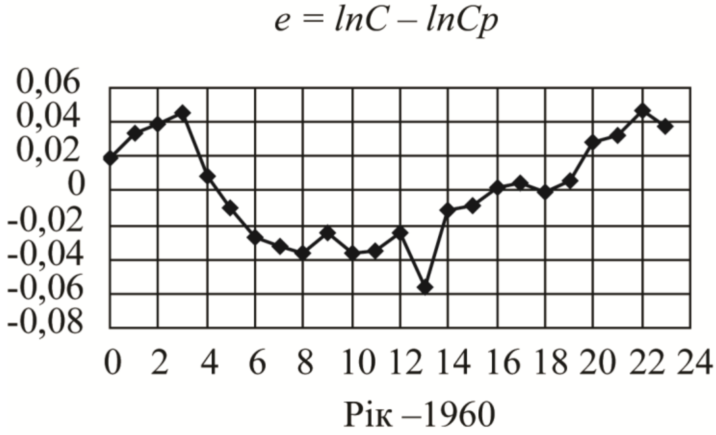


Рис. 24.3. **Графік залишків моделі**

У даному прикладі на рис. 24.3 явно видно відсутність випадковості в поведінці залишків і спостерігається ефект, схожий на наявність сезонних коливань. Обчислений коефіцієнт автокореляції 1-го порядку $\rho = 0{,}83$ свідчить про наявність великої автокореляції, це підтверджується також статистикою Дарбіна– Уотсона $DW \approx 2(1-\rho) = 0{,}33$.

Приклад. У табл. 24.3 представлені вихідні дані

$$Y = \ln C,\ X_1 = \ln W,\ X_2 = \ln P$$

і перетворені значення $V_t = Y_t - \rho Y_{t-1},\ U_t = X_t - \rho X_{t-1}$ для $\rho = 0{,}832$.

Таблиця 24.3

Дані Y, X_1, X_2 для $\rho = 0$ і V, U_1, U_2 для $\rho = 0{,}832$

Рік	***Y***	***X*₀**	***X*₁**	***X*₂**	**Yp**	***e***	***V***	***U*₀**	***U*₁**	***U*₂**	***Ypp***	***Ee***
1	2	3	4	5	6	7	8	9	10	11	12	13
1960	4,159	1	6,194	4,649	4,140	0,019						
1961	4,205	1	6,222	4,655	4,171	0,033	0,743	0,168	1,066	0,785	0,726	3,479
1962	4,258	1	6,263	4,654	4,220	0,039	0,758	0,168	1,084	0,779	0,750	3,508
1963	4,304	1	6,296	4,652	4,259	0,045	0,759	0,168	1,082	0,778	0,749	3,555
1964	4,349	1	6,364	4,649	4,341	0,008	0,766	0,168	1,123	0,776	0,795	3,554
1965	4,402	1	6,424	4,644	4,412	-0,010	0,782	0,168	1,126	0,774	0,800	3,602
1966	4,446	1	6,472	4,631	4,474	-0,027	0,782	0,168	1,124	0,765	0,803	3,643
1967	4,49	1	6,512	4,627	4,523	-0,033	0,789	0,168	1,124	0,772	0,799	3,691
1968	4,538	1	6,553	4,614	4,575	-0,037	0,800	0,168	1,132	0,762	0,813	3,725
1969	4,589	1	6,583	4,605	4,613	-0,024	0,811	0,168	1,128	0,764	0,808	3,781
1970	4,625	1	6,622	4,601	4,661	-0,036	0,805	0,168	1,142	0,768	0,821	3,804
1971	4,667	1	6,658	4,605	4,702	-0,035	0,817	0,168	1,146	0,775	0,821	3,846

Закінчення таблиці 24.3

1	2	3	4	5	6	7	8	9	10	11	12	13
1972	4,723	1	6,697	4,605	4,748	-0,025	0,838	0,168	1,155	0,772	0,833	3,890
1973	4,772	1	6,763	4,596	4,828	-0,056	0,840	0,168	1,188	0,763	0,875	3,897
1974	4,822	1	6,755	4,555	4,833	-0,011	0,850	0,168	1,125	0,729	0,826	3,996
1975	4,854	1	6,775	4,536	4,864	-0,009	0,840	0,168	1,152	0,744	0,846	4,008
1976	4,905	1	6,810	4,540	4,903	0,002	0,864	0,168	1,170	0,764	0,854	4,051
1977	4,951	1	6,849	4,549	4,946	0,005	0,868	0,168	1,180	0,770	0,862	4,089
1978	5,001	1	6,896	4,551	5,001	-0,001	0,880	0,168	1,195	0,764	0,881	4,120
1979	5,042	1	6,923	4,541	5,036	0,006	0,879	0,168	1,182	0,753	0,875	4,167
1980	5,074	1	6,929	4,533	5,046	0,028	0,877	0,168	1,166	0,753	0,857	4,217
1981	5,105	1	6,956	4,545	5,073	0,032	0,881	0,168	1,188	0,772	0,870	4,235
1982	5,121	1	6,964	4,572	5,074	0,047	0,871	0,168	1,174	0,789	0,843	4,278
1983	5,143	1	6,999	4,597	5,105	0,038	0,880	0,168	1,202	0,791	0,873	4,270

За даними Y, X_1, X_2 знайдені коефіцієнти регресії ($a = -1{,}783; b_1 = 1{,}190; b_2 = -0{,}312$), обчислені розрахункові значення Y_p і залишки моделі *e*. За цими залишками отримана початкова оцінка коефіцієнта автокореляції $\rho = 0{,}832$.

За перетвореними даними V, U_0, U_1, U_2 знайдені нові оцінки коефіцієнтів регресії, обчислені нові розрахункові значення Y_{pp} і нові залишки моделі *ee*, за якими отримана нова оцінка коефіцієнта автокореляції ρ^*.

У табл. 24.4 представлені результати розрахунків за різних значеннях ρ. В останньому стовпці значення ρ= 0,999 знайдено ітераційно.

Таблиця 24.4

Параметри моделі за різних значень ρ

ρ	0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,999
a	-1,783	-1,787	-1,751	-1,663	-1,510	-1,281	-0,973	-0,588	-0,096	1,263	65,120
b_1	1,190	1,191	1,191	1,190	1,188	1,184	1,177	1,166	1,135	0,896	0,286
b_2	-0,312	-0,313	-0,321	-0,339	-0,369	-0,412	-0,470	-0,537	-0,598	-0,529	-0,410
ρ^*	0,822	0,914	0,963	0,980	0,987	0,991	0,993	0,995	0,996	0,998	0,999

Маємо, що з урахуванням ρ модель лише погіршилася. Наприклад, для $\rho = 0{,}832$ залишки виявилися одного знака і зростають практично лінійно (табл. 24.3 стовпець *ee*).

Дані обчислення не підтвердили ефективності ітераційної процедури Кохрана−Оркатта. Рекомендується уточнити специфікацію моделі. Є думка, що автокореляція залишків є наслідком неправильної специфікації моделі. Наприклад, під час спроб описати нелінійну залежність більш простою лінійною моделлю часто з'являється позитивна автокореляція залишків. На рис. 24.4 зображена лінійна апроксимація гіперболічної залежності $y = a + \frac{b}{x}$. У цьому випадку знаки залишків залишаються однаковими для кількох послідовних спостережень, після чого знову змінюються. Така поведінка залишків не є випадковою.

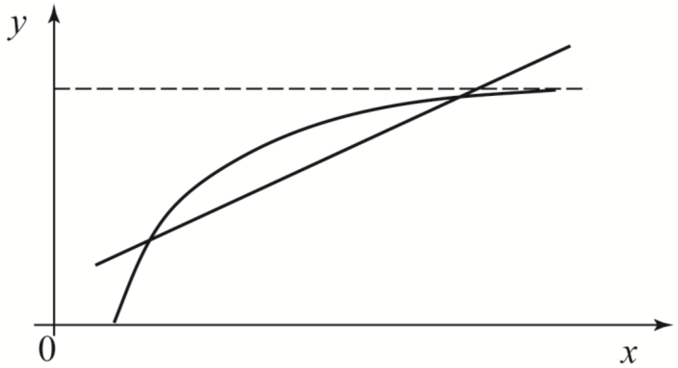


Рис. 8.4. **Неправильна форма зв'язку**

Графік залишків цього випадку дуже нагадує графік рис. 24.3. Основним джерелом появи небажаних ефектів є відсутність в моделі важливої змінної, що пояснює, і поява автокореляції. Отже, необхідно не корегувати оцінки параметрів, а звернути увагу на можливі помилки специфікації цієї моделі. При цьому, вибір неправильної форми зв'язку також є помилкою специфікації.

Приклад. Продовжимо попередній приклад. Спроба врахувати автокореляцію 1-го порядку фактично призвела до моделі з лаговими змінними, тобто зі змінними за попереднім періодом часу:

$$y_t - \rho y_{t-1} = a(1-\rho) + b_1\left(x_{1t} - \rho x_{1,t-1}\right) + b_2\left(x_{2t} - \rho x_{2,t-1}\right) + \varepsilon_t ,$$

або $$y_t = a' + b_1 x_{1t} + b_2 x_{2t} + c_0\, y_{t-1} + c_1 x_{1,t-1} + c_2 x_{2,t-1} + \varepsilon_t ,$$

де, згідно з традиційною методикою, коефіцієнти a, b_1, b_2, c_0, c_1, c_2 пов'язані між собою і ще з одним параметром ρ нелінійними зв'язками:

$$c_0 = \rho;\ c_1 = -\rho\, b_1;\ c_2 = -\rho\, b_2;\ a' = a(1-\rho).$$

Не будемо враховувати ці обмеження. У такому випадку формально отримаємо узагальнену лінійну модель з лаговими змінними, коефіцієнти якої просто обчислюються звичайним МНК:

$$y_{pt} = 0{,}74 + 0{,}22x_{1t} - 0{,}20x_{2t} + 0{,}87\,y_{t-1} - 0{,}11x_{1,t-1} + 0{,}01x_{2,t-1}.$$

|tb| (1,5) (2,5) (1,4) (14,0) (0,9) (0,1)

У дужках наведені значення статистик Стьюдента на кожну оцінку. Послідовно вибракуємо незначущі за критерієм Стьюдента члени і отримаємо таку авторегресійну модель:

$$y_{pt} = 0{,}50 + 0{,}15x_{1t} - 0{,}16x_{2t} + 0{,}84\,y_{t-1},$$

$$\ln C_t = 0{,}50 + 0{,}15\ln W_t - 0{,}16\ln P_t + 0{,}84\ln C_{t-1}.$$

|tb| (1,3) (3,1) (2,4) (22,4)

Ця модель значуща в цілому і складається лише зі значущих за Стьюдентом членів. Коефіцієнт детермінації $R^2 = 0{,}9996$ дуже високий, модель майже точно відтворює дані. Автокореляції залишків більше немає (рис. 24.5).

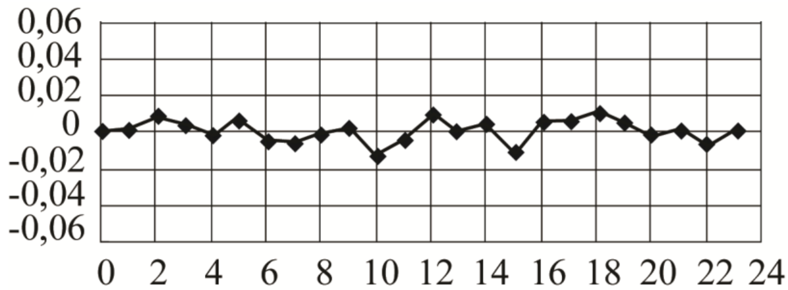


Рис. 24.5. **Графік залишків для нової авторегресійної моделі**

Нова оцінка коефіцієнта автокореляції $\rho = 0{,}03$ також свідчить про відсутність автокореляції. У авторегресійній моделі з'явився член $0{,}84lnC_{t-1}$, що позначає залежність витрат поточного року lnC_t від витрат попереднього періоду. Таким чином, була знайдена правильна специфікація моделі.

Отже, була розглянута одна з найбільш можливих причин появи позитивної автокореляції залишків – це помилки специфікації моделі у вигляді відсутності важливих змінних, що пояснюють. Щодо від'ємної автокореляції, то вона не призводить до суттєвих помилок, оскільки від'ємна автокореляція проявляється у тенденції до строгого чергування знаків залишків моделі. Але при виведенні формул для обчислення дисперсій і коваріацій коефіцієнтів регресії передбачається випадкове, а не строге чергування знаків залишків. Строге чергування знаків – це також суттєва особливість і вона повинна бути врахована в правильній за специфікацією моделі.

Запитання для самоперевірки

1. Що таке автокореляція?
2. Які види автокореляції розрізняють?
3. У яких даних найчастіше зустрічається автокореляція?
4. Які основні причини автокореляції?
5. Які передумови МНК порушуються при автокореляції?
6. Які наслідки автокореляції?
7. Перерахуйте основні методи виявлення автокореляції.
8. Опишіть схему використання статистики Дарбіна−Уотсона.
9. Перерахуйте методи усунення автокореляції?
10. У чому полягає суть ітераційного методу Кохрана−Оркатта

25. Проблеми інтерпретації параметрів багатофакторної моделі

25.1. Інтерпретація β-коефіцієнтів.

25.2. Інтерпретація параметрів моделей без вільного члена.

25.1. Інтерпретація β-коефіцієнтів

Для проведення факторного аналізу складних характеристик об’єктів в економіці найчастіше використовують лінійну множинну регресію, мета якої – побудувати модель залежності результативної ознаки від багатьох факторів, що впливають на неї. Рівняння лінійної множинної регресії має вигляд:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + \ldots + b_mx_m, \quad (25.1)$$

де $\hat{y}$ – результативна ознака (залежна змінна);

$x_1, x_2, \ldots, x_m$ – фактори (незалежні змінні);

$b_0, b_1, b_2, ..., b_m$ – параметри регресії.

Зупинимося на інтерпретації параметрів регресії. Фахівці в області кореляційно-регресійного аналізу вважають, що стала b_0 не має економічної інтерпретації і геометрично визначає точку перетину гіперповерхні регресії з віссю ординат [11, с. 101]. Параметр b_i вказує середню величину змінення y при зміненні x_i на одну одиницю за умови, що інші фактори залишаються без зміни і закріплені на середньому рівні. У реальних моделях від'ємні значення коефіцієнтів викликають здивування в їх інтерпретації, оскільки часто є протиріччя чинним в економіці закономірностям взаємозв'язку між показниками. Поясненням такого факту є мале число спостережень. У будь-якому випадку рекомендується доповнити рівняння чистої регресії (25.1) рівнянням регресії в стандартизованих величинах:

$$t_y = \beta_1 t_1 + \beta_2 t_2 + ... + \beta_m t_m, \tag{25.2}$$

де $t_y, t_1, t_2, ..., t_m$ – стандартизовані змінні;

$\beta_1, \beta_2, \beta_m$ – стандартизовані коефіцієнти регресії.

Тут з'являється можливість порівняння дії факторів на результативну ознаку. Порівняння відбувається за оцінкою інтенсивності впливу факторів на результативну ознаку, тобто за величиною β_i. Стандартизовані коефіцієнти β_i множинної регресії характеризують швидкість зміни середнього значення результативної ознаки по кожному фактору при незмінних значеннях решти факторів, включених в аналіз. Іншими словами, на яку частину стандартного відхилення змінилося б середнє значення t_y, якщо б значення відповідного t_i збільшилося на стандартне відхилення, а інші фактори залишилися без зміни. При цьому β_i приймають значення в інтервалі [-1, 1]. Однак у розв’язанні реальних економічних задач вони іноді виходять за межі цього інтервалу. Традиційно вважається, що цього не може бути, і коли на практиці зустрічається зазначений ефект, його пояснюють можливими, ще не виявленими помилками в комп'ютерній програмі.

Вперше думка про неможливість появи бета-коефіцієнтів, більших за одиницю за модулем, була висловлена в словнику [35]. Підстава для такого висновку була приведена всього одна – в одновимірному випадку (одна пояснювальна змінна) стандартизований коефіцієнт регресії збігається з парним коефіцієнтом кореляції, який, звичайно, не може бути більшим за одиницю за абсолютною величиною. Наведене обґрунтування не є достатнім для того, щоб робити які-небудь висновки про поведінку β коефіцієнтів у багатовимірному випадку. Однак така думка існує до цих пір.

Розглянемо цю проблему для двовимірного випадку, тобто з двома факторами.

У роботі [5] показано, що при деяких поєднаннях коефіцієнтів кореляції β можуть бути більшими за одиницю за абсолютною величиною. Наприклад, для таких коефіцієнтів кореляції $r_{x_1x_2}=0{,}7;$ $r_{x_1y}=-0{,}35;$ $r_{x_2y}=0{,}35$ в результаті розв'язання системи нормальних рівнянь:

$$\begin{cases} r_{x_1y}=\beta_1+\beta_2 r_{x_1x_2} \\ r_{x_2y}=\beta_1 r_{x_1x_2}+\beta_2 \end{cases} \qquad \begin{cases} -0{,}35=\beta_1+0{,}7\beta_2 \\ 0{,}35=0{,}7\beta_1+\beta_2 \end{cases}$$

виходить $\beta_1=-1{,}167;$ $\beta_2=1{,}167$ (тобто обидва стандартизованих коефіцієнта регресії виявляються більшими за одиницю за модулем). Додатково обчислюємо коефіцієнт детермінації і коефіцієнт множинної кореляції:

$$R^2=\beta_1 r_{x_1y}+\beta_2 r_{x_2y}=(-1{,}167)\cdot(-0{,}35)+1{,}167\cdot 0{,}35=0{,}817\,;$$

$$R=\sqrt{0{,}817}=0{,}904.$$

Два фактори x_1 і x_2 пояснюють близько 82 % мінливості результативної змінної y. Можна обчислити ще коефіцієнти частинної кореляції, коли один з факторів фіксується на своєму середньому значенні:

$$r_{yx_1.x_2}=\beta_1\cdot\sqrt{\frac{1-r^2_{x_1x_2}}{1-r^2_{yx_2}}}=-1{,}167\cdot\sqrt{\frac{1-0{,}7^2}{1-0{,}35^2}}=-0{,}889;$$

$$r_{yx_2.x_1}=\beta_2\cdot\sqrt{\frac{1-r^2_{x_1x_2}}{1-r^2_{yx_1}}}=1{,}167\cdot\sqrt{\frac{1-0{,}7^2}{1-0{,}35^2}}=0{,}889.$$

Пояснення значень β, що вийшли за межі інтервалу [-1, 1], таке. Після стандартизації значення всіх змінних утворюють систему багатовимірних векторів однакові «довжини» – їх скалярні квадрати однакові і рівні n (числу спостережень). Можна нормувати всі ці вектори так, щоб їх довжини дорівнювали одиниці:

$$t_y = \frac{y - \bar{y}}{s_y \sqrt{n}}, t_1 = \frac{x_1 - \bar{x}_1}{s_{x_1} \sqrt{n}}, t_2 = \frac{x_2 - \bar{x}_2}{s_{x_2} \sqrt{n}},$$

де s_y, s_{x_1}, s_{x_2} – середні квадратичні відхилення:

$$s_y = \sqrt{\frac{1}{n}\sum\left(y - \overline{y}\right)^2}; \quad s_{x_1} = \sqrt{\frac{1}{n}\sum\left(x_1 - \overline{x_1}\right)^2}; \quad s_{x_2} = \sqrt{\frac{1}{n}\sum\left(x_2 - \overline{x_2}\right)^2}.$$

Тепер всі скалярні квадрати рівні одиниці $(t_y, t_y) = 1$; $(t_1, t_1) = 1$; $(t_2, t_2) = 1$, а скалярні добутки різних векторів рівні коефіцієнтам кореляції

$$(t_1, t_2) = r_{x_1 x_2};\ (t_1, t_y) = r_{yx_1};\ (t_2, t_y) = r_{yx_2}.$$

Рівняння регресії в цих змінних

$$t_y = t_{y\,p} + e = \beta_1 t_1 + \beta_2 t_2 + e$$

можна трактувати як розкладання одиничного вектора *Y* за системою координат (t_1, t_2, e), де вектор e – вектор похибок (нев'язок), який ортогональний площині одиничних векторів (t_1, t_2): $e \perp t_1, e \perp t_2, e \perp t_{y\,p}$, так як згідно з методом найменших квадратів (МНК) коефіцієнти регресії визначаються із системи нормальних рівнянь, в якій записані умови нормальності, ортогональності вектора похибок до кожного члена моделі (тобто до кожного вектора факторів). Вектор розрахункових значень $t_{y\,p} = \beta_1 t_1 + \beta_2 t_2$ є ортогональна проекція одиничного вектора t_y на площину векторів (t_1, t_2). Скалярний квадрат цього вектора дорівнює коефіцієнту детермінації $(t_{y\,p}, t_{y\,p}) = R^2$, а його довжина – коефіцієнту множинної кореляції $|t_{y\,p}| = R \le 1|$. При такій інтерпретації стандартизовані коефіцієнти регресії β_1, β_2 будуть косокутними проекціями в розкладанні вектора $t_{y\,p}$ (його довжина не більше одиниці) за системою одиничних, але не ортогональних векторів t_1, t_2. Кут між базисними векторами t_1, t_2 дорівнює $arccos(r_{x_1 x_2})$. Наприклад, для $r_{x_1 x_2} = 0{,}7$ цей кут дорівнює 0,795 радіан або 45,57°. Виникає питання, чи можуть проекції одиничного вектора в розкладанні за одиничним базисом бути більше одиниці за абсолютною величиною? Ні, якщо базис ортогональний і можливо, якщо базис не ортогональний.

На рис. 25.1 зображено розкладання вектора $t_{y\,p}$ за базисом t_1, t_2 для $r_{x_1x_2} = 0{,}7$; $r_{yx_1} = -0{,}35$; $r_{yx_2} = 0{,}35$. Для прийнятих значень коефіцієнтів кореляції кут між базисними векторами дорівнює 45,57°; довжина вектора розрахункових значень дорівнює R = 0,904; косокутні проекції цього вектора за одиничним базисом t_1, t_2 опинилися за модулем більшими за одиницю: β_1 = –1,167; β_2 = 1,167.

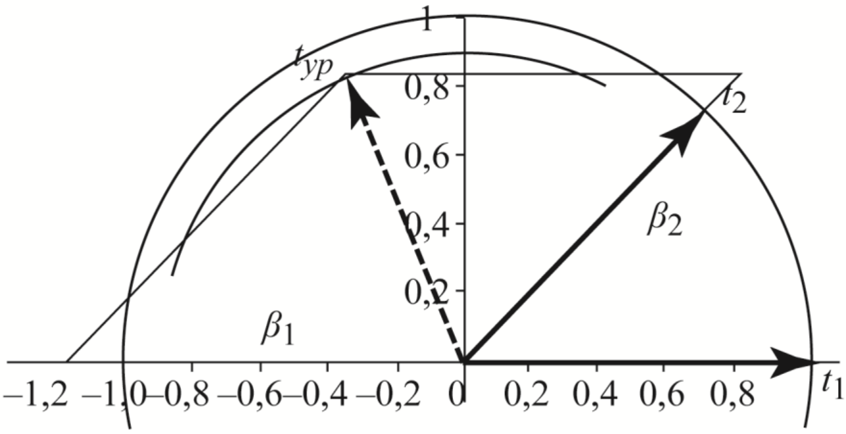


Рис. 25.1. **Косокутні координати вектора $t_{y\,p}$ за базисом t_1, t_2**

Математично тут все правильно. Розглянемо реальний приклад з економічного аналізу залежності ефекту фінансового важеля (y) від коефіцієнта зносу основних засобів (x_1), коефіцієнта оновлення основних засобів (x_2), коефіцієнта вибуття основних засобів (x_3), коефіцієнта обіговості основних засобів (x_4), коефіцієнта загальної (поточної) ліквідності (x_5), коефіцієнта термінової ліквідності (x_6), коефіцієнта абсолютної ліквідності (x_7), коефіцієнта обіговості активів (x_8), коефіцієнта обіговості дебіторської заборгованості (x_9), періоду обіговості дебіторської заборгованості (x_{10}), коефіцієнта обіговості запасів (x_{11}), тривалості обіговості запасів (x_{12}) , коефіцієнта обіговості сукупного капіталу (x_{13}). У табл. 25.1 наведені значення цих показників.

Таблиця 25.1

Значення фінансових показників

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	y
1	2	3	4	5	6	7	8	9	10	11	12	13	14
0,61	0,12	0,14	3,02	9,65	5,59	2,32	1,07	4,23	86,37	3,73	97,91	5,85	0,21
0,61	0,14	0,10	3,23	10,65	6,62	3,14	0,91	3,67	99,59	3,35	109,04	5,40	0,10

Закінчення табл. 25.1

1	2	3	4	5	6	7	8	9	10	11	12	13	14
0,59	0,17	0,08	2,98	12,79	8,31	3,59	0,78	2,99	122,08	3,66	99,76	4,69	0,08
0,66	0,01	0,10	1,49	2,39	1,56	0,21	1,10	7,32	49,85	9,22	39,58	2,01	0,06
0,72	0,01	0,00	1,47	13,89	8,79	1,62	0,95	4,38	83,31	4,49	81,21	2,20	0,03
0,52	0,17	0,10	5,69	0,92	0,43	0,02	2,16	21,72	16,81	9,56	38,17	5,21	0,24
0,91	0,01	0,00	7,58	1,76	0,91	0,01	2,62	11,51	31,72	9,26	39,42	9,12	0,11
0,89	0,04	0,00	8,97	1,27	0,51	0,00	3,38	17,96	20,32	10,74	33,99	9,06	0,12
0,37	0,13	0,02	2,08	2,61	1,98	0,06	1,39	8,69	42,00	-42,19	-8,65	3,10	0,25
0,41	0,09	0,01	4,03	2,46	1,57	0,12	1,94	13,31	27,43	-67,74	-5,39	5,64	0,27
0,43	0,13	0,00	4,98	3,26	2,68	0,29	2,29	12,63	28,91	-58,31	-6,26	7,32	0,48
0,44	0,15	0,07	8,18	4,44	3,75	0,89	2,87	15,15	24,09	-28,54	-12,79	11,29	0,32
0,71	0,02	0,00	1,11	5,30	2,34	1,42	1,08	21,42	17,04	5,85	62,38	1,95	0,13
0,76	0,02	0,23	1,31	6,23	4,80	0,16	0,76	5,26	69,40	-6,17	-59,14	2,43	0,11
0,74	0,02	0,16	1,95	4,65	3,19	0,25	0,92	4,98	73,35	-21,85	-16,71	3,09	0,17
0,73	0,05	0,08	2,76	5,74	3,98	0,05	1,13	4,47	81,68	9,52	38,33	3,23	0,07
0,63	0,02	0,01	0,73	2,53	1,44	0,36	0,58	4,19	87,13	2,29	159,38	1,17	0,03

Побудувавши залежності результативної ознаки t_y від факторів отримали моделі, у яких β коефіцієнти вийшли за межі інтервалу [-1, 1]:

$$t_y = \beta_4 x_4 + \beta_8 x_8;$$

$$t_y = \beta_4 x_4 + \beta_{13} x_{13};$$

$$t_y = \beta_5 x_5 + \beta_6 x_6;$$

$$t_y = \beta_4 x_4 + \beta_8 x_8 + \beta_{13} x_{13}.$$

У табл. 25.2 наведено коефіцієнти парної кореляції для змінних y, x_4, x_8, x_{13}, x_5, x_6.

Таблиця 25.2

Кореляційна матриця

	y	x_4	x_8	x_{13}	x_5	x_6
y	1	0,409861	0,510128	0,517883	-0,34474	-0,27375
x_4	0,409861	1	0,94595	0,945351	-0,37805	-0,351
x_8	0,510128	0,94595	1	0,861387	-0,548	-0,5232

x_{13}	0,517883	0,945351	0,861387	1	-0,21969	-0,17953
x_5	-0,34474	-0,37805	-0,548	-0,21969	1	0,982608
x_6	-0,27375	-0,351	-0,5232	-0,17953	0,982608	1

Для першої моделі маємо $r_{x_4x_8} = 0{,}946;\quad r_{yx_4} = 0{,}410;\quad r_{yx_8} = 0{,}510$.

У результаті розрахунків отримуємо $\beta_4 = -\ 0{,}691$; $\beta_2 = 1{,}164$; $R = 0{,}557$. Один з β-коефіцієнтів виявився більшим за одиницю. На рис. 25.2 наведена геометрична інтерпретація цієї ситуації.

Для другої моделі маємо $r_{x_4x_{13}} = 0{,}945;\quad r_{yx_4} = 0{,}410;\quad r_{yx_{13}} = 0{,}518$. Ці значення мало відрізняються від аналогічних значень попередньої моделі. У результаті розрахунків отримуємо $\beta_4 = -\ 0{,}750$; $\beta_{13} = 1{,}227$; $R = 0{,}573$. Тут також один з β-коефіцієнтів виявився більшим за одиницю. Геометрична інтерпретація дуже схожа на рис. 25.2.

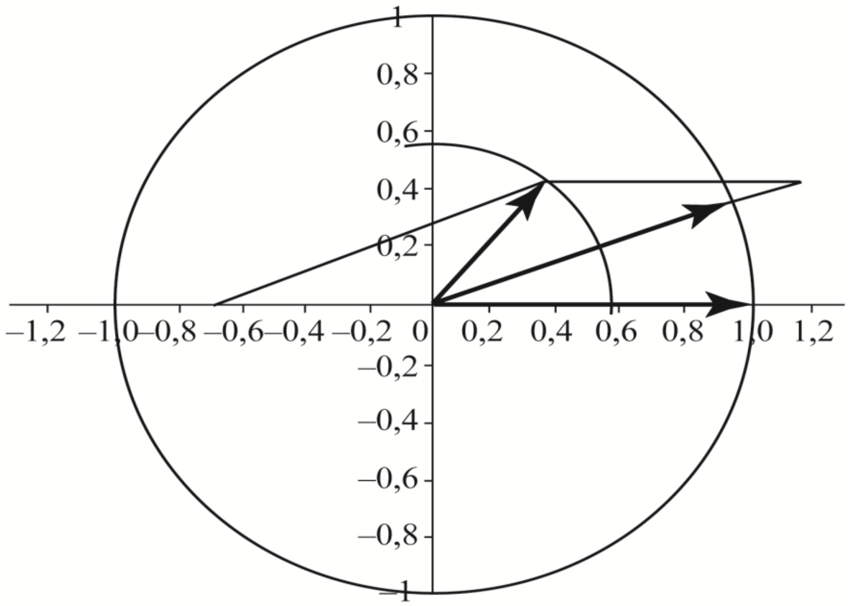


Рис. 25.2. **Геометрична інтерпретація для першої моделі**

Для третьої моделі $r_{x_5x_6} = 0{,}983;\quad r_{yx_5} = -0{,}345;\quad r_{yx_6} = -0{,}274$, звідки отримуємо $\beta_5 = -2{,}197$; $\beta_6 = 1{,}885$; $R = 0{,}491$. Тут обидва β-коефіцієнти виявилися значно більше одиниці за абсолютною величиною. Геометрична інтерпретація наведена на рис. 25.3.

Тіснота кореляційних зв'язків для всіх наведених двовимірних моделей була невисокою з коефіцієнтами множинної кореляції $R = 0{,}49 \div 0{,}57$ (моделі пояснювали всього від $R^2 \approx 24$ % до $R^2 \approx 32$ % повної мінливості).

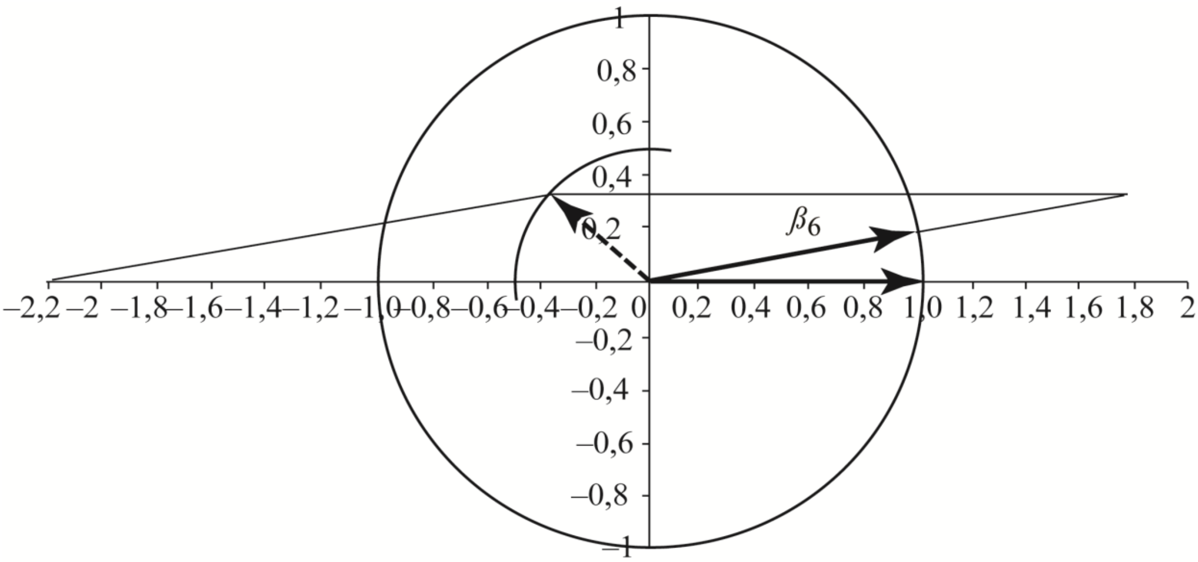


Рис. 25.3. **Геометрична інтерпретація для третьої моделі**

Крім аналізу двовимірних моделей була складена багатовимірна модель, в яку були включені всі зазначені вище фактори. Алгоритм покрокової регресії вибракував як незначущі змінні x_5, x_6 і в результаті була отримана тривимірна модель $t_y = \beta_4 x_4 + \beta_8 x_8 + \beta_{13} x_{13}$ з коефіцієнтом множинної кореляції $R = 0{,}779$ (іншими словами, тривимірна модель пояснила близько $R^2 \approx 61$ % повної мінливості даних). Модель значима в цілому і складається лише із значущих членів. Усі $\beta -$коефіцієнти і для цього випадку виявилися більшими за одиницю за абсолютною величиною.

У табл. 25.3 наведено фрагмент кореляційної матриці з табл. 25.2 для змінних y, x_4, x_8, x_{13}.

Таблиця 25.3

Кореляційна матриця для y, x_4, x_8, x_{13}

	$y = x_{14}$	x_4	x_8	x_{13}
x_4	0,409861	1	0,94595	0,945351
x_8	0,510128	0,94595	1	0,861387
x_{13}	0,517883	0,945351	0,861387	1

Стандартизовані коефіцієнти регресії визначаються із системи нормальних рівнянь:

$$0{,}410 = \beta_4 + 0{,}946\cdot\beta_8 + 0{,}945\cdot\beta_{13},$$
$$0{,}510 = 0{,}946\cdot\beta_4 + \beta_8 + 0{,}861\cdot\beta_{13},$$
$$0{,}518 = 0{,}945\cdot\beta_4 + 0{,}861\cdot\beta_8 + \beta_{13}.$$

Розв'язком даної системи є числа $\beta_4 = -2{,}870$; $\beta_8 = 1{,}713$; $\beta_{13} = 1{,}756$, тобто всі $\beta-$коефіцієнти виявилися істотно більшими за одиницю за абсолютною величиною. Коефіцієнт детермінації і коефіцієнт множинної кореляції:

$$R^2 = -2{,}870\cdot0{,}410 + 1{,}713\cdot0{,}510 + 1{,}756\cdot0{,}518 = 0{,}606;$$
$$R = 0{,}779.$$

Значимість моделі в цілому:

$F = \dfrac{R^2}{1-R^2}\cdot\dfrac{m}{n-m-1} = \dfrac{0{,}606}{1-0{,}606}\cdot\dfrac{3}{17-3-1} = 6{,}685$, значення більше табличного $F_{0,01}(3;13) = 5{,}74$, звідки випливає висновок про значущість моделі в цілому. Рівняння залежності ефекту фінансового важеля (y) від коефіцієнта обіговості основних засобів (x_4), коефіцієнта обіговості активів (x_8), коефіцієнта обіговості сукупного капіталу (x_{13}) (у кожного коефіцієнта регресії в дужках наведені значення статистик Стьюдента):

$$t_y = -2{,}871x_4 + 1{,}713x_8 + 1{,}756x_{13}.$$
$$(-3{,}3) \qquad (3{,}0) \qquad (3{,}1)$$

Таким чином, модель значуща в цілому (за критерієм Фішера) і складається тільки із значущих (за критерієм Стьюдента) членів, однак залишаються $\beta-$коефіцієнти більші за одиницю за модулем.

У сучасних статистичних пакетах зазвичай відразу отримують рівняння чистої регресії:

$$y = -\,0{,}0733 - 0{,}133\cdot x_4 + 0{,}242\cdot x_8 + 0{,}0717\cdot x_{13}.$$
$$(-1{,}3) \quad (-3{,}3) \qquad (3{,}0) \qquad (3{,}1)$$

Відповідно до моделі при збільшенні коефіцієнта обіговості основних засобів (x_4) значення ефекту фінансового важеля зменшується на 0,133 одиниці. У цьому прикладі явно видно помилковість традиційного трактування коефіцієнтів регресії. Дійсно, тут всі три фактори пов'язані дуже тісними кореляційними зв'язками $r_{x_4x_8} = 0{,}946$; $r_{x_4x_{13}} = 0{,}945$; $r_{x_8x_{13}} = 0{,}861$ і тому вони не можуть варіювати незалежно один від одного. Фактично це різні форми однієї і тієї ж причини з додатною кореляцією з результативною ознакою $r_{yx_4} = 0{,}410$; $r_{yx_8} = 0{,}510$ $r_{yx_{13}} = 0{,}518$. Таким чином, традиційна інтерпретація справедлива лише при незалежності змінних, що пояснюють, коли кожен аргумент може змінюватися незалежно від інших. У регресійному аналізі, як правило, фактори не є незалежними, між ними майже завжди є кореляційні зв'язки, саме тому для визначення коефіцієнтів регресії складається і розв’язується система нормальних рівнянь.

Рівняння регресії становить компактний опис вихідних даних. Методом найменших квадратів отримано вираз з найменшими похибками, який призначений для найкращого відтворення даних. Будуючи ж моделі в економічному аналізі, відбувається використання даного виразу не зовсім за призначенням і при цьому робляться висновки при варіюванні факторів в інших діапазонах порівняно з вихідними даними. Так, тут з'являється аргумент − відсутність альтернативних моделей.

У задачі визначенні залежності ефекту фінансового важеля (y) від факторів з тісними мультиколінеарними зв'язками проблема вирішується переходом у систему головних компонент. У табл. 25.4 обчислені власні числа і власні вектори для кореляційної матриці 3-го порядку щодо x_4, x_8, x_{13}:

Таблиця 25.4

Власні числа λ і власні вектори U_1, U_2, U_3

№	U_1	U_2	U_3
1	0,589	–0,002	0,808
2	0,571	–0,706	–0,418
3	0,571	0,708	–0,414
λ	2,836	0,139	0,026
%	94,5%	4,6%	0,9%
Σ	94,5%	99,1%	100%

Компоненти складаються як лінійні комбінації вихідних стандартизованих змінних з коефіцієнтами, записаними у власних векторах. Якщо округлити цифри в табл. 25.4, то виходить, що компоненти в цьому прикладі будуть близькі до таких виразів:

$$F_1 = t_4 + t_8 + t_{13},$$

$$F_2 = -t_8 + t_{13},$$
$$F_1 = 2t_4 - t_8 - t_{13}.$$

Компоненти F_1, F_2, F_3 взаємно ортогональні (незалежні), центровані, але не нормовані. Нормовані (стандартизовані) компоненти позначимо Φ_1, Φ_2, Φ_3. При цьому перша (головна) компонента Φ_1 буде близька до середнього арифметичного вихідних стандартизованих змінних.

У модель слід включати не всі, а тільки головні компоненти з дисперсіями (власними числами), більшими за одиницю. Крім того, в сумі головні компоненти повинні пояснювати близько 80 % загальної мінливості. У нашій задачі цим вимогам задовольняє одна перша (головна) компонента, яка пояснює
94,5 % повної мінливості. Інші дві компоненти практично не варіюють і становлять малі помилки.

При врахуванні в моделі тільки головних компонент відкидаються малі помилки, крім всіх подальших вибраковок, які можуть бути зроблені на підставі статистичних критеріїв.

Маємо наступне рівняння:

$$t_y = 0{,}492\Phi_1 = 0{,}172t_1 + 0{,}167t_2 + 0{,}167t_3,$$

яке пояснює всього $R^2 = 0{,}492^2 \approx 24{,}2$ % мінливості $У$. Порівняно з одномірними моделями:

$$t_y = 0{,}410t_1,$$
$$t_y = 0{,}510t_2,$$
$$t_y = 0{,}518t_3,$$

в яких весь сукупний ефект приписується одному фактору, в трьохфакторній моделі (яка отримана за допомогою методу головних компонент) сукупний ефект розділений на три приблизно однакові частини.

Існує небажаний ефект включення в модель малих похибок, після чого різко змінюються значення коефіцієнтів регресії («проплескування»), що іноді можна запобігти, якщо при послідовному підключенні чергової змінної аналізувати залишкові дисперсії кожної змінної, яка не включена в модель [8].

Однак алгоритм послідовного підключення-виключення не доведений до завершення, оскільки він ніяк не враховує можливе вичерпування мінливості пояснювальних змінних. Існуюче стандартне програмне забезпечення не гарантує запобігання безглуздих результатів.

Припущення, що факт $|\beta| > 1$ пов'язаний з мультиколінеарністю доповнюється припущенням про неправильне визначення напрямку причинно-наслідкових зв'язків, коли як результативна обрана не та змінна, іншими словами прийнята неправильна специфікація моделі.

Якщо допустити, що результативною ознакою може бути будь-який з факторів (табл. 25.1), то можна знайти 66 двохфакторні моделі, для яких $\beta-$коефіцієнти будуть більше одиниці за модулем. У табл. 25.5 знайдені комбінації (y, x_1, x_2), розташовані за стовпцями в порядку зростання коефіцієнта множинної кореляції R (від 0,29 для першої комбінації до 0,99 для останньої 66-ї комбінації). При цьому зменшувався коефіцієнт кореляції між чинниками $r_{x_1x_2}$ (від 0,98 до 0,25). Виділено комбінації, для яких хоча б один з $\beta-$коефіцієнтів виявився більше 2 (за модулем). Для комбінації (12, 5, 6) обидва $\beta-$коефіцієнти виявилися більшими 3 (за модулем).

Таблиця 25.5

Комбінації (y, x_1, x_2) з $|\beta| > 1$

(2, 5, 6)	(11, 9, 10)	(12, 9, 10)	(9, 8, 13)	**(10, 4, 8)**	(4, 8, 9)
(13, 5, 6)	**(11, 5, 6)**	(6, 4, 8)	(5, 9, 10)	(9, 10, 12)	(13, 4, 6)
(1, 5, 6)	(3, 4, 8)	(5, 4, 8)	(2, 1, 11)	(13, 8, 10)	(4, 5, 8)
(1, 4, 13)	(7, 4, 13)	(7, 4, 8)	(7, 9, 10)	(13, 7, 8)	(4, 6, 8)
(3, 5, 6)	(14, 4, 8)	**(12, 5, 6)**	(10, 8, 13)	(9, 7, 10)	(4, 7, 8)
(2, 4, 8)	(5, 4, 13)	(5, 8, 13)	(7, 5, 6)	(4, 13, 15)	(4, 8, 10)
(11, 4, 13)	(14, 4, 13)	(6, 8, 13)	(13, 8, 9)	(13, 4, 10)	(5, 6, 9)
(2, 4, 13)	(6, 4, 13)	(7, 8, 13)	(9, 6, 10)	(13, 4, 8)	(6, 5, 15)
(11, 4, 8)	(10, 4, 13)	(9, 4, 8)	(9, 5, 10)	(8, 4, 13)	(6, 5, 7)
(14, 5, 6)	(9, 4, 13)	(6, 9, 10)	(13, 5, 8)	(13, 4, 9)	(6, 5, 11)
(12, 4, 8)	(9, 5, 6)	(11, 1, 2)	(13, 6, 8)	(13, 4, 5)	(6, 5, 12)

Серед цих комбінацій інтерес становлять x_6, x_5, x_{11}, де як результативна обрана змінна $y = x_6$. Мультиколінеарності тут немає, так як $r_{x_5x_{11}} = 0{,}252$. У табл. 25.5 є 15 комбінацій з малими коефіцієнтами кореляції $|\ r_{x_ix_j}\ | < 0{,}6$, і в усіх цих моделях присутній ефект, що досліджується. Таким чином, встановлено, що поява $\beta-$коефіцієнтів, більших за модулем одиниці, не пов'язана з наявністю тісних мультиколінеарних зв'язків.

У всіх знайдених комбінаціях неприємний ефект зникає, якщо замінити результативну змінну на і t_{15} з цього ж списку. Так, на рис. 25.3 зображена геометрична інтерпретація для t_5, t_6, t_{14}, де як результативна обрана остання змінна $y = x_{14}$. Обидва стандартизованих коефіцієнта регресії виявилися істотно більшими за одиницю (за модулем): $\beta_5 = -2{,}197$; $\beta_6 = 1{,}885$. Тіснота зв'язку для цієї моделі невисока – R = 0,491. Якщо ж прийняти як результативну змінну X_5, виходить доброякісна модель $\beta_6 = 0{,}960$; $\beta_{15} = -\ 0{,}082$ з високою тіснотою зв'язку R = 0,986. На рис. 25.4 зображена геометрична інтерпретація результатів цих розрахунків.

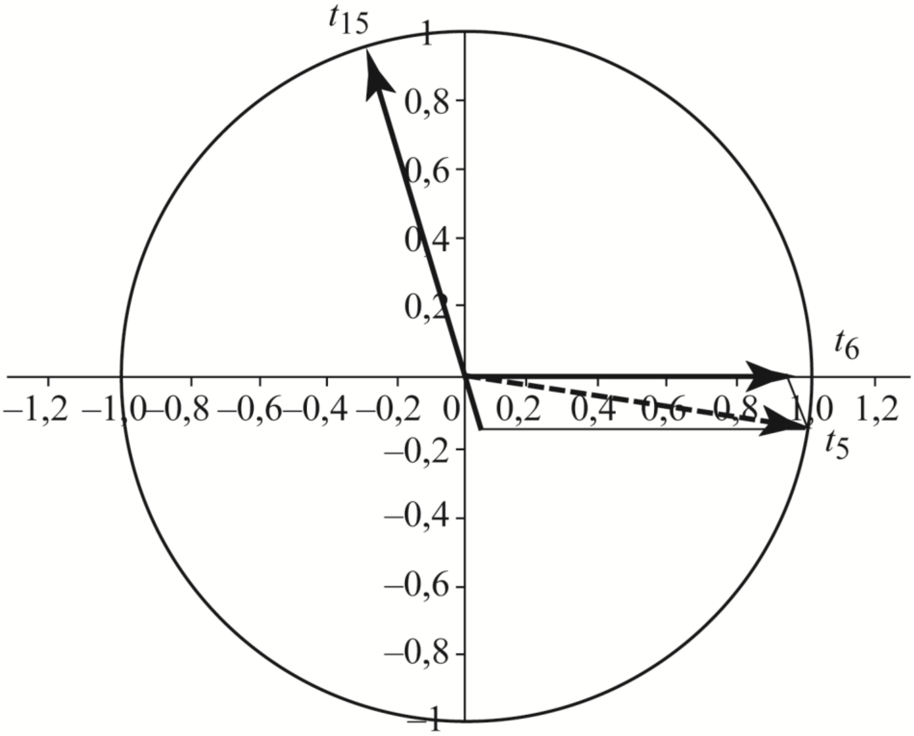


Рис. 25.4. **Модель** $t_5 = \beta_6 t_6 + \beta_{15} t_{15}$

Метод найменших квадратів є основним інструментом для опису кореляційних зв'язків. Альтернативи цьому методу поки не запропоновано. Однак, у регресійному аналізі за дослідником залишається вибір результативної ознаки. Помилкове уявлення структури моделі при н t_6 вильному прийнятті деякої змінної як результативної може мати різні наслідки. Поява значень $\beta-$коефіцієнтів, більших за одиницю за модулем, – t_5 ін з таких наслідків, і наявність цього ефекту треба розцінювати як явну ознаку неправильного визначення структури (специфікації) моделі.

Таким чином, обчислюючи регресійні моделі для економічного аналізу слід знати, що може вийти рівняння з $\beta-$коефіцієнтами за модулем, більшими за одиницю. Причому статистичні критерії значимості окремих членів моделі підтвердять її якість; вони не допоможуть запобігти включення в модель сторонніх перешкод, у результаті чого може бути отримано рівняння регресії, позбавлене будь-якого сенсу. За наявності мультиколінеарності можна рекомендувати обов'язкове застосування методу головних компонент для запобігання небажаних викривлень результатів регресійного аналізу. Однак при отриманні моделі з $\beta-$коефіцієнтами за модулем більшими за одиницю рекомендується ретельно вивчити напрямок причинно-наслідкових зв'язків досліджуваних явищ і уточнити специфікацію моделі.

25.2. Інтерпретація параметрів моделей без вільного члена

Розглянемо одну з проблем множинної регресійної моделі, пов'язаної зі змістом вільного члена. Вільний член лінійної багатофакторної моделі $y = b_0 + b_1 x_1 + b_2 x_2$ відіграє ту ж роль, що і в однофакторному випадку – він потрібен у моделі для того, щоб при середніх значеннях $(\bar{x}_1, \bar{x}_2)$ змінних, що пояснюють, було отримане середнє значення результативної ознаки ($\bar{y}$).

Прикладів регресійних моделей, у яких вільний член дорівнює нулю, багато в економіці: гіпотеза безперервного (постійного) прибутку Мілтона−Фрідмана, згідно з якою постійне споживання пропорційно постійному прибутку; в теорії аналізу витрат стверджується, що змінні витрати виробництва пропорційні випуску продукції; в монетарній теорії рівень зміни цін тобто рівень інфляції пропорційний рівню зміни пропозиції грошей [1]. З позиції математики моделі без вільного члена з'являються у випадках, коли відомо, що лінія регресії обов'язково повинна проходити через фіксований вузол, найчастіше через початок координат. Наприклад, квадратична залежність $y = b_1 x + b_2 x^2$ завжди проходить через початок координат при будь-яких значеннях параметрів.

Моделі без вільного члена також з'являються у процесі використання зваженого методу найменших квадратів, коли для подолання наслідків гетероскедастичності всі рівняння спочатку множаться на вагову функцію.

Наприклад, нелінійну модель $y = \dfrac{1}{b_0 + b_1 x}$ можна привести до лінійного вигляду (щодо параметрів) функціональним перетворенням залежної змінної $1/y = b_0 + b_1 x$. Однак відомо, що при функціональних перетвореннях залежної змінної майже завжди з'являється істотна гетероскедастичність залишків моделі, що здатна викривити значення оцінок параметрів моделі. Вагова функція призначена надати більшу вагу надійним спостереженням і меншу нестабільним даним з великою мінливістю. У даному прикладі вагова функція дорівнює $g = y^2$. Помножимо лінійне рівняння (в координатах x, $\dfrac{1}{y}$) на цю вагову функцію і отримаємо модель без вільного члена $y = b_0 y^2 + b_1 x y^2$, яка лінійна відносно параметрів і не має гетероскедастичності. МНК-оцінки параметрів такої моделі – доброякісні і не мають систематичних похибок (зміщень).

Однак відсутність в моделі вільного члена призводить до того, що сума залишків не буде дорівнювати нулю. Всі статистичні характеристики якості моделі – коефіцієнт детермінації, статистики Фішера і Стьюдента, обчислені за стандартними формулами будуть тепер неправильними. Моделі з відсутнім або нульовим перетином можна використовувати в деяких випадках, однак тут необхідно пам'ятати декілька специфічних моментів. По-перше, $\sum e_i$, яка завжди дорівнює нулю в моделі з наявною величиною перетину (загальноприйнята модель), не завжди повинна дорівнювати нулю для регресії, що проходить через початок координат. По-друге, R^2 – коефіцієнт детермінації, який завжди додатний у загальноприйнятій моделі, в деяких випадках може перетворитися у від’ємний для регресії, що проходить через початок координат. Цей випадок отримаємо тому, що R^2 неоднозначно передбачає, що перетин включається в модель. Тому R^2, отриманий за правилами, може не відповідати регресійній моделі, яка проходить через початок координат.

Саме через ці специфічні особливості моделі регресії, що проходить через початок координат, користуватися слід обережно. Якщо немає достатньої впевненості в її використанні, краще використовувати загальноприйняту модель з наявним перетином. Це має дві переваги. По-перше, якщо включаємо в модель величину перетину, але вона статистично не значима для всіх практичних цілей, маємо регресію, що проходить через початок координат. По-друге, і це головне, якщо фактично перетин існує, але необхідно наполягти на тому, щоб використовувати регресію, що проходить через початок координат, то допущена помилка специфікації, тим самим порушуючи передумову 5 класичної моделі лінійної регресії [15, с. 133 – 134].

Розглянемо детально модель без вільного члена. Для оцінки параметрів моделей без вільного члена $y = b_1 x_1 + b_2 x_2 + e$ пропонується додати в метод найменших квадратів обов'язкову умову $\sum e = 0$. Маємо задачу на умовний екстремум: необхідно знайти мінімум суми квадратів залишків моделі за умови рівності нулю їх суми:

$$\begin{cases} \sum e^2 \to \min, \\ \sum e = 0. \end{cases}$$

Складемо функцію Лагранжа, де множник Лагранжа позначено 2λ:

$$F = \sum e^2 - 2\lambda \sum e,$$

і прирівняємо до нуля її частинні похідні за всіма змінними λ, b_1, b_2:

$$\begin{cases} \frac{\partial F}{\partial \lambda} = 2\sum e = 0 \\ \frac{\partial F}{\partial b_1} = 2\sum e \frac{\partial e}{\partial b_1} - 2\lambda \sum \frac{\partial e}{\partial b_1} = 0 \\ \frac{\partial F}{\partial b_2} = 2\sum e \frac{\partial e}{\partial b_2} - 2\lambda \sum \frac{\partial e}{\partial b_2} = 0 \end{cases}$$

Необхідно врахувати, що

$$e = y - b_1 x_1 - b_2 x_2,$$

$$\frac{\partial e}{\partial b_1} = -x_1; \quad \frac{\partial e}{\partial b_2} = -x_2 .$$

Звідки отримаємо таку систему зв'язків за залишками моделі:

$$\begin{cases} \sum e = 0, \\ \sum e x_1 = \lambda \sum x_1, \\ \sum e x_2 = \lambda \sum x_2 . \end{cases}$$

Помножимо рівняння $y = b_1 x_1 + b_2 x_2 + e$ на кожну змінну (1, x_1, x_2, y, e) і підсумуємо ці вирази за всіма спостереженнями:

$$\sum y = b_1 \sum x_1 + b_2 \sum x_2 + \sum e,$$

$$\sum y x_1 = b_1 \sum x_1 x_1 + b_2 \sum x_1 x_2 + \sum e x_1,$$

$$\sum y x_2 = b_1 \sum x_1 x_2 + b_2 \sum x_2 x_2 + \sum e x_2,$$

$$\sum yy = b_1\sum yx_1 + b_2\sum yx_2 + \sum ye\ ,$$
$$\sum ye = b_1\sum ex_1 + b_2\sum ex_2 + \sum ee\ .$$

Врахуємо зв'язки, покладені на залишки, і з перших трьох сумарних виразів отримаємо таку систему рівнянь для оцінки параметрів моделі без вільного члена:

$$\begin{cases} \sum y = b_1\sum x_1 + b_2\sum x_2\ , \\ \sum yx_1 = \lambda\sum x_1 + b_1\sum x_1x_1 + b_2\sum x_1x_2\ , \\ \sum yx_2 = \lambda\sum x_1 + b_1\sum x_1x_2 + b_2\sum x_2x_2\ . \end{cases}$$

Порівняємо отриману систему рівнянь з системою нормальних рівнянь для оцінки параметрів моделі з вільним членом $y = b_0 + b_1x_1 + b_2x_2$:

$$\begin{cases} \sum y = nb_0 \quad b_1\sum x_1 + b_2\sum x_2\ , \\ \sum yx_1 = b_0\sum x_1 + b_1\sum x_1x_1 + b_2\sum x_1x_2\ , \\ \sum yx_2 = b_0\sum x_1 + b_1\sum x_1x_2 + b_2\sum x_2x_2\ . \end{cases}$$

Єдина відмінність нової системи – в першому рівнянні відсутній член nb_0 .

З останніх двох сумарних виразів можна отримати:

$$\sum y^2 = b_1\sum yx_1 + b_2\sum yx_2 + b_1\sum ex_1 + b_2\sum ex_2 + \sum e^2 =$$
$$= b_1\sum yx_1 + b_2\sum yx_2 + \lambda(b_1\sum x_1 + b_2\sum x_2) + \sum e^2 =$$
$$=\lambda\sum y + b_1\sum yx_1 + b_2\sum yx_2 + \sum e^2\ ,$$

що дуже нагадує аналогічний вислів для моделі з вільним членом (з заміною λ на b_0).

З історії науки відомо, що в 1927 році економіст Пол Дуглас, розглядаючи діаграми логарифмів капітальних витрат (K), обсягів випуску продукції (Y) і витрат на працю (L), виявив, що відстані від точок графіка $\ln Y$ до точок графіків $\ln K$, $\ln L$ складають однакову пропорцію для всіх спостережень з 1900 по 1922 рр. На підставі цього факту математик Чарльз Кобб показав, що така особливість має місце для залежності $Y = K^bL^{1-b}$, що має зараз назву Кобба−Дугласа.

Своє відкриття П. Дуглас зробив лише тому, що для 1989 року він прийняв значення всіх показників K, Y, L за 1 (100 %), тобто задавши фіксований вузол, де перетинаються графіки залежностей $\ln K$, $\ln Y$, $\ln L$ за часом ($x = t - 1989$).

Фактично він побачив лінійність трендів логарифмів показників (рис. 25.5) і цього факту було достатньо, щоб зробити такі висновки.

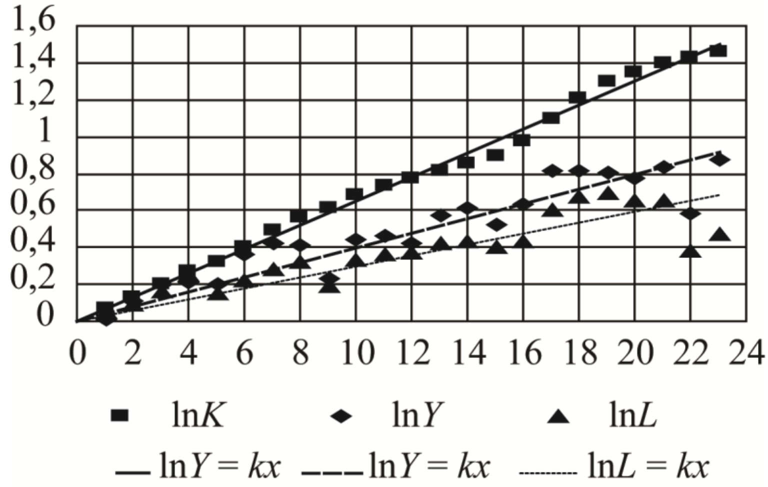


Рис. 25.5

Розглянемо детально апроксимацію *lnL* = *f*(*x*). Дані П. Дугласа представлені в табл. 25.6.

Таблиця 25.6

Вихідні дані *K, Y, L*

№	*t*	*K*	*Y*	*L*	*lnK*	*lnY*	*lnL*
1	1900	1,07	1,01	1,05	0,0677	0,0100	0,0488
2	1901	1,14	1,12	1,10	0,1310	0,1133	0,0953
3	1902	1,22	1,22	1,18	0,1989	0,1989	0,1655
4	1903	1,31	1,24	1,29	0,2700	0,2151	0,2546
5	1904	1,38	1,22	1,16	0,3221	0,1989	0,1484
6	1905	1,49	1,43	1,25	0,3988	0,3577	0,2231
7	1906	1,63	1,52	1,33	0,4886	0,4187	0,2852
8	1907	1,76	1,51	1,38	0,5653	0,4121	0,3221
9	1908	1,85	1,26	1,21	0,6152	0,2311	0,1906
10	1909	1,98	1,55	1,40	0,6831	0,4383	0,3365
11	1910	2,08	1,59	1,44	0,7324	0,4637	0,3646
12	1911	2,16	1,53	1,45	0,7701	0,4253	0,3716
13	1912	2,26	1,77	1,52	0,8154	0,5710	0,4187
14	1913	2,36	1,84	1,54	0,8587	0,6098	0,4318
15	1914	2,44	1,69	1,49	0,8920	0,5247	0,3988
16	1915	2,66	1,89	1,54	0,9783	0,6366	0,4318
17	1916	2,98	2,25	1,82	1,0919	0,8109	0,5988
18	1917	3,35	2,27	1,96	1,2090	0,8198	0,6729
19	1918	3,66	2,23	2,00	1,2975	0,8020	0,6931
20	1919	3,87	2,18	1,93	1,3533	0,7793	0,6575
21	1920	4,07	2,31	1,93	1,4036	0,8372	0,6575
22	1921	4,17	1,79	1,47	1,4279	0,5822	0,3853
23	1922	4,31	2,40	1,61	1,4609	0,8755	0,4762

Якщо апроксимувати залежність звичайною лінійною моделлю:

$$lnL = b_0 + b_1 x + e,$$

то будуть отримані такі МНК-оцінки параметрів $b_0 = 0{,}08299$; $b_1 = 0{,}02435$. Сума залишків моделі дорівнює нулю $\sum e = 0$, сума квадратів залишків $\sum e^2 = 0{,}1829$. Виконується співвідношення $s_y^2 = s_p^2 + s_e^2$, де позначено $y = lnL$, $y_p = b_0 + b_1 x$, s_p^2, – дисперсія обчислених значень. У числах маємо розкладання загальної дисперсії на дві складові $0{,}03404 = 0{,}02608 + 0{,}00796$, звідки знаходимо коефіцієнт детермінації $R^2 = 0{,}7664$. Через фіксований вузол лінія регресії не проходить (рис. 25.5).

Якщо прийняти модель без вільного члена $lnL = kx + e$, то отримаємо $k = 0{,}02965$. Графік залежності проходить через фіксований вузол (рис. 25.5). Сума залишків цієї моделі не дорівнює нулю $\sum e = 0{,}4467$, сума квадратів залишків $\sum e^2 = 0{,}2200$ (це більше порівняно з попередньою моделлю). Не виконується співвідношення $s_y^2 = s_p^2 + s_e^2$, тому не можна обчислити коефіцієнт детермінації, що показує частину мінливості, яка пояснюється моделлю. Також не можна обчислити інші статистичні характеристики.

Оцінимо параметри моделі без вільного члена іншим методом. Слід скласти систему рівнянь для визначення k і λ:

$$\sum y = k\sum x,$$

$$\sum yx = \lambda\sum x + k\sum x^2,$$

$$\sum y^2 = \lambda\sum y + k\sum yx + \sum e^2.$$

З першого рівняння цієї системи, яке еквівалентне умові $\sum e = 0$, отримаємо нову оцінку параметра k:

$$k = \frac{\sum y}{\sum x} = \frac{8{,}6289}{276} = 0{,}03126.$$

Для однофакторної моделі за даним методом пропонується визначати єдиний параметр моделі k з обов'язковою умовою $\sum e = 0$, а не з $\sum e^2 \to min$. З другого рівняння додатково знаходимо $\lambda = -\ 0{,}02536$. Третє співвідношення можна використовувати для контролю всіх обчислень. Сума залишків цієї моделі дорівнює нулю $\sum e = 0$, сума квадратів залишків $\sum e^2 = 0{,}2313$ (це більше порівняно з попередніми моделями). Співвідношення $s_y^2 = s_p^2 + s_e^2$ не виконується. Графіки усіх трьох моделей представлені на рис. 25.6.

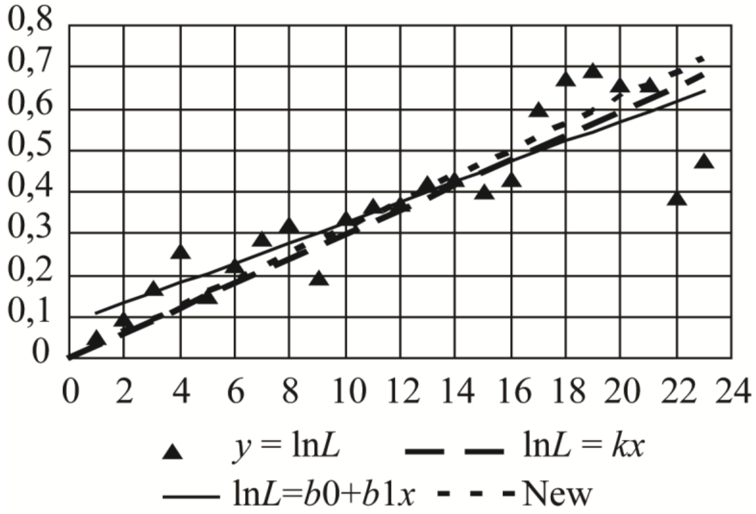


Рис. 25.6. Апроксимація часового тренду для *lnL*

Отже, даний метод не має особливих переваг порівняно зі стандартним МНК, єдине, що досягнуто – це виконання умов рівності нулю суми залишків моделі.

Запитання для самоперевірки

1. Для яких цілей рекомендується додатково до «чистої регресії» будувати регресійну модель з β-коефіцієнтами?

2. У яких межах змінюються β-коефіцієнти?

3. Чи можуть в економіці зустрічатися регресійні моделі в стандартизованих змінних з $|\beta| \geq 1$?

4. Які причини зумовлюють $|\beta| \geq 1$?

5. Які існують рекомендації у процесі побудови лінійної регресійної моделі, щоб $|\beta| < 1$?

6. Наведіть приклад класичних економетричних моделей в економіці без вільного члена.

7. Через який вузол проходить лінія регресії без вільного члена?

8 . Які математичні проблеми існують у регресійній моделі без вільного члена?

9. Які існують рекомендації у вирішенні математичних проблем у регресійній моделі без вільного члена?

26. Узагальнені схеми регресійного аналізу

26.1. Деякі альтернативні схеми регресійного аналізу.

26.2. Моделі з *dummy*-змінними.

26.3. Новітні (*Advanced*) методи регресійного аналізу.

26.1. Деякі альтернативні схеми регресійного аналізу

Удосконалення регресійного аналізу в аспекті методики оцінки параметрів моделей пішло за двома напрямками. По-перше, ускладнюється стандартна процедура МНК так, щоб вона була в змозі долати ті чи інші труднощі (метод покрокової регресії, метод псевдоберненої матриці, метод SVD, двокрокова і трикрокова процедура МНК, узагальнений метод НК). По-друге, робляться спроби взагалі відмовитися від МНК і замінити його на зовсім іншу процедуру оцінки параметрів. У деяких пропозиціях вноситься стільки змін і доповнень в стандартний МНК, що, по суті, такі пропозиції слід також розглядати як нові (альтернативні) схеми регресійного аналізу.

Метод ортогональної (нормальної) регресії

Назва методу не має нічого спільного ні з ортогоналізацією змінних (подібно методу ортогоналізації Грамма−Шмідта або методу головних компонент), ні з класичним визначенням регресійної (кореляційної) залежності.

Згідно з першою передумовою регресійного аналізу в стандартному МНК необхідно одну зі змінних обов'язково позначити результативною ознакою і всі випадкові похибки приписати тільки цій змінній. У результаті виходить одна з декількох (абсолютно різних) спряжених регресійних моделей. Ця перша передумова МНК неприйнятна у випадку рівноправності всіх змінних, коли сумісна їх мінливість визначається тим, що вони є наслідками однієї і тієї ж причини. У методі ортогональної регресії пропонується випадкову помилку вимірювати ортогонально (по нормалі) до поверхні регресії, а не уздовж якої-небудь координатної осі.

Усі подальші міркування проводимо в просторі стандартизованих змінних, так як при зміні масштабів змінних умова ортогональності порушується.

Щоб знайти відстань ε від гіперплощини:

$$a_0 x_0 + a_1 x_1 + \ldots + a_m x_m = 0$$

до довільної точки, треба нормувати всі коефіцієнти (привести рівняння площини до нормальної форми) і підставити в це рівняння координати точки, що нас цікавить:

$$\begin{cases} a_0 x_0 + a_1 x_1 + \ldots + a_m x_m = \varepsilon, \\ a_0^2 + a_1^2 + \ldots + a_m^2 = 1. \end{cases}$$

Тут змінна y позначена через x_0, щоб не виділяти її.

Отримаємо задачу на умовний екстремум – необхідно знайти мінімум суми квадратів похибок $[\varepsilon^2]$ за умови нормування коефіцієнтів:

$$\sum_{i=0}^{m} a_i^2 = 1.$$

Перетворимо суму квадратів похибок, враховуючи, що для стандартизованих змінних $\lfloor x_i x_j \rfloor = n \cdot r_{x_i x_j}$:

$$\left[\varepsilon^2\right] = \left[\left(\sum_{i=0}^{m} a_i\, x_i\right)^2\right] = \left[\sum_{i=0}^{m}\sum_{j=0}^{m} a_i\, a_j x_i x_j\right] = n\sum_{i=0}^{m}\sum_{j=0}^{m} a_i\, a_j r_{x_i x_j} \to \min.$$

Складаємо функцію Лагранжа:

$$F = \sum_{i=0}^{m}\sum_{j=0}^{m} r_{x_i x_j} a_i\, a_j - \lambda\left(\sum_{i=0}^{m} a_i^2 - 1\right),$$

і прирівнюємо до нуля її частинні похідні по a_i. Отримуємо задачу на власні значення і власні вектори (розширеної) кореляційної матриці (включаючи кореляції з y):

$$\begin{cases} \sum_{j=0}^{m} r_{x_i x_j} a_j = \lambda a_i, \\ i = \overline{0, m}. \end{cases}$$

Таким чином отримали коефіцієнти a_i.

Гребнева регресія

Оцінки МНК параметрів моделі є спроможними, незміщеними та ефективними; їм відповідає мінімальне значення суми квадратів похибок $[e^2]$. Але за наявності мультиколінеарності МНК-оцінки стають нестійкими. Запропоновано (Гоерл) перед оберненням матриці $X^T X$ додати до її діагональних елементів малі числа k, що запобігає виродженню матриці, збільшує її визначник і зменшує дисперсії параметрів.

Нові оцінки (ридж-оцінки) параметрів обчислюються за формулами:

$$\boldsymbol{b} = (X^T X + kI)^{-1} X^T Y .$$

Від такої операції (збільшення діагональних елементів матриці X^TX на мале число k) сума квадратів похибок збільшується, але типовий характер залежності залишкової дисперсії від k подібний графіку на рис. 26.1 – до якоїсь межі $k < k^*$ залишкова дисперсія зростає незначно.

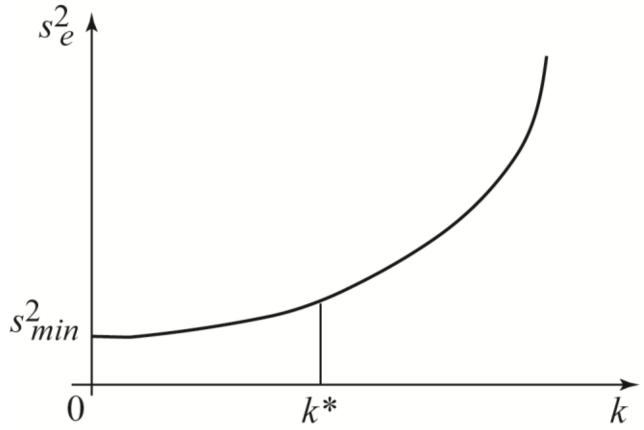


Рис. 26.1. **Зростання залишкової дисперсії при збільшенні діагональних елементів матриці X^TX на k**

Стійкість оцінок параметрів характеризується сумою їх дисперсій:

$$L(b) = \sum_{i=1}^{m} s_{b_i}^2 .$$

Цю формулу можна перетворити до більш ясного виду, якщо знайти всі власні числа λ_i і власні вектори U_i матриці $X^TX = UD_\lambda U^T$. Тут через D_λ позначена діагональна матриця власних чисел, а через U – ортогональна матриця власних векторів.

Обчислюємо $U^T \boldsymbol{b} = \boldsymbol{g}$.

Тепер сума дисперсій параметрів моделі може бути представлена у формі:

$$L(b) = \sum_{i=1}^{m} \frac{\lambda_i s_e^2(k) + k^2 g_i}{(\lambda_i + k)^2} .$$

При мультиколінеарності (виродженні матриці X^TX) одне або кілька власних чисел λ_i близькі до нуля і при $k=0$ виходить дуже велике (якщо не нескінченне) значення $L(\boldsymbol{b})$. Введення навіть дуже малого k відразу виправляє становище.

Існує кілька рекомендацій з вибору величини *k*. Вибирають таке найменше значення параметра *k*, за якого поведінка всіх коефіцієнтів регресії вже стабілізувалася, але ще не почалося інтенсивне зростання залишкової дисперсії.

Для порівняння різних альтернативних схем регресійного аналізу були зроблені численні перевірки на даних з різними особливостями. Перше місце у всіх таких перевірках впевнено зайняли ридж-оцінки. Деякі методи були спеціально розроблені для подолання конкретних особливостей, у цих випадках ридж-оцінки ділили перше-друге місце з одним із спеціалізованих методів. МНК-оцінки у всіх перевірках посіли останнє місце за всіма критеріями якості. Загальний висновок – якщо мультиколінеарності немає, то ридж-оцінки близькі до МНК-оцінок, а в разі мультиколінеарності вони значно кращі, причому перевага ридж-оцінок збільшується з ростом числа змінних *m*.

26.2. Моделі з *dummy*-змінними

Dummy у перекладі з англійської мови означає «макет», тому іноді зустрічається назва цих змінних «макетні змінні», а іноді й неправильна назва «фіктивні змінні».

Dummy-змінні призначені для обліку в моделі якісних змінних, які виміряні на номінальних шкалах. Наприклад, стать: жіноча, чоловіча; стан економіки: до кризи і після; сезон: зимовий, весняний, літній, осінній (це категорії якісної змінної). Якщо в модель необхідно ввести змінну, що має кілька категорій, вводиться відповідна кількість *dummy*-змінних, вона дорівнює одиниці, якщо спостерігається відповідна категорія і нулю, якщо вона відсутня в спостереженні. У сумі всі *dummy*-змінні утворюють $X_0 \equiv 1$, тому не можна включати в модель відразу всі *dummy*-змінні або ж не треба включати в модель вільний член. Звичайно, одну з категорій беруть еталонною і її в модель не включають. Тому отримуємо модель для еталонної категорії з виправленням для всіх інших категорій. Значимість цих виправлень за критерієм Стьюдента є значимістю відмінностей кожної категорії від еталонної.

Приклад. Відомі витрати на газ та електроенергію в США протягом 1977 – 1982 рр. Маємо типову ситуацію, коли бажаючи збільшити обсяг вибірки, замість середньорічних використовують квартальні дані. Не дивлячись на те, що вибірка при цьому збільшилася в чотири рази, значимість моделі різко погіршилася, а коефіцієнт детермінації зменшився до нуля. Причина полягає в тому, що при переході до квартальних даних з'явилися сезонні коливання, на інтенсивному фоні яких зовсім загубився досліджуваний ефект – часовий лінійний тренд (рис. 26.2).

Отже, необхідно врахувати в моделі цю особливість процесу, і тому найбільш економним способом є введення *dummy*-змінних для опису міжквартальних відмінностей. Тому необхідно ввести в модель чотири додаткові *dummy*-змінні: $Z_1 = 1$ – для 1-го кварталу і $Z_1 = 0$ – для всіх інших кварталів; $Z_2 = 1$ – для 2-го кварталу і $Z_2 = 0$ – для всіх інших кварталів; $Z_3 = 1$ – для 3-го кварталу і $Z_3 = 0$ – для всіх інших кварталів; $Z_4 = 1$ – для 4-го кварталу і $Z_4 = 0$ – для всіх інших кварталів. Якщо за еталон взяти 1-й квартал, то змінні Z_1 не слід включати в модель. У результаті обчислень буде отримано рівняння регресії для еталонної категорії з виправленнями для вільного члена для інших категорій.

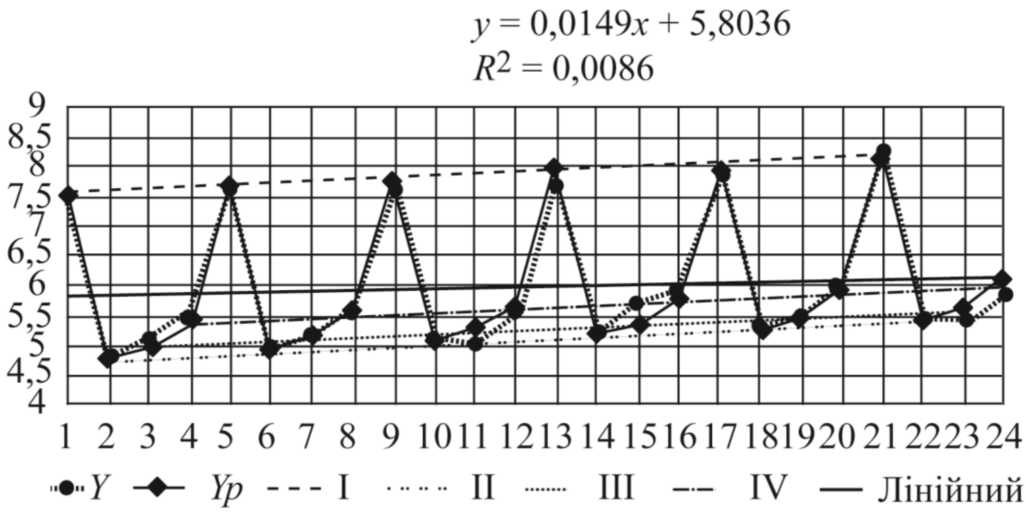


Рис. 26.2. **Сезонні коливання даних за кварталами**

Можна підключити в модель всі відразу *dummy*-змінні, але тоді в модель слід включити вільний член, оскільки $Z_1 + Z_2 + Z_3 + Z_4 = 1$. У результаті таких обчислень будуть отримані рівняння регресії для кожної категорії окремо і правильно підраховані всі статистичні характеристики за винятком статистики Фішера, яка буде занижена в $\frac{m+1}{m} = \frac{5}{4}$. При цих різних підходах за критерієм Стьюдента оцінюються різні ефекти якісного фактора. За іншою методикою оцінюється значимість ефекту кожної категорії, а за першою, рекомендованою, значимість розбіжностей кожної категорії від еталонної. Зараз неявно передбачається, що для кожної сукупності (для кожної категорії) зберігаються незмінними всі закономірності в залежностях від кількісних змінних, а вплив якісної ознаки проявляється тільки в управліннях вільного члена. Бувають і більш складні ситуації, коли для різних категорій виявляються різні ефекти кількісних чинників. Тоді в моделі слід врахувати також члени зі взаємозв'язками. Вихідні дані разом з введеними *dummy*-змінними представлені в табл. 26.1, де T – номер кварталу, Y – витрати (результативна ознака).

Таблиця 26.1

Дані для побудови моделі

Z_1	Z_2	Z_3	Z_4	T	Y	Y_p
1	0	0	0	1	7,33	7,5141
0	1	0	0	2	4,7	4,7691
0	0	1	0	3	5,1	4,9991
0	0	0	1	4	5,46	5,4158
1	0	0	0	5	7,65	7,6405
0	1	0	0	6	4,92	4,8955
0	0	1	0	7	5,15	5,1255
0	0	0	1	8	5,55	5,5421
1	0	0	0	9	7,96	7,7668
0	1	0	0	10	5,01	5,0218
0	0	1	0	11	5,05	5,2518
0	0	0	1	12	5,59	5,6685
1	0	0	0	13	7,74	7,8932
0	1	0	0	14	5,1	5,1482
0	0	1	0	15	5,67	5,3782
0	0	0	1	16	5,92	5,7948
1	0	0	0	17	8,04	8,0195
0	1	0	0	18	5,27	5,2745
0	0	1	0	19	5,51	5,5045
0	0	0	1	20	6,04	5,9212
1	0	0	0	21	8,26	8,1459
0	1	0	0	22	5,51	5,4009
0	0	1	0	23	5,41	5,6309
0	0	0	1	24	5,83	6,0476

Значущі коефіцієнтів регресії перед Z_2, Z_3, Z_4 в цій моделі показують, що витрати на газ та електроенергію в еталонному 1-му кварталі істотно відрізняються від інших кварталів. Наведемо рівняння регресії для кожного кварталу:

$$Yp = 7{,}4825 – 2{,}777\cdot Z2 – 2{,}278\cdot Z3 – 2{,}193\cdot Z4 + 0{,}0316\cdot T;\quad R^2 = 0{,}9867$$

(tb) (98,5) (33,2) (30,7) (25,9) (7,3)

Y_p(I) = 7,4825 + 0,0316·T (Z_1 = 1),

Y_p(II) = 7,4825 – 2,777 + 0,0316·T = 4,7059 + 0,0316·T (Z_2 = 1),

Y_p(III) = 7,4825 – 2,278 + 0,0316·T = 4,9034 + 0,0316·T (Z_3 = 1),

Y_p(IV) = 7,4825 – 2,193 + 0,0316·T = 5,2894 + 0,0316·T (Z_4 = 1).

На тлі інтенсивних сезонних коливань значимість досліджуваного лінійного тренда виявилася заниженою до нуля (R^2 = 0,0086), оцінка кутового коефіцієнта занижена більш ніж у два рази.

За допомогою трьох додаткових *dummy*-змінних Z_2, Z_3, Z_4 дуже добре описана вся залежність з сезонними коливаннями. Слід зазначити, що спроби описати ці коливання аналітичним виразом у вигляді трьох гармонік (з такими періодами, як рік, півроку, квартал) будуть неекономними і вимагають 6-ти додаткових параметрів (по два параметри на кожну гармоніку). Можна в модель включити всі відразу *dummy*-змінні, але тоді модель не повинна мати вільний член. Обчислюємо параметри такої лінійної моделі:

$$Y_p = (b_1Z_1 + b_2Z_2 + b_3Z_3 + b_4Z_4) + b_5T \text{ (константа відсутня).}$$

У такому випадку будуть отримані всі рівняння для кожного кварталу (не треба перераховувати вільний член). Однак тепер статистики Стьюдента оцінюватимуть відхилення середнього рівня витрат у кварталі від нуля, а не від обраного еталона. Всі характеристики, крім F, вийдуть такі ж, тільки статистику Фішера слід збільшити в $(m + 1) / m = 5 / 4$ разів: F = 280,93·5/4 = 351,16.

26.3. Новітні (*Advanced*) методи регресійного аналізу

У новітніх методах регресійного аналізу результативна ознака може бути дискретною і навіть якісною, вираженою категоріями або номінаціями. У цих методах вивчається ймовірність цих категорій залежно від значень пояснювальних змінних.

У загальному регресійному аналізі на основі зовнішньої поведінки результативної ознаки, заданої деякою кількістю випадкових спостережень, визначають форму зв'язку і далі методом найменших квадратів оцінюють його параметри або коефіцієнти регресії. У лінійному регресійному аналізі форма зв'язку приймається лінійною відносно параметрів моделі, іншими словами, можливі попередні перетворення змінних. У результаті цього виходить, що результативна ознака завжди буде розподілена асимптотично нормально зі своїми змінними характеристиками (середнім і дисперсією) для кожного набору пояснювальних змінних.

В узагальненій лінійній моделі вид закону розподілу результативної ознаки приймається на основі змістовних обґрунтувань, виконується апроксимація не кінцевої залежності $y(x)$, а деякого параметра функції розподілу. Далі, знаючи розподіл результативної ознаки, для кожного набору пояснювальних змінних обчислюють середнє (очікуване) значення результативної ознаки і її дисперсію. Вважається, що один з параметрів функції розподілу, а саме лінійний предиктор, залежить від пояснювальних змінних і приймається його апроксимація у вигляді лінійної комбінації пояснювальних змінних лінійної відносно її коефіцієнтів, які оцінюються методом найменших квадратів. Так, для опису залежності від однієї змінної, що пояснює в традиційному регресійному аналізі беруть лінійну модель, що має вигляд: $\hat{y}(x) = b_0 + b_1 x$, де $\hat{y}(x) = y_p(x)$ – очікуване (середнє) значення *y* для кожного *x*; можливі попередні функціональні перетворення початкових змінних (X, Y). Коефіцієнти регресії a, b оцінюють методом найменших квадратів, після чого визначають незміщену оцінку залишкової $s_e^2 = Const$ і дисперсії обчислених значень для кожного *x*: $\sigma^2(\hat{y}) = \frac{s_e^2}{N}\left(1 + \frac{(x - \bar{x})^2}{s_x^2}\right)$, де *N* – число спостережень, s_x^2 – дисперсія пояснювальної змінної *x*. Межі 95– відсоткового довірчого інтервалу на кожне обчислене значення $\hat{y}(x)$ обчислюється за формулами:

$$y_d = \hat{y} - t_{0,05}(N-2)\cdot\sigma(\hat{y}); \quad y_u = \hat{y} + t_{0,05}(N-2)\cdot\sigma(\hat{y}),$$

де y_d – нижня межа (*down*), y_u – верхня межа (*up*), $t_{0,05}$**(***dfe***)** – квантиль розподілу Стьюдента, *dfe* **= (***N – 2***)** – залишкове число ступенів свободи.

В узагальненій лінійній моделі вважається, що результативна ознака *Y* є випадковою величиною з відомим видом закону розподілу, який описується функцією щільності ймовірності $f(y \mid \mu, \nu)$, де μ і ν – параметри функції розподілу (імовірності – для дискретної або щільність ймовірності – для неперервної випадкової величини). Приймається, що один з цих параметрів, а саме лінійний предиктор μ, залежить від пояснювальних змінних, другий параметр, якщо закон розподілу двопараметричний, від пояснювальних змінних не залежить – ν = *Const*. Лінійний предиктор є комбінацією пояснювальних змінних і ця комбінація лінійна відносно параметрів: $\mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m$ (допускаються також попередні функціональні перетворення початкових змінних). За експериментальними даними $\left(x_i\,, y_i\right)$ методом найменших квадратів оцінюються коефіцієнти лінійного предиктора β_j і константа ν. Передбачається, що функціональна залежність між предиктором і очікуваними значеннями результативної змінної $\mu = g\left(M\left(y\right)\right) = g\left(\hat{y}\right)$. Функція g вважається відомою і називається функцією зв'язку. Традиційна модель регресійного аналізу є окремим випадком узагальненої лінійної моделі при тотожній формі зв'язку $\hat{y} = \mu = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_m x_m$ і нормальному законі розподілу $Y \sim N\left(\mu, \sigma\right)$. Вкажемо, що навіть у цьому випадку регресійна модель і узагальнена лінійна модель не еквівалентні, оскільки в них різні методи визначення ширини довірчої смуги на обчислені значення.

Біноміальна модель («логіт-» і «пробіт-» аналізи)

Приклад. Маємо дослідження впливу витрат на рекламу на обсяги реалізації мобільних телефонів. У табл. 26.2 наведено дані про питому частку реалізованих телефонів у філіях фірми, що мають різні витрати на рекламу. Кількість реалізованих телефонів m в партіях товару з n одиниць для кожного рівня витрат на рекламу x розподілено за біноміальним законом $P\left(Y = m\right) = P_n(m) = C_n^m p^m (1-p)^{n-m}$ з математичним сподіванням (середнім) $y_p = M\left(m\right) = np$ і дисперсією $D\left(m\right) = np\left(1-p\right)$.

Таблиця 26.2

Витрати на рекламу в філіях

Філія	Витрати на рекламу (x)	Обсяг партії товару (n)	Кількість проданих телефонів (m)	Відносна частина реалізованих телефонів (p)
1	10,2	50	44	0,8800
2	7,7	49	42	0,8571
3	5,1	46	24	0,5217

4	3,8	48	16	0,3333
5	2,6	50	6	0,1200

Ймовірність $P_n(m)$ залежить від двох параметрів: n – кількості товару в партії і $p(x)$ – питомої відносної частини реалізованих телефонів. Відносна частина реалізованих телефонів $0 \le p \le 1$ буде великою (близькою до 1) при великих витратах на рекламу і маленькою (близькою до нуля) при невеликих.

Необхідно апроксимувати залежність $p(x)$ досить простим виразом. За даними табл. 26.2 побудовано графік p від x (рис. 26.3). Експериментальних точок мало, хоча кожна точка є результатом $n \approx 50$ спостережень. З очевидних міркувань ясно, що ця сигмоїдальна залежність повинна бути схожою на інтегральну функцію розподілу (не обов'язково нормального закону) або на логістичну залежність. Забігаючи наперед слід зазначити, що обидві апроксимації призвели до практично однакових результатів (на рис. 26.3 суцільна і пунктирна лінії).

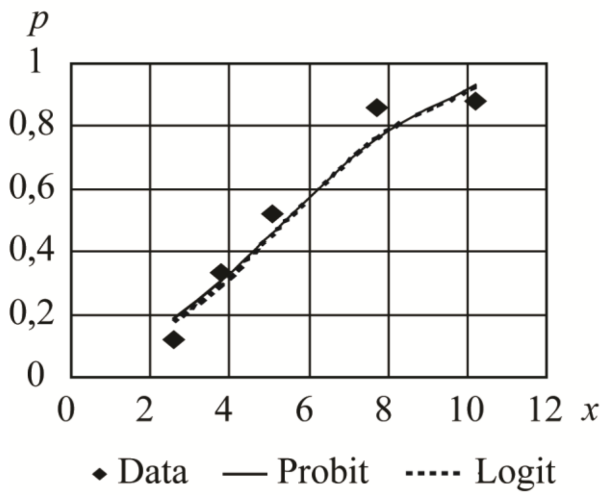


Рис. 26.3. **Апроксимація відносної частини реалізованого товару залежно від витрат на рекламу**

Якщо застосовується апроксимація функцією нормального розподілу, використовується назва «пробіт-аналіз»; якщо застосовується апроксимація логістичною функцією, використовується назва «логіт-аналіз». Порівняємо обидва види біноміальної моделі.

Пробіт-аналіз. Апроксимуємо залежність параметра $p(x)$ біноміального розподілу інтегральною функцією стандартного нормального закону:

$$p(x) = \Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{z^2}{2}} dz \leftrightarrow \quad t = \Phi^{-1}(p),$$

де Φ^{-1} – функція, обернена до функції розподілу нормального закону,

а $t = \dfrac{x-a}{\sigma} = b_0 + b_1 x$ – лінійний предиктор (звідки і йде назва «лінійна» модель).

Для кожного відомого значення p з табл. 26.2 знаходимо квантилі стандартного нормального розподілу $t = \Phi^{-1}(p)$ (табл. 26.3) і позначаємо їх у вигляді точок (x–t) на рис. 3 а. Далі методом найменших квадратів оцінюємо коефіцієнти регресії лінійного предиктора $t = 0{,}3113x - 1{,}6922$ і в табл. 3 обчислюємо розрахункові (згладжені) значення:

$$p_n = \Phi(0{,}3113x - 1{,}6822).$$

Таблица 26.3

Розрахункові значення $p(x)$ в пробіт-(p_n) і логіт-(p_s) аналізі

№	x	p	t	p_n	s	p_s
1	10,2	0,8800	1,174987	0,930978	1,99243	0,924514
2	7,7	0,8571	1,067571	0,759548	1,81529	0,767224
3	5,1	0,5217	0,054519	0,458368	0,080043	0,457004
4	3,8	0,3333	-0,43073	0,305291	-0,70819	0,298392
5	2,6	0,1200	-1,17499	0,188669	-1,99243	0,184671

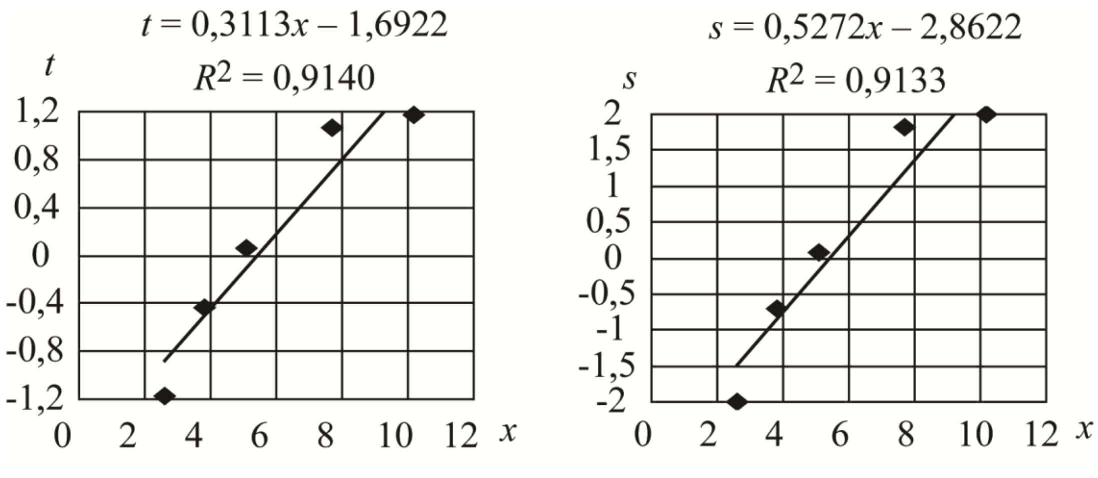
Для порівняння відразу наведемо обчислення за логіт-аналізу.

Логіт-аналіз. Апроксимуємо залежність параметра $p(x)$ біноміального розподілу логістичною функцією:

$$p(x) = \frac{1}{1+e^{-s}} = \frac{1}{1+e^{-(a+bx)}} \quad \leftrightarrow \quad s = \ln\frac{p}{1-p},$$

де $s = a + bx$ – лінійний предиктор.

Для кожного відомого значення p з табл. 26.2 знаходимо значення $s = ln\dfrac{p}{1-p}$ (табл. 26.3) і відмічаємо їх у вигляді точок (x–s) на графіку рис. 26.4б). Далі методом найменших квадратів оцінюємо коефіцієнти регресії лінійного предиктора $s = 0{,}5272x - 2{,}8622$ і в табл. 26.3 обчислюємо розрахункові (згладжені) значення $p_s = \dfrac{1}{1+e^{-(0{,}5272\,x - 2{,}8622)}}$.



а) б)

Рис. 26.4. **Апроксимація залежності $p(x)$ в функціональних змінних**

а) пробіт-аналіз: $t = \Phi^{-1}(p(x))$; б) логіт-аналіз: $s = p(x) / (1 - p(x))$

На рис. 26.3 побудовані графіки $p_n(x)$ – суцільною лінією, $p_s(x)$ – пунктирною лінією. Бачимо, що обчислені значення p_n, p_s за двома видами аналізу (пробіт- і логіт-) близькі між собою, тобто обидва види аналізу еквівалентні. Це не випадково, оскільки логістичною залежністю можна добре апроксимувати функцію розподілу нормального закону. Як певну перевагу логіт-аналізу відзначимо більш простий вид математичних формул, записаних за допомогою елементарних функцій.

Якщо правильно прийнято припущення про логістичну форму зв'язку, точки на графіку (x–s) повинні групуватися біля деякої прямої, однак тут (як і для пробіт-аналізу), є явні відхилення від прямої. В узагальнених лінійних моделях передбачаються попередні функціональні перетворення змінних. Так у наведеному прикладі доцільно перейти до логарифмів змінної, що пояснює. При цьому буде істотно покращена якість апроксимації як при пробіт-, так і при логіт-аналізі (рис. 26.5).

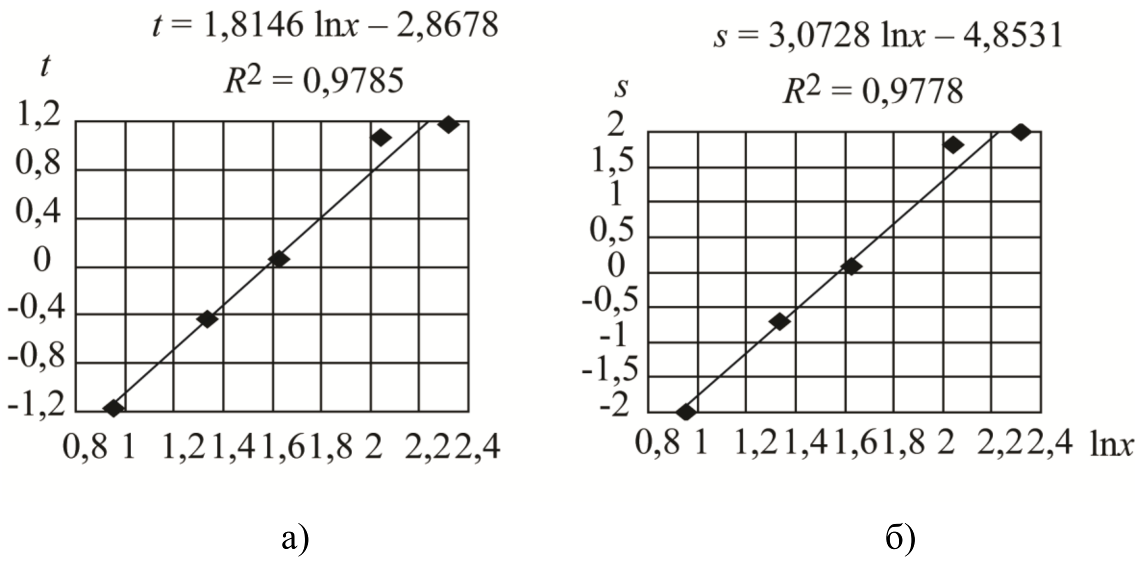


Рис. 26.5. **Апроксимація залежності $p(\ln x)$ в функціональних змінних а) пробіт-аналіз: $t=\Phi^{-1}(p(\ln x))$, б) логіт-аналіз: $s=\ln(p(\ln x)/(1-p(\ln x)))$**

З рис. 26.5 видно, що коефіцієнт детермінації після переходу від логарифмів збільшився від 0,91 до 0,98. Тепер можна назвати узагальнену лінійну модель узагальненою логарифмічною моделлю. Отже, функція розподілу результативної змінної (y – кількість реалізованих телефонів) визначена повністю. Вона залежить від двох параметрів n= $const$ і $p = f(x)$:

$$p=\frac{1}{1+e^{-s}}=\frac{1}{1+e^{-(3{,}0728\ \ln x-4{,}8531)}}\approx\frac{x^3}{x^3+100}.$$

Для відомого n = 50 і для кожного x = 0, 1, 2, 3 ..., 12 в табл. 26.4 обчислені середні значення (математичні сподівання) результативної змінної $Y=np(x)$ і середні квадратичні відхилення $\sigma=\sqrt{np(x)(1-p(x))}$.

При n = 50 біноміальний розподіл вже асимптотично нормальний (розподіл Лапласа), тому межі 95–відсоткових довірчих інтервалів визначаємо за правилом: $Y_d = Y - t_{0,05}\cdot\sigma$ – нижня межа (*down*), $y_u = Y + t_{0,05}\cdot\sigma$ – верхня (*up*) межа, $t_{0,05}$ = 2,01 знаходимо з таблиць Стьюдента при $dfe = n - 2 = 48$.

На рис. 26.6 зображено графік залежності $y(x)$ разом з 95–відсотковою довірчою смугою. Точками позначені значення, що спостерігаються m (скориговані на однаковий обсяг партій n = 50).

Таблиця 26.4

Обчислені значення $У$ з 95–відсотковими довірчими смугами

x	*s*	*p*	σ	*Y* = *np*	-95%	+95%
1	-4,853	0,008	0,620	0,387	-0,859	1,633
2	-2,723	0,062	1,700	3,081	-0,337	6,499
3	-1,477	0,186	2,750	9,292	3,763	14,820
4	-0,593	0,356	3,385	17,794	10,989	24,599
5	0,092	0,523	3,532	26,154	19,055	33,253
6	0,653	0,658	3,355	32,880	26,136	39,624
7	1,126	0,755	3,041	37,758	31,646	43,869
8	1,537	0,823	2,699	41,149	35,724	46,574
9	1,899	0,870	2,380	43,486	38,702	48,270
10	2,222	0,902	2,100	45,112	40,890	49,333
11	2,515	0,925	1,860	46,260	42,521	49,999
12	2,783	0,942	1,657	47,086	43,757	50,416

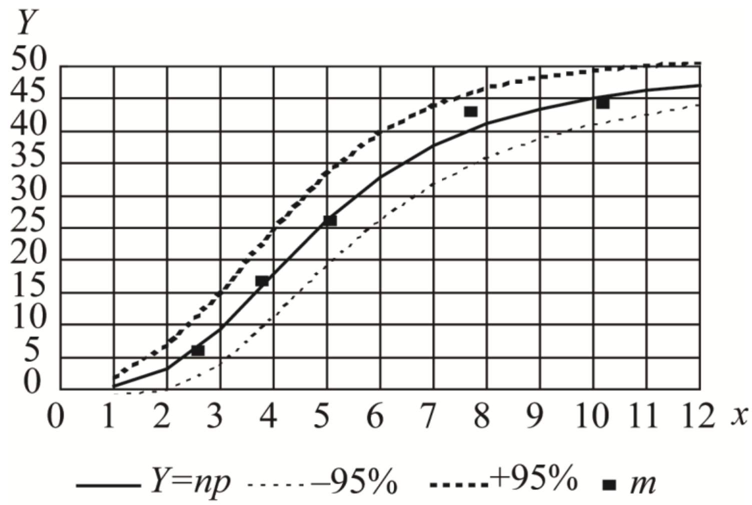


Рис. 26.6. **Результати аналізу за узагальненою моделлю**

На цьому ж прикладі порівняємо результати проведених обчислень (за узагальненою лінійною моделлю) з результатами обчислень за нелінійним регресійним аналізом. Нехай відносна частка $p(x)$ є результативною ознакою. Використовуючи попереднє дослідження, можна отримати ту ж саму форму зв'язку для опису вибраної результативної ознаки:

$$y = p = \frac{1}{1+e^{-s(x)}} \leftrightarrow s(x) = \ln\frac{p}{1-p},$$

$$\hat{s} = 3{,}0728 \cdot \ln x - 4{,}8531.$$

У функціональних координатах ($X = lnx$, s) залежність, яка вивчається, є лінійною. Отже, до цього моменту методики узагальненої лінійної моделі та регресійного аналізу збігалися (звичайно, тільки для даного простого прикладу). Однак, далі починаються істотні розбіжності.

В узагальненій лінійній моделі $p(x)$ вважається параметром біноміального розподілу результативної ознаки (другий параметр $n = 50$). Знаючи функцію розподілу, для кожного значення аргументу x обчислюємо сподіване значення $\hat{y} = M\left(\frac{m}{n}\right) = \frac{n\hat{p}}{n} = \hat{p}$ і дисперсію $\sigma^2(\hat{y}) = D\left(\frac{m}{n}\right) = \frac{n\hat{p}(1-\hat{p})}{n^2} = \frac{\hat{p}(1-\hat{p})}{n}$. Графік рівняння регресії для $y = p(x)$ разом з межами 95−відсоткового інтервалу (рис. 26.7а)) тільки масштабом осі ординат відрізняється від графіка на рис. 26.6.

У традиційному регресійному аналізі прийнята зовсім інша методика визначення меж довірчої смуги. Параметр $n = Const$ (число об'єктів у групі для обчислення відносної частки $p = {}^{m}/_{n}$) взагалі не входить в модель. Вважається, що існує всього $N = 5$ спостережень (5 точок) у функціональних координатах (X, s). Для цих 5-ти точок знаходимо незміщену оцінку дисперсії залишків лінійної регресійної моделі $MSE = Const$, де $MSE = \frac{\sum (s-\hat{s})^2}{N-2}$ = 0,331079.

Далі для кожного значення аргументу обчислюється дисперсія розрахункових значень за формулою:

$$\sigma^2(\hat{s}) = \frac{MSE}{N}\left(1 + \frac{\left(X - \overline{X}\right)^2}{s_X^2}\right) = \frac{0{,}331079}{5}\left(1 + \frac{\left(ln\,x - 5{,}88\right)^2}{7{,}5336}\right).$$

У табл. 26.5 для x = 0, 1, 2 ..., 12 обчислені розрахункові значення $\hat{s}$, їх дисперсії $\sigma^2(\hat{s})$ і межі 95−відсоткової довірчої смуги s_d, s_u:

$$s_d = \hat{s} - t_{0,05}(N-2)\cdot\sigma(\hat{s}); \quad s_u = \hat{s} + t_{0,05}(N-2)\cdot\sigma(\hat{s});$$

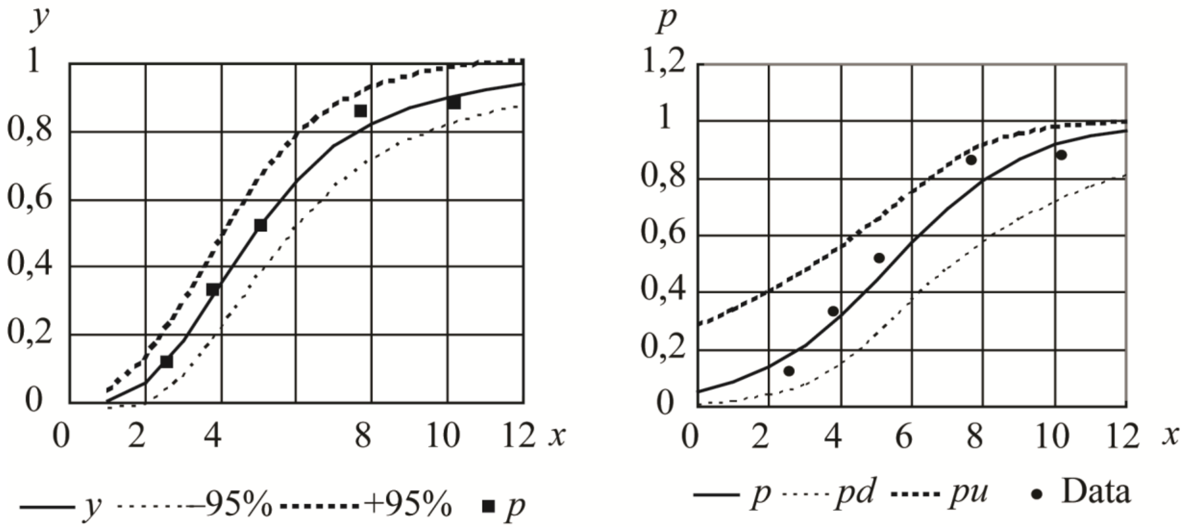
де $t_{0,05}(3)$ = 3,1824 знайдено за таблицею Стьюдента.

Таблиця 26.5

Обчислення нелінійної регресійної залежності

X	s	σ^2	s_d	s_u
0	-2,862	0,370104	-4,79808	-0,92592
1	-2,3348	0,27553	-4,0053	-0,6643
2	-1,8076	0,198535	-3,22561	-0,38959
3	-1,2804	0,139119	-2,46741	-0,09339
4	-0,7532	0,097281	-1,7458	0,239402
5	-0,226	0,073022	-1,08598	0,633981
6	0,3012	0,066342	-0,5185	1,120903
7	0,8284	0,077241	-0,05608	1,712875
8	1,3556	0,105719	0,320845	2,390355
9	1,8828	0,151775	0,642971	3,122629
10	2,41	0,215411	0,932952	3,887048
11	2,9372	0,296625	1,203936	4,670464
12	3,4644	0,395417	1,463207	5,465593

Далі виконано зворотне перетворення $p = \dfrac{1}{1+e^{-s}}$ і на підставі цього побудовано графік (рис. 26.7б)). Порівнюючи рис. 26.7а) і рис. 26.7б) відзначимо, що хоча лінії регресії однакові (для даного простого прикладу), але є істотні відмінності в ширині довірчої смуги (смуги невизначеності). Метод нелінійного регресійного аналізу привів до менш певних результатів порівняно з методикою узагальненої лінійної моделі.



а) б)

Рис. 26.7. **Порівняння узагальненої лінійної моделі (а) з нелінійним регресійним аналізом (б)**

Таким чином, узагальнені лінійні моделі передбачають більш реалістичну постановку проблеми опису випадкових залежностей. Не треба більше візуально вгадувати вид нелінійної регресійної моделі, що не просто навіть в однофакторних моделях. Виконується апроксимація не кінцевої нелінійної залежності, а одного з параметрів функції розподілу, а саме лінійного предиктора, тобто з обчислювальної точки зору залишаємося в межах лінійних моделей. Вид функції розподілу результативної ознаки вважається відомим завідомо з змістовних умов завдання.

Запитання для самоперевірки

1. У чому суть методу ортогональної (нормальної) регресії?
2. Які призначення гребневої регресії?
3. Яке призначення *dummy*-змінних?
4. У яких випадках рекомендується будувати моделі з dummy-змінними?
5. Які особливості мають новітні методи регресійного аналізу?
6. У яких випадках рекомендується будувати біноміальну модель?
7. У чому ідея пробіт-аналізу?
8. У чому ідея логіт-аналізу?

27. Системи одночасних рівнянь

27.1. Складові систем одночасних рівнянь.

27.2. Проблема ідентифікації.

27.3. Методи оцінювання параметрів систем рівнянь. Непрямий метод найменших квадратів (НМНК).

27.4. Двокроковий метод найменших квадратів (ДМНК).

27.1. Складові систем одночасних рівнянь

Використання одного окремого рівняння регресії для моделювання залежності результативної ознаки від факторів є грубим, реально в економіці існують механізми взаємозв’язків між факторами і результатами. Більш того в одному випадку ознака є фактором, в іншому є результатом. В одних рівняннях певна змінна розглядається як та, що пояснює (незалежна), в інші рівняння системи вона входить як залежна змінна. Тобто на практиці для моделювання економічних явищ, процесів слід будувати таку модель:

$$\begin{cases} y_1 = b_{12}y_2 + b_{13}y_3 + \ldots + b_{1n}y_n + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m + \varepsilon_1, \\ y_2 = b_{21}y_1 + b_{23}y_3 + \ldots + b_{2n}y_n + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m + \varepsilon_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = b_{n1}y_1 + b_{n2}y_2 + \ldots + b_{nn-1}y_{n-1} + a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nm}x_m + \varepsilon_n. \end{cases} \qquad (27.1)$$

Для опису структури зв'язків між змінними використовують систему так званих одночасних рівнянь, які ще називаються структурними рівняннями (27.1). В економетриці ця система рівнянь називається структурною формою моделі. Кожне рівняння системи одночасних рівнянь не може розглядатись самостійно. Для знаходження параметрів системи одночасних рівнянь класичний МНК не застосовується.

Одним з найпростіших прикладів систем одночасних рівнянь є модель «попит-пропозиція», що містить функції попиту, пропозиції лінійні від ціни і умову рівноваги:

$$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \varepsilon_{t1}, \alpha_1 < 0, \\ q_t^S = \beta_0 + \beta_1 p_t + \varepsilon_{t2}, \beta_1 > 0, \\ q_t^D = q_t^S, \end{cases} \qquad (27.2)$$

де перше рівняння – функція попиту, друге рівняння – функція пропозиції, третє рівняння – умова рівноваги, p_t – ціна товару в момент часу, t, ε_{t1} і ε_{t2} – випадкові складові.

Прикладом систем одночасних рівнянь є кейнсіанська модель формування доходів, розглядається закрита економіка без державних витрат:

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t + \varepsilon_t, \\ y_t = c_t + i_t, \end{cases} \qquad (27.3)$$

де перше рівняння – функція споживання, друге рівняння – макроекономічна тотожність, y_t і i_t – значення сукупного випуску та інвестицій в момент часу t, ε_t – випадкова складова.

Однією з можливих нестохастичних форм моделі рівноваги на ринку товарів є така модель:

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t, & (1) \\ \tau_t = \alpha_0 + \alpha_1 y_t, & (2) \\ i_t = \gamma_0 + \gamma_1 r_t, & (3) \\ y_{(d)t} = y_t + \tau_t, & (4) \\ g_t = \bar{g}, & (5) \\ y_t = c_t + i_t + g_t, & (6) \end{cases} \quad (27.4)$$

де рівняння (1) – функція споживання; (2) – функція податків; (3) – функція інвестицій; (4) – функція наявного доходу; (5) – державні витрати; (6) – макроекономічна тотожність; y_t – значення в момент часу t національного доходу; c_t – значення в момент часу t споживання; i_t – значення в момент часу t бажаного обсягу приватних інвестицій; g_t – державних витрат (у даному випадку $g_t = \bar{g} = const$); τ_t – обсягу податків; $y_{(d)t}$ – наявного доходу; r_t – відсоткової ставки.

Змінні в системах одночасних рівнянь діляться на два великі класи: ендогенні змінні, значення яких визначаються всередині моделі, та екзогенні – зовнішні стосовно моделі, їх значення визначаються поза моделлю і тому вважаються фіксованими. Наприклад, всі змінні в системі (27.1) є ендогенними, оскільки вони визначаються всередині системи.

Якщо розглянути найпростішу структурну форму моделі:

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

то тут y – ендогенні змінні, x – екзогенні змінні і коефіцієнти b_i і a_j називають структурними коефіцієнтами моделі. Всі змінні в моделі виражені у відхиленнях від середнього рівня, тобто за x прийнято $x - \bar{x}$, а за y прийнято $y - \bar{y}$, тому вільний член у кожному рівнянні системи відсутній.

У кейнсіанській моделі (27.2) c_t і y_t оцінюються всередині моделі (ендогенні змінні), i_t задається поза моделлю, отже, вона є екзогенною змінною. Зауважимо, що змінні c_t і y_t можуть бути виражені через i_t і ε_t, а саме:

$$\begin{cases} y_t = \dfrac{\beta_0}{1-\beta_1} + \dfrac{1}{1-\beta_1} i_t + \dfrac{\varepsilon_t}{1-\beta}, \\ c_t = \dfrac{\beta_0}{1-\beta_1} + \dfrac{\beta_1}{1-\beta_1} i_t + \dfrac{\varepsilon_t}{1-\beta}, \end{cases} \tag{27.5}$$

де коефіцієнт $\dfrac{1}{1-\beta_1}$ становить грошовий мультиплікатор, що визначає, на яку величину збільшується сукупний дохід при збільшенні обсягу інвестицій на одиницю.

Рівняння, що складають вихідну модель, називаються структурними рівняннями моделі. У свою чергу їх підрозділяють на поведінкові рівняння (описують взаємодію між змінними) та рівняння-тотожності (співвідношення, які повинні виконуватися в усіх випадках, і вони не містять параметрів і випадкових складових).

Якщо використати МНК для оцінювання структурних коефіцієнтів моделі, то оцінки будуть зміщеними і неспроможними. Тому для визначення структурних коефіцієнтів моделі структурну форму моделі перетворюють у приведену форму моделі. Рівняння, в яких відображена схема визначення ендогенних змінних, називаються рівняннями в приведеній формі (приведеними рівняннями). Приведеними рівняннями називаються рівняння, в яких ендогенні змінні виражені тільки через екзогенні або зумовлені змінні, а також випадкові складові. Прикладами таких рівнянь є рівняння системи (27.5). Зумовленими змінними називаються лагові ендогенні змінні, значення яких визначені до розгляду співвідношення.

Система лінійних функцій ендогенних змінних від екзогенних є приведеною формою моделі:

$$\begin{cases} \hat{y}_1 = \delta_{11}x_1 + \delta_{12}x_2 + \ldots + \delta_{1m}x_m + \varepsilon_1, \\ \hat{y}_2 = \delta_{21}x_1 + \delta_{22}x_2 + \ldots + \delta_{2m}x_m + \varepsilon_1, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \hat{y}_n = \delta_{n1}x_1 + \delta_{n2}x_2 + \ldots + \delta_{nm}x_m + \varepsilon_n. \end{cases}$$

Застосувавши МНК для оцінки δ, можна оцінити значення ендогенних змінних через екзогенні, а далі оцінити значення ендогенних змінних через екзогенні. Слід відмітити, що коефіцієнти приведеної форми моделі є нелінійними функціями коефіцієнтів структурної форми моделі. Це покажемо на простому прикладі. Для структурної моделі вигляду, в якій для спрощення відсутні похибки ε:

$$\begin{cases} y_1 = b_{12} y_2 + a_{11} x_1, \\ y_2 = b_{21} y_1 + a_{22} x_2, \end{cases}$$

приведена форма моделі має вигляд:

$$\begin{cases} y_1 = \delta_{11} x_1 + \delta_{12} x_2, \\ y_2 = \delta_{21} x_1 + \delta_{22} x_2, \end{cases}$$

коли з першого рівняння структурної моделі виражається y_2:

$$\begin{cases} y_2 = \dfrac{y_1 - a_{11} x_1}{b_{12}}, \\ y_2 = b_{21} y_1 + a_{22} x_2, \end{cases}$$

маємо $\dfrac{y_1 - a_{11} x_1}{b_{12}} = b_{21} y_1 + a_{22} x_2$. Після елементарних перетворень, виразимо y_1:

$y_1 = \dfrac{a_{11}}{1 - b_{12} b_{21}} x_1 + \dfrac{a_{22} b_{12}}{1 - b_{12} b_{21}} x_2$. Позначивши $\delta_{11} = \dfrac{a_{11}}{1 - b_{12} b_{21}}$, $\delta_{12} = \dfrac{a_{22} b_{12}}{1 - b_{12} b_{21}}$, отримали перше рівняння приведеної форми моделі $y_1 = \delta_{11} x_1 + \delta_{12} x_2$, в якій коефіцієнти є нелінійні співвідношення коефіцієнтів структурної форми моделі. Виконавши аналогічні перетворення будемо мати:

$$y_2 = \frac{a_{11} b_{21}}{1 - b_{12} b_{21}} x_1 + \frac{a_{22}}{1 - b_{12} b_{21}} x_2,$$

позначивши $\delta_{21} = \dfrac{a_{11} b_{21}}{1 - b_{12} b_{21}}$ і $\delta_{22} = \dfrac{a_{22}}{1 - b_{12} b_{21}}$ отримаємо друге рівняння приведеної форми моделі: $y_2 = \delta_{21} x_1 + \delta_{22} x_2$.

Приведена форма моделі дозволяє отримати значення ендогенної змінної через значення екзогенних змінних, але в ній відсутні оцінки взаємозв’язку між ендогенними змінними.

27.2. Проблема ідентифікації

У процесі переходу від приведеної форми моделі до структурної виникає проблема ідентифікації. Під проблемою ідентифікації розуміється можливість чисельної оцінки параметрів структурних рівнянь за оцінками коефіцієнтів приведених рівнянь.

Розглянемо проблему ідентифікації для структурної моделі з двома ендогенними змінними:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m, \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + \ldots + a_{2m}x_m. \end{cases}$$

З другого рівняння виразимо y_1:

$$y_1 = \frac{y_2}{b_{21}} - \frac{a_{21}}{b_{21}}x_1 - \frac{a_{22}}{b_{21}}x_2 - \ldots - \frac{a_{2m}}{b_{21}}x_m.$$

Тоді в системі маємо два рівняння для ендогенної змінної y_1 з одним і тим же складом екзогенних змінних, але з різними коефіцієнтами перед ними:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2 + \ldots + a_{1m}x_m, \\ y_1 = \dfrac{y_2}{b_{21}} - \dfrac{a_{21}}{b_{21}}x_1 - \dfrac{a_{22}}{b_{21}}x_2 - \ldots - \dfrac{a_{2m}}{b_{21}}x_m. \end{cases}$$

Маємо два варіанта для обчислення структурних коефіцієнтів однієї і тієї ж моделі, це пов’язано з неповною її ідентифікацією. Структурна модель у повному вигляді, що має в кожному рівнянні системи n ендогенних і m екзогенних змінних, містить $n(n-1+m)$ параметрів. Наприклад, при $n=2$ і $m=3$ повний вид структурної моделі буде:

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + a_{12}x_2 + a_{13}x_3, \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + a_{23}x_3, \end{cases}$$

де маємо вісім структурних коефіцієнтів.

Зведена форма моделі в повному виді містить nm параметрів. Для наведеного прикладу має бути шість коефіцієнтів зведеної форми моделі:

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + \delta_{13}x_3, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + \delta_{23}x_3. \end{cases}$$

На основі шести коефіцієнтів зведеної форми моделі необхідно визначити вісім структурних коефіцієнтів, що не може призвести до єдності розв’язку. В повному вигляді структурна модель містить більшу кількість параметрів, ніж зведена форма моделі. Відповідно $n(n-1+m)$ параметрів структурної моделі не може бути однозначно визначеними з nm параметрів зведеної форми моделі. Щоб отримати єдино можливий розв’язок для структурної моделі, слід передбачити, що деякі структурні коефіцієнти моделі дорівнюють нулю, оскільки окремі ознаки слабко зв'язані з ендогенною змінною. Таким чином зменшується кількість структурних коефіцієнтів моделі. В прикладі положимо $a_{13} = 0, a_{21} = 0$, структурна модель буде мати вигляд:

$$\begin{cases} y_1 = b_{12} y_2 + a_{11} x_1 + a_{12} x_2, \\ y_2 = b_{21} y_1 + a_{22} x_2 + a_{23} x_3, \end{cases}$$

де кількість структурних коефіцієнтів не перевищує кількості коефіцієнтів зведеної моделі.

Є й інші рекомендації: прирівняти деякі коефіцієнти один до одного, передбачаючи що їх вплив на ендогенну змінну однаковий. Також можна використати обмеження вигляду $b_{ij} + a_{ij} = 0$.

Висхідна система рівнянь називається:

а) ідентифікованою, якщо за оцінками коефіцієнтів зведених рівнянь можна однозначно визначити коефіцієнти структурних рівнянь (зазвичай це вдається зробити, коли кількість рівнянь для визначення коефіцієнтів структурних рівнянь в точності дорівнює кількості цих коефіцієнтів);

б) неідентифікованою (недовизначеною), якщо за коефіцієнтами зведених рівнянь можна отримати кілька варіантів значень коефіцієнтів структурних рівнянь (зазвичай це відбувається коли кількість рівнянь для визначення коефіцієнтів структурних рівнянь менше числа визначених коефіцієнтів);

в) надідентифікованою (перевизначеною), якщо за оцінками коефіцієнтів зведених рівнянь неможливо визначити коефіцієнти структурних рівнянь. У даному випадку система, що зв'язує коефіцієнти структурних рівнянь з коефіцієнтами зведених рівнянь, є несумісною (зазвичай у таких випадках число рівнянь для оцінки коефіцієнтів структурних рівнянь більше числа визначених коефіцієнтів).

Для визначення ідентифікованих структурних рівнянь застосовуються необхідні і достатні умови. Нехай система одночасних рівнянь включає в себе N рівнянь відносно N ендогенних змінних. Система містить M екзогенних або зумовлених змінних. Нехай n і m – кількість відповідно ендогенних і екзогенних змінних в рівнянні, яке перевіряється на ідентифікованість. Змінні, що не входять в дане рівняння, але входять в інші рівняння системи, називають виключеними змінними (їх кількість дорівнює $N - n$ для ендогенних і $M - m$ для екзогенних змінних відповідно).

Перша необхідна умова ідентифікованості. Рівняння ідентифіковане, якщо воно виключає принаймні $N-1$ змінну (ендогенну або екзогенну), присутню в моделі:

$$(N-n)+(M-m)\geq N-1.$$

Друга необхідна умова ідентифікованості. Рівняння ідентифіковане, якщо кількість виключених з рівняння екзогенних змінних не менше кількості ендогенних змінних у цьому рівнянні, зменшеного на одиницю:

$$M-m\geq n-1.$$

Знаки рівностей в обох необхідних умовах відповідають точній ідентифікації рівнянь.

Необхідна і достатня умова ідентифікованості. У моделі, що містить N рівнянь відносно N ендогенних змінних, умова ідентифікованості виконується тоді і тільки тоді, коли ранг матриці, складеної з виключених з даного рівняння змінних, але які входять в інші рівняння системи, дорівнює $N-1$.

Приклад. Дослідити модифіковану модель Кейнса.

$$\begin{cases} c_t = \beta_0 + \beta_1 y_t + \varepsilon_t, \\ i_t = \gamma_0 + \gamma_1 y_t + \gamma_2 y_{t-1} + \nu_t, \\ y_t = c_t + i_t + g_t. \end{cases}$$

Тут g_t – обсяг державних витрат. Ендогенні змінні: c_t, i_t, y_t (N = 3). Екзогенні змінні: g_t, y_{t-1}(M = 2).

Перше рівняння містить ендогенні змінні c_t, y_t (n = 2) і не містить екзогенні змінні (m = 0).

(N - n) + (M - m)$\geq$ N – 1, тобто (3-2)+(2-0)$\geq$3-1 і 3$\geq$2. Тому перше рівняння перевизначено.

Друге рівняння містить ендогенні змінні i_t, y_t (n = 2) і екзогенну змінну (m = 1).

(N - n) + (M - m)$\geq$ N – 1, тобто (3-2)+(2-1)$\geq$3-1 і 2 = 2. Тому друге рівняння ідентифіковано.

Третє рівняння містить ендогенні змінні, c_t, i_t, y_t (n = 3) і екзогенну змінну g_t (m = 1).

(N - n) + (M - m)$\geq$ N – 1, тобто (3-3)+(2-1)$\geq$3-1 і 1$\geq$2 (помилково). Тому третє рівняння надідентифіковано.

27.3. Методи оцінювання параметрів систем рівнянь. Непрямий метод найменших квадратів (НМНК)

Безпосереднє використання МНК для оцінки параметрів кожного з рівнянь системи одночасних рівнянь призводить до отримання зміщених і неспроможних оцінок. Як правило, це відбувається внаслідок корельованості однієї або декількох пояснювальних змінних з випадковим відхиленням. Тому застосовують інші методи.

Найбільшого поширення набули такі методи оцінювання коефіцієнтів структурної моделі: непрямий метод найменших квадратів; двокроковий метод найменших квадратів; трикроковий метод найменших квадратів; метод максимальної правдоподібності з повною інформацією; метод максимальної правдоподібності з обмеженою інформацією.

Непрямий і двокроковий методи найменших квадратів розглядаються як традиційні методи оцінки коефіцієнтів структурної моделі. Ці методи не складно реалізувати. Непрямий метод найменших квадратів застосовується для ідентифікованої системи одночасних рівнянь, а двокроковий метод найменших квадратів використовується для оцінки коефіцієнтів надідентифікованої моделі. Інші наведені методи також використовуються для надідентифікованих систем рівнянь.

Метод максимальної правдоподібності розглядається як найбільш загальний метод оцінювання, результати якого при нормальному розподілі ознак збігаються з МНК. Однак за великого числа рівнянь системи цей метод призводить до досить складних обчислювальних процедур. Тому як модифікація служить метод максимальної правдоподібності з обмеженою інформацією (метод найменшого дисперсійного відношення). Однак не зважаючи на популярність даного методу в 60-х роках, він був практично витіснений двокроковим методом найменших квадратів (ДМНК), у зв'язку з великою простотою останнього.

Подальшим розвитком ДМНК є трикроковий метод найменших квадратів (ТМНК). Цей метод оцінювання придатний для всіх видів рівнянь структурної моделі. Однак за деяких обмежень на параметри ефективнішим є ДМНК.

Непрямий метод найменших квадратів (НМНК) заснований на використанні приведених рівнянь. Така назва методу обумовлена обчисленням оцінок b_1 і b_2 через оцінки приведених рівнянь. Оцінки b_1 і b_2, які одержані з НМНК, є спроможними. Слід також зазначити, що оцінки b_1 і b_2 визначаються однозначно, а відповідне рівняння називається ідентифікованим.

Непрямий метод найменших квадратів включає в себе такі етапи: 1) виходячи із структурних рівнянь, будуються рівняння в приведеній формі; 2) оцінюються за МНК параметри рівнянь в приведеній формі; 3) на основі знайдених оцінок, оцінюються параметри структурних рівнянь.

Приклад. Розглянемо модель «попит-пропозиція»:

$$\begin{cases} q_t^D = \alpha_0 + \alpha_1 p_t + \varepsilon_{t1}, \\ q_t^S = \beta_0 + \beta_1 p_t + \beta_2 \varpi_t + \nu_t, \\ q_t^D = q_t^S, \end{cases}$$

де перше рівняння – функція попиту, друге рівняння – функція пропозиції, третє рівняння – умова рівноваги, p_t і ϖ_t – ціна товару і зарплата в момент часу t відповідно, ε_t і ν_t – випадкові складові.

Є наступні результати спостережень.

p	10	15	5	8	4
q	6	6	18	12	8
ϖ	2	6	2	7	4

Знайдемо оцінки параметрів цієї системи за допомогою НМНК.

Позначимо $q_t^D = q_t^S$ через q_t. Тоді:

$$\begin{cases} q_t = \alpha_0 + \alpha_1 p_t + \varepsilon_{t1}, \\ q_t = \beta_0 + \beta_1 p_t + \beta_2 \varpi_t + \nu_t, \end{cases}$$

Прирівняємо праві частини цих рівнянь:

$\alpha_0 + \alpha_1 p_t + \varepsilon_{t1} = \beta_0 + \beta_1 p_t + \beta_2 \varpi_t + \nu_t$, звідси

$$p_t = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} + \frac{\beta_2 \varpi_t}{\alpha_1 - \beta_1} + \frac{\nu_t - \varepsilon_t}{\alpha_1 - \beta_1}.$$

Введемо позначення: $\pi_{10} = \dfrac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$, $\pi_{11} = \dfrac{\beta_2}{\alpha_1 - \beta_1}$, $\psi_t = \dfrac{\nu_t - \varepsilon_t}{\alpha_1 - \beta_1}$.

Тоді $p_t = \pi_{10} + \pi_{11}\varpi_t + \psi_t$. Підставляємо цей вираз для p_t в перше рівняння, розкриємо дужки і введемо наступне позначення:

$\pi_{20} = \alpha_0 + \alpha_1\pi_{10}$, $\pi_{21} = \alpha_1\pi_{11}$, $\phi_t = \alpha_1\psi_t + \varepsilon_t$.

Тоді $q_t = \pi_{20} + \pi_{21}\varpi_t + \phi_t$.

Отримуємо систему з приведених рівнянь:

$$\begin{cases} p_t = \pi_{10} + \pi_{11}\varpi_t + \psi_t, \\ q_t = \pi_{20} + \pi_{21}\varpi_t + \phi_t. \end{cases}$$

Оцінимо параметри кожного з цих рівнянь за допомогою МНК, маємо $\pi_{11} = 0{,}75$; $\pi_{10} = 5{,}25$; $\pi_{21} = -0{,}46$; $\pi_{20} = 11{,}53$.

Тоді $\alpha_1 = \dfrac{\pi_{21}}{\pi_{11}} = -\dfrac{0{,}46}{0{,}75} = -0{,}61, \alpha_0 = \pi_{20} - \alpha_1\pi_{10} = 11{,}53 - (-0{,}61)5{,}25 = 14{,}73$.

Не можна оцінити коефіцієнти β_i на підставі отриманих результатів. Виникає проблема ідентифікації.

27.4. Двокроковий метод найменших квадратів (ДМНК)

Цей метод застосовується до перевизначених рівнянь.

Продовження прикладу. Перше рівняння висхідної системи $c_t = \beta_0 + \beta_1 y_t + \varepsilon_t$ перевизначено, тобто за оцінками коефіцієнтів приведених рівнянь неможливо визначити оцінки коефіцієнтів β_0 і β_1.

Для змінної y_t будуємо приведене рівняння $y_t = \pi_{10} + \pi_{11}y_{t-1} + \pi_{12}g_t + \varpi_t$ (ϖ_t – випадкове відхилення), знаходимо за допомогою МНК оцінки коефіцієнтів π_{10}; π_{11}; π_{12} і з рівняння отримуємо оцінки $\hat{y}_t$ за екзогенними змінними g_t, y_{t-1}.

З рівняння $c_t = \beta_0 + \beta_1\hat{y}_t$ знаходимо оцінки коефіцієнтів β_0 і β_1 за допомогою МНК.

Приклад. Розглянемо економетричну модель Клейна, складену для опису економіки США в період з 1921 по 1941 р. р.

Структурна модель містить три регресійних рівняння і три тотожності:

$$\begin{cases} C_t = b_0 + b_1 \cdot P_t + b_2 \cdot P_{t-1} + b_3 \cdot (W_t + W_t') + e_t, \\ I_t = b_0' + b_1' \cdot P_t + b_2' \cdot P_{t-1} + b_3' \cdot K_{t-1} + e_t', \\ W_t = b_0'' + b_1'' \cdot X_t + b_2'' \cdot X_{t-1} + b_3'' \cdot (t - t_{cp}) + e_t'', \\ X_t = C_t + I_t + G_t, \\ P_t = X_t - W_t - T_t, \\ K_t = K_{t-1} + I_t, \end{cases}$$

де *C* – фонд споживання (витрати на споживчі товари); *I* – капіталовкладення; *W* – особисті доходи в приватному секторі; *W'* – особисті доходи в державному секторі (зарплата); *X* – обсяг виробництва; *T* – податки; *K* – запаси капіталу; *P* – прибуток; *G* – державні витрати на товари та послуги; *t* – календарний час (рік). Всі змінні вимірюються в мільйонах доларів США у цінах 1931 р. Потрібно оцінити параметри моделі b_0, b_1, b_2, b_3; b_0, b'_1, b'_2, b'_3; b''_0, b''_1, b''_2, b''_3.

Висхідні дані наведені в табл. 27.1. Для змінних P_t, K_t, X_t у дужках наведені дані за 1920 р., так як у модель входять ці ж змінні з запізненням P_{t-1}, K_{t-1}, X_{t-1}. Значення t_{cp}=1931.

Таблиця 27.1

Зведені економічні показники США за 1921 – 1941 рр.

t	*C*	*P*	*W*	*W'*	*I*	*K*	*X*	*G*	*T*
1	2	3	4	5	6	7	8	9	10
(1920)		(12,7)				(182,8)	(44,9)		
1921	41,9	12,4	25,5	2,7	-0,2	182,6	45,6	3,9	7,7
1922	45,0	16,9	29,3	2,9	1,3	184,5	50,1	3,2	3,9
1923	49,2	18,4	34,1	2,9	5,2	189,7	57,2	2,8	4,7
1924	50,6	19,4	33,9	3,1	3,0	192,7	57,1	3,5	3,8
1925	52,6	20,1	35,4	3,2	5,1	197,8	63,0	3,3	5,5
1926	55,1	19,6	37,4	3,3	5,6	203,4	64,0	3,3	7,0
1927	56,2	19,8	37,9	3,6	4,2	207,6	64,4	4,0	6,7
1928	57,3	21,1	39,2	3,7	3,0	210,6	64,5	4,2	4,2
1929	57,8	21,7	41,3	4,0	5,1	215,7	67,0	4,1	4,0
1930	55,0	15,6	37,9	4,2	1,0	216,7	61,2	5,2	7,7
1931	50,9	11,4	34,5	4,8	-3,4	219,3	53,4	5,9	7,5
1932	45,6	7,0	29,0	5,3	-6,2	207,1	44,3	4,9	8,3
1933	46,5	11,2	28,5	5,6	-5,1	202,0	45,1	3,7	5,4
1934	48,7	12,3	30,6	6,0	-3,0	199,0	49,7	4,0	6,8
1935	51,3	14,0	33,2	6,1	-1,3	197,3	54,4	4,4	7,2
1936	57,7	17,6	36,8	7,4	2,1	199,8	62,7	2,9	8,3
1937	58,7	17,3	41,0	6,7	2,0	201,8	65,0	4,3	6,7

1938	57,5	15,3	38,2	7,7	-1,9	199,9	60,9	5,3	7,4
1939	61,6	19,0	41,6	7,8	1,3	201,2	60,5	6,6	8,9
1940	65,0	21,1	45,0	8,0	3,3	204,5	75,7	7,4	9,6
1941	69,7	23,5	53,3	8,5	4,9	209,4	88,4	13,8	11,6

Спочатку були оцінені параметри кожного рівняння окремо стандартним МНК:

$$C_t = 16{,}24 + 0{,}193 \cdot P_t + 0{,}090 \cdot P_{t-1} + 0{,}796 \cdot (W_t + W'_t) + e_t\,,$$

$$I_t = 10{,}13 + 0{,}480 \cdot P_t + 0{,}333 \cdot P_{t-1} - 0{,}112 \cdot K_{t-1} + e'_t\,,$$

$$W_t = 1{,}50 + 0{,}439 \cdot X_t + 0{,}146 \cdot X_{t-1} + 0{,}130 \cdot (t - 1931) + e''_t\ .$$

Аналіз отриманих моделей показав суттєві проблеми. Коефіцієнти регресії при змінній P_t і P_{t-1} (прибуток) є нормами відрахувань. Маємо, що коефіцієнти перед P_t значно перевищують коефіцієнти перед P_{t-1}, що є протиріччям, оскільки існує деяка тимчасова затримка між отриманням прибутку та його інвестиціями. Отримали зміщені значення параметрів – їх реальні значення в кілька разів відрізняються від обчислених МНК-оцінок. Економетрична модель з такими зміщеними параметрами – неадекватна, вона неправильно описує дійсний стан економіки країни.

Проаналізувавши систему регресійних рівнянь, виявляємо, що в даному прикладі порушена перша висхідна передумова, згідно з якою випадкові відхилення повинні відноситися тільки до результативних ознак і не повинні корелювати зі змінними, що пояснюють. Але в системі одночасних рівнянь виходить, що спільно залежні змінні в кінцевому підсумку корелюють з відхиленнями того ж рівняння (наприклад, у ланцюзі рівнянь відносно C_t, X_t, P_t випливає $e_t \rightarrow C_t \rightarrow X_t \rightarrow P_t$, тобто кореляція e_t і P_t , чого не повинно бути; аналогічно, ланцюг рівнянь відносно I_t , X_t , P_t призводить до непередбачуваної кореляції між e'_t і P_t).

Для оцінки параметрів системи одночасних рівнянь запропонована процедура двокрокового регресійного аналізу. Замінимо всі залежні змінні, що пояснюють (в правих частинах рівнянь системи) на їх розрахункові значення. Розрахункові значення, як відомо, не корелюють з випадковими похибками, тому відповідна передумова регресійного аналізу вже не буде порушена.

У моделі є змінні: випадкові відхилення e_t, e'_t, e''_t; екзогенні змінні, відомі в кожен момент часу t, W'_t , G_t , T_t; ендогенні змінні, на які впливають позасистемні змінні C_t , I_t , W_t , X_t , P_t , K_t . Усього таких змінних 6, для їх визначення є 6 рівнянь, саме ці показники необхідно назвати залежними пояснювальними змінними, коли вони входять в рівняння як аргументи. Ще є лагові ендогенні змінні: X_{t-1}, P_{t-1}, K_{t-1}, значення лагових змінних відомі до поточного моменту часу. Позасистемні (екзогенні) і лагові змінні називаються зумовленими, вони відомі до поточного моменту часу. До них відносяться t, W'_t , G_t , T_t , X_{t-1}, P_{t-1}, K_{t-1}.

Для отримання розрахункових значень C^R, I^R, W^R складаються лінійні моделі відносно всіх зумовлених змінних; розрахункові значення X^R і P^R виходять з відповідних тотожностей:

$$C_t^R = a_0 + a_1 t + a_2 W_t' + a_2 G' + a_3 T_t + a_4 X_{t-1} + a_5 P_{t-1} + a_6 K_{t-1},$$

$$I_t^R = b_0 + b_1 t + b_2 W_t' + b_2 G' + b_3 T_t + b_4 X_{t-1} + b_5 P_{t-1} + b_6 K_{t-1},$$

$$W_t^R = c_0 + c_1 t + c_2 W_t' + c_2 G' + c_3 T_t + c_4 X_{t-1} + c_5 P_{t-1} + c_6 K_{t-1},$$

$$X_t^R = C_t^R + I_t^R + G_t,$$

$$P_t^R = X_t^R - W_t^R - T_t,$$

$$K_t^R = K_{t-1} + I_t^R.$$

Слід обчислити розрахункові значення і запам'ятати їх у вигляді нових змінних (пропонується виконати ці операції на ЕОМ). На другому кроці оцінюються параметри структурної моделі:

$$\begin{cases} C_t = 16{,}55 + 0{,}017 P_t^R + 0{,}216 P_{t-1} + 0{,}819\left(W_t^R + W_t'\right) + e_t, \\ I_t = 20{,}28 + 0{,}150 P_t^R + 0{,}616 P_{t-1} - 0{,}158 K_{t-1} + e_t', \\ W_t = 1{,}50 + 0{,}439 X_t^R + 0{,}147 X_{t-1} + 0{,}130(t - 1931) + e_t'. \end{cases}$$

Двокроковим методом найменших квадратів отримані пояснені з погляду економіки оцінки спірних коефіцієнтів регресії (порівняйте їх з раніше знайденими). Великі зміни у розв'язанні отримані, переважно, за рахунок заміни P на P^R. На рис. 27.1 показана відповідність між реальними і згладженими за МНК даними P і P^R.

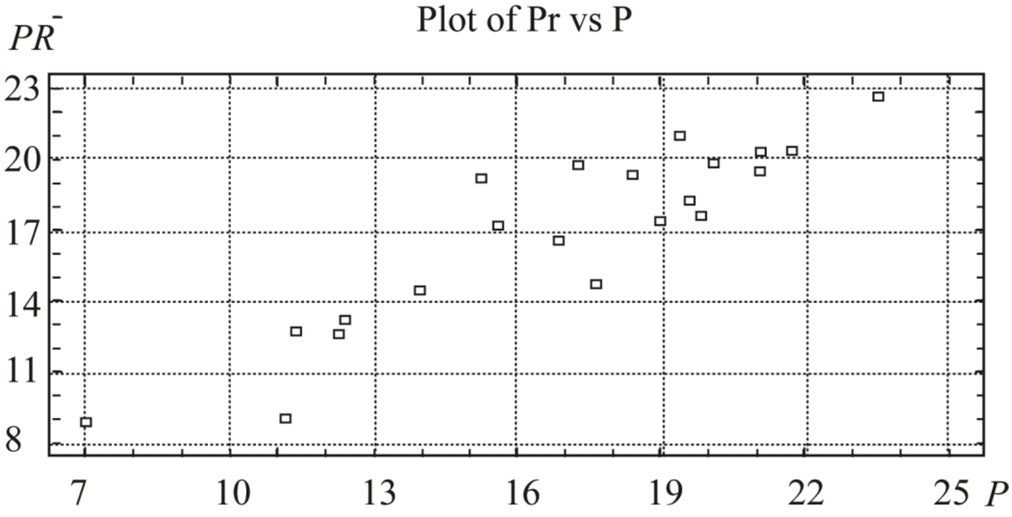


Рис. 27.1. **Відповідність між *P* і *P*^{R}**

Коефіцієнт кореляції між ними дорівнює 0.909. Отже, незначні поправки у вихідних даних призвели до таких великих зміщень в параметрах моделі.

Запитання для самоперевірки

1. Які основні причини використання систем одночасних рівнянь?

2. Наведіть основні типи моделей застосування систем одночасних рівнянь.

3. Які види змінних розрізняють в системах одночасних рівнянь? Дайте їм визначення.

4. У чому полягає відмінність між структурними рівняннями системи і рівняннями у приведеній формі?

5. Чому звичайний МНК практично не використовується для оцінки систем одночасних рівнянь?

6. Які найбільш поширені методи оцінювання коефіцієнтів структурної моделі?

7. У чому полягає суть непрямого методу найменших квадратів?

8. Назвіть причини неідентифікованості і надідентифікованості систем одночасних рівнянь.

9. Наведіть необхідні і достатні умови ідентифікованості систем.

10. У чому полягає суть двокрокового методу найменших квадратів?

28. Динамічні економетричні моделі

28.1. Загальна характеристика динамічних економетричних моделей.
28.2. Інтерпретація параметрів моделей з розподіленим лагом.
28.3. Інтерпретація параметрів авторегресійних моделей.
28.4. Вивчення структури лага і вибір виду моделі з розподіленим лагом.

28.1. Загальна характеристика динамічних економетричних моделей

В економічному аналізі використовуються щорічні, щоквартальні, щомісячні, щоденні дані, що становлять значення показників. Впорядковані за часом дані називаються часовими рядами.

Нехай досліджується показник, наприклад, прибуток підприємства. Його значення в поточний момент часу t позначають y_t, в наступні моменти позначають $y_{t+1}, y_{t+2}, ..., y_{t+k}, ...$, а в попередні моменти позначаються $y_{t-1}, y_{t-2}, ..., y_{t-k}, ...$. У процесі вивчення залежностей між такими показниками або під час аналізу їх розвитку в часі як пояснювальні змінні використовуються не тільки поточні значення змінних, а й деякі попередні за часом значення, а також час T. Моделі такого типу називаються динамічними.

Виділяють два типи динамічних економетричних моделей. До моделей першого типу належать моделі авторегресії і моделі з розподіленим лагом, в яких значення змінної за минулі періоди часу (лагові змінні) безпосередньо включені в модель. Моделі другого типу враховують динамічну інформацію в неявному вигляді. У ці моделі включені змінні, що характеризують очікуваний або бажаний рівень результату, або одного з чинників у момент часу t.

Наприклад, на виручку від реалізації або прибуток компанії поточного періоду можуть впливати витрати на рекламу або проведення маркетингових досліджень, зроблених компанією в попередні моменти часу. Величину l, що характеризує запізнення у впливі фактора на результат, називають лагом, а часові ряди самих факторних змінних, які зміщені на один або більше моментів часу – лаговими змінними.

Лагові моделі з розподіленим лагом мають вигляд:

$$y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \varepsilon_t .$$

Моделі авторегресії мають вигляд:

$$y_t = a + b_0 x_t + c_1 y_{t-1} + \varepsilon_t .$$

Причин наявності лагів в економіці багато, серед них найчастіше виділяють такі: психологічні причини, які обумовлені інерцією в поведінці людей; технологічні причини, обумовлені інерцією із заміною обладнання, технічного переозброєння, інерцією з переходом на нові технології; інституційні причини, наприклад, контракти між фірмами, трудові договори вимагають певної сталості в перебіг часу; механізми формування економічних показників, наприклад, інфляція є інерційним процесом, грошовий мультиплікатор (накопичення грошей у банківській системі) також проявляється на певному інтервалі часу.

Побудова моделей з розподіленим лагом і моделей авторегресії мають декілька особливостей. Першою особливістю виокремлюють те, що оцінка параметрів авторегресії та більшість моделей з розподіленим лагом не може бути виконана за допомогою звичайного методу найменших квадратів, оскільки порушують його передумови і вимагається застосування спеціальних статистичних методів. Необхідність вирішення проблеми вибору оптимальної величини лага та визначення його структури є другою особливістю. Третьою особливістю є існування взаємозв’язку між моделями з розподіленим лагом та моделями авторегресії. Дана особливість допускає перехід від одного типу моделей до іншого, що є необхідним в окремих випадках.

28.2. Інтерпретація параметрів моделей з розподіленим лагом

Модель з розподіленим лагом у загальному вигляді:

$$y_t = a + b_0 x_t + b_1 x_{t-1} + \ldots + b_l x_{t-l} + \varepsilon_t .$$

Дана модель пояснює те, що коли в деякий момент часу t відбувається змінення незалежної змінної x, то воно буде впливати на значення змінної y протягом l наступних моментів часу.

Коефіцієнт регресії b_0 характеризує середню абсолютну зміну y_t при змінені x_t на 1 одиницю свого виміру в певний фіксований період часу t, без урахування впливу лагових значень фактора x. Цей коефіцієнт називають короткостроковим мультиплікатором.

Будь-яку суму коефіцієнтів $\sum_{j=0}^{h} b_j$, $h < p$ називають проміжним мультиплікатором: у момент $t+1$ сукупний вплив факторної змінної x_t на результат y_t складе $(b_0 + b_1)$; в момент часу $t+2$ відповідно $(b_0 + b_1 + b_2)$ і т. д. Суму всіх b_j називають довгостроковим мультиплікатором $b_0 + b_1 + b_2 + ... + b_l = b$ і він характеризує абсолютну зміну в довгостроковому періоді $t+l$ результату y під впливом зміни на 1 одиницю фактора x. Розраховуються відносні коефіцієнти в моделі з розподіленим лагом: $\beta_j = \frac{b_j}{b}$. Якщо всі коефіцієнти b_j мають однакові знаки, то для будь-якого j $0 < \beta_j < 1$; $\sum_{j=0}^{l} \beta_j = 1$. Відносні коефіцієнти β_j є вагою для відповідних коефіцієнтів b_j, кожен з них вимірює частку загальної зміни результативної ознаки в момент часу $t + j$. Знаючи величину β_j можна визначити середній лаг і медіанний лаг.

Середній лаг: $\bar{l} = \sum_{j=0}^{l} j \cdot \beta_j$ становить середній період, протягом якого буде відбуватися зміна результату під впливом зміни фактора в момент часу t. Невелика величина середнього лага свідчить про відносно швидке реагування результатів на зміну фактора, а велика величина – про те, що вплив фактора на результат відбудеться через тривалий період часу. Медіанний лаг – це величина лага, для якого $\sum_{j=0}^{l_{Me}} \beta_j \approx 0{,}5$. Це період, протягом якого з моменту часу t буде реалізована половина загального впливу фактора на результат.

Слід відмітити, що в застосуванні методу найменших квадратів до таких моделей виникають труднощі, пов’язані з тим, що поточні і лагові значення незалежної змінної тісно зв’язані один з одним, тому оцінка параметрів моделі відбувається в умовах великої мільтиколінеарності факторів. Другою проблемою є те, що при великій величині лага знижується кількість спостережень, за якими будується модель та збільшується кількість факторних ознак. Це, як відомо, призводить до зменшення степенів свободи в моделі.

Третя проблема складається в тому, що в моделях з розподіленим лагом часто виникає проблема автокореляції залишків. Наведені проблеми в застосуванні методу найменших квадратів до оцінювання параметрів моделей з розподіленим лагом призводить до значної невизначеності оцінок параметрів моделі, зниженню їх точності й отриманню неефективних оцінок.

28.3. Інтерпретація параметрів авторегресійних моделей

У моделі авторегресії $y_t = a + b_0 x_t + c_1 y_{t-1} + \varepsilon_t$, як і в моделі з розподіленим лагом, b_0 характеризує короткострокову зміну y_t під впливом зміни x_t на 1 од. До моменту часу $t+1$ результат y_t зміниться під впливом зміни чинника в момент часу t на b_0, а y_{t+1} під впливом своєї зміни в попередній момент часу на c_1. Абсолютна зміна результату в момент часу $t+1$ складе $(b_0 c_1)$, а в момент часу $t+2$ абсолютна зміна складе $\left(b_0 c_1^2\right)$ і т. д. Довгостроковий мультиплікатор у моделі авторегресії розраховується як сума короткострокового і проміжного мультиплікатора:

$$b = b_0 + b_0 c_1 + b_0 c_1^2 + b_0 c_1^3 + \ldots .$$

Враховуючи $|c_1| < 1$, маємо $b = b_0 + b_0 c_1 + b_0 c_1^2 + b_0 c_1^3 + \ldots = \dfrac{b_0}{1 - c_1}$.

Така інтерпретація коефіцієнтів моделі авторегресії і розрахунок довгострокового мультиплікатора базуються на передумові про наявність нескінченного лага у впливі поточного значення залежної змінної на її майбутні значення.

Розглянемо моделі адаптивних сподівань та неповного коригування, оскільки оцінку параметрів цих моделей можна здійснювати, застосувавши звичайну модель авторегресії.

Модель адаптивних сподівань має вигляд:

$$y_t = a + b x_{t+1}^{*} + \varepsilon_t , \qquad (28.1)$$

де y_t – значення результативної ознаки;

x_{t+1}^{*} – очікуване значення факторної ознаки.

Дана модель називається довгостроковою функцією моделі адаптивних сподівань.

Очікування формується як:

$$x_{t+1}^* - x_t^* = \alpha\left(x_t - x_t^*\right) \text{ або } x_{t+1}^* = \alpha\, x_t + (1-\alpha)x_t^*,$$

де $0 < \alpha < 1$ − коефіцієнт сподівання. Чим ближче його значення до 1, тим більше реалізується очікуване значення в попередній період. Таким чином, у кожний період часу $t+1$ сподівань коригується на деяку частку α різниці між фактичним значенням факторної ознаки і його очікуваним значенням у попередній період.

Врахувавши в моделі (28.1) вираз $x_{t+1}^* = \alpha\, x_t + (1-\alpha)x_t^*$:

$$y_t = a + b\left(\alpha\, x_t + (1-\alpha)x_t^*\right) + \varepsilon_t = a + \alpha\, b x_t + (1-\alpha)b x_t^* + \varepsilon_t. \qquad (28.2)$$

Якщо модель (28.1) має місце для періоду t, то вона буде мати місце і для періоду $t-1$:

$$y_{t-1} = a + b x_t^* + \varepsilon_{t-1}. \qquad (28.3)$$

Далі, помноживши даний вираз на $(1-\alpha)$ та віднявши від (28.2), отримаємо:

$$y_t - (1-\alpha)y_{t-1} = a - (1-\alpha)a + \alpha\, b\, x_t + \varepsilon_t - (1-\alpha)\varepsilon_{t-1} \text{ або}$$

$$y_t = \alpha\, a + \alpha\, b\, x_t + (1-\alpha)y_{t-1} + u_t, \qquad (28.4)$$

де $u_t = \varepsilon_t - (1-\alpha)\varepsilon_{t-1}$. Отримали модель авторегресії, для якої можна знайти параметри, а потім перейти до моделі (28.1). Модель (28.4) включає тільки фактичні значення змінних, тому її параметри можна визначити на основі наявної статистичної інформації за допомогою стандартних статистичних методів. Модель (28.4) називається короткостроковою функцією моделі адаптивних сподівань.

Відмінною особливістю моделі неповного коригування є те, що емпірично спостерігається результативна ознака. Загальний вигляд моделі неповного коригування такий:

$$y_t^* = a + b x_t + \varepsilon_t. \qquad (28.5)$$

Дану модель називають ще моделлю неповного коригування або довгостроковою функцією моделі неповного коригування.

Формування очікувань y_t^* здійснюється за схемою:

$$y_t - y_{t-1} = \beta\left(y_t^* - y_{t-1}\right) + v_t, \qquad (28.6)$$

де $0 < \beta < 1$ та є коригуючим коефіцієнтом. Чим ближче він до 1, тим у більшій степені реальна динаміка показника відповідає очікуванню. Перетворивши (28.6) до вигляду

$$y_t = \beta y_t^* + (1 - \beta) y_{t-1} + v_t, \qquad (28.7)$$

та підставивши сюди $y_t^* = a + bx_t + \varepsilon_t$, отримаємо:

$$y_t = \beta(a + bx_t + \varepsilon_t) + (1 - \beta) y_{t-1} = \beta a + \beta bx_t + (1 - \beta) y_{t-1} + u_t, \quad (28.8)$$

де $u_t = \beta \varepsilon_t + v_t$.

Наведений вираз є основним рівнянням моделі неповного коригування та є короткостроковою функцією моделі. Знаючи оцінки параметрів рівняння, знаходять β, а після чого обчислюють параметри a та b моделі (28.5), що описує залежність очікуваного значення результата від значень факторної ознаки.

28.4. Вивчення структури лага і вибір виду моделі з розподіленим лагом

Якщо побудувати графік залежності коефіцієнтів регресії від величини лага, можна отримати графічне зображення структури лага або розподіл у часі впливу факторної змінної на результат. Структура лага може бути різною (рис. 28.1). Аналогічно графічний аналіз структури лага можна проводити і за допомогою відносних коефіцієнтів регресії β_j. Основна складність у виявленні структури лага полягає в тому, як отримати значення параметрів b_j або β_j. Звичайний МНК рідко буває корисним у цих цілях. Тому в більшості випадків припущення про структуру лага базуються на загальних положеннях економічної теорії.

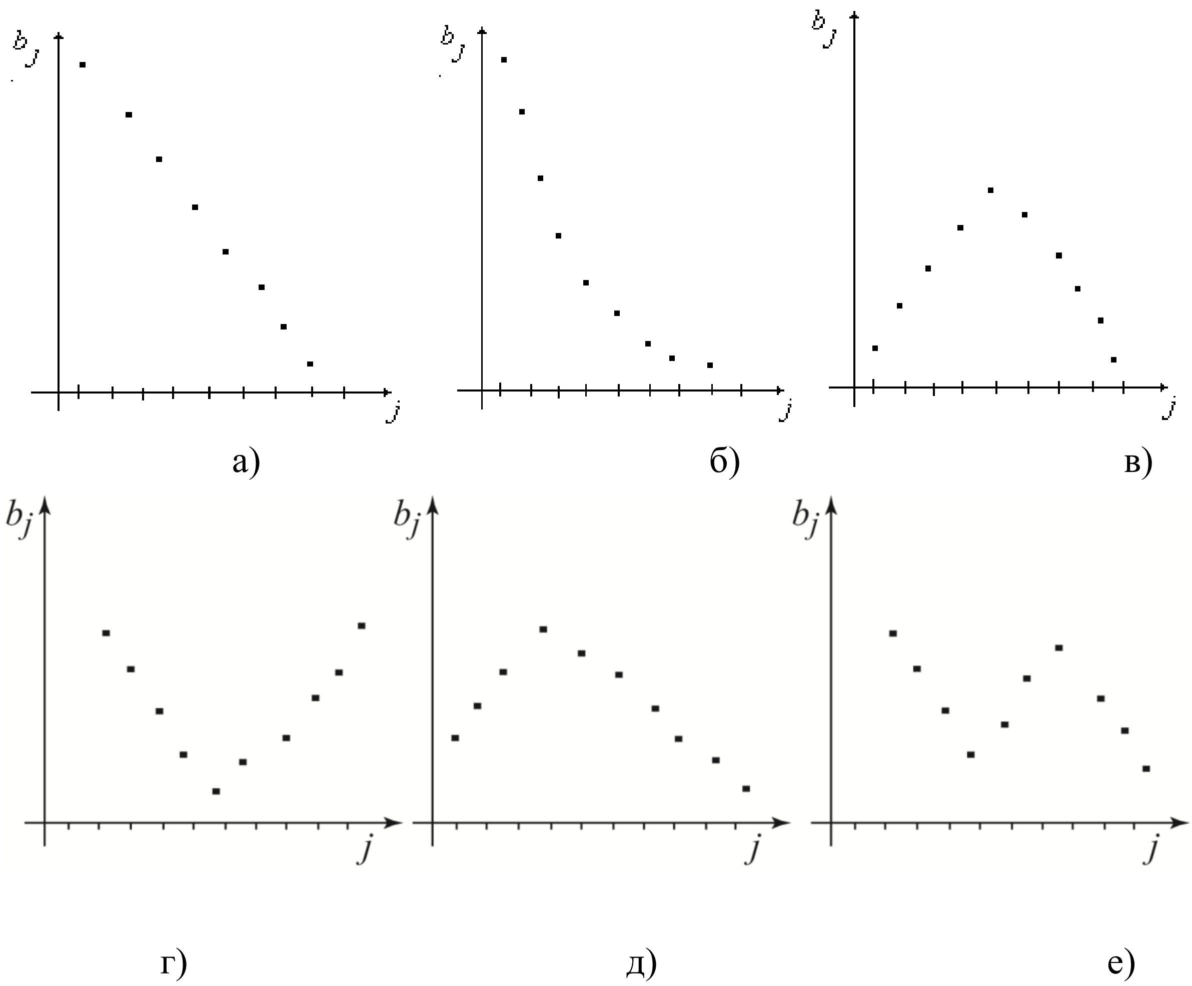


Рис. 28.1. **Графічне зображення структури лага: а) лінійна, б) геометрична, в) перевернута V-подібна, г), д), е) поліноміальна**

Запитання для самоперевірки

1. Які моделі називаються динамічними?
2. Які типи динамічних моделей розрізняють?
3. Що називається лагом? Які змінні називаються лагові?
4. Які причини в економіці існування лагів?
5. Що називається короткостроковим мультиплікатором у моделях з розподіленим лагом?
6. Що називається проміжним і довгостроковим мультиплікаторами в моделях з розподіленим лагом?
7. Що характеризує середній лаг у моделях з розподіленим лагом?
8. Що характеризує медіанний лаг у моделях з розподіленим лагом?
9. Що називається короткостроковим мультиплікатором в моделі авторегресії?

10. Що характеризує середній лаг в моделі авторегресії?
11. Що характеризує медіанний лаг в моделі авторегресії?
12. У чому полягає особливість моделі адаптивних сподівань?
14. У чому полягає особливість моделі неповного коригування?
15. Як вивчити структуру лага в моделях з розподіленим лагом?

29. Методи розробки динамічних економетричних моделей

29.1. Метод Алмон.

29.2. Метод Койка.

29.1. Метод Алмон

Метод Алмон заснований на теоремі Вейєрштрасса, згідно з якою передбачається, що вагові коефіцієнти підпорядковані поліноміальному розподілу.

На рис. 29.1 представлений розподіл вагових коефіцієнтів.

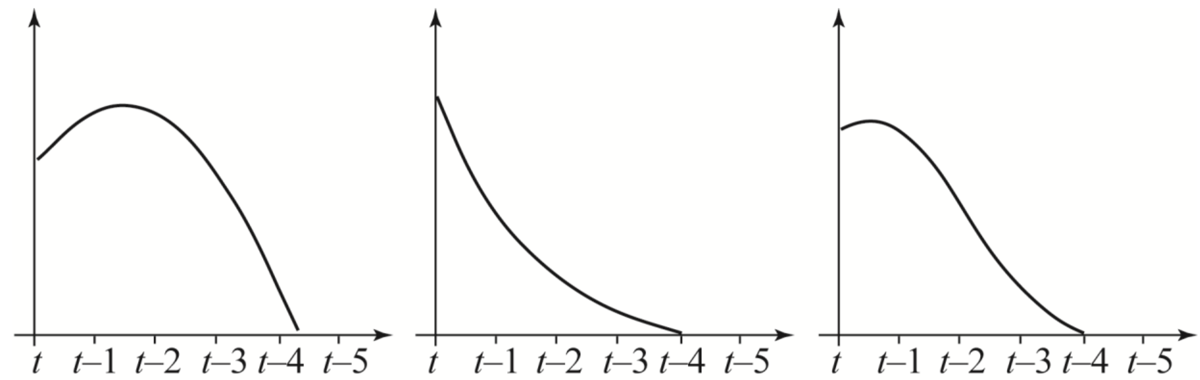


Рис.29.1. **Апріорні уявлення про розподіл вагових коефіцієнтів**

Передбачається, що подібні залежності можна апроксимувати поліномом невеликого ступеня другого або третього. Вибір функції залишається за дослідником і може бути розроблений на основі експериментів. Далі дослідник повинен вибрати число лагових значень змінної, що пояснює. І це число, визначене в результаті експерименту, спрямоване на одержання гарного опису наявних даних. Іншими словами, замість використання звичайного МНК оцінювання параметрів моделі

$$y_t = a + b_0 \cdot x_t + b_1 \cdot x_{i-1} + b_2 \cdot x_{t-2} + \dots + b_k \cdot x_{t-k},$$

пропонується спочатку апроксимувати поведінку коефіцієнтів регресії поліномом невеликого ступеня:

$$b_i = a_0 + a_1 \cdot i + a_2 \cdot i^2 + \dots a_m \cdot i^m, \quad (m \le k)$$

і тимчасово перейти до нових змінних:

$$\begin{aligned} Z_0 &= x_t + x_{t-1} + x_{t-2} + \dots + x_{t-k}, \\ Z_1 &= \qquad x_{t-1} + 2 \cdot x_{t-2} + \dots + k \cdot x_{t-k}, \\ Z_2 &= \qquad x_{t-1} + 2^2 \cdot x_{t-2} + \dots + k^2 \cdot x_{t-k}, \\ &\dots\dots\dots\dots\dots\dots\dots\dots \\ Z_m &= \qquad x_{t-1} + 2^m \cdot x_{t-2} + \dots + k^m \cdot x_{t-k}. \end{aligned}$$

Далі необхідно обчислити параметри моделі щодо нових змінних

$$y_t = a + a_0 \cdot Z_0 + a_1 \cdot Z_1 + a_2 \cdot Z_2 + \dots + a_m \cdot Z_m$$

і виконати зворотний перехід до початкової моделі

$$\begin{aligned} b_0 &= a_0, \\ b_1 &= a_0 + a_1 + a_2 + \dots + a_m, \\ b_2 &= a_0 + 2 \cdot a_1 + 2^2 \cdot a_2 + \dots + 2^m \cdot a_m, \\ &\dots\dots\dots\dots\dots\dots\dots\dots \\ b_k &= a_0 + k \cdot a_1 + k^2 \cdot a_2 + \dots + k^m \cdot a_m. \end{aligned}$$

Якщо $m = k$ (ступінь апроксимуючого полінома дорівнює максимальному лагу), то в кінці будуть отримані МНК-оцінки параметрів початкової моделі, які можна було отримати простіше – безпосереднім використанням МНК.

Якщо $m < k$, то можна отримати непередбачені результати. Для того, щоб запобігти нереальних висновків, рекомендується розглянути графіки залежностей $b_k = f(k)$ і задати найбільш відповідні значення m і k. Однак тоді втрачається об'єктивність методу. Зазвичай вказують про переваги методу Алмон у той час, коли маємо суттєві недоліки. Обсяг обчислювальної роботи значно більший, нові змінні пов'язані тісними мультиколінеарними зв'язками, допускаються не контрольовані помилки невдалої апроксимації, втрачається наукова об'єктивність, відсутні оцінки значущості b_i.

Приклад. У табл. 29.1 наведено за 24 періоди обсяги реалізації (y) і ціни за одиницю (x) продукції заводу «Укрелектромаш» (м. Харків). Складемо лагову модель

$$y_t = a + b_0 \cdot x_t + b_1 \cdot x_{i-1} + b_2 \cdot x_{t-2} + \dots + b_k \cdot x_{t-k},$$

де максимальний лаг k не може бути більшим 25 % від числа періодів, тобто $k \le 6$.

Для ілюстрації, звичайним МНК були обчислені параметри моделей для $k = 4$, $k = 5$, $k = 6$ разом з усіма їх статистичними характеристиками. На рис. 29.2 наведені графіки поведінки параметрів моделі b_i залежно від прийнятого максимального лага k. Бачимо, що представлені графіки відрізняються від графіків рис. 29.1.

Методом Алмон були обчислені всі комбінації моделей для $k = 4$, $k = 5$, $k = 6$ і $2 \le m \le k$.

У всіх комбінаціях де $m = k$ були отримані звичайні МНК-оцінки параметрів, але без статистичних оцінок їх значущості. Для прикладу розглянемо результати обчислень при $k = 4$ і $m = 2$, $m = 3$ (до речі, це найпоширеніший варіант обчислень за методом Алмон).

У табл. 29.1 наведені значення лагових змінних, на підставі яких обчислені значення нових змінних z_0, z_1, z_2, z_3. Втрачено $k = 4$ спостережень (17,7%). Значення z_j дуже швидко зростають з номером j. Між цими змінними існує дуже сильна кореляція, наприклад, $r_{z_2 z_3} = 0{,}999$.

Таблиця 29.1

Обсяги реалізації (y) і ціни за одиницю (x) продукції «Укрелектромаш» (м. Харків)

Період	y_t	x_t	x_{t-1}	x_{t-2}	x_{t-3}	x_{t-4}	Z_0	Z_1	Z_2	Z_3
1	2141	187,74								
2	2934	180,63	187,74							
3	3371	173,79	180,63	187,74						
4	5448	158,34	173,79	180,63	187,74					
5	7200	150,66	158,34	173,79	180,63	187,74	851,16	1 798,77	5 483,01	18 441,03
6	5497	157,82	150,66	158,34	173,79	180,63	1 008,98	2 649,93	9 931,71	41 137,53
7	5410	159,35	157,82	150,66	158,34	173,79	1 168,33	3 658,91	16 240,55	79 891,43
8	4089	171,39	159,35	157,82	150,66	158,34	1 151,98	3 513,06	15 527,44	76 363,32
9	1492	200,87	171,39	159,35	157,82	150,66	1 172,22	3 400,63	14 854,67	72 680,71
10	934	199,53	200,87	171,39	159,35	157,82	1 197,96	3 356,32	14 312,44	69 008,86
11	907	198,51	199,53	200,87	171,39	159,35	1 238,13	3 445,9	14 464,38	68 902,48
12	2630	183,20	198,51	199,53	200,87	171,39	1 270,67	3 629,41	15 211,97	72 195,07
13	3033	179,06	183,20	198,51	199,53	200,87	1 291,91	3 795,34	16 008,28	75 857,62
14	4192	170,12	179,06	183,20	198,51	199,53	1 302,68	3 971,8	17 082,72	81 903,34
15	4226	164,34	170,12	179,06	183,20	198,51	1 295,63	4 074,75	17 930,89	87 582,81
16	5047	160,63	164,34	170,12	179,06	183,20	1 255,39	3 964,29	17 533,39	85 996,95

17	1067	191,39	160,63	164,34	170,12	179,06	1 247,25	3 822,97	16 940,39	83 306,59
18	1008	196,10	191,39	160,63	164,34	170,12	1 244,84	3 680,65	16 106,59	78 754,99
19	9572	151,75	196,10	191,39	160,63	164,34	1 213,39	3 643,09	15 735,93	76 523,95
20	4259	167,87	151,75	196,10	191,39	160,63	1 202,2	3 603,06	15 461,56	74 456,82
21	1490	201,98	167,87	151,75	196,10	191,39	1 234,06	3 614,42	15 534	74 501,72
22	446	200,32	201,98	167,87	151,75	196,10	1 270,04	3 698,1	15 944,24	76 812,42
23	560	208,51	200,32	201,98	167,87	151,75	1 317,92	3 843,73	16 739,61	81 913,39
24	1256	205,55	208,51	200,32	201,98	167,87	1 332,08	3 821,92	16 366,88	79 334,56

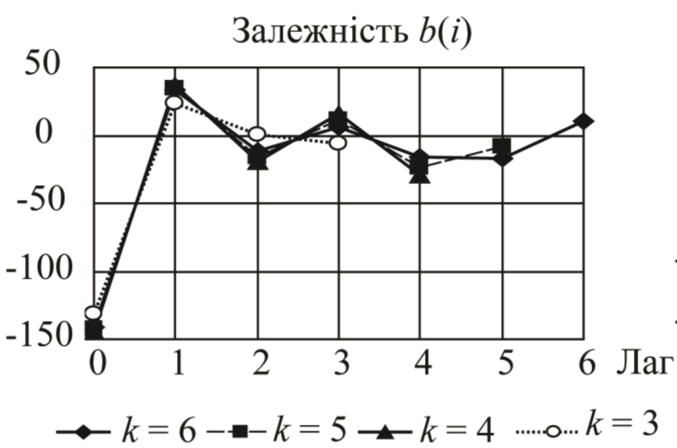


Рис. 29.2. **Залежність параметрів моделі від лага**

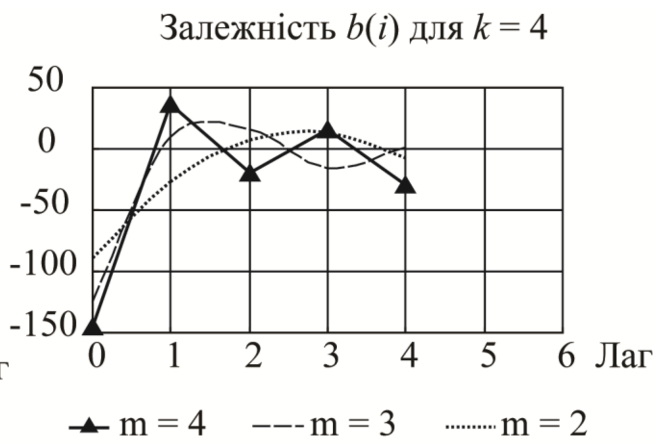


Рис. 29.3. **Якість апроксимації параметрів моделі**

На рис. 29.3 показана якість апроксимації залежності b_i поліномами 2-го і 3-го ступенів (поліном 4-го ступеня точно відтворює 5 заданих точок). Як бачимо, апроксимація не надто вдала. Для порівняння наведемо рівняння регресій для $k = 4$ і $m = 4$, $m = 3$, $m = 2$.

$m = 4$: $y_t = 28\,537 - 144{,}7x_t + 36{,}4x_{t-1} - 19{,}3x_{t-2} + 16{,}3x_{t-3} - 29{,}1x_{t-4}$

$m = 3$: $y_t = 23\,887 - 124{,}9x_t + 9{,}5x_{t-1} + 16{,}3x_{t-2} - 16{,}1x_{t-3} + 0{,}9x_{t-4}$

$m = 2$: $y_t = 21\,863 - 89{,}6x_t - 27{,}1x_{t-1} + 7{,}4x_{t-2} + 13{,}8x_{t-3} - 7{,}9x_{t-4}$

У цих моделях збігаються навіть знаки параметрів.

Це ще раз підтверджує перебільшене значення методу Алмон. З обчислювальної точки зору ніяких переваг не маємо, оскільки поліноми Алмон також тісно корелюють між собою, як і початкові лагові змінні.

29.2. Метод Койка

Метод Алмон використовується для побудови моделі з кінцевою величиною лага *l*. Для оцінки параметрів моделі з нескінченною величиною лага рекомендується метод Койка.

Койк Л. М. припустив, що існує деякий постійний темп λ $(0<\lambda<1)$ зменшення в часі лагових впливів фактора на результат:

$$b_j = b_0 \cdot \lambda; \quad j = 0,1,2,...$$

Чим ближче λ до 0, тим вище темп зниження впливу фактора на результат у часі і тим більше частка впливу на результат припадає на поточні значення чинника.

Виразимо за допомогою $b_j = b_0 \cdot \lambda; \quad j = 0,1,2,...$ всі коефіцієнти b_j у моделі $y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + ... + \varepsilon_t$ через b_0, λ:

$$y_t = a + b_0 x_t + b_0 \lambda x_{t-1} + b_0 \lambda^2 x_{t-2} + ... + \varepsilon_t .$$

Тоді для періоду $(t-1)$ модель можна записати наступним чином:

$$y_{t-1} = a + b_0 x_{t-1} + b_0 \lambda x_{t-2} + b_0 \lambda^2 x_{t-3} + ... + \varepsilon_{t-1} .$$

Помножимо обидві частини моделі на λ:

$$\lambda \cdot y_{t-1} = \lambda \cdot a + b_0 \lambda \cdot x_{t-1} + b_0 \lambda^2 x_{t-2} + b_0 \lambda^3 x_{t-3} + ... + \lambda \cdot \varepsilon_{t-1} .$$

Віднімемо даний вираз з $y_t = a + b_0 x_t + b_0 \lambda x_{t-1} + b_0 \lambda^2 x_{t-2} + ... + \varepsilon_t$:
$y_t - \lambda \cdot y_{t-1} = a - \lambda \cdot a + b_0 x_t + \varepsilon_t - \lambda \cdot \varepsilon_{t-1}$, після перетворення отримуємо модель Койка $y_t = a(1-\lambda) + b_0 x_t + \lambda \cdot y_{t-1} + u_t$, де $u_t = \varepsilon_t - \lambda \cdot \varepsilon_{t-1}$.

Отримана модель є модель двохфакторної лінійної регресії, точніше авторегресії. Визначивши її параметри, знайдемо λ і оцінки параметрів a і b_0 вихідної моделі. Далі за допомогою співвідношень $b_j = b_0 \cdot \lambda; \quad j = 0, 1, 2,...$ визначають параметри $b_1, b_2, ...$ моделі $y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + ... + \varepsilon_t$.

Відзначимо, що застосування звичайного МНК до оцінки параметрів моделі $y_t = a(1-\lambda) + b_0 x_t + \lambda \cdot y_{t-1} + u_t$ призведе до отримання зміщених оцінок її параметрів зважаючи на наявність в цій моделі як фактора лагової результативної змінної y_{t-1}.

Описаний алгоритм отримав назву перетворення Койка. Це перетворення дозволяє перейти від моделі з нескінченними розподіленими лагами до моделі авторегресії, що містить дві незалежні змінні x_t і y_{t-1}.

Незважаючи на нескінченне число лагових змінних у моделі $y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \ldots + \varepsilon_t$, геометрична структура лага дозволяє визначити величини середнього і медіанного лагів у моделі Койка. Оскільки сума коефіцієнтів регресії в даній моделі є сума геометричної прогресії, тобто

$$\sum_{j=0}^{\infty} b_j = b_0 + b_0\lambda + b_0\lambda^2 + b_0\lambda^3 + \ldots = b_0(\lambda + \lambda^2 + \lambda^3 + \ldots) = \frac{b_0}{1-\lambda},$$

то середній лаг визначається як:

$$\bar{l} = \frac{\sum_{j=0}^{\infty} j \cdot b_j}{\sum_{j=0}^{\infty} b_j} = \frac{b_0\lambda \cdot (\lambda + \lambda^2 + \lambda^3 + \ldots)}{\frac{b_0}{1-\lambda}} = \frac{b_0 \cdot \lambda \frac{1}{(1-\lambda)^2}}{\frac{b_0}{1-\lambda}} = \frac{\lambda}{1-\lambda}.$$

Зауважимо, що при $\lambda = 0{,}5$ середній лаг $T = 1$, а при $l < 0{,}5$ середній лаг $T < 1$, тобто вплив фактора на результат у середньому займає менше одного періоду часу. Величину $(1-\lambda)$ інтерпретують звичайно як швидкість, з якою відбувається адаптація результату в часі до зміни факторної ознаки. Для розрахунку медіанного лага необхідно виконання наступної умови:

$$\sum_{j=0}^{l_{Me}-1} b_j = \sum_{j=0}^{l_{Me}-1} \frac{b_j}{\sum_{j=0}^{\infty} b_j} = \sum_{j=0}^{l_{Me}-1} \frac{b_0 \cdot \lambda^j}{\frac{b_0}{1-\lambda}} = \sum_{j=0}^{l_{Me}-1} \lambda^j (1-\lambda) = 0{,}5.$$

Тому медіанний лаг в моделі Койка дорівнює: $l_{Me} = \frac{\ln 0{,}5}{\ln \lambda}$.

Приклад. Використавши дані економіки Канади протягом 1961 – 1981 рр., вчені Джеффрі Сакс і Майкл Бруно побудували модель:

$$U_t = \delta_0 + \delta_1 \cdot U_{t-1} + \delta_2 \cdot t + \delta_3 \cdot w_t + \varepsilon_t,$$

де U_t, U_{t-1} – рівень безробіття в періоди t і t – 1;

w_t – перевищення реальної заробітної плати порівняно з її рівнем у умовах повної зайнятості (значення отримані за допомогою обчислень);

t – час;

$\delta_0, \delta_1, \delta_2, \delta_3$ – параметри моделі;

ε_t – похибка;

$$U_t = \delta_0 + 0{,}63 \cdot U_{t-1} + 0{,}07 \cdot t + 15{,}72 \cdot w_t, \quad R^2 = 0{,}85.$$
$$t = 5{,}46 \qquad t = 2{,}82 \quad t = 2{,}23$$

Змінна w_t у моделі означає попит на роботу. Якщо передбачити, що змінна w_t здійснює вплив на рівень безробіття з нескінченним лагом в умовах геометричної структури лага, то відповідно до методу Койка отримаємо таку модель з розподіленим лагом:

$$U_t = a + b_0 w_t + b_0 \cdot \lambda \cdot w_{t-1} + b_0 \cdot \lambda^2 \cdot w_{t-2} + \ldots + c \cdot t + \varepsilon_t.$$

Ця модель відрізняється від $y_t = a + b_0 x_t + b_1 x_{t-1} + b_2 x_{t-2} + \ldots + \varepsilon_t$ тим, що, крім поточного та лагового значень факторної ознаки, вона враховує фактор часу *t*. Відповідні алгебраїчні перетворення згідно з методом Койка призводять до моделі авторегресії виду:

$$U_t = (a \cdot (1-\lambda) + \lambda \cdot c) + (1-\lambda) \cdot U_{t-1} + c \cdot (1-\lambda) \cdot t + b_0 w_t + u_t,$$

тобто $\delta_0 = a \cdot (1-\lambda) + \lambda \cdot c; \quad \delta_1 = 1-\lambda; \quad \delta_2 = c \cdot (1-\lambda); \ \delta_3 = b_0$.

У моделі Джеффрі Сакса і Майкла Бруно $\lambda = 0{,}63$. При цьому параметри моделі Койка:

$$c = \frac{0{,}07}{1-0{,}63} = 0{,}189;$$

$$a = \frac{\delta_0}{1-0{,}63} + 0{,}189 \cdot 0{,}63 = 0{,}119 + 2{,}703 \cdot \delta_0;$$

$$b_0 = 15{,}72;$$

$$b_1 = 15{,}72 \cdot 0{,}63 = 9{,}904;$$

$$b_2 = 15{,}72 \cdot (0{,}63)^2 = 6{,}239;$$

$$b_3 = 15{,}72 \cdot (0{,}63)^3 = 3{,}931 \text{ і т. д.}$$

Модель Койка набуває такого вигляду:

$$U_t = (0{,}119 + 2{,}703 \cdot \delta_0) + 15{,}72 \cdot w_t + 9{,}904 \cdot w_{t-1} + 6{,}239 \cdot w_{t-2} + 3{,}931 \cdot w_{t-3} + \\ + \ldots + 0{,}189t.$$

Середній лаг дорівнює $T = \dfrac{0{,}63}{1-0{,}63} = 1{,}703$.

Величина медіанного лага: $l_{Me} = \dfrac{\ln 0{,}5}{\ln 0{,}63} = \dfrac{-0{,}69314}{-0{,}46203} = 1{,}500203$.

Звідси можна зробити висновки, що в середньому вплив відхилень між реальною заробітною платою в економіці Канади і її величиною в умовах повної зайнятості проявляється протягом відносно короткого періоду часу, це 1,7 року, при цьому половина впливу реалізується протягом перших 1,5 років з моменту зміни.

Приклад. Обчислити модель з розподіленими лагами для опису витрат на житло (y) залежно від рівня доходів (x) і відносних цін поточного (p) і кількох попередніх періодів (табл. 29.2); перетворити цю модель в авторегресійну методом Койка; оцінити її параметри і зробити висновки щодо коротко і довгострокового впливу змінних, що пояснюють.

Таблиця 29.2

Значення змінних

Рік	y_t	x_t	p_t	$ln\, y_t$	$ln\, y_{t-1}$	$ln\, x_t$	$ln\, p_{t-1}$	$ln\, x_{t-1}$	$ln\, p_{t-1}$	$ln\, x_{t-2}$	$ln\, p_{t-2}$
1960	64,0	489,7	104,5	4,1589	4,1092	6,1938	4,6492	6,1732	4,6492	6,1327	4,6501
1961	67,0	503,8	105,1	4,2047	4,1589	6,2222	4,6549	6,1938	4,6492	6,1732	4,6492
1962	70,7	524,9	105,0	4,2584	4,2047	6,2632	4,6540	6,2222	4,6549	6,1938	4,6492
1963	74,0	542,3	104,8	4,3041	4,2584	6,2958	4,6521	6,2632	4,6540	6,2222	4,6549
1964	77,4	580,8	104,5	4,3490	4,3041	6,3644	4,6492	6,2958	4,6521	6,2632	4,6540
1965	81,6	616,3	104,0	4,4018	4,3490	6,4237	4,6444	6,3644	4,6492	6,2958	4,6521
1966	85,3	646,8	102,6	4,4462	4,4018	6,4720	4,6308	6,4237	4,6444	6,3644	4,6492
1967	89,1	673,5	102,2	4,4898	4,4462	6,5125	4,6269	6,4720	4,6308	6,4237	4,6444
1968	93,5	701,3	100,9	4,5380	4,4898	6,5529	4,6141	6,5125	4,6269	6,4720	4,6308
1969	98,4	722,5	100,0	4,5890	4,5380	6,5827	4,6052	6,5529	4,6141	6,5125	4,6269
1970	102,0	751,6	99,6	4,6250	4,5890	6,6222	4,6012	6,5827	4,6052	6,5529	4,6141
1971	106,4	779,2	100,0	4,6672	4,6250	6,6583	4,6052	6,6222	4,6012	6,5827	4,6052
1972	112,5	810,3	100,0	4,7230	4,6672	6,6974	4,6052	6,6583	4,6052	6,6222	4,6012
1973	118,2	865,3	99,1	4,7724	4,7230	6,7631	4,5961	6,6974	4,6052	6,6583	4,6052
1974	124,2	858,4	95,1	4,8219	4,7724	6,7551	4,5549	6,7631	4,5961	6,6974	4,6052
1975	128,3	875,8	93,3	4,8544	4,8219	6,7751	4,5358	6,7551	4,5549	6,7631	4,5961
1976	134,9	906,8	93,7	4,9045	4,8544	6,8099	4,5401	6,7751	4,5358	6,7551	4,5549
1977	141,3	942,9	94,5	4,9509	4,9045	6,8490	4,5486	6,8099	4,5401	6,7751	4,5358
1978	148,5	988,8	94,7	5,0006	4,9509	6,8965	4,5507	6,8490	4,5486	6,8099	4,5401
1979	154,8	1015,5	93,8	5,0421	5,0006	6,9231	4,5412	6,8965	4,5507	6,8490	4,5486
1980	159,8	1021,6	93,0	5,0739	5,0421	6,9291	4,5326	6,9231	4,5412	6,8965	4,5507
1981	164,8	1049,3	94,2	5,1047	5,0739	6,9559	4,5454	6,9291	4,5326	6,9231	4,5412
1982	167,5	1058,3	96,7	5,1210	5,1047	6,9644	4,5716	6,9559	4,5454	6,9291	4,5326
1983	171,3	1095,4	99,2	5,1434	5,1210	6,9989	4,5971	6,9644	4,5716	6,9559	4,5454

Для демонстрації були обчислені моделі, де $x_{t-1}, p_{t-1}, x_{t-2}, p_{t-2}$ лагові змінні, що враховують зрушення в один і два періоди:

1) $\ln y_p = -1{,}51 + 1{,}81 \ln x - 0{,}35 \ln p$; $R^2 = 0{,}991$;

2) $ln\, y_p = 0{,}86 + 1{,}09\, ln\, x_{t-1} - 0{,}73\, ln\, p_{t-1};\; R^2 = 0{,}993;$

$$t_{b_0} = 0{,}5; \quad t_{b_{xt-1}} = 19{,}4; \quad t_{b_{pt-1}} = 2{,}2;$$

3) $ln\, y_p = 2{,}31 + 1{,}02\, ln\, x_{t-2} - 0{,}95\, ln\, p_{t-2};\; R^2 = 0{,}996;$

$$t_{b_0} = 1{,}4; \quad t_{b_{xt-2}} = 20{,}6; \quad t_{b_{pt-2}} = 3{,}3.$$

Таким чином, обчислені моделі майже не мають відмінностей: значення еластичностей витрат за доходами майже однакові; всі рівняння значущі в цілому і мають майже однакові коефіцієнти детермінації. Тому не має ніяких переваг та чи інша моделі. Це не випадково, оскільки лагові змінні звичайно тісно корелюють. Тому коефіцієнти регресії в будь-якому з цих рівнянь фактично відображають сумарні ефекти за кількома часовими періодами. Правильна модель повинна враховувати ефекти кожного періоду окремо, тобто необхідно розробити модель з розподіленим лагом:

4) $ln\, y_p = -1{,}51 + 1{,}81\, ln\, x_t - 0{,}35\, ln\, p_t;$

(t_b) (0,8) (21,0) (1,1)

$$t_{b_0} = 0{,}8; \quad t_{b_x} = 21{,}0; \quad t_{b_p} = 1{,}1$$

5) $ln\, y_p = 0{,}17 + 0{,}22\, ln\, x_t + 1.00\, ln\, p_t + 0{,}88\, ln\, x_{t-1} - 1{,}60\, ln\, p_{t-1};$

$$t_{b_0} = 0{,}1; \quad t_{b_{ln\,x}} = 0{,}8; \quad t_{b_{ln\,p}} = 2{,}8; \quad t_{b_{ln\,xt-1}} = 2{,}9; \quad t_{b_{ln\,pt-1}} = 4{,}0;$$

6) $$\begin{aligned} \ln y_p &= 2{,}14 + 0{,}31 \ln x_t - 0{,}16 \ln p_t + 0{,}18 \ln x_{t-1} - 0{,}65 \ln p_{t-1} + \\ &+ 0{,}55 \ln x_{t-2} - 1{,}44 \ln p_{t-2}; \end{aligned}$$

$$t_{b_0} = 1{,}2; \quad t_{b_{ln\,x}} = 1{,}2; \quad t_{b_{ln\,p}} = 0{,}3; \quad t_{b_{ln\,xt-1}} = 0{,}4; \quad t_{b_{ln\,pt-1}} = 0{,}7;$$

$$t_{b_{ln\,xt-2}} = 1{,}6; \quad t_{b_{ln\,pt-2}} = 2{,}5.$$

Усі три моделі значущі в цілому і мають майже однакові коефіцієнти детермінації $R^2 = 0{,}991 \div 0{,}997$. Проте зі збільшенням членів коефіцієнти регресії стають все менш значимі і все більш не стійкими. Це пояснюється мультиколінеарністю – тісними кореляційними зв'язками між лаговими змінними.

За допомогою перетворень Л. Койка отримаємо модель:

$$ln\, y_p = 0{,}50 + 0{,}15\, ln\, x_t - 0{,}16\, ln\, p_t + 0{,}845\, ln\, y_{t-1};\; R^2 = 0{,}9996$$

$$t_{b_0} = 1{,}3; \quad t_{b_{ln\,x}} = 3{,}1; \quad t_{b_{ln\,p}} = 2{,}4; \quad t_{b_{ln\,yt-1}} = 22{,}4.$$

За допомогою даної моделі можна оцінити коротко- і довгострокові ефекти. У короткостроковому аспекті (для поточного періоду) значення y_{t-1} необхідно розглядати як фіксоване, тоді еластичність витрат за доходами і ціни будуть дорівнювати коефіцієнтам регресії: $b = 0{,}15$; $c = -0{,}16$. У довгостроковому аспекті, коли *x*, *p*, *y* прирівнюються до своїх рівноважових значень (тоді $y_{t-1} = y_t$), виявляється, що довгостроковий вплив *x* на *y* дорівнює величині $\frac{b}{1-q} = \frac{0{,}15}{1-0{,}845} = 0{,}96$, а довгостроковий вплив *p* на *y* дорівнює $\frac{c}{1-q} = \frac{-0{,}16}{1-0{,}845} = -1{,}02$. Ці числа близькі до значимих коефіцієнтів регресії моделі (29.1) − (29.3).

Перетворення Койка ґрунтується на декількох обтяжуючих припущеннях, що коефіцієнти регресії при лагових змінних експоненціально зменшуються, починаючи з першого члена. Іноді слід передбачити, що змінення залежної змінної *y* відповідь на змінення незалежної змінної спочатку невелика, а потім зростає з часом і тільки потім починає зменшуватися. Щоб врахувати цей можливий ефект, приймемо гіпотезу про експоненціальне зменшення вагових коефіцієнтів не з першого, а з другого члена. Тоді перетворення Койка призведе до моделі:

$$ln\,y = a + (b\,ln\,x + c\,ln\,p) + \left(b_0\,ln\,x_t + c_0\,ln\,p_t\right) + qy_t + u,$$

де $b_0 = b_1 - bq;\ c_0 = c_1 - cq$.

Вагові коефіцієнти для цін минулих років обчислюються аналогічно. Параметри b_0, c_0 уточнюють поведінку першого члена (b_1, c_1).

Запитання для самоперевірки

1. У чому суть методу Алмон?

2. При якій структурі лага застосуємо метод Алмон?

3. У чому відмінність застосування методу Алмон при побудові реальної моделі з розподіленим лагом?

4. У чому суть перетворення Койка?

5. При якій структурі лага застосуємо метод Койка?

6. У чому відмінність застосування методу Койка при побудові реальної моделі з розподіленим лагом?

30. Моделювання одновимірних часових рядів

30.1. Основні елементи часового ряду.

30.2. Автокореляція рівнів часового ряду і виявлення його структури.

30.3. Згладжування часових рядів за допомогою ковзних середніх.

30.1. Основні елементи часового ряду

В аналізі багатьох економічних показників використовують щорічні, щоквартальні, щомісячні, щоденні дані. Прикладом цього є річні дані про валовий національний продукт (ВНП), випуск валової продукції (ВВП), обсяг чистого експорту і т. д., місячні дані про обсяг продажу продукції, щоденні дані виробництва продукції на конкретному підприємстві. Для об’єктивності аналізу дані слід систематизувати в часі. Упорядковані дані за часом їх отримання є часовим рядом.

Моделі, побудовані на основі даних, що характеризують один об'єкт за ряд послідовних моментів (періодів) часу, називаються моделями часових рядів.

Кожен рівень часового ряду формується під впливом великої кількості факторів, які умовно можна поділити на три групи:

1) фактори, що формують тенденцію ряду;

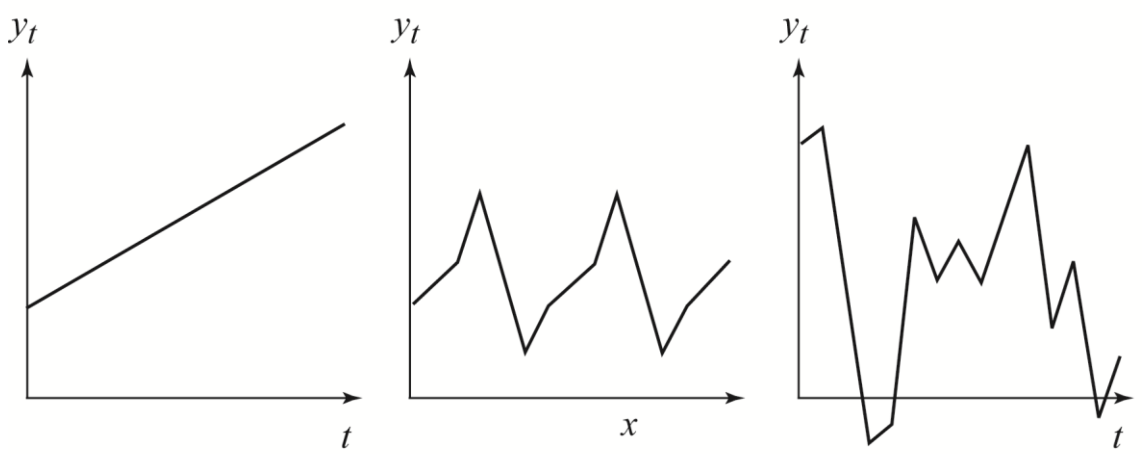
2) фактори, що формують циклічні коливання ряду;

3) випадкові чинники.

Кожен рівень часового ряду формується під впливом тенденції, сезонних коливань і випадкової компоненти. Процедура ідентифікації називається декомпозицією. Кожна компонента ідентифікується окремо. Потім вклади кожної компоненти комбінуються з метою отримання прогнозів майбутніх значень часових рядів. Методи декомпозиції використовуються як для короткочасних, так і для довготривалих прогнозів. З їх допомогою можна відображати зростання або спад, що лежить в основі ряду, або коригувати значення ряду, виключаючи з них одну або кілька компонент.

Слід зазначити, що останнім часом до прогнозів, зроблених на основі методу декомпозиції, відносяться як до проміжних, а сам метод розглядають як інструмент досягнення розуміння часових рядів.

Метод декомпозиції припускає виділення компонент: трендової, циклічної, сезонної і випадкової. Тренд – це компонента, що представляє основне зростання (спад) у часовому ряді. Трендова компонента утворюється під впливом постійної зміни факторів і позначається буквою T. Циклічна компонента − послідовність хвилеподібних флуктуацій або цикли більше одного року. Зміна економічних умов зазвичай відбувається циклічно. Позначається циклічна компонента буквою C. На практиці ідентифікувати цикл складно, він часто здається частиною тренда. У цьому випадку компоненту називають трендово-циклічною і позначають буквою T. Сезонні зміни зазвичай присутні в квартальних, місячних і тижневих даних. Під сезонними варіаціями розуміються зміни з більш-менш стабільною структурою, що мають річну циклічність. Сезонна компонента позначається буквою S. Випадкова компонента обумовлена впливом безліччю різноманітних подій, які самі по собі несуттєві, але спільно можуть дати значний ефект E. При різних поєднаннях у досліджуваному явищі або процесі ці компоненти можуть приймати різні форми. На рис. 30.1а) представлений гіпотетичний часовий ряд, що містить зростаючу тенденцію; на рис. 30.1б) – гіпотетичний часовий ряд містить лише сезонну компоненту; на рис. 30.1в) – гіпотетичний часовий ряд, який містить лише випадкову компоненту.



а) б) в)

Рис. 30.1. **Основні компоненти часового ряду: а) зростаюча тенденція; б) сезонна компонента; в) випадкова компонента**

У більшості випадків фактичний рівень часового ряду можна представити як суму або добуток трендової, циклічної і випадкової компонент. Двома найпростішими моделями, які зв'язують величину часового ряду, що спостерігається y_t з компонентами тренду T, сезонності S та випадковості E є модель адитивних компонент і модель мультиплікативних компонент. Модель, в якій часовий ряд представлений як сума перерахованих компонент, називається адитивною моделлю часового ряду: $Y_t = T_t + S_t + E_t$. Модель, в якій часовий ряд представлений як добуток перерахованих компонент, називається мультиплікативною моделлю часового ряду: $Y_t = T_t \cdot S_t \cdot E_t$. Модель адитивних компонент застосовується в тих випадках, коли аналізований часовий ряд має приблизно однакові зміни протягом всієї тривалості ряду або іншими словами: всі значення ряду істотно зменшуються в межах смуги постійної ширини, центрованої на рівні ряду. Модель мультиплікативних компонент ефективніша в тих випадках, коли зміна часової послідовності збільшується зі зростанням рівня, тобто значення ряду розходяться як такі, що мають тренд, а послідовність значень, що спостерігається нагадує рупор або лійку.

Основна задача економетричного дослідження окремого часового ряду – виявлення і додання кількісного вираження кожної з перерахованих вище компонент з тим, щоб використовувати отриману інформацію для прогнозування майбутніх значень ряду або при побудові моделей взаємозв'язку двох або більше часових рядів.

30.2. Автокореляція рівнів часового ряду і виявлення його структури

За наявності у часовому ряді тенденції і 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 кореляції між рівнями вихідного часового ряду і рівнями цього ряду, зсунутими на кілька кроків у часі.

Приклад. Розрахунок коефіцієнтів автокореляції рівнів для часового ряду витрат на кінцеве споживання.

Нехай є такі умовні дані (табл. 30.1) про середні витрати на кінцеве споживання (y_t , ден. од.) за 8 років.

Таблиця 30.1

Розрахунок коефіцієнта автокореляції першого порядку

t	y_t	y_{t-1}	y_t - $\overline{y_1}$	y_{t-1}- $\overline{y_2}$	$(y_t - \overline{y_1})\cdot(y_{t-1}$-$\overline{y_2})$	$(y_t - \overline{y_1})^2$	$(y_{t-1} - \overline{y_2})^2$
1	7	-	-	-	-	-	-
2	8	7	-3,29	-3	9,87	10,8241	9
3	8	8	-3,29	-2	6,58	10,8241	4
4	10	8	-1,29	-2	2,58	1,6641	4
5	11	10	-0,29	0	0	0,0841	0
6	12	11	0,71	1	0,71	0,5041	1
7	14	12	2,71	2	5,42	7,3441	4
8	16	14	4,71	4	18,84	22,1841	16
Усього	86	70	-0,03	0	44,0	53,4287	38

$$r_1 = \frac{\sum_{t=2}^{n}\left(y_t - \overline{y_1}\right)\left(y_{t-1} - \overline{y_2}\right)}{\sqrt{\sum_{t=2}^{n}\left(y_t - \overline{y_1}\right)^2\left(y_{t-1} - \overline{y_2}\right)^2}} = \frac{44}{\sqrt{53,42 \cdot 38}} = 0,976\,,$$

$$\overline{y_1} = \frac{\sum_{t=2}^{n} y_t}{n-1} = \frac{8+8+10+11+12+14+16}{7} = 11,29\,,$$

$$\overline{y_2} = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1} = \frac{7+8+8+10+11+12+14}{7} = 10\,.$$

Отримане значення свідчить про дуже тісну залежність між витратами на кінцеве споживання поточного і що безпосередньо передує років і, отже, про наявність у часовому ряді витрат на кінцеве споживання сильної лінійної тенденції.

Аналогічно можна визначити коефіцієнти автокореляції другого і більш високих порядків. Так, коефіцієнт автокореляції другого порядку характеризує тісноту зв'язку між рівнями (табл. 30.2).

Таблиця 30.2

Розрахунок коефіцієнта автокореляції другого порядку

t	y_t	y_{t-2}	$y_t - \overline{y_3}$	$y_{t-2} - \overline{y_4}$	$(y_t - \overline{y_3}) \cdot (y_{t-2} - \overline{y_4})$	$(y_t - \overline{y_3})^2$	$(y_{t-2} - \overline{y_4})^2$
1	7	-	-	-	-	-	-
2	8	-	-	-	-	-	-
3	8	7	-3,83	-2,33	8,9239	14,6689	5,4289
4	10	8	-1,83	-1,33	2,4339	3,3489	1,7689
5	11	8	-0,83	-1,33	1,1039	0,6889	1,7689
6	12	10	0,17	0,67	0,1139	0,0289	0,4489
7	14	11	2,17	1,67	3,6239	4,7089	2,7889
8	16	12	4,17	2,67	11,1339	17,3889	7,1289
Усього	86	56	0,02	0,02	27,3334	40,8334	19,3334

$$r_2 = \frac{\sum_{t=2}^{n}\left(y_t - \overline{y_3}\right)\left(y_{t-2} - \overline{y_4}\right)}{\sqrt{\sum_{t=2}^{n}\left(y_t - \overline{y_3}\right)^2\left(y_{t-2} - \overline{y_4}\right)^2}} = \frac{27,3334}{\sqrt{40,8334 \cdot 19,3334}} = 0,973,$$

$$\overline{y_3} = \frac{\sum_{t=3}^{n} y_t}{n-2} = \frac{8+10+11+12+14+16}{6} = 11,83,$$

$$\overline{y_4} = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1} = \frac{7+8+8+10+11+12}{6} = 9,33.$$

Отримані результати підтверджують висновок про те, що ряд витрат на кінцеве споживання містить лінійну тенденцію.

Число періодів, за якими розраховується коефіцієнт автокореляції, називається лагом. Є думка, що максимальний лаг повинен бути не більше $\frac{n}{4}$. Коефіцієнт автокореляції характеризує тісноту тільки лінійного зв'язку поточного та попереднього рівнів ряду. Для деяких часових рядів, що мають сильну нелінійну тенденцію, коефіцієнт автокореляції може наближатися до нуля. За знаком коефіцієнта автокореляції не можна робити висновок про зростаючу або спадаючу тенденції в рівнях ряду.

Послідовність коефіцієнтів автокореляції рівнів першого, другого і так далі порядків називають автокореляційною функцією часового ряду. Графік залежності її значень від величини лага (порядку коефіцієнта автокореляції) називається корелограмою. За допомогою аналізу автокореляційної функції і корелограми можна виявити структуру ряду.

30.3. Згладжування часових рядів за допомогою ковзних середніх

Під час аналізу часового ряду існує основне завдання – визначення основної тенденції у розвитку досліджуваного явища. У деяких випадках загальна тенденція простежується в динаміці показника, в інших ситуаціях вона не може проглядатися через існуючі випадкові коливання. Наприклад, в окремі моменти часу сильні коливання в курсах акцій можуть заступити наявність тенденції до зростання або зниження цього показника. На практиці простим методом виявлення загальної тенденції є укрупнення інтервалів. Наприклад, ряд тижневих даних можна перетворити в ряд місячних даних, ряд квартальних даних − у річні. Таким перетворенням може бути підсумовування рівнів вихідного ряду або знаходження середніх значень. Цей метод називається згладжуванням часового ряду. Суть цього методу полягає в заміні фактичних рівнів часового ряду розрахунковими, які меншою мірою схильні до коливань і сприяють більш чіткому прояву тенденції розвитку.

Згідно з різними підходами методи згладжування поділяються на дві групи: аналітичний та алгоритмічний. Аналітичний підхід припускає завдання загального вигляду функції, що описує регулярну, невипадкову складову, а потім проводиться статистичне оцінювання невідомих коефіцієнтів моделі і визначаються згладжені значення рівнів часового ряду шляхом підстановки відповідного значення в отримане рівняння. У алгоритмічному підході передбачається алгоритм розрахунку невипадкової складової в будь-який заданий момент часу. До алгоритмічного підходу відносяться ковзні середні, які дозволяють згладити як випадкові, так і періодичні коливання, виявити наявну тенденцію у розвитку процесу.

Алгоритм згладжування за простою ковзною середньою реалізується за такими етапами: 1) визначають довжину інтервалу згладжування l, що включає в себе l послідовних рівнів ряду ($l < n$).Чим ширше інтервал згладжування, тим більшою мірою поглинаються коливання, і тенденція розвитку носить більш плавний характер. Чим сильніше коливання, тим ширше повинен бути інтервал згладжування; 2) розбивають весь період спостережень на частини й інтервал згладжування переходить по ряду з кроком, що дорівнює 1; 3) розраховують середні арифметичні з рівнів ряду, що утворюють кожну частину; 4) замінюють фактичні значення ряду, що стоять в центрі кожної частини, на відповідні середні значення. Рекомендують довжину інтервалу згладжування l вибирати непарним числом $l = 2p + 1$, оскільки в цьому випадку отримані значення ковзної середньої припадають на середній член інтервалу. Спостереження, які беруться для розрахунку середнього значення, називаються активною частиною. При непарному значенні $l = 2p + 1$ всі рівні активної частини можуть бути представлені у вигляді:

$$y_{t-p}, y_{t-p+1}, ..., y_{t-1}, y_t, y_{t+1}, ..., y_{t+p-1}, y_{t+p},$$

де y_t – центральний рівень активної частини;

$y_{t-p}, y_{t-p+1}, ..., y_{t-1}$ – послідовність з p рівнів активної частини, попередньої центральної;

$y_{t+1}, ..., y_{t+p-1}, y_{t+p}$ – послідовність з рівнів активної частини, що йде за центральною.

У цьому випадку ковзна середня визначається за формулою:

$$\overline{y_t} = \frac{\sum_{i=t-p}^{t+p} y_i}{2p+1} = \frac{y_{t-p} + y_{t-p+1} + ... + y_{t+p-1}, y_{t+p}}{2p+1},$$

де y_i – фактичне значення i-го рівня; $\overline{y_t}$ – значення ковзної середньої в момент t; $2p+1$ – довжина інтервалу згладжування.

У ході використання простої ковзної середньої вирівнювання в кожній активній частині проводиться по прямій і апроксимація невипадкової складової здійснюється за допомогою лінійної функції $\overline{y_t} = a_0 + a_1 t$.

Метод згладжування призводить до усунення періодичних коливань в часовому ряді, якщо довжина інтервалу згладжування дорівнює або кратна періоду коливань. Рекомендується для усунення сезонних коливань використовувати ковзні середні з довжиною інтервалу згладжування, що дорівнює 4 або 12, проте при цьому не буде виконуватися умова непарності. У цьому випадку ковзна середня розраховується за формулою:

$$\overline{y_t} = \frac{\frac{1}{2}y_{t-p} + y_{t-p+1} + \ldots + y_{t-1} + y_t + y_{t+1} + \ldots + y_{t+p-1} + \frac{1}{2}y_{t+p}}{2p} =$$

$$= \frac{\frac{1}{2}y_{t-p} + \sum\limits_{i=t-p+1}^{t+p-1} y_i + \frac{1}{2}y_{t+p}}{2p}.$$

Тому для згладжування сезонних коливань з квартальним часовим рядом або місячним часовим рядом використовують 4-членну і 12-членну ковзну середню:

$$\overline{y_t} = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4};$$

$$\overline{y_t} = \frac{\frac{1}{2}y_{t-6} + y_{t-51} + \ldots + y_t + \ldots + y_{t+5} + \frac{1}{2}y_{t+6}}{12}.$$

Метод простої ковзної середньої рекомендується застосовувати, якщо графічне зображення часового ряду нагадує пряму. Якщо ж даний процес розвивається нелінійно, то проста ковзна середня призводить до істотних спотворень і в цьому випадку рекомендують використовувати зважену ковзну середню:

$$\overline{y_t} = \frac{\sum_{i=t-p}^{t+p} y_i w_i}{\sum_{i=t-p}^{t+p} w_i},$$

де w_i – вагові коефіцієнти.

Вагові коефіцієнти мають властивості: 1) симетричні щодо центрального рівня; 2) сума ваги з урахуванням загального множника, винесеного за дужки, дорівнює одиниці; 3) є як позитивна, так і негативна вага, що дозволяє згладженій кривій зберігати різні згини кривої тренду.

Приклад. За даними про врожайність пшениці (y_t) за 16 років розрахувати трирічні і семирічні ковзні середні і графічно порівняти результати (табл. 30.3).

Таблиця 30.3

Врожайність пшениці, ц / га

t	1	2	3	4	5	6	7	8
y_t	10,3	14,3	7,7	15,8	14,4	16,7	15,3	20,2
t	9	10	11	12	13	14	15	16
y_t	17,1	7,7	15,3	16,3	19,9	14,4	18,7	20,7

Розв'язання. Результати розрахунків наведено в табл. 30.4

Таблиця 30.4

Розрахунок ковзних середніх

t	y_t	$l=3$	$l=7$
1	10,3	–	–
2	14,3	10,8	–
3	7,7	12,6	–
4	15,8	12,6	13,5
5	14,4	15,6	14,9
6	16,7	15,5	15,3
7	15,3	17,4	15,3
8	20,2	17,5	15,2
9	17,1	15,0	15,5
10	7,7	13,4	16,0
11	15,3	13,1	15,8

12	16,3	17,2	15,6
13	19,9	16,9	16,1
14	14,4	17,7	–
15	18,7	17,9	–
16	20,7	–	–

За трирічною ковзною середньою:

$$\bar{y}_2 = \frac{10,3 + 14,3 + 7,7}{3} = 10,8;\quad \bar{y}_3 = \frac{14,3 + 7,7 + 15,8}{3} = 12,6;$$

$$\bar{y}_4 = \frac{7,7 + 15,8 + 14,4}{3} = 12,6;\quad \bar{y}_5 = \frac{15,8 + 14,4 + 16,7}{3} = 15,6;$$

$$\bar{y}_6 = \frac{14,4 + 16,7 + 15,3}{3} = 15,5;\quad \bar{y}_7 = \frac{16,7 + 15,3 + 20,2}{3} = 17,4;$$

$$\bar{y}_8 = \frac{15,3 + 20,2 + 17,1}{3} = 17,5;\quad \bar{y}_9 = \frac{20,2 + 17,1 + 7,7}{3} = 15,0;$$

$$\bar{y}_{10} = \frac{17,1 + 7,7 + 15,3}{3} = 13,4;\quad \bar{y}_{11} = \frac{7,7 + 15,3 + 16,3}{3} = 13,1;$$

$$\bar{y}_{12} = \frac{15,3 + 16,3 + 19,9}{3} = 17,2;\quad \bar{y}_{13} = \frac{16,3 + 19,9 + 14,4}{3} = 16,9;$$

$$\bar{y}_{14} = \frac{19,9 + 14,4 + 18,7}{3} = 17,7;\quad \bar{y}_{15} = \frac{14,4 + 18,7 + 20,7}{3} = 17,9.$$

За семирічною ковзною середньою:

$$\bar{y}_4 = \frac{10,3 + 14,3 + 7,7 + 15,8 + 14,4 + 16,7 + 15,3}{7} = 13,5;$$

$$\bar{y}_5 = \frac{14,3 + 7,7 + 15,8 + 14,4 + 16,7 + 15,3 + 20,2}{7} = 14,9;$$

$$\bar{y}_6 = \frac{7,7 + 15,8 + 14,4 + 16,7 + 15,3 + 20,2 + 17,1}{7} = 15,3;$$

$$\bar{y}_7 = \frac{15,8 + 14,4 + 16,7 + 15,3 + 20,2 + 17,1 + 7,7}{7} = 15,3;$$

$$\bar{y}_8 = \frac{14,4 + 16,7 + 15,3 + 20,2 + 17,1 + 7,7 + 15,3}{7} = 15,2;$$

$$\bar{y}_9 = \frac{16,7 + 15,3 + 20,2 + 17,1 + 7,7 + 15,3 + 16,3}{7} = 15,5;$$

$$\bar{y}_{10} = \frac{15,3 + 20,2 + 17,1 + 7,7 + 15,3 + 16,3 + 19,9}{7} = 16,0;$$

$$\bar{y}_{11} = \frac{20,2 + 17,1 + 7,7 + 15,3 + 16,3 + 19,9 + 14,4}{7} = 15,8;$$

$$\bar{y}_{12} = \frac{17,1 + 7,7 + 15,3 + 16,3 + 19,9 + 14,4 + 18,7}{7} = 15,6;$$

$$\bar{y}_{13} = \frac{7,7 + 15,3 + 16,3 + 19,9 + 14,4 + 18,7 + 20,7}{7} = 16,1.$$

На рис. 30.2 представлені фактичний і згладжені ряди врожайності.

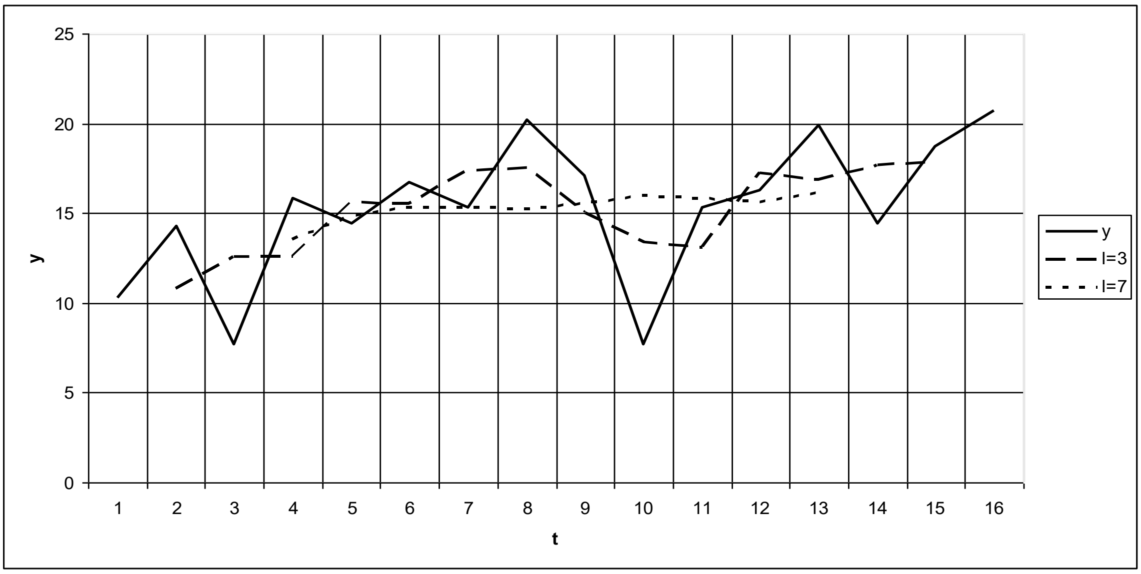


Рис. 30.2. **Згладжування ряду врожайності за допомогою ковзних середніх, де —— фактичні рівні y_t; — — $l=3$; ········ $l=7$**

Таким чином, ряд, згладжений за семирічною ковзною середньою ($l=7$), має гладкий характер порівняно з рядом, згладженим за трирічною ковзною середньою ($l=3$). Чим більше довжина інтервалу згладжування, тим більш гладким згладжений ряд виходить.

Запитання для самоперевірки

1. З яких компонентів складається часовий ряд?
2. Дати визначення кожної з компонент часового ряду.
3. Яка модель часового ряду називається адитивною?
4. Яка модель часового ряду називається мультиплікативною?
5. Які існують рекомендації про застосування адитивної і мультиплікативної моделей часового ряду?

6. Що характеризує коефіцієнт автокореляції?

7. Що називається автокорреляційною функцією часового ряду?

8. Що називається коррелограммою?

9. Як можна виявити структуру лага?

10. Яка основна задача аналізу часового ряду?

11. Які основні підходи в методах згладжування?

12. Які основні етапи алгоритму згладжування за простою ковзною середньою?

31. Моделювання тенденції часового ряду

31.1. Застосування моделей кривих росту в прогнозуванні основної тенденції розвитку.

31.2. Методи вибору кривих росту й оцінка адекватності і точності обраних моделей.

31.1. Застосування моделей кривих росту в прогнозуванні основної тенденції розвитку

Для моделювання тенденції розвитку процесу або явища в реальних економічних задачах найчастіше використовують моделі кривих росту. Це є функції часу $y = f(x)$, при цьому вважається, що вплив інших факторів несуттєво чи непрямо враховується через фактор часу.

Прогнозування на основі моделей кривих росту грунтується на екстраполяції, тобто на продовження на наступні періоди тенденції, яка встановлена за попередніми періодами. Оскільки даний метод прогнозування передбачає побудову аналітичної функції, що характеризує залежність рівнів ряду від часу або тренду, то він називається аналітичним вирівнюванням часового ряду.

У процедурі прогнозування на основі кривих росту виділяють такі етапи:

1) вибір однієї або декількох кривих, форма яких відповідає характеру зміни часового ряду;

2) оцінка параметрів обраних кривих;

3) перевірка адекватності обраних кривих прогнозованого процесу чи явища, оцінка точності моделей і остаточний вибір кривої росту;

4) розрахунок точкового та інтервального прогнозів.

Моделі кривих росту рекомендують розділяти на три групи. До першої групи належать функції, що використовуються для опису процесів з монотонним характером тенденції розвитку і відсутності меж зростання. Це характерно для тенденцій зміни багатьох економічних показників промислових підприємств. До другої групи належать криві, що описують процес, який має межу зростання в досліджуваному періоді. Такі процеси найчастіше демографічні, хоча зустрічаються і в дослідженні економічних процесів на промислових підприємствах. Функції, які належать до другого класу, називаються кривими з насиченням. Якщо криві насичення мають точки перегину, то вони відносяться до третього класу – до S-подібних кривих. За кривими третьої групи прогнозують процеси науково-технічного прогресу, нового виробництва продукції.

У прогнозуванні економічних показників за допомогою кривих росту найчастіше застосовуються такі функції: лінійний тренд: $y = a + bt$; гіпербола: $y = a + \frac{b}{t}$; експоненціальний тренд: $y = e^{a+bt}$; тренд у формі степеневої функції: $y = a \cdot t^b$; парабола другого та високих порядків: $y = a + b_1 t + b_2 t^2 + ... + b_k t^k$. У переважній більшості випадків розрахунок оцінок параметрів моделей здійснюється за допомогою методу найменших квадратів у формі регресійних моделей, в яких як залежна змінна слугують значення показників, а фактором є час. Для нелінійних трендових моделей застосовують процедуру лінеаризації.

31.2. Методи вибору кривих росту й оцінка адекватності і точності обраних моделей

Для визначення типу тенденції використовують кілька способів: якісний аналіз досліджуваного процесу, графічний метод, а також методи, що використовують коефіцієнти автокореляції рівнів ряду.

Найбільш простим методом вибору кривих зростання вважається візуальний, основний на графічному зображенні часового ряду. Якщо на графічному зображенні чітко не простежується тенденція, то рекомендують провести деякі стандартні перетворення ряду (наприклад, згладжування) і підібрати відповідну функцію. За допомогою сучасних програмних засобів, наприклад, статистичних пакетів процедуру, вибору кривої зростання здійснити дуже легко.

Вибір найкращого рівняння для побудови тренду можна здійснити шляхом перебору основних форм тренду, розрахунку за кожним рівнянням скоригованого коефіцієнта детермінації і вибору рівняння тренду з максимальним значенням скоригованого коефіцієнта детермінації.

Статистична якість побудованих моделей кривих росту для прогнозу перевіряється за критеріями перевірки якості побудованих регресійних моделей: критерію Стьюдента, критерію Фішера, критерію Дарбіна−Уотсона. Існування автокореляції залишків може істотно спотворювати прогнозні значення.

Приклад. Для проведення економічного аналізу зовнішньоекономічної діяльності промислового підприємства важливо прогнозувати значення показника ефективності реалізації продукції на внутрішньому ринку (y). У табл. 31.1 наведені квартальні значення показника протягом 5 років.

Таблиця 31.1

Квартальні значення показника

Y	1,225	1,242	1,264	1,205	1,277	1,289	1,241	1,235	1,236	1,234
t	1	2	3	4	5	6	7	8	9	10
Y	1,193	1,182	1,236	1,201	1,199	1,236	1,182	1,184	1,168	1,201
t	11	12	13	14	15	16	17	18	19	20

Побудувати модель кривої росту показника ефективності реалізації продукції промислового підприємства на внутрішньому ринку і прогнозувати його значення на два наступні квартали.

Розв'язок. За допомогою статистичного пакету Statgraphics Plus V5.1 International Professional спочатку була обчислена модель кривої зростання в лінійній формі: $y = 1{,}2605 - 0{,}0037t$, при цьому статичну якість моделі було перевірено за допомогою критеріїв Стьюдента, Фішера, скоригованого коефіцієнта детермінації і критерію Дарбіна-Уотсона: $t_a = 105{,}9$; $t_b = -3{,}74$; $F = 14{,}0$; $R^2 = 0{,}4062$; $DW = 1{,}999$. Модель відносно якісна. На наступні квартали прогноз значення показника ефективності реалізації продукції на внутрішньому ринку такий: 1,18245; 1,17873. Бачимо, що ефективність реалізації продукції на внутрішньому ринку у промислового підприємства буде трохи знижуватися в наступних двох кварталах.

У табл. 31.2 представлені розраховані альтернативні моделі за основними кривими зростання за рейтингом значень коефіцієнтів детермінації.

Таблиця 31.2

Альтернативні моделі

Альтернативні моделі	Скоригований коефіцієнт детермінації
1	2
$y=\dfrac{1}{0{,}7931+0{,}0025t}$	0,4124
$y=e^{0{,}2317-0{,}003t}$	0,4094
$y=(1{,}1228-0{,}0017t)^2$	0,4078
$y=1{,}2605-0{,}0037t$	0,4062
$y=1{,}248-0{,}00029t-0{,}000163t^2$	0,3961
$y=1{,}2829-0{,}0199\sqrt{t}$	0,3407
$y=1{,}2672t^{-0{,}0175}$	0,2348
$y=1{,}2667-0{,}0213\,ln\,t$	0,2320
$y=\dfrac{1}{0{,}8242-\dfrac{0{,}0275}{t}}$	0,0256
$y=e^{0{,}1937+\frac{0{,}033}{7}}$	0,0238
$y=1{,}2142+\dfrac{0{,}0403}{t}$	0,0219

З табл. 31.2 видно, що виходячи із значення скоригованого коефіцієнта детермінації найкращою моделлю є $y=\dfrac{1}{0{,}7931+0{,}0025t}$. Прогноз по цій моделі ефективності реалізації продукції на внутрішньому ринку такий: 1,18282; 1, 17935. Розбіжність прогнозних значень, розрахованих за двома моделями, різняться в тисячних одиницях.

Показники точності моделі прогнозування описують значення випадкових похибок, отриманих у процесі використання моделі. Відомо, що похибка прогнозу – це величина, що характеризує відхилення між фактичним і прогнозним значенням показника.

Абсолютна похибка прогнозу визначається за формулою:

$$\Delta_t = \hat{y}_t - y_t,$$

де $\hat{y}_t$ – прогнозне значення показника,

y_t – фактичне значення показника.

Абсолютну похибку прогнозу доповнюють обчисленням відносної похибки, яка виражається у процентах відносно фактичного значення показника:

$$\delta_t = \frac{\hat{y}_t - y_t}{y_t} \cdot 100\ \%.$$

Вважається, що коли абсолютна і відносна похибки більше 0, то це свідчить про «завищену» прогнозну оцінку, а коли менше 0, то прогнозне значення було занижено.

Також рекомендують використовувати середні похибки за модулем:

$$\left|\overline{\Delta}_t\right| = \frac{\sum_{t=1}^{n}\left|\hat{y}_t - y_t\right|}{n},$$

$$\left|\overline{\delta}_t\right| = \frac{1}{n}\sum_{t=1}^{n}\frac{\hat{y}_t - y_t}{y_t} \cdot 100\ \%,$$

де n – число рівнів часового ряду, для яких визначалось прогнозне значення.

Характеристику $\left|\overline{\delta}_t\right|$ (Mean Absolute Percentage Error – MAPE) широко використовують для порівняння точності прогнозів різнорідних об’єктів прогнозування. Фахівці вважають, якщо значення $\left|\overline{\delta}\ \right|$<10 % це свідчить про високу точність моделі, якщо 10 %<$\left|\overline{\delta}\ \right|$<20 % , то прогноз визнається добрим, а при

20 %<$\left|\overline{\delta}\ \right|$<50 % – задовільним. Проте такий підхід не враховує специфіку часових рядів і характеру динаміки, призначення прогнозів та інших характеристик, що обумовлює «механічність» підходу.

Також під час порівняння моделей прогнозування використовується середня квадратична похибка:

$$s=\sqrt{\frac{\sum_{t=1}^{n}\left(\hat{y}_t-y_t\right)^2}{n}},$$

іноді в знаменнику замість довжини часового ряду n використовується величина $(n-k)$, де k – число коефіцієнтів, які оцінюються в моделі. У ході вибору кращого методу прогнозування використовують суми квадратів похибок (Sums of Squares Errors – SSE), середні квадрати похибок (Mean Squares Errors – MSE). І чим менші значення всіх розглянутих характеристик, тим вища точність моделі прогнозу.

Фахівці за статистичними методами прогнозування вважають, що не може бути чисто формальних підходів до вибору методів і моделей прогнозування. Успішне застосування статистичних методів прогнозування на практиці можливо лише при поєднанні знань в області самих методів з глибоким знанням об'єкта дослідження, із змістовним аналізом досліджуваного явища.

Запитання для самоперевірки

1. З яких етапів складається процедура прогнозування на основі кривих росту?
2. Які типи моделі кривих росту розрізняють?
3. Які функції найчастіше використовуються для прогнозування економічних показників?
4. Перерахуйте методи вибору кривих росту.
5. Як здійснити вибір найкращого рівняння для побудови тренду?
6. Наведіть показники точності моделі прогнозування.

32. Моделювання сезонних і циклічних коливань

32.1. Загальна характеристика методів моделювання сезонних і циклічних коливань.

32.2. Статистичні методи оцінки рівня сезонності.

32.3. Приклади побудови адитивної і мультиплікативної моделей.

32.1. Загальна характеристика методів моделювання сезонних і

циклічних коливань

Проблемами періодичних коливань в економіці займаються давно. Відомі вчені К. Жюгляр, С. Кітчин, С. Кузнець, Н. Кондратьєв, пізніше У. Персонс, У. Мітчелл, Слуцький Є. Є., пізніше –Четвериков Н. С., Г. Ферстер та інші внесли істотний внесок у вирішення цих проблем.

Як відомо, моделювання циклічних коливань здійснюється майже за тією ж схемою, що й моделювання сезонних коливань, хоча моделювання циклів з довгими періодами має свої особливості [5, с. 84 – 101].

В процесі моделювання сезонних і циклічних коливань використовують кілька підходів:

1) на основі методу ковзних середніх;

2) із застосуванням *dummy*-змінних;

3) гармонічний аналіз;

4) на основі індексів сезонності;

5) сезонні різницеві оператори;

6) адаптивні моделі.

Модель сезонних коливань із застосуванням *dummy*-змінних розглянута в темі 10 «Узагальнені схеми регресійного аналізу». Слід зазначити, що число *dummy*-змінних повинно бути на одиницю менше числа різних періодів часу всередині одного сезонного циклу (при моделюванні поквартальних коливань таких змінних три, помісячних – одинадцять). Кожна *dummy*-змінна дорівнює одиниці для певного періоду і нулю для інших.

Моделювання сезонних коливань методами гармонійного аналізу ґрунтується на використанні гармонік ряду Фур'є:

$$y_t = a_0 + \sum_{k=1}^{m}\left(a_k \cos kt + b_k \sin kt\right),$$

де k – номер гармоніки (зазвичай ціле число від 1 до 4).

Параметри даного рівняння визначаються за допомогою МНК:

$$a_0 = \frac{\sum_t y_t}{n}; \; a_t = \frac{2}{n}\sum_t y_t \cos kt; \; b_t = \frac{2}{n}\sum_t y_t \sin kt.$$

Значення t змінюються від 0 з приростом $\frac{2\pi}{n}$, $n -$ число рівнянь часового ряду всередині сезонного циклу.

Розрахунок індексів сезонності ґрунтується на декількох підходах, виходячи з динаміки досліджуваних процесів.

Якщо середньорічні значення рівня ряду протягом досліджуваного періоду залишаються порівняно незмінними, то індекси сезонності обчислюють за формулами (спосіб постійної середньої):

$$i_t = \frac{\overline{y_k}}{\overline{y}} \cdot 100,$$

де $\overline{y_k}$ – середня арифметична фактичних рівнів одночасних місяців (кварталів);

$k = \overline{1,12}$ – для місяців, $k = \overline{1,4}$ – для кварталів;

$\overline{y}$ – середня арифметична рівнів за досліджуваний період.

Якщо протягом кожного року досліджуваного періоду тренд, що підвищується або знижується, відсутній або не значний, то індекси сезонності можуть бути обчислені за формулою:

$$i_t = \frac{\sum_{l=1}^{m} \frac{y_{lk}}{\overline{y_l}}}{m},$$

де y_{lk} – рівень часового ряду для k-го місяця (кварталу) l — го року;

$\overline{y_l}$ – середня арифметична фактичних рівнів ряду l — го року;

m – число років у досліджуваному періоді.

Застосування сезонних різницевих операторів призначене для виключення сезонних компонент з часового ряду і припускає перехід від вихідного часового ряду $y_t, i = \overline{1,n}$ до ряду вигляду $\Delta y_t = y_t - y_{t-\tau}, i = \tau + 1, ..., n$, де τ – період сезонності. Іноді використовують сезонні різницеві оператори більш високих порядків, наприклад, другого порядку $\Delta^2 y_t = y_t - y_{t-\tau}$.

Адаптивні методи і моделі враховують інформаційну нерівноцінність вихідних даних, наприклад, при прогнозуванні ретроспективна інформація не рівнозначна і слід більше враховувати дані, що розташовані ближче до теперішнього моменту часу.

У джерелі адаптивних методів лежить модель експоненціального згладжування.

Адаптивні моделі налаштовуються на нову інформацію, що відображає зміни умов. До адаптивних моделей належать моделі Брауна, Хольта, Хольта−Уінтерса, моделі авторегресії та ін.

Рекомендована наступна схема побудови адаптивної моделі:

1) за декількома першими рівняннями часового ряду необхідно знайти оцінки параметрів моделі;

2) за побудованою моделлю знайти прогноз на один крок вперед;

3) обчислити помилку прогнозування;

4) скорегувати моделі з урахуванням знайденої помилки;

5) за скоригованою моделлю розрахувати прогноз ще на один крок і т. д.

Відзначається найважливіша перевага адаптивних методів – побудова самокоректуючих моделей, здатних враховувати результат прогнозу, зробленого на попередньому кроці. Тому адаптивні методи особливо вдало використовуються під час оперативного, короткострокового прогнозування.

32.2. Статистичні методи оцінки рівня сезонності

У даний часу процесі опису та прогнозування тренд-сезонних процесів використовуються комбіновані підходи, що пов'язані із застосуванням індексів сезонності разом з кривими росту, процедури, що спираються на адаптивні моделі, сезонний варіант моделі ARIMA, спеціалізовані підходи, що враховують особливості конкретних часових рядів.

Нехай динаміка ряду характеризується незмінними в часі сезонними ефектами. Процедура розрахунку сезонної складової залежить від прийнятої моделі часового ряду – адитивної або мультиплікативної. Вибір форми моделі залежить від структури сезонних коливань. Якщо амплітуда коливань майже постійна, то розробляють адитивну модель часового ряду, в якій значення сезонної компоненти вважаються постійними для різних циклів. Якщо амплітуда сезонних коливань зростає або зменшується, то розробляють мультиплікативну модель часового ряду. У адитивній моделі характеристики сезонності вимірюються в абсолютних величинах, в мультиплікативній – у відносних.

Алгоритм розрахунку для адитивної сезонності складається з чотирьох етапів [5, с. 84 – 87]. На першому етапі передбачається опис тенденції за допомогою ковзної середньої при парній довжині інтервалів згладжування. Ковзна середня для часових рядів місячної динаміки визначається за формулою:

$$y_t^{'} = \frac{\frac{1}{2}y_{t-6} + y_{t-5} + \ldots + y_t + \ldots + y_{t+5} + \frac{1}{2}y_{t+6}}{12};$$

для рядів квартальної динаміки:

$$y_t^{'} = \frac{\frac{1}{2}y_{t-2} + y_{t-1} + y_t + y_{t+1} + \frac{1}{2}y_{t+2}}{4}.$$

На другому етапі розраховується відхилення фактичних значень від рівнів згладженого ряду: $x_t = y_t - y_t^{'}$. Рівні знову отриманого ряду відображають ефект сезонності та випадковості.

На третьому етапі для елімінування впливу випадкових факторів визначаються попередні значення сезонної складової як середні значення з рівнів x_t для одноіменних місяців (кварталів).

На четвертому етапі проводиться коригування початкових значень сезонної складової. Для адитивного випадку сума значень сезонної складової для повного сезонного циклу дорівнює нулю. Тому скориговані оцінки сезонної компоненти визначаються так:

$$S_i = \overline{x_i} - \overline{x},$$

де $\overline{x} = \frac{1}{m}\sum_{i=1}^{m}\overline{x_i}$, m – число фаз у сезонному циклі (для місячної динаміки $m = 12$, для квартальних даних $m = 4$).

В алгоритмі розрахунку для мультиплікативної сезонності змінюється другий та четвертий етапи, а перший і третій залишаються без зміни. На другому етапі розраховується відхилення фактичних значень від рівнів згладженого ряду як $x_t = \frac{y_t}{y_t^{'}}$. Взаємопогашення сезонних коливань в мультиплікативній формі виражається в тому, що середня арифметична значень коефіцієнтів сезонності для сезонного циклу дорівнює одиниці.

Остаточні оцінки коефіцієнтів сезонності будуть:

$$S_i = \overline{x_i}\,\overline{x}, \quad i = \overline{1,m},$$

де $\overline{x} = \dfrac{m}{\sum\limits_{i=1}^{m} \overline{x_i}}$, m – число фаз у сезонному циклі.

Процедуру побудови тренд-сезонних моделей рекомендують проводити в такій послідовності.

1. Оцінювання сезонної складової проводити за розглянутими алгоритмами з урахуванням характеру сезонності: адитивної або мультиплікативної.

2. Сезонне коригування вихідних даних.

3. Розрахунок параметрів тренду на основі часового ряду, отриманого на попередньому кроці.

4. Моделювання динаміки вихідного ряду з урахуванням трендової і сезонної складових.

5. Оцінка точності й адекватності отриманої моделі.

6. Застосування обчисленої моделі для прогнозування.

32.3. Приклади побудови адитивної і мультиплікативної моделей

Як було вже сказано, для обчислення адитивної і мультиплікативної моделей необхідно знайти значення T, S, E для кожного рівня ряду.

Процес побудови моделі складається з таких етапів.

1. Вирівнювання висхідного ряду методом ковзної середньої.

2. Обчислення значень сезонної компоненти S.

3. Вилучення сезонної компоненти з висхідних рівнів ряду і отримання вирівняних даних ($T+E$) в адитивній або ($T \cdot E$) мультиплікативній моделі.

4. Аналітичне вирівнювання рівнів ($T+E$) або ($T \cdot E$) і обчислення значень T з використанням отриманого рівняння тренду.

5. Обчислення отриманих за моделлю значень ($T+S$) або ($T \cdot S$).

6. Обчислення абсолютних і (або) відносних похибок.

Якщо отримані значення похибок не мають автокореляції, ними можна замінити висхідні рівні ряду і в подальшому використовувати часовий ряд похибок E для аналізу взаємозв'язку висхідного ряду й інших часових рядів.

Приклад. Відомо умовні дані про обсяги споживання електроенергії жителями регіону за 16 кварталів (табл. 32.1), а також встановлено, що даний часовий ряд має сезонні коливання періодичністю 4. Визначити доцільно побудови адитивної або мультиплікативної моделей та побудувати її.

Таблиця 32.1

Споживання електроенергії жителями регіону, млн. кВт/год

№ кварталу t	Споживання електроенергії y_t	Всього за чотири квартала	Ковзна середня за чотири квартали	Центрована ковзна середня	Оцінка сезонної компоненти
1	2	3	4	5	6
1	6,0	–	–	–	–
2	4,4	24,4	6,10	–	–
3	5,0	25,6	6,40	6,250	-1,250
4	9,0	26,0	6,50	6,450	2,550
5	7,2	27,0	6,75	6,625	0,575
6	4,8	28,0	7,00	6,875	-2,075
7	6,0	28,8	7,20	7,100	-1,100
8	10,0	29,6	7,40	7,300	2,700
9	8,0	30,0	7,50	7,450	0,550
10	5,6	31,0	7,75	7,625	-2,025
11	6,4	32,0	8,00	7,875	-1,475
12	11,0	33,0	8,25	8,125	2,875
13	9,0	33,6	8,40	8,325	0,675
14	6,6	33,4	8,33	8,375	-1,775
15	7,0				
16	10,8				

Розв'язання. За висхідними даними побудуємо графік споживання електроенергії жителями регіону за 16 кварталів (рис. 32.1).

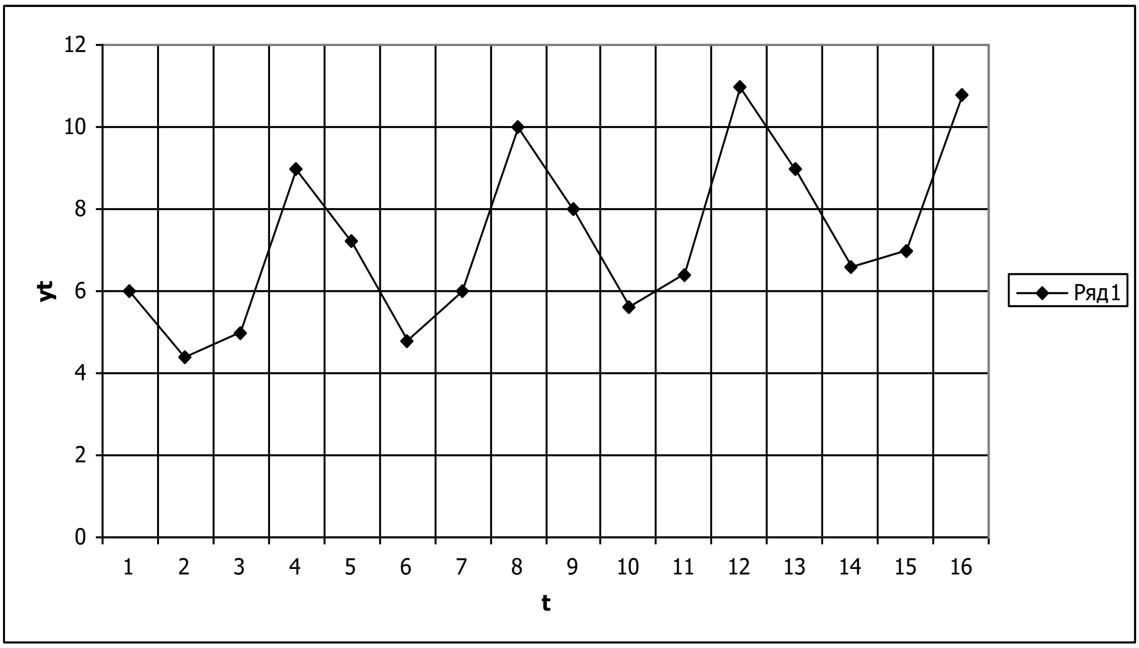


Рис. 32.1. **Споживання електроенергії жителями регіону**

На рис. 32.1 спостерігаються сезонні коливання, причому амплітуди коливань майже рівні, тому доцільно розробити адитивну модель.

Спочатку вирівняємо висхідні рівні ряду методом ковзної середньої. Для цього послідовно знайдемо суми рівнів ряду за кожні чотири квартали зі зсувом в один термін і визначимо умовні річні обсяги споживання електроенергії (табл. 32.1). Розділимо кожне отримане значення на 4. Маємо ковзні середні за чотири квартали, що не містять сезонні компоненти. Далі знайдемо середні значення з двох послідовних ковзних середніх, що є центрованими ковзними середніми, які використаємо для обчислення оцінки сезонної компоненти (різниця між фактичними рівнями ряду та центрованими ковзними середніми). На основі оцінок обчислимо значення сезонної компоненти S (табл. 32.2), знайшовши середні за кожний квартал оцінки сезонної компоненти. При цьому вважається, що сезонні впливи протягом періоду взаємно гасяться. В адитивних моделях сума значень сезонної компоненти за всіма кварталами повинна дорівнювати нулю.

Таблиця 32.2

Обчислення значень сезонної компоненти в адитивній моделі

Показники	Рік	Квартали, i			
		I	II	III	IV
	1	–	–	-1,250	2,550
	2	0,575	-2,075	-1,100	2,700

	3	0,550	-2,025	-1,475	2,875
	4	0,675	-1,775	–	–
Всього		1,800	-5,875	-,825	8,125
Середні оцінки сезонної компоненти $\overline{S}_i$		0,600	-1,958	-1,275	2,708
Зкорегована сезонна компонента S_i		0,581	-1,977	-1,294	2,690

Отже, $\sum_{i=1}^{4}\overline{S}_i = 0{,}6 - 1{,}958 - 1{,}275 + 2{,}708 = 0{,}075$. Обчислимо коригувальний коефіцієнт: $k = \frac{0{,}075}{4} = 0{,}01875$ та зкориговані значення сезонної компоненти як різниця між середньою оцінкою і коригувальним коефіцієнтом $S_i = \overline{S}_i - k$, при цьому перевіримо умову рівності нулю суми значень сезонної компоненти: 0,581−1,977−1,294+2,690=0.

Таким чином, маємо такі значення сезонної компоненти: I квартал $S_1 = 0{,}581$, II квартал $S_2 = -1{,}979$, III квартал $S_3 = -1{,}294$, IV квартал $S_4 = 2{,}690$ (табл. 32.2).

Далі елімінуємо вплив сезонної компоненти, віднявши її значення від кожного рівня висхідного часового ряду: $T + E = Y - S$ (табл. 32.3), тобто маємо тенденцію і випадкову величину.

Таблиця 32.3

Обчислення вирівняних значень T і похибок E в адитивній моделі

t	y_t	S_i	$T + E = Y - S$	T	$T + S$	$E = y_t - (T + S)$	E^2
1	6,0	0,581	5,419	5,902	6,483	-0,483	0,2333
2	4,4	-1,977	6,337	6,088	4,111	0,289	0,0835
3	5,0	-1,294	6,294	6,275	4,981	0,019	0,0004
4	9,0	2,690	6,310	6,461	9,151	-0,151	0,0228
5	7,2	0,581	6,619	6,648	7,229	-0,029	0,0008
6	4,8	-1,977	6,777	6,834	4,857	-0,057	0,0032

7	6,0	-1,294	7,294	7,020	5,727	0,273	0,0745
8	10,0	2,690	7,310	7,207	9,896	0,104	0,0108
9	8,0	0,581	7,419	7,393	7,974	0,026	0,0007
10	5,6	-1,977	7,577	7,580	5,603	-0,030	0,0009
11	6,4	-1,294	7,694	7,766	6,472	-0,072	0,0052
12	11,0	2,690	8,310	7,952	10,642	0,358	0,1382
13	9,0	0,581	8,419	8,139	8,720	0,280	0,0784
14	6,6	-1,977	8,577	8,325	6,348	0,252	0,0635
15	7,0	-1,294	8,294	8,519	7,218	-0,218	0,0475
16	10,8	2,690	8,110	8,698	11,388	-0,588	0,3457

Для визначення компоненти T в адитивній моделі, вирівняємо ряд $T+E$ за допомогою лінійного тренду: $T = 5{,}7065 + 0{,}187176t$, при цьому стандартна похибка коефіцієнта регресії дорівнює 0,0150373, $R^2 = 0{,}91713$, число степенів свободи 14. Обчислимо значення T за моделлю $T = 5{,}7065 + 0{,}187176t$, підставивши t. Графік рівняння тренда представлений на рис. 32.2.

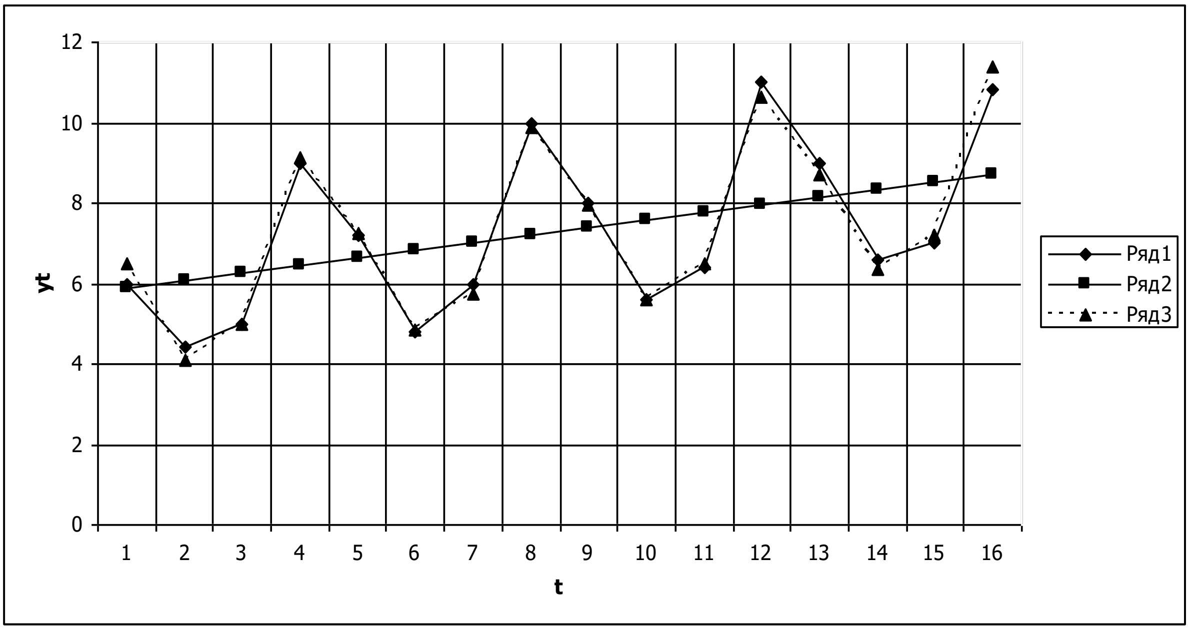


Рис. 32.2. **Споживання електроенергії жителями регіону, де ряд 1 – фактичні значення рівнів ряду, ряд 2 – вирівняні за трендом, ряд 3 – отримані за адитивною моделлю**

Для знаходження значень рівнів ряду за адаптивною моделлю прибавимо до рівнів T значення сезонної компоненти для відповідних кварталів, графічно це зображено на рис. 32.2. Абсолютна похибка обчислень адитивної моделі обчислюється як $E = Y - (T + S)$ (табл. 32.3), $\sum_{i=1}^{16} E^2 = 1{,}10$. Маємо, що $\left(1 - \frac{1{,}10}{71{,}59}\right) \cdot 100 = 98{,}46$, де 71,59 це сума квадратів відхилень рівнів ряду від його середнього рівня. Отже, адитивна модель пояснює на 98,46 % загальної варіації рівнів часового ряду споживання електроенергії за останні 16 кварталів.

Приклад. Щоквартальні дані про прибуток компанії за останні чотири роки представлені в табл. 32.4. Перевірити можливість побудови мультиплікативної моделі даного часового ряду та побудувати її.

Таблиця 32.4

Прибуток компанії, ум. од.

квартал / рік	I	II	III	IV
1	72	100	90	64
2	70	92	80	58
3	62	80	68	48
4	52	60	50	30

За даними, що в табл. 32.4 побудуємо графік часового ряду (рис. 32.3).

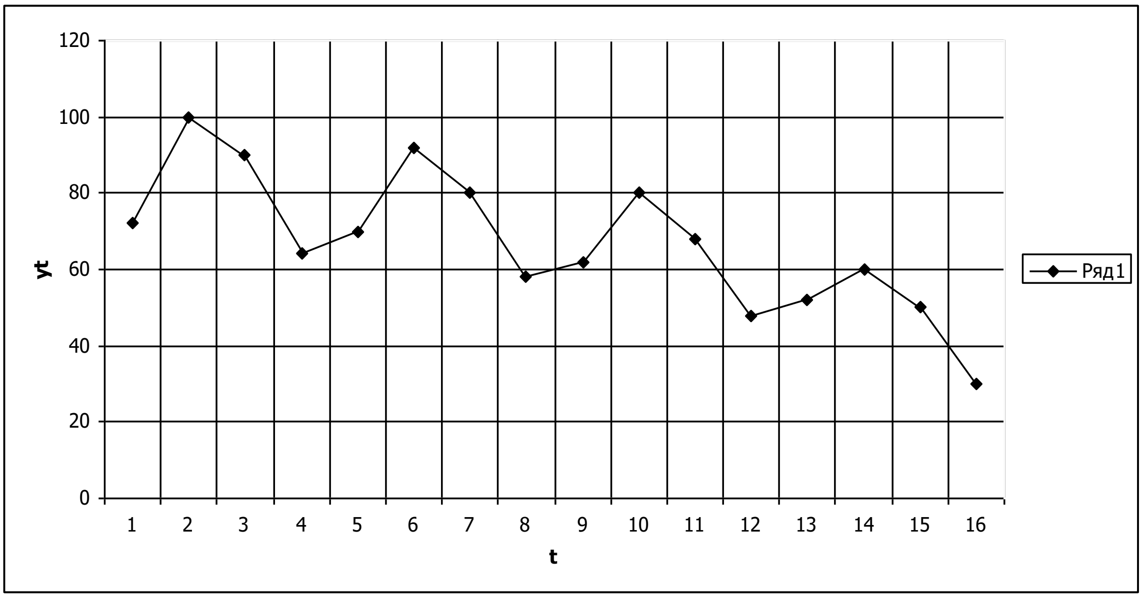


Рис. 32.3. **Прибуток компанії**

Графік часового ряду прибутку компанії демонструє наявність сезонних коливань та загальної спадної тенденції рівнів ряду, при цьому період коливань дорівнює 4. Отже робимо висновок про можливість побудови мультиплікативної моделі. Для визначення компонент даної моделі виконаємо вирівнювання висхідних рівнів ряду методом ковзної середньої. Методика, що застосовується на цьому етапі повністю співпадає з методикою адитивної моделі. В табл. 32.5 представлені результати обчислень оцінок сезонної компоненти. Оцінки сезонної компоненти знаходяться діленням фактичних рівнів ряду на центровані ковзні середні. Далі ці оцінки використовують для обчислення значень сезонної компоненти S.

Таблиця 32.5

Обчислення оцінок сезонної компоненти в мультиплікативній моделі

№ кварталу t	Прибуток компанії y_t	Всього за чотири квартала	Ковзна середня за чотири квартали	Центрована ковзна середня	Оцінка сезонної компоненти
1	2	3	4	5	6
1	72	–	–	–	–
2	100	326	81,5	–	–
3	90	324	81,0	81,25	1,108
4	64	316	79,0	80,00	0,800
5	70	306	76,5	77,75	0,900
6	92	300	75,0	75,75	1,215
7	80	292	73,0	74,00	1,081
8	58	280	70,0	71,50	0,811
9	62	268	67,0	68,50	0,905
10	80	258	64,5	65,75	1,217
11	68	248	62,0	63,25	1,075
12	48	228	57,0	59,50	0,807
13	52	210	52,5	54,75	0,950
14	60	192	48,0	50,25	1,194
15	50				
16	30				

Закінчення табл. 32.5

Для обчислення значень сезонної компоненти S спочатку знайдемо середні за кожний квартал S_i (табл. 32.6). В мультиплікативній моделі взаємне погашення сезонних впливів виражається в тому, що сума значень сезонної компоненти за всіма кварталами повинна рівнятись числу періодів у циклі, в прикладі це дорівнює 4.

Таблиця 32.6

Обчислення значень сезонної компоненти в мультиплікативній моделі

Показники	Рік	Квартали, i			
		I	II	III	IV
	1	–	–	1,108	0,800
	2	0,900	1,215	1,081	0,817

	3	0,905	1,217	1,075	0,807
	4	0,950	1,194	–	–
Усього		2,755	3,626	3,264	2,424
Середні оцінки сезонної компоненти $\overline{S_i}$		0,918	1,209	1,088	0,808
Зкорегована сезонна компонента S_i		0,913	1,202	1,082	0,803

Отже, $\sum_{i=1}^{4}\overline{S_i} = 0{,}918 + 1{,}209 + 1{,}088 + 0{,}808 = 4{,}023.$ Обчислимо коригувальний коефіцієнт: $k = \frac{4}{4{,}023} = 0{,}9943$ та зкориговані значення сезонної компоненти як добуток середньої оцінки і коригувального коефіцієнта $S_i = \overline{S_i} \cdot k$, при цьому перевіримо умову рівності чотирьом суми значень сезонної компоненти: 0,913 + +1,202 + 1,082 + 0,803 = 4.

Таким чином, маємо такі значення сезонної компоненти: I квартал $S_1 = 0{,}913$, II квартал $S_2 = 1{,}202$, III квартал $S_3 = 1{,}082$, IV квартал $S_4 = 0{,}803$ (табл. 32.7).

Розділивши кожний рівень висхідного ряду на відповідні значення сезонної компоненти, отримаємо $T \cdot E = Y : S$ (табл. 32.7), що враховують тільки тенденцію і випадкову компоненту.

Таблиця 32.7

Обчислення вирівняних значень T і похибок E в мультиплікативній моделі

t	y_t	S_i	$T \cdot E = Y / S_i$	T	$T \cdot S$	$E = y_t : (T \cdot S)$	$E = y_t - (T \cdot S)$	E^2
1	72	0,913	78,86	87,80	80,16	0,898	-8,16	66,66
2	100	1,202	83,19	85,03	102,20	0,978	-2,20	4,86
3	90	1,082	83,18	82,25	89,00	1,011	1,00	1,00
4	64	0,803	79,70	79,48	63,82	1,003	0,18	0,03
5	70	0,913	76,67	76,70	70,03	1,000	-0,03	0,00
6	92	1,202	76,54	73,93	88,86	1,035	3,14	9,85
7	80	1,082	73,94	71,15	76,99	1,039	3,01	9,08
8	58	0,803	72,23	68,38	54,91	1,056	3,09	9,57
9	62	0,913	67,91	65,60	59,90	1,035	2,10	4,43
10	80	1,202	66,56	62,83	75,52	1,059	4,48	20,08

11	68	1,082	62,85	60,05	64,98	1,047	3,02	9,14
12	48	0,803	59,78	57,28	45,99	1,044	2,01	4,03
13	52	0,913	56,96	54,50	49,76	1,045	2,24	5,02
14	60	1,202	49,92	51,73	62,18	0,965	-2,18	4,73
15	50	1,082	46,21	48,95	52,97	0,944	-2,97	8,79
16	30	0,803	37,36	46,18	37,08	0,809	-7,08	50,12

Для визначення компоненти T в мультиплікативній моделі, вирівняємо ряд $T \cdot E$ за допомогою лінійного тренду: $T = 90{,}59 - 2{,}773t$, при цьому стандартна похибка коефіцієнта регресії дорівнює 0,225556, $R^2 = 0{,}915239$, число степенів свободи 14. Обчислимо значення T за моделлю $T = 90{,}59 - 2{,}773t$, підставивши t. Графік рівняння тренду представлений на рис. 32.4.

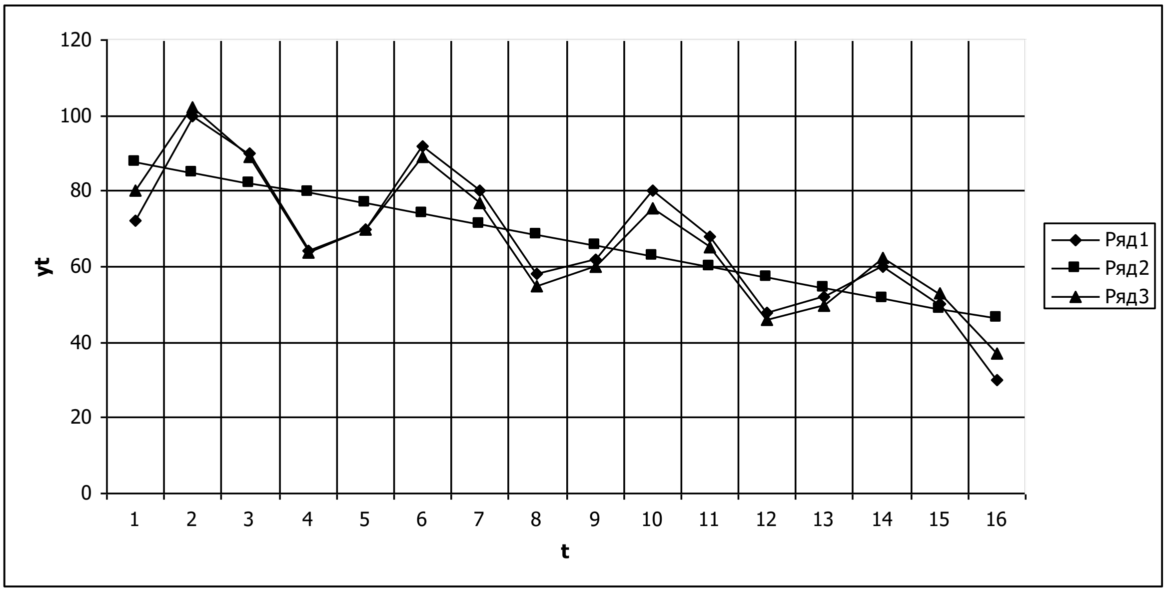


Рис. 32.4. **Прибуток компанії, де ряд 1 – фактичні значення рівнів ряду, ряд 2 – вирівняні за трендом, ряд 3 – отримані за мультиплікативною моделлю**

Для знаходження значень рівнів ряду за мультиплікативною моделлю помножимо рівні T на значення сезонної компоненти для відповідних кварталів, графічно це зображено на рис. 32.4. Похибка обчислень мультиплікативної моделі обчислюється як $E = y_t : (T \cdot S)$ (табл. 32.7). Якщо даний часовий ряд похибок не містить автокореляції, то його можна використовувати замість висхідного ряду у вивченні взаємозв'язку з іншими часовими рядами.

Для порівняння обчисленої мультиплікативної моделі з іншими моделями часового ряду визначається сума квадратів абсолютних похибок, де абсолютні похибки: $E = y_t - (T \cdot S)$, тобто $\sum_{i=1}^{16} E^2 = 207{,}40$. Маємо, що:

$$\left(1 - \frac{207{,}40}{5\,023}\right)100 = 95{,}87,$$

де 5 023 – сума квадратів відхилень рівнів ряду від його середнього рівня. Отже, мультиплікативна модель пояснює на 95,87 % загальної варіації рівнів часового ряду прибутку компанії за останні 16 кварталів.

Запитання для самоперевірки

1. Перерахуйте підходи в моделюванні сезонних і циклічних коливань.

2. У чому полягає відмінність кожного з підходів у моделюванні сезонних і циклічних коливань.

3. Від чого залежить процедура розрахунку сезонної складової?

4. Приведіть етапи алгоритму розрахунку для адитивної сезонності.

5. У чому полягає відмінність алгоритму розрахунку для мультиплікативної сезонності?

6. Розкрийте зміст процедури побудови тренд-сезонних моделей.

7. Розкрийте зміст процедури побудови адитивної і мультиплікативної моделей.

Використана література

1. Айвазян С. А. Прикладная статистика и основы эконометрики : учебник для вузов / С. А. Айвазян, В. С. Мхитарян. – М. : ЮНИТИ, 1998. – 1022 с.

2. Бородич С. А. Эконометрика : учебн. пособ. / С. А. Бородич. – Мн. : Новое знание, 2001. – 408 с.

3. Вильямс Дж. Совершенный стратег или Букварь по теории стратегических игр / Дж. Вильямс; пер. с англ. – М. : Советское радио, 1960. – 270 с.

4. Доугерти К. Введение в эконометрику / К. Доугерти; пер. с англ. – М. : ИНФРА-М, 1999. – 402 с.

5. Дуброва Т. А. Статистические методы прогнозирования : учебн. пособ. для вузов / Т. А. Дубова. – М.: ЮНИТИ-ДАНА, 2003. – 206 с.

6. Єгоршин О. О. Лабораторний практикум з навчальної дисципліни «Економіко-математичні методи та моделі: економетрика» : навч.-практ. посібн. / О. О. Єгоршин, Л. М. Малярець. – Х. : Вид. ХНЕУ, 2011. – 148 с.

7. Єгоршин О. О. Математичне програмування : підручник для студ. вищ. навч. закл. / О. О. Єгоршин, Л. М. Малярець. – Х. : ВД «ІНЖЕК», 2006. – 384 с.

8. Єгоршин О. О. Тексти лекцій з курсу "Математичне програмування" курсу "Математика для економістів." Ч. ІІІ / О. О. Єгоршин, Л. М. Малярець. – Х. : Вид. ХДЕУ, 2001. – 88 с.

9. Егоршин А. А. Корреляционно-регрессионный анализ : учебн. пособ. для студ. экон. спец. вузов / А. А. Егоршин, Л. М. Малярец. – Х. : Основа, 1998. – 208 с.

10. Егоршин А. А. Математическое программирование : учебн. пособ. / А. А. Егоршин, Л. М. Малярец. – Х. : ИД «ИНЖЭК», 2003. – 240 с.

11. Елисеева И. И. Практикум по эконометрике : учебн. пособ. / И. И. Елисеева, С. В. Курышева, Н. М. Гордеенко; под ред. И. И. Елисеевой. – 2-е изд., перераб. и доп. – М. : Финансы и статистика, 2006. – 344 с.

12. Кузнецов Ю. Н. Математическое программирование: учебн. пособ. / Ю. Н. Кузнецов, В. И. Кузубов, А. Б. Волощенко. – 2-е изд., перераб. и доп. – М. : Высшая школа, 1980. – 300 с.

13. Лабораторний практикум з навчальної дисципліни «Економіко-математичне моделювання» : навч.-практ. посібн. / Л. М. Малярець, П. М. Куликов, І. Л. Лебедєва та ін. – Х. : Вид. ХНЕУ, 2009. – 136 с.

14. Лугін О. Є. Економетрія : навч. посібн. / О. Є. Лугін, С. В. Білоусова, О. М. Білоусов. – К. : Центр навчальної літератури, 2005. – 252 с.

15. Лук'яненко І. Г. Економетрика : підручник / І. Г. Лук'яненко, Л. І. Краснікова. – К. : Тов. «Знання», КОО, 1998. – 494 с.

16. Магнус Я. Р. Економетрика : начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. – М. : Дело, 1997. – 248 с.

17. Маленво Э. Статистические методы эконометрии. Вып. 1 / Э. Маленво. – М. : Статистика, 1975. – 424 с.

18. Маленво Э. Статистические методы эконометрии. Вып. 2 / Э. Маленво. – М. : Статистика, 1975. – 325 с.

19. Малярець Л. М. Вимірювання ознак обєктів в економіці: методологія та практика. Наукове видання. – Х. : Вид. ХНЕУ, 2006. – 384 с.

20. Малярец Л. М. Экономико-математические методы и модели : учебн. пособ. для иностранных студентов / Л.М. Малярец. – Х. : Изд. ХНЭУ, 2013. – 288 с.

21. Математичні методи в сучасних економічних дослідженнях : монографія / заг. ред. докт. екон. наук, проф. Л. М. Малярець. – Х. : Вид. ХНЕУ, 2011. – 272 с.

22. Наконечний С. І. Економетрія : підручник / С. І. Наконечний, Т. О. Терещенко, Т. П. Романюк. – 3-тє вид., доп. та перероб. – К. : КНЕУ, 2005. – 520 с.

23. Наконечний С. І. Математичне програмування : навч. посібн. / С. І. Наконечний, С. С. Савіна. – К. : КНЕУ, 2005. – 452 с.

24. Пономаренко В. С. Багатовимірний аналіз соціально-економічних систем : навч. посібн. / В. С. Пономаренко, Л. М. Малярець. – Х. : Вид. ХНЕУ, 2009. – 384 с.

25. Практикум по эконометрике : учебн. пособ. / И. И. Елисеева, С. В. Курышева, Н. М. Гордеенко и др.; под ред. И. И. Елисеевой. – 2-е изд., перераб. и доп. – М. : Финансы и статистика, 2007. – 344 с.

26. Просветов Г. И. Эконометрика. Задачи и решения : учебн.-метод. пособ. / Г. И. Просветов. – М. : Издательство РДЛ, 2004. – 104 с.

27. Таха Хемди А. Введение в исследование операций / Таха Хемди А.; пер. с англ. – 7-е изд. – М.: Издательский дом «Вильямс», 2005. – 912 с.

28. Христиановский В. В. Прикладная эконометрия : учебник для экон. вузов / В. В. Христиановский, Н. Г. Гузь, О. Г. Кривенчуг. – Донецк : ДонГУ, 1998. – 172 с.

29. Ферстер Э. Методы корреляционного и регрессионного анализа: руководство для экономистов / Э. Ферстер, Б. Ренц; пер. с нем. и предисл. В. М. Ивановой. – М. : Финансы и статистика, 1983. – 302 с.

30. Фролькис В. А. Введение в теорию и методы оптимизации для экономистов / В. А. Фролькис. – 2-е изд. – СПб. : Питер, 2002. – 320 с.

31. Эзекиэл М. Методы корреляции и регрессии / М. Эзекиэл, К. А. Фокс. – М. : Статистика, 1966. – 324 с.

32. Эконометрика : учебник / под ред. И. И. Елисеевой. – М. : Финансы и статистика, 2004. – 344 с.

33. Экономико-математические методы и модели: учебн. пособ. / под общ. ред. А. В. Кузнецова. – Мн. : БГЭУ, 1999. – 416 с.

34. Экономико-математические методы и прикладные модели : учебн. пособ. для вузов / под ред. В. В. Федосеева. – М. : ЮНИТИ, 1999. – 392 с.

35. Экономико-математический словарь: словарь современной экономической науки / авт.-состав. Л. И. Лопатников. – М. : Изд. «ABF», 1996. – 704 с.

36. Экономико-математический энциклопедический словарь / [гл. ред. В. И. Данилов-Данильян]. – М.: Большая Российская энциклопедия; изд. дом «ИНФРА-М», 2003. – 688 с.

38. Энциклопедия. Математическая энциклопедия: Гл. ред. И.М. Виноградов, т. 3 Коо – Од – М. : «Советская Энциклопедия», 1982. – 1184 стб., ил.

Зміст

НАВЧАЛЬНЕ ВИДАННЯ

Малярець Людмила Михайлівна

ЕКОНОМІКО-МАТЕМАТИЧНІ МЕТОДИ ТА МОДЕЛІ

Навчальний посібник

Відповідальний за випуск **Малярець Л. М.**

Відповідальний редактор **Сєдова Л. М.**

Редактор **Лященко О. Г.**

Коректор

План 2014 р. Поз. №

Підп. до друку Формат 60×90 1/16. Папір MultiCopy. Друк Riso.

Ум.-друк. арк. Обл.-вид. арк. Тираж прим. Зам. №

Видавець і виготівник – видавництво ХНЕУ, 61001, м. Харків, пр. Леніна, 9а

Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи

Дк № 481 від 13.06.2001 р.