

МІЖРЕГІОНАЛЬНА
АКАДЕМІЯ УПРАВЛІННЯ ПЕРСОНАЛОМ



МАУП

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

ПРАКТИКУМ З ЕКОНОМЕТРІЇ

За редакцією О. О. Юнькової

Рекомендовано

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

Київ

ДП «Видавничий дім «Персонал»
2009

ББК 65в6я73
П69

Рецензенти: *А. А. Кадієвський*, д-р екон. наук, проф.
А. В. Бессалов, д-р техн. наук, проф.
С. С. Карташова, канд. фіз.-мат. наук, доц.

*Схвалено Вченою радою Міжрегіональної Академії
управління персоналом (протокол № 3 від 29.03.06)*

*Рекомендовано Міністерством освіти і науки України
(лист № 14/18.2-1014 від 21.04.06)*

Практикум з економетрії: навч. посіб. / О. Л. Лещинський,
П69 В. В. Рязанцева, О. О. Юнькова, І. І. Юртин; за ред. О. О. Юнь-
кової. — К. : ДП «Вид. дім «Персонал», 2009. — 256 с. — Біб-
ліогр.: с. 251–252.

ISBN 978-966-608-841-6

У посібнику розглянуто основні задачі економетричних досліджень і методи їх розв'язання. Основним методом оцінювання параметрів регресійних рівнянь обрано метод найменших квадратів (МНК). Описано особливості оцінювання параметрів у разі порушення основних умов застосування класичного МНК і альтернативні методи оцінювання параметрів моделей (метод головних компонентів, узагальнений МНК тощо). Динаміку економічних процесів розглянуто на прикладах моделей часових рядів. Для моделювання економічних процесів з прямими та оберненими зв'язками використано системи одночасних рівнянь.

Посібник містить розділи з вищої математики, алгебри, теорії ймовірностей і математичної статистики, необхідні для засвоєння курсу економетрії.

Призначений для студентів економічних спеціальностей.

ББК 65в6я73

ISBN 978-966-608-841-6

© О. Л. Лещинський, В. В. Рязанцева,
О. О. Юнькова, І. І. Юртин, 2009
© Міжрегіональна Академія управління
персоналом (МАУП), 2009
© ДП «Видавничий дім «Персонал», 2009

ВСТУП

Характерною рисою сучасної економічної науки є широке застосування математичних методів, що разом з глибоким економічним аналізом дає змогу ґрунтовно досліджувати процеси та явища, які відбуваються в економіці, передбачати і прогнозувати їхній розвиток.

Особливе місце серед економічних дисциплін посідає *економетрія* (в інших джерелах “економетрика”) – порівняно новий напрям економічної науки, що утворився поєднанням трьох галузей знань:

- теоретичної економіки, яка дає якісне описання економічних явищ чи процесів;
- математики, яка забезпечує кількісне описання залежностей у вигляді математичних моделей – співвідношень між економічними показниками;
- математичної статистики, яка визначає методи досліджень, насамперед методи кореляційно-регресійного аналізу.

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

Теоретична економіка, виявляючи зв'язок між окремими економічними показниками, визначає ті, що найщільніше пов'язані між собою, і вказує лише напрям зв'язку між ними. Наприклад, у законі попиту чи законі пропозиції формулюється залежність обсягу попиту і обсягу пропозиції довільного нормального товару від його ціни. Однак для прийняття управлінських рішень потрібно мати кількісну оцінку залежності, наприклад відповідний коефіцієнт еластичності. Отримати ті чи інші конкретні кількісні оцінки можливо, лише базуючись на певному досвіді, який відображається у статистичних даних.

На підставі цих даних потрібно підтвердити чи спростувати припущення про зв'язок між показниками і кількісно його оцінити. У свою чергу, для застосування кількісних методів потрібно мати математичну модель, яка формалізує цей зв'язок, тобто подає його у вигляді певних математичних співвідношень — рівнянь чи систем рівнянь, що відображають залежність між економічними показниками абстрактною мовою.

При побудові моделі керуються припущенням про випадковий характер багатьох економічних показників: здебільшого їх можна кількісно виміряти, але точні значення неможливо передбачити заздалегідь (через впливи багатьох неконтрольованих чинників).

Зв'язки, що існують між економічними показниками, можуть бути цілком визначеними — детермінованими в разі, коли один із показників є функцією від іншого. Однак строга функціональна залежність реалізується в економіці рідко. Частіше спостерігається так звана статистична, зокрема кореляційна, залежність.

Кореляційна залежність проявляється в тому, що зі змінюванням однієї випадкової величини змінюється середнє значення іншої.

Наприклад, витрати на споживання залежать насамперед від доходу споживача. Однак для окремих людей з однаковими доходами витрати є різними, отже, не існує функції, яка б дала змогу за відомим доходом точно визначити обсяг витрат. Але, як показує досвід, середні витрати зі збільшенням доходу зазвичай зростають, а зі зменшенням — спадають, і це дає підстави говорити про кореляційну залежність між ними, тобто про залежність у середньому.

Кореляційні зв'язки між економічними показниками (чи іншими випадковими величинами) описують за допомогою *регресійних* рівнянь.

Регресійні рівняння дають змогу вивчати змінювання певного економічного показника під впливом найголовніших факторів — інших економічних показників. Показник, що вивчається, називається результативною, залежною чи ендогенною змінною. Показники, що впливають на нього, — факторними, незалежними чи екзогенними змінними. Коефіцієнти, що поєднують змінні в рівняннях, є незмінними і називаються параметрами моделі.

Модель з довільними коефіцієнтами є економічною моделлю, тобто застосовною до будь-яких ситуацій, а економетрична модель “прив'язана” до певної ситуації через відповідну статистичну базу даних. Саме це визначає місце економетрії в системі економічних наук:

економетрія є тією ланкою, що поєднує загальну економічну теорію з практикою економічної діяльності.

Оскільки будь-яка економетрична модель відображає не лише теоретичний, якісний взаємозв'язок між показниками, а й містить емпіричну (статистичну) інформацію, то в ній, на відміну від економічної моделі, завжди присутні стохастичні залишки — розбіжності між заданими та модельними значеннями результативного показника. Саме ймовірнісні характеристики залишків моделі зумовлюють якість тієї чи іншої аналітичної форми моделі.

Основне завдання економетричного дослідження — *побудова та перевірка моделей*. Побудова моделі передбачає вибір форми залежності між показниками (специфікація моделі) та обчислення на підставі статистичних даних конкретних значень параметрів — їхніх оцінок (параметризація). Оскільки ці результати отримують за вибірковими даними, то перевірка моделі зводиться до перевірки статистичних гіпотез. При цьому перевіряється адекватність моделі загалом, а також значущість окремих її елементів (параметрів). Однак перевірка моделі буде неповною без аналізу залишків, оскільки якість моделі зумовлюють стохастичні характеристики залишків.

Побудована і перевірена економетрична модель може бути застосована не лише для прогнозування значень окремих показників, а й для аналізу зв'язку між ними та імітації певних економічних явищ і процесів.

Зауважимо, що в багатьох економетричних моделях є такі екзогенні змінні, що можуть бути змінені керівними органами (державним регулюванням чи керівництвом фірми). Ці керовані змінні, наприклад державні витрати та податки, є політичними інструментами. Якщо відома структура економічного процесу, то державні органи, змінюючи значення таких змінних, могли б робити заданими ендогенні змінні, тобто впливати на подальший розвиток процесу. І це зумовлює ще один із напрямів застосування економетричних моделей.

Математичні моделі, які застосовують в економетрії, поділяють на певні типи, адже кожний тип є визначальним при виборі дослідницької процедури і конкретного методу дослідження.

За рівнем пізнавальної цінності економетричні моделі поділяють на причинно-наслідкові, симптоматичні, авторегресійні та моделі тенденції розвитку.

Причинно-наслідковими називаються моделі, в яких між результативними й факторними змінними існує причинно-наслідкова за-

лежність. У цих моделях досліджувана змінна є наслідком, а факторні змінні — причиною її змін. Симптоматичні моделі характеризуються тим, що змінні моделі неможливо інтерпретувати в категоріях причин і наслідків. Пояснювальними змінними в таких моделях є ті, що найсильніше корелюють з результативною змінною. В авторегресійних моделях у ролі факторних змінних виступають значення результативної змінної, зміщені за часом, тобто визначені в попередні часові періоди чи моменти часу. Моделі тенденції розвитку описують розвиток явища в часі. У моделях такого типу результативна змінна є функцією лише однієї факторної змінної — змінної часу, яка набуває значень з натурального ряду чисел — послідовних номерів досліджуваних інтервалів чи моментів часу.

Помилки, що трапляються в економетричних дослідженнях, поділяють на три типи.

1. Помилки вимірювань. Найчастіше вони виникають на макрорівні, коли показники формуються як сума індивідуальних значень. Деякі із показників (наприклад, доходи фізичних чи юридичних осіб) визначаються на підставі даних, які подають самі особи, і можуть не відповідати дійсності (зокрема, приховані доходи). Чим більше неперевірених даних потрапляє до інформаційної бази, тим більшими є помилки вимірювань. Усунути їх практично неможливо.

2. Помилки вибірки. Вони зумовлені неоднорідністю та випадковістю даних початкової вибіркової сукупності, яка формується для вивчення реальних економічних процесів. Для отримання задовільного результату із сукупності вилучають одиниці з аномальними значеннями досліджуваних ознак. Однак і в цьому разі результати досліджень залежать від вибірки. Для того щоб збільшити їхню вірогідність, потрібно збільшити обсяг спостережень.

3. Помилки специфікації. Від правильно підібраної специфікації (форми) моделі залежить величина випадкових похибок. Остання буде тим меншою, чим менші розбіжності між заданими і розрахованими значеннями результативної змінної. До помилок специфікації відносять не лише неправильний вибір математичної функції, що відображає зв'язок між показниками, а й відсутність у рівнянні суттєвого фактора, який впливає на результативну змінну. Помилки специфікації дослідник може і має звести до мінімуму, змінюючи форму рівняння (замінивши лінійний зв'язок нелінійним, парну регресію — множинною тощо).

При проведенні економетричних досліджень вважається, що помилки вимірювань і помилки вибірки мінімальні або взагалі відсутні і основна увага приділяється зменшенню помилок специфікації.

Детальний алгоритм повного дослідження регресійної моделі наведено у першому розділі практикуму.

Про застосування цього алгоритму до моделей парної регресії йтиметься у другому розділі.

Методи вибору факторних змінних і форми залежності між показниками для моделей множинної (багатофакторної) регресії наведено у третьому розділі.

Приклад повного дослідження множинної моделі розглядається в четвертому розділі.

Особливості дослідження моделей часових рядів розглянуто у п'ятому розділі, а методи оцінювання параметрів систем одночасних рівнянь — у шостому.

Математичний інструментарій, необхідний в економетричних дослідженнях, зосереджений у сьомому розділі. Крім того, цей розділ містить також перелік деяких убудованих функцій EXCEL, які можна застосувати у відповідних розрахунках.

ПОРЯДОК ПОБУДОВИ І ДОСЛІДЖЕННЯ РЕГРЕСІЙНОЇ МОДЕЛІ

Незалежно від рівня чи галузі застосування економетричних моделей можна виокремити кілька основних етапів їх побудови й дослідження.

1.1. Основні етапи економетричного дослідження

I етап. *Постановка задачі та попередня обробка статистичних даних:*

- 1) початковий аналіз економічного процесу, що вивчається, його теоретичний (якісний) опис;
- 2) визначення мети дослідження, для досягнення якої потрібно застосувати модель;
- 3) вибір найсуттєвіших факторів, що впливають на досягнення мети (специфікація змінних);
- 4) збирання і попередній аналіз статистичних даних — спостережень змінних, що увійшли до моделі.

II етап. *Побудова моделі:*

- 5) вибір математичної моделі з фіксованою формою усіх структурних рівнянь (специфікація моделі);
- 6) вибір методу оцінювання невідомих параметрів моделі з урахуванням припущень щодо ймовірнісних характеристик випадкових відхилень;
- 7) реалізація процедури оцінювання параметрів, яка забезпечує найкраще наближення модельних значень результативної змінної до її спостережених значень.

III етап. *Дослідження моделі:*

- 8) перевірка якості моделі загалом і її окремих параметрів;

- 9) аналіз залишків моделі;
 - 10) інтерпретація отриманих результатів і встановлення їх відповідності поставленим цілям;
 - 11) аналіз неузгодженості й коригування моделі;
 - 12) прийняття рішень щодо подальших досліджень з урахуванням альтернативних можливостей;
- IV етап. *Застосування моделі:*

- 13) застосування моделі згідно з поставленою метою її побудови:
 - аналіз коефіцієнтів регресії, обчислення коефіцієнтів еластичності;
 - імітація певних економічних ситуацій і визначення засобів керування ними;
 - прогнозування значень досліджуваних показників (точковий та інтервальний прогноз).

Розглянемо детальніше перелічені етапи.

1.2. Постановка задачі та статистична база даних

Потреби дослідження зв'язків між економічними показниками виникають на різних рівнях економічної діяльності, тому при побудові якісної моделі спираються на різні розділи економічної теорії.

Аналіз економічної діяльності окремого підприємства та прогнозування його подальшого розвитку потребують дослідження:

- 1) виробничій функції, яка визначає залежність між обсягом виробленої продукції та витраченими для цього ресурсами, наприклад основним капіталом та працею;
- 2) функції ціни, що дає змогу дослідити, як зміниться ціна товару, якщо зміниться обсяг поставок і ціни конкуруючих товарів;
- 3) функції попиту, що дає змогу встановити, як зміниться попит на продукцію, якщо змінюватимуться ціна товару, ціни товарів-конкурентів і доходи споживачів;
- 4) функції витрат, яка описує залежність середніх витрат на виробництві від ціни та кількості виробничих ресурсів;
- 5) функції чутливості ринку, яка визначає залежність обсягу збуту продукції від витрат на рекламу та індексу "чистоти" виробленої продукції ("екологічного індикатора");
- 6) рівняння стратегії підприємства, в якому відображається, як залежить рентабельність підприємства від питомої ваги на

ринку товарів, подібних до тих, що їх виробляє підприємство, а також від якості товарів, витрат на маркетинг та наукові дослідження, від інвестиційних витрат тощо.

Залежність між доходом та споживанням, ціною та попитом, процентною ставкою та інвестиціями, обмінним курсом валюти та обсягом чистого експорту — проблеми, які вивчають мікро- та макроекономіка.

Як зазначалося, економетричне дослідження завжди поєднує теорію (математичні моделі) та практику (статистичні дані). Моделі застосовують для опису та пояснення процесів, що вивчаються, а статистичні дані — для побудови та обґрунтування моделей.

Економетричні моделі розрізняють за рівнем агрегування змінних (мікро- чи макроекономічні показники), за способом відображення змінних (у постійних чи поточних цінах, в абсолютних значеннях чи приростах показників), за кількістю змінних (одно- чи багатофакторні моделі), за кількістю рівнянь (одне чи декілька), за часом спостережень (річні, квартальні чи місячні дані). Класифікують моделі також за призначенням і метою використання (аналітичні, імітаційні, прогностичні).

Метою збирання статистичних даних є побудова інформаційної бази для прийняття рішень. Природно, що аналіз даних і прийняття рішень проводяться на підставі деякої інтуїтивної (неявної) або кількісної (явної) економічної моделі. Тому збирають саме ті дані, що стосуються певної моделі. Дані можна отримати через опитування, анкетування, інтерв'ювання або із джерел офіційної статистичної звітності. Кожний показник, отриманий одним із цих способів, називають спостереженням.

Економічні дані зазвичай поділяють на два види: перехресні або просторові дані (*cross-section data*) та часові ряди (*time series*).

Перехресні дані — це дані про економічні показники, отримані для різних однотипних об'єктів (фірм, регіонів) одночасно або їхня часова належність є несуттєвою. Наприклад, дані бюджетних досліджень населення в певний момент часу, чисельність працівників на підприємстві тощо. Просторові дані зазвичай незалежні між собою й мають однаковий (частіше нормальний) закон розподілу. Такі дані застосовують при дослідженні структури економічних зв'язків.

Часові ряди, навпаки, характеризують той самий об'єкт, але в різні моменти часу. На підставі часових рядів вивчають динаміку економіч-

них процесів. Наприклад, динаміка рівня інфляції за певний період відображається часовими рядами.

Послідовні значення часових рядів можуть бути пов'язані між собою певними залежностями: спостерігаються деякі закономірності у відхиленнях від загальної тенденції розвитку чи виявляються затримки показників (часові лаги). Тому методи обробки таких даних дещо відрізняються від методів, що їх застосовують для обробки перехресних даних.

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

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

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

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

1.3. Специфікація моделі

Залежно від аналітичної форми опису моделі поділяють на лінійні та нелінійні. У лінійній моделі результативна змінна є лінійною функцією факторних змінних, параметрів і випадкової складової рів-

няння. Серед нелінійних виокремлюють моделі, нелінійні відносно факторних змінних, але лінійні відносно параметрів моделі, а також моделі, нелінійні відносно і факторних змінних, і параметрів. Багато нелінійних моделей, що застосовуються в економетрії, можна звести до лінійного вигляду. Тому саме лінійні моделі розглядаються найдокладніше. Економетричні методи, що застосовуються у складніших ситуаціях, є зазвичай модифікацією методів, розроблених для найпростіших ситуацій.

Способи відбору найсуттєвіших впливових факторів і вибір форми залежності між показниками детально розглядаються в розд. 3.

У загальному випадку

$$y = f(x_1, x_2, \dots, x_m, a_0, a_1, \dots, a_m) + u,$$

або

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

де y – показник, що досліджується (результативна, залежна, ендогенна змінна); x_1, x_2, \dots, x_m – показники, що впливають на нього (факторні, незалежні, екзогенні змінні); $a_0, a_1, a_2, \dots, a_m$ – невідомі параметри моделі, які потрібно оцінити на підставі спостережень залежної змінної y та спостережень незалежних змінних x_1, x_2, \dots, x_m ; u – випадкова складова регресійного рівняння.

Однак частіше модель задають у лінійному вигляді

$$y = a_0 + a_1x_1 + \dots + a_mx_m + u, \quad (1.1)$$

Якщо відомо оцінки параметрів $\hat{a}_0, \hat{a}_1, \hat{a}_2, \dots, \hat{a}_m$ то модельні (розрахункові, теоретичні, аналітичні) значення результативного показника обчислюють за формулою

$$\hat{y}_i = \hat{a}_0 + \hat{a}_1x_{i1} + \hat{a}_2x_{i2} + \dots + \hat{a}_mx_{im}, \quad i = 1, 2, \dots, n.$$

чи за відповідною іншою формулою.

Розбіжності між модельними та відомими значеннями результативного показника $u_i = y_i - \hat{y}_i$ ($i = 1, 2, \dots, n$) називають залишками моделі.

Випадкова складова u є результативною дією всіх неконтрольованих випадкових факторів, що спричинюють відхилення реальних значень досліджуваного показника y від його аналітичних (обчислених на підставі обраної регресійної залежності) значень.

1.4. Оцінювання параметрів за МНК

Для оцінювання параметрів лінійних моделей зазвичай застосовують метод найменших квадратів (МНК). Ідея методу полягає в тому, щоб на підставі статистичних даних обчислити такі значення параметрів моделі, за яких розбіжності між заданими та модельними значеннями результативного показника (залишки моделі) якнайменші. Критерієм вибору параметрів у цьому методі є мінімізація суми квадратів залишків за всією сукупністю спостережень:

$$\sum_{i=1}^n u_i^2 \rightarrow \min.$$

Обчислення оцінок параметрів за цим методом зводиться до розв'язання системи нормальних рівнянь, усі коефіцієнти якої отримують на підставі статистичних даних.

Відомо n спостережень результативного показника — век-

тор $y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}$ і відповідно n спостережень факторних змінних — матрицю

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{pmatrix},$$

індекси елементів x_{ij} якої позначають відповідно номер спостереження ($i = 1, 2, \dots, n$) і номер факторної змінної ($j = 1, 2, \dots, m$).

Оператор оцінювання за МНК параметрів множинної моделі (1.1), тобто розрахункова формула методу, має вигляд

$$A = (X^T X)^{-1} (X^T y), \quad (1.2)$$

де A — вектор невідомих параметрів $A = \begin{pmatrix} a_0 \\ a_1 \\ \dots \\ a_m \end{pmatrix}$;

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1m} \\ 1 & x_{21} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nm} \end{pmatrix} \text{ — матриця спостережень незалежних змінних,}$$

доповнена стовпцем x_0 — вектором, що містить n одиниць;
 X^T — транспонована матриця X ,

$$X^T = \begin{pmatrix} 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & \dots & x_{n1} \\ \dots & \dots & \dots & \dots \\ x_{1m} & x_{2m} & \dots & x_{nm} \end{pmatrix};$$

$(X^T X)$, $(X^T Y)$ — добутки матриць; $(X^T X)^{-1}$ — обернена матриця.

1.5. Основні припущення методу найменших квадратів

Для застосування МНК необхідно виконання таких передумов:

1) кожне значення випадкової складової рівняння u , $i = 1, 2, \dots, n$ є випадковою величиною з математичним сподіванням, яке дорівнює нулю:

$$M(u) = \frac{1}{n} \sum_{i=1}^n u_i = 0;$$

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

3) змінні моделі не корельовані із залишками;

4) факторні (незалежні) змінні не корельовані між собою.

Крім того, для перевірки значущості моделі загалом і окремих її параметрів бажано, щоб залишки моделі були нормально розподіленими випадковими величинами з нульовим математичним сподіванням і сталою дисперсією. Однак навіть у разі порушення цієї вимоги МНК-оцінки параметрів моделі можуть мати задовільні властивості.

Для перевірки випадкового характеру залишків користуються насамперед графічним методом: будують графік залежності залишків моделі від модельних значень результативної ознаки — діаграму розсіювання.

Якщо точки діаграми розташовані приблизно рівномірно в деякій горизонтальній смужі, то залишки є випадковими величинами, отже,

застосування МНК виправдане. Якщо в розташуванні точок виявляється інша закономірність, то це може означати, що:

- а) залишки не випадкові;
- б) залишки не мають сталої дисперсії;
- в) залишки не мають систематичного характеру.

Виявлення систематичного впливу на залежну змінну, який не враховано в моделі, можна трактувати як помилку специфікації, але наявність вільного члена моделі часто дає змогу скоригувати модель у такий спосіб, щоб забезпечити виконання першої умови. Крім того, щоб досягти незалежності залишків і модельних значень результативної ознаки, потрібно збільшити обсяг спостережень або ввести до моделі додаткові фактори, або змінювати форму залежності між змінними, поки залишки не стануть випадковими.

Якість оцінок параметрів моделі значною мірою залежить також від співвідношення між залишками моделі й факторними змінними. Для виявлення залежностей між ними також можна будувати графіки — діаграми розсіювання залишків моделі залежно від усіх факторних змінних, які включені до моделі. Як і в попередньому випадку, якщо точки рівномірно розташовані в горизонтальній смужці, то залежності немає. Нерівномірність у розташуванні точок може свідчити про несталість дисперсії залишків чи неправильну специфікацію моделі; скупчення точок у певних місцях свідчить про наявність систематичної похибки моделі.

Друга умова означає, що залишки моделі є похибками вимірювання. Причиною цього може бути автокореляція чи гетероскедастичність. Якщо між компонентами вектора залишків існує кореляційна залежність, то таке явище називається автокореляцією. Наявність автокореляції в моделі свідчить про існування кореляції між послідовними значеннями деякої незалежної змінної або про неврахований суттєвий фактор, що впливає на залежну змінну і не може бути усунутий за рахунок вільного члена моделі. Загальний вплив неврахованих у моделі змінних може виявитись також у тому, що дисперсія для окремих груп спостережень буде змінюватись, тобто виникає явище гетероскедастичності. У будь-якому разі порушення другої передумови впливає на методи оцінювання параметрів моделі.

Існування залежності між залишками та змінними моделі часто пов'язане з тим, що в моделі присутні лагові (затримані в часі) змінні або модель складається з одночасних структурних рівнянь. Для

оцінювання параметрів і в цьому разі потрібно застосовувати інші методи.

На якість оцінок, отриманих за МНК, може впливати залежність між незалежними змінними. Якщо між незалежними змінними моделі існують щільні лінійні зв'язки, то це явище називають мультиколінеарністю. Моделі, в яких спостерігається мультиколінеарність, стають надзвичайно чутливими до конкретного набору даних, до специфікації моделі, мають значні відхилення від дійсних значень параметрів узагальненої моделі.

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

При виконанні перелічених передумов оцінки параметрів регресійного рівняння, отримані за допомогою МНК, є незміщеними, обґрунтованими, ефективними та інваріантними.

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

Зокрема, якщо порушується друга передумова МНК (за наявності автокореляції чи гетероскедастичності), то отримані за цим методом оцінки втрачають властивість ефективності, хоча залишаються незміщеними та обґрунтованими.

При порушенні третьої передумови, яке найчастіше трапляється в системах рівнянь або в лагових моделях (моделях з часовим зрушенням показників), МНК-оцінки виявляються зміщеними і необґрунтованими.

Якщо порушується четверта передумова, тобто між змінними існують мультиколінеарні зв'язки, то це призводить до зміщення МНК-оцінок. Застосування моделей, що мають зміщені, необґрунтовані чи неефективні оцінки, втрачає сенс.

1.6. Показники якості регресійної моделі

У класичному регресійному аналізі вважається, що функція регресії відома до оцінювання параметрів, тобто регресійна модель специфікована вірно. Однак в емпіричних економічних і соціальних дослідженнях не завжди відомо, скільки факторів має бути введено до моделі і яка залежність між ними існує. Щоб забезпечити адекватне відтворення досліджуваного явища чи процесу, необхідно вибирати регресійну функцію серед багатьох варіантів. А для цього існують спеціальні критерії якості моделі.

Одним із критеріїв якості побудованої моделі є похибка апроксимації – абсолютне значення різниць між відповідними заданими і розрахованими значеннями результативної змінної. Однак порівнювати між собою такі різниці неможливо, тому для всіх спостережень визначають відносні похибки апроксимації A_i , які вимірюються у процентах. Такі похибки можна порівнювати між собою і визначити точність наближення для кожного спостереження. Для оцінки моделі загалом обчислюють середнє значення відносних похибок $\bar{A} = \frac{1}{n} \sum_{i=1}^n A_i$,

де $A_i = \left| \frac{\hat{y}_i - y_i}{y_i} \right| \cdot 100\%$, $i = 1, 2, \dots, n$.

Моделю вважається вдало підбраною, якщо середня похибка апроксимації не перевищує 8–10%. При порівнянні різних моделей, побудованих на одній статистичній базі, перевагу надають моделям з меншою похибкою \bar{A} .

Для перевірки якості моделі застосовують також узагальнені показники:

- 1) стандартну похибку рівняння;
- 2) коефіцієнт детермінації;
- 3) коефіцієнт множинної кореляції;
- 4) стандартну похибку параметрів.

Стандартна похибка рівняння характеризує абсолютну величину розкиду випадкової складової рівняння відносно лінії регресії і обчислюється за формулою $S_u = \sqrt{S_u^2}$, де $S_u^2 = \frac{1}{n} \sum_{i=1}^n u_i^2$ – точкова оцінка емпіричної (вибіркової) дисперсії залишків $u_i = \hat{y}_i - y_i$, тобто розбіжностей між модельними \hat{y}_i та заданими y_i значеннями результатив-

ного показника, $i = 1, 2, \dots, n$, n – кількість спостережень, за якими обчислюють оцінки параметрів.

Поправка степенів вільності (на кількість обчислених параметрів $m + 1$) дає незміщену оцінку дисперсії залишків:

$$\hat{\sigma}_u^2 = \frac{1}{n - m - 1} \sum_{i=1}^n u_i^2. \quad (1.3)$$

Зрозуміло, що перевагу надають моделям, в яких стандартна похибка рівняння менша порівняно з іншими моделями. Але така оцінка якості має суттєвий недолік: через те, що для неї не визначена верхня межа, порівняння різних моделей за цим критерієм досить проблематичне.

Коефіцієнт детермінації показує, яка частина руху залежної змінної описується даним регресійним рівнянням, тобто яку частину змінування показника відтворює дана модель:

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

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

Загальну суму квадратів відхилень значень результативної змінної від її середнього значення можна розкласти на суму квадратів відхилень, які пояснює регресія $\sum_{i=1}^n (\hat{y}_i - \bar{y})^2$, і суму квадратів відхилень, які не пояснює регресія $\sum_{i=1}^n (y_i - \hat{y}_i)^2$, тобто

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

отже, коефіцієнт детермінації визначається як відношення дисперсій:

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

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

Кількість факторів, які враховано в моделі, впливає на значення коефіцієнта детермінації: уведення до моделі кожної нової змінної збільшує значення коефіцієнта детермінації. Тому щоб запобігти невиправданому розширенню моделі й мати змогу порівнювати моделі з різною кількістю факторів, вводять спеціальний оцінений коефіцієнт детермінації, який має вигляд

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

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

Неважко помітити, що обидва коефіцієнти пов'язані залежністю:

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

Обчислений у такий спосіб коефіцієнт детермінації називається скоригованим за Тейлом і позначається \bar{R}_T^2 . Інше коригування, запропоноване іншим автором – Аемією, виконується за формулою

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

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

Обидва коефіцієнти \bar{R}_T^2 і \bar{R}_A^2 ураховують той факт, що введення до моделі кожного нового фактора зменшує число степенів вільності – різницю між кількістю даних і кількістю отриманих на їх підставі результатів, яке дорівнює $n - (m + 1)$. А для застосування статистичних критеріїв перевірки якості отриманих результатів бажано мати якомога більшу кількість степенів вільності.

Очевидно, $\bar{R}^2 \leq R^2$, тобто зі збільшенням кількості факторів моделі оцінений коефіцієнт детермінації зростає повільніше, ніж R^2 . Крім того, коли $R^2 = 1$, то і $\bar{R}^2 = 1$. Але коли R^2 прямує до нуля, оцінений коефіцієнт стає від'ємним. Така властивість скоригованого коефіцієнта детермінації дає змогу об'єктивніше оцінювати якість моделей з різною кількістю факторів, причому при застосуванні коефіцієнта \bar{R}_A^2 (скоригованого за Аемією) перевага однозначно надається рівнянню з меншою кількістю факторів.

Коефіцієнт множинної кореляції R (\bar{R}) визначає міру зв'язку залежної змінної з усіма незалежними факторами: $R = \sqrt{R^2}$ ($\bar{R} = \sqrt{\bar{R}^2}$). За аналогією з коефіцієнтом кореляції парної регресії він змінюється в межах від -1 до $+1$, але в разі множинної моделі не застосовується для інтерпретації напряму зв'язку.

Якщо R близький до нуля, то це означає, що рівняння регресії погано описує зв'язок між змінними або фактори мало впливають на результат.

Для лінійної множинної регресії коефіцієнт кореляції обчислюють за формулою

$$R = \sqrt{\sum_{i=1}^n \beta_i r_i},$$

де β_i — параметри регресії, обчислені для стандартизованих даних (див. розд. 7); r_i — парні коефіцієнти кореляції залежної змінної з i -ю незалежною змінною X_i . Цей самий коефіцієнт можна обчислити за допомогою визначників матриць парних коефіцієнтів міжфакторної кореляції (див. розд. 3).

Стандартна похибка рівняння, коефіцієнт детермінації та множинної кореляції є характеристиками, за якими перевіряється правильність вибору форми моделі та впливових незалежних змінних моделі. При порівнянні регресійних рівнянь, що описують даний процес чи явище, вирішальними критеріями є стандартна похибка рівняння (найменша) та коефіцієнт детермінації і множинної кореляції (якого ближчі до 1).

1.7. Перевірка надійності регресійного рівняння

Оцінки параметрів і показники якості моделі, обчислені на підставі статистичних даних, потребують перевірки на статистичну значущість (вірогідність). Тому перевірка регресійного рівняння зводиться до перевірки статистичних параметричних гіпотез. Адекватність моделі (значущість коефіцієнта детермінації) перевіряється за критерієм Фішера, значущість коефіцієнта кореляції та окремих параметрів моделі — за критерієм Стьюдента.

Для перевірки статистичної значущості коефіцієнта детермінації R^2 висувається нульова гіпотеза $H_0 : R_2 = 0$. Це означає, що досліджуване рівняння не пояснює змінювання результативного показника під

впливом відповідних факторів, отже, всі коефіцієнти при незалежних змінних мають дорівнювати нулю. Нульова гіпотеза в разі лінійної за параметрами моделі набуває вигляду

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

Альтернативною до неї є гіпотеза $H_1 : R^2 > 0$.

Це означає, що модель відображає певний вплив факторів на результативний показник.

Оскільки коефіцієнт детермінації визначається як співвідношення дисперсій, то для перевірки основної гіпотези застосовують F -критерій Фішера з m та $n - m - 1$ степенями вільності. За отриманими в моделі значеннями коефіцієнта детермінації R^2 обчислюють експериментальне значення F -статистики:

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

яке необхідно порівняти з табличним значенням розподілу Фішера $F_{\text{табл}} = F(\alpha, m, n - m - 1)$, де α – заданий рівень значущості (часто $\alpha = 1\%$ або $\alpha = 5\%$), n – кількість спостережень, m – кількість факторів, і зробити статистичний висновок. Якщо $F_{\text{табл}} < F_{\text{експ}}$, то нульова гіпотеза відхиляється. Відхилення нуль-гіпотези свідчить про адекватність побудованої моделі. Якщо приймається H_0 , то модель вважається неадекватною.

Коефіцієнт кореляції, як вибіркова характеристика, перевіряється на значущість за допомогою t -критерію Стьюдента. При цьому фактичне значення t -статистики обчислюють за формулою

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

і порівнюють з табличним значенням t -розподілу з $n - m - 1$ степенями вільності та при заданому рівні значущості α для двосторонньої критичної області. Якщо абсолютна величина експериментального значення t -статистики перевищує табличне:

$$|t_{\text{експ}}| > t_{\text{табл}},$$

то роблять висновок про достовірність (значущість) коефіцієнта кореляції, отже, зв'язок між залежною змінною та всіма незалежними факторами є суттєвим.

Крім загальних показників адекватності моделі існують також оцінки, що дають змогу встановити якість окремих частин рівняння, зокрема одного чи кількох коефіцієнтів регресії.

Статистичну значущість кожного параметра моделі перевіряють за допомогою t -критерію Стьюдента. Нульова гіпотеза для кожного параметра a_j ($j = 0, 1, \dots, m$) при цьому має вигляд

$$H_0 : a_j = 0, j = 0, 1, m,$$

тобто маємо $m + 1$ гіпотез

$$a_0 = 0, a_1 = 0, \dots, a_m = 0.$$

Щоб прийняти або відхилити ці гіпотези, обчислюють $m + 1$ експериментальне значення для кожного параметра моделі:

$$t_{0 \text{ експ}} = \frac{a_0}{\sqrt{\sigma_u^2 c_{00}}} = \frac{a_0}{S_{a_0}}, \quad t_{1 \text{ експ}} = \frac{a_1}{\sqrt{\sigma_u^2 c_{11}}} = \frac{a_1}{S_{a_1}}, \dots, \quad t_{m \text{ експ}} = \frac{a_m}{\sqrt{\sigma_u^2 c_{mm}}} = \frac{a_m}{S_{a_m}},$$

де $c_{00}, c_{11}, \dots, c_{mm}$ — діагональні елементи оберненої матриці $(X^T X)^{-1}$.

Кожне експериментальне значення t_j критерію Стьюдента порівнюється з табличним значенням $t_{\text{табл}} = t(\alpha/2, n - m - 1)$. Якщо значення t_j -статистики потрапляє до критичної області (більше за абсолютним значенням від $t_{\text{табл}}$), то приймають альтернативну гіпотезу про значущість відповідного параметра. Інакше роблять висновок про статистичну незначущість параметра a_j , а це означає, що відповідна незалежна змінна суттєво не впливає на змінювання результативного показника і, можливо, може бути виключена з моделі.

Зауваження. На практиці частіше застосовують грубішу оцінку значущості параметра, а саме припускають, що стандартні похибки мають становити 45–50 % значення параметра, аби можна було стверджувати про його статистичну значущість.

Довірчі інтервали для кожного окремого параметра a_j лінійної моделі обчислюють через стандартну похибку параметра (середньоквадратичне відхилення) і табличне значення розподілу Стьюдента:

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

Експериментальні значення t_j -статистик застосовують також для розрахунку часткових коефіцієнтів детермінації ΔR_j^2 , які визначають граничний внесок j -го фактора в загальний коефіцієнт детермінації. Величина ΔR_j^2 показує, на яку величину зменшиться коефіцієнт детермінації R^2 , якщо j -й фактор (і лише він) буде вилучений з групи пояснювальних змінних. Частковий коефіцієнт детермінації визначають за формулою

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

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

Перевірка значущості отриманих результатів не вичерпується переліченими діями. Перед подальшим застосуванням моделі потрібно пересвідчитись, чи мають обчислені оцінки параметрів необхідні властивості — незміщеність, ефективність та обґрунтованість.

Параметри лінійної моделі мають відтворювати вплив кожного фактора на результативний показник, а саме показувати, наскільки зміниться значення ендогенної змінної, якщо деяка факторна змінна (і лише вона) зміниться на одиницю. Наявність мультиколінеарності (кореляційної залежності між факторними змінними) призводить до зміщення оцінок параметрів, отже, не дає змоги виявити вплив кожного фактора окремо і тим самим правильно проаналізувати ситуацію, відтворену в моделі.

Для тестування мультиколінеарності застосовують різні методи, однак найповніше її досліджують за алгоритмом Феррара–Глобера, який дає змогу виявити мультиколінеарність усього масиву незалежних змінних, залежність кожної незалежної змінної від сукупності інших незалежних змінних, а також попарну залежність між факторними змінними (п. 7.7.1, 7.7.2).

Якщо виявлено мультиколінеарність, то для її усунення застосовують один із способів:

- виводять із моделі “зайву” змінну, тобто ту, що має найбільшу кількість кореляційних зв’язків з іншими незалежними змінними;
- переходять до іншого способу вимірювання показників (замість абсолютних значень беруть відносні або відхилення від середнього тощо);
- збільшують обсяг спостережень чи використовують додаткову інформацію;
- оцінюють параметри іншим методом, зокрема методом головних компонентів.

Наявність автокореляції чи гетероскедастичності залишків вказує на їх не випадковий характер, а це, у свою чергу, свідчить про неправильну специфікацію моделі (не враховано важливі фактори) чи про

невідповідну статистичну базу (невиправдано великий розкид даних) і кореляційну залежність між рівнями даних. Неefективність оцінок параметрів, що виникає внаслідок автокореляції чи гетероскедастичності залишків, проявляється у великих стандартних похибках параметрів, що призводить до неправильних висновків при перевірці статистичних гіпотез і до неточних прогнозів за такою моделлю.

Аналіз залишків на автокореляцію та гетероскедастичність потрібен для коригування даних, моделі та методу оцінювання параметрів.

Для виявлення автокореляції та гетероскедастичності залишків застосовують спеціальні тести. Тестування гетероскедастичності виконується за допомогою: μ -критерію, параметричного і непараметричного тестів Гольдфельда–Квандта, тесту Глейсера та ін. (п. 7.7.3). Для тестування автокореляції застосовують: тест Дарбіна–Уотсона, тест фон Неймана, циклічний і нециклічний коефіцієнти кореляції (п. 7.7.6).

За позитивного результату, отриманого за деяким із тестів, відповідне явище (автокореляцію чи гетероскедастичність) намагаються усунути.

Зокрема, якщо виявлено гетероскедастичність, а при цьому:

- дисперсії залишків відомі (що взагалі рідкість), то параметри регресійної моделі обчислюють за МНК;
- дисперсії залишків невідомі, то необхідно трансформувати початкову модель (п. 7.7.4);
- відомий вигляд залежності між дисперсією залишків та однією із незалежних змінних x_j , хоча самі дисперсії невідомі, то параметри регресійної моделі обчислюють за узагальненим МНК (п. 7.7.5).

Якщо причиною гетероскедастичності залишків є неправильна специфікація моделі, то введення до моделі неврахованого фактора також може сприяти усуненню цього явища.

Автокореляцію залишків можна усунути, змінивши специфікацію моделі (увівши до моделі додатковий фактор) чи перетворивши початкові дані. Ефeктивність оцінок параметрів значно зросте, якщо їх оцінити іншим методом (узагальненим МНК, методом Кочрена–Оркатта, методом Дарбіна тощо).

1.8. Прогнозування за багатofакторною регресійною моделлю

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

Прогнозними значеннями факторних змінних називають такі, що не потрапили до вибірки спостережень (у разі просторових даних) чи перебувають за межами періоду спостережень (у разі часових рядів). Прогнозом результативного показника є таке його значення, що визначається прогнозними значеннями факторних змінних.

Прогнозні значення \hat{y}_p обчислюють за формулою регресії (1.1), коли відомі оцінки параметрів $\hat{a} = (\hat{a}_0, \hat{a}_1, \hat{a}_2, \dots, \hat{a}_m)$ і задано вектор прогнозних значень незалежних змінних $x_p = (1, x_{1p}, x_{2p}, \dots, x_{mp})$.

Єдине значення результативного показника y_p , обчислене за цією формулою, називають точковим прогнозом. При цьому розрізняють середній прогноз (оцінку математичного сподівання результативної змінної) та індивідуальний прогноз (оцінку певної реалізації цієї змінної y_p). Дисперсію прогнозу обчислюють відповідно за формулами

$$\hat{\sigma}_e^2 = \hat{\sigma}_u^2 x_p^T (X^T X)^{-1} x_p \quad (\text{для математичного сподівання})$$

і

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

Зрозуміло, що в багатьох випадках реальне значення результативного показника y_t не збігатиметься зі значенням його математичного сподівання, але якщо розглядати велику кількість вибірок, на підставі яких визначатиметься прогноз, то можна гарантувати, що приблизно $(1 - \alpha) \cdot 100\%$ результатів потраплять в інтервал (окіл математичного сподівання), межі якого визначають за формулами

$$(\hat{y}_p - t_{\text{табл}} \sqrt{\hat{\sigma}_e^2}; \quad \hat{y}_p + t_{\text{табл}} \sqrt{\hat{\sigma}_e^2})$$

і

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

де $t_{\text{табл}}$ — табличне значення критерію Стьюдента з $n - m - 1$ степенями вільності та із заданим рівнем значущості $\alpha/2$ (для двосторонньої критичної області). Ці проміжки визначають інтервальний прогноз результативного показника, а значення $\sqrt{\hat{\sigma}_e^2}$ і $\sqrt{\hat{\sigma}_{e(i)}^2}$ — похибки прогнозу.

Якість прогнозу буде тим кращою, чим повніше виконуються вимоги до статистичних даних, надійніше оцінено параметри моделі, точніше визначено прогнозні значення факторних змінних.

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

Висновки

1. Економетричне дослідження поєднує економічну теорію (математичну модель) і економічну практику (статистичні дані).
2. Математична модель, що відображає кореляційний зв'язок між економічними показниками, є регресійним рівнянням.
3. Параметри регресійного рівняння оцінюють за допомогою методу найменших квадратів.
4. Для оцінювання якості моделі обчислюють стандартну похибку рівняння, коефіцієнт детермінації та множинної кореляції.
5. Верифікація моделі полягає у статистичній перевірці моделі загалом, значущості коефіцієнта кореляції та окремих параметрів моделі.
6. Перевірка адекватності моделі виконується за допомогою критерію Фішера, значущість коефіцієнта кореляції та окремих параметрів моделі перевіряється за допомогою критерію Стьюдента.
7. Виконання передумов застосування МНК забезпечує якісні характеристики оцінок параметрів (незміщеність, ефективність, обґрунтованість).
8. За допомогою регресійної моделі можна визначати два типу прогнозу: точковий та інтервальний. Перший із них — це найвірогідніше значення результативної змінної, обчислене за моделлю для прогнозованих значень факторних змінних; другий — це проміжок, до якого з високою ймовірністю (близькою до одиниці) можуть потрапити реальні значення показника, що прогнозується.

ПАРНА ЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ

Однією з найпростіших економетричних моделей є модель парної регресії. Рівняння парної регресії характеризує зв'язок між двома змінними, що виявляється як певна закономірність лише в середньому для всієї сукупності спостережених даних, а для кожного окремого спостереження може не виконуватись. Хоча в чистому вигляді залежність лише двох показників виявляється рідко, використання цієї моделі дає змогу зрозуміти сутність багатьох процесів і дослідити їх. Зазвичай цю модель застосовують тоді, коли з усіх факторів, які впливають на результат, можна виокремити найвпливовіший. Саме його обирають за факторну змінну.

2.1. Вибір форми залежності

Нехай відомо n спостережень двох економічних показників X і Y :

$$(x_1, x_2, \dots, x_n) \qquad (y_1, y_2, \dots, y_n),$$

де X є факторною (незалежною) змінною, а Y – результативною (залежною) змінною.

Для вибору форми залежності між двома показниками будують діаграму розсіювання (кореляційне поле). Для цього кожне спостереження (x_i, y_i) ($i = 1, 2, \dots, n$) відображають точкою на числовій площині xu . Розміщення точок на графіку може бути “підказкою” в разі вибору форми залежності між змінними. Найкращою вважають функцію, графік якої проходить якомога ближче до точок кореляційного поля і відображає основну тенденцію в їхньому розміщенні.

Якщо точки кореляційного поля гуртуються навколо деякої прямої лінії, залежність між змінними можна описати лінійним рівнянням

$$Y = a_0 + a_1X + u,$$

де a_0, a_1 — параметри регресії, які відображають її структуру і які потрібно оцінити; u — випадкова складова, яка не може бути оцінена моделлю, що розглядається.

Якщо прямої лінії не існує, то відповідну форму залежності потрібно шукати серед нелінійних функцій. Зокрема, це можуть бути такі функції:

- гіперболічна

$$Y = a_0 + a_1 \frac{1}{X};$$

- степенева

$$Y = a_0 X^{a_1};$$

- логарифмічна

$$Y = a_0 + a_1 \log X;$$

- поліноміальна

$$Y = a_0 + a_1 X + a_2 X^2 + \dots + a_k X^k.$$

Саме ці функції найчастіше відображають залежності, що існують між економічними показниками.

Нескладними перетвореннями більшість з них можна звести до лінійного відносно структурних параметрів вигляду. Наприклад, після логарифмування степеневі функції отримуємо

$$\log Y = \log a_0 + a_1 \log X$$

або

$$V = b_0 + b_1 Z,$$

де $V = \log Y$; $b_0 = \log a_0$; $b_1 = a_1$; $Z = \log X$.

Заміна змінних в інших функціях дає змогу отримати лінійну не лише за структурними параметрами моделі, а й за змінними функцію:

$$Y = a_0 + a_1 Z,$$

де $Z = \frac{1}{X}$ для гіперболічної функції чи $Z = \log X$ для логарифмічної.

Якщо покласти $Z_1 = X$; $Z_2 = X^2, \dots, Z_k = X^k$, дістанемо множину регресію $Y = a_0 + a_1 Z_1 + a_2 Z_2 + \dots + a_k Z_k$, лінійну і за параметрами, і за змінними.

Пропонований підхід можна застосувати для побудови моделей множинної регресії. У цьому разі будують n графіків, що відобража-

ють залежність Y від кожної незалежної змінної, визначають аналітичну форму залежностей, причому кожного разу з'ясовують, чи є ці функції лінійними відносно структурних параметрів моделі, і в разі потреби перетворюють їх до лінійної форми.

Зауваження. Деякі моделі є нелінійними за структурними параметрами, які неможливо перетворити до лінійного вигляду, тому оцінювати їх параметри потрібно спеціальними методами.

У пропонованому практикумі вважатимемо, що, яким би не був істинний зв'язок між змінними, його вдалося звести до лінійного вигляду. Отже, постає завдання оцінювання параметрів лінійних моделей.

2.2. Оцінювання параметрів

Нехай рівняння регресії записано у вигляді

$$Y = a_0 + a_1 X + u.$$

Для парної лінійної моделі

$$y = a_0 + a_1 x$$

за відомих оцінок параметрів \hat{a}_0 і \hat{a}_1 розрахункові значення обчислюють за формулою

$$\hat{y}_i = \hat{a}_0 + \hat{a}_1 x_i,$$

де x_i — задані значення незалежної (факторної) змінної.

Значення x_i і y_i відомі й незмінні, тобто для даної функції вони є константами, а параметри a_0 і a_1 невідомі й вважаються змінними.

Згідно з методом найменших квадратів потрібно дібрати такі значення параметрів моделі, за яких сума квадратів відхилень між заданими y_i і розрахунковими \hat{y}_i значеннями результативної змінної буде найменшою, тобто

$$S = \sum (\hat{y}_i - y_i)^2 \rightarrow \min.$$

Функція S залежить від двох змінних: $S = S(a_0, a_1)$. Для визначення екстремуму функції S обидві її частинні похідні прирівнюють до нуля:

$$\frac{\partial S}{\partial a_0} = -2 \sum_{i=1}^n (a_0 + a_1 x_i - y_i) = 0;$$

$$\frac{\partial S}{\partial a_1} = -2 \sum_{i=1}^n (a_0 + a_1 x_i - y_i) x_i = 0;$$

у результаті отримують систему нормальних рівнянь:

$$a_0 n + a_1 \sum x_i = \sum y_i;$$

$$a_0 \sum x_i + a_1 \sum x_i^2 = \sum x_i y_i.$$

Якщо кожне з них поділити на n – кількість спостережень, то система набуде вигляду

$$\begin{cases} a_0 + a_1 \bar{x} = \bar{y}; \\ a_0 \bar{x} + a_1 \bar{x}^2 = \overline{xy}. \end{cases}$$

і матиме розв'язок

$$\begin{cases} \hat{a}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{x^2 - (\bar{x})^2} = r_{xy} \cdot \frac{S_y}{S_x}; \\ \hat{a}_0 = \bar{y} - \hat{a}_1 \bar{x}, \end{cases}$$

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

середні арифметичні значення; $S_x^2 = \bar{x}^2 - (\bar{x})^2$, $S_y^2 = \bar{y}^2 - (\bar{y})^2$ – вибіркові дисперсії.

Параметр a_0 – це значення результату при нульовому значенні фактора. Якщо ознака-фактор не має чи не може мати нульового значення, відповідно параметр a_0 не матиме економічного змісту. Лише знак цього параметра можна інтерпретувати: додатне його значення вказує, що результат змінюється повільніше, ніж фактор.

Рівняння регресії завжди доповнюють показником щільності зв'язку. Для лінійної моделі таким показником є лінійний коефіцієнт кореляції

$$r_{xy} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y},$$

де $\text{cov}(x, y)$ – коваріація випадкових величин X, Y ; σ_x, σ_y – їхні середньоквадратичні відхилення. Вибірковий коефіцієнт кореляції обчислюють за формулою

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{S_x S_y},$$

де S_x, S_y – вибіркові значення середньоквадратичних відхилень.

Знак лінійного коефіцієнта кореляції r_{xy} визначає знак коефіцієнта регресії a_1 , який, у свою чергу, визначає напрям впливу факторної змінної на результативну (прямий чи обернений). Якщо абсолютне значення коефіцієнта a_1 близьке до нуля, це означає, що між змінними не існує лінійного зв'язку (однак можливий будь-який інший).

2.3. Дослідження парної регресійної моделі

Після оцінювання параметрів a_0 , a_1 маємо дослідити побудовану модель. Для цього виконуємо такі кроки.

1. Для перевірки адекватності отриманої моделі обчислюємо:

- а) залишки моделі — розбіжності між спостереженими та розрахованими значеннями залежної змінної $u_i = y_i - \hat{y}_i$, $i = 1, 2, \dots, n$;
- б) залишкове математичне сподівання (середнє значення залишків)

$$\bar{u} = \frac{1}{n} \sum_{i=1}^n u_i \text{ і вибірккову залишкову дисперсію}$$

$$S_u^2 = \frac{1}{n} \sum (u_i - \bar{u})^2 = \sigma_u^2 \frac{1}{n-1} \sum_{i=1}^n (u_i - \bar{u})^2,$$

а також незміщену дисперсію

$$\sigma_u^2 = \frac{1}{n-1} \sum_{i=1}^n u_i^2;$$

в) відносну похибку апроксимації

$$A_i = \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\% \text{ або } A_i = \left| \frac{u_i}{y_i} \right| \cdot 100\%;$$

г) коефіцієнт детермінації

$$R^2 = 1 - \frac{S_u^2}{S_y^2} \text{ або } R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2};$$

д) вибіркковий коефіцієнт кореляції $R = \sqrt{R^2}$.

2. Перевіряємо статистичну значущість отриманих результатів:

- а) адекватність моделі загалом: за допомогою F -критерію Фішера перевіряємо гіпотезу

$$H_0 : a_1 = 0,$$

проти альтернативної

$$H_1 : a_1 \neq 0;$$

б) істотність (значущість) кожного коефіцієнта регресії: за допомогою t -критерію Стьюдента перевіряємо гіпотезу

$$H_0^j : a_j = 0 \text{ для } j = 0, 1,$$

тобто маємо дві гіпотези $H_0^0 : a_0 = 0$ та $H_0^1 : a_1 = 0$

проти відповідних альтернативних гіпотез

$$H_1^0 : a_0 \neq 0 \text{ та } H_1^1 : a_1 \neq 0.$$

3. Визначаємо довірчі зони регресії при рівні значущості α .

4. Будемо довірчі інтервали для параметрів регресії.

5. Обчислюємо прогностичні значення y_p за значеннями x_{1p}, x_{2p}, \dots , що лежать за межами базового періоду, і знаходимо межі довірчих інтервалів індивідуальних прогнозованих значень і межі довірчих інтервалів середнього прогнозу.

Числові характеристики стохастичних залишків моделі (математичне сподівання і дисперсія) є визначальними в дослідженні якості моделі. Математичне сподівання залишків (їх середнє значення) має дорівнювати нулю, дисперсія має бути якомога меншою.

Квадрат коефіцієнта кореляції, тобто коефіцієнт детермінації $R^2 = (r_{xy})^2$, відображає частку варіації результативної ознаки, що пояснюється регресійним рівнянням, у загальній варіації ознаки. Іншими словами, R^2 показує, яка частина змінювання результативної змінної описується зазначеним регресійним рівнянням. Величина $(1 - R^2)$ характеризує частку варіації, спричиненої впливом усіх не врахованих у моделі факторів (залишкову дисперсію). Якщо $R^2 \rightarrow 1$, це означає, що змінювання результату значною мірою пояснюється змінюванням факторної змінної й лише незначна частина змінювання — іншими факторами. Якщо $R^2 \rightarrow 0$, рівняння не відображає змінювання результату під впливом даного фактора, що, у свою чергу, означає, що факторна змінна не є впливовою, отже, модель не адекватна явищу.

Наскільки близьким до нуля є коефіцієнт детермінації R^2 , можна з'ясувати за допомогою критерію Фішера (F -критерію). Його зазвичай застосовують для порівняння дисперсій. У цьому разі потрібно встановити, чи суттєвими (значущими) є відмінності між залишковою дисперсією й дисперсією, яка пояснюється регресійним рівнянням. При цьому висувають нульову гіпотезу

$$H_0 : R^2 = 0$$

і альтернативну до неї

$$H_1 : R^2 \neq 0,$$

де $R^2 = \frac{S_{\hat{y}}^2}{S_y^2}$, $S_{\hat{y}}^2 = \frac{1}{n} \sum (\hat{y}_i - \bar{y})^2$ – дисперсія, що пояснюється рівнянням; $S_y^2 = \frac{1}{n} \sum (y_i - \bar{y})^2$ – загальна дисперсія результату.

Якщо перша дисперсія суттєво менша від другої, це рівнозначно тому, що рівняння не пояснює змінювання результату, тобто вплив факторної змінної несуттєвий. А це те саме, що параметр a_1 , який визначається через коефіцієнт кореляції, дорівнює нулю. Тому гіпотези формулюють ще так:

$$H_0 : a_1 = 0,$$

$$H_1 : a_1 \neq 0.$$

Експериментальне значення F -критерію обчислюють за формулою

$$F_{\text{експ}} = \frac{R^2}{1 - R^2} (n - 2) \quad \text{або} \quad F_{\text{експ}} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\frac{1}{n-2} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

і порівнюють з табличним значенням $F_{\text{табл}} = F(\alpha, k_1, k_2)$, де α – рівень значущості, $k_1 = 1$, $k_2 = n - 2$ – степені вільності. Величина $F_{\text{табл}}$ є межею критичних (малоймовірних) значень F -розподілу. Якщо $F_{\text{експ}} > F_{\text{табл}}$, тобто $F_{\text{експ}}$ потрапляє до області критичних значень, нульову гіпотезу відхиляють. Це означає, що відмінність між дисперсіями несуттєва, отже, модель пояснює змінювання результату впливом певного фактора.

У лінійній регресії потрібно оцінювати не лише значущість рівняння загалом, а й значущість окремих параметрів, а також коефіцієнта кореляції.

Для того щоб здійснити таке оцінювання, потрібно визначити стандартні похибки параметрів і похибку коефіцієнта кореляції. Для парної регресії ці значення відповідно такі:

$$S_{a_0} = \sqrt{\frac{\sigma_u^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}; \quad S_{a_1} = \sqrt{\frac{\sigma_u^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}; \quad S_R = \sqrt{\frac{1 - R^2}{n - 2}}.$$

Значущість параметрів визначають за допомогою критерію Стюдента (t -критерію). На підставі вибірових даних потрібно зробити висновок про суттєву відмінність від нуля кожного параметра моделі й коефіцієнта кореляції. Нульові гіпотези для параметрів мають вигляд

$$H_0^0 : a_0 = 0; \quad H_0^1 : a_1 = 0;$$

альтернативні

$$H_1^0 : a_0 \neq 0; \quad H_1^1 : a_1 \neq 0$$

і

$$H_0 : R = 0,$$

$$H_1 : R \neq 0$$

для коефіцієнта кореляції.

Експериментальне значення t -критерію $t_{\text{експ}}$ обчислюють як відношення значення параметра до його стандартної похибки:

$$t_{\text{експ}}^j = \frac{a_j}{S_{a_j}}, \quad j = 0, 1.$$

Тобто

$$t_{\text{експ}}^0 = \frac{a_0}{S_{a_0}}, \quad t_{\text{експ}}^1 = \frac{a_1}{S_{a_1}}.$$

Якщо значення параметра порівняно з похибкою досить мале, відмінність параметра від нуля несуттєва. Щоб зробити такий висновок, потрібно $t_{\text{експ}}^j$ порівняти з табличним значенням $t_{\text{табл}} = t(\alpha/2, n - 2)$, де $\alpha/2$ – рівень значущості для двосторонньої критичної області. Якщо $|t_{\text{експ}}^j| > t_{\text{табл}}$, нульову гіпотезу відхиляють, отже, значення відповідного параметра чи коефіцієнта кореляції можна вважати значущим.

Стандартні похибки S_{a_1} , S_{a_2} , S_R (середньоквадратичні відхилення) застосовують не лише для перевірки значущості параметрів, а й для розрахунку їхніх довірчих інтервалів:

$$\gamma_{a_0} = a_0 \pm \Delta_{a_0}; \quad \gamma_{a_1} = a_1 \pm \Delta_{a_1}, \quad \gamma_R = R \pm \Delta_R,$$

де $\Delta_{a_j} = t_{\text{табл}} S_{a_j}$, ($j = 0, 1$) $\Delta_R = t_{\text{табл}} S_R$ – граничні похибки відповідно параметрів регресії і коефіцієнта кореляції.

Зауваження. Оскільки коефіцієнт кореляції має певний економічний зміст, його довірчі інтервали не повинні містити суперечливих результатів, наприклад одночасно охоплювати додатні та від’ємні значення. Якщо все ж виявиться такий факт, це означатиме, що модель неправильно описує зв’язок між змінними, отже, потрібно шукати іншу форму залежності між ними.

2.4. Прогнозування за лінійною моделлю

Після побудови регресійного рівняння й перевірки його значущості модель можна застосувати для прогнозування.

Значення результативної змінної, обчислене для заданого прогнозного значення факторної змінної, визначає точковий прогноз результату. Однак через розсіювання показника потрібен інтервальный прогноз. Для визначення інтервального прогнозу потрібно обчислити його стандартну похибку. Для парної регресії ця похибка

$$S_{y_p} = \sigma_u \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}.$$

Така формула стандартної похибки прогнозованого середнього значення результату при заданому прогнозному значенні фактора визначає похибку розміщення лінії регресії. Похибка досягає мінімального значення при $x_p = \bar{x}$ і зростає з віддаленням від \bar{x} у будь-який бік. Чим більша різниця між x_p і \bar{x} , то більша похибка прогнозу. Найкращого результату прогнозування можна досягти, якщо x_p міститься у середині спостережених значень фактора. Якщо x_p перебуває за межами спостережень, то розмах довірчого інтервалу зростає.

Довірчий інтервал для прогнозного значення результативного показника обчислюють за формулою

$$\Upsilon_{y_p} = y_p \pm \Delta_{y_p},$$

де $\Delta_{y_p} = t_{\text{табл}} S_{y_p}$ – гранична похибка прогнозу.

У разі прогнозування на основі рівняння регресії слід пам’ятати, що величина прогнозу залежить не лише від стандартної похибки

індивідуального значення результативної змінної, а й від точності прогнозу факторної змінної. Це значення можна визначити за допомогою інших моделей залежно від конкретної ситуації, а також за результатами аналізу динаміки окремого фактора, тобто за допомогою моделей часових рядів.

2.5. Приклад побудови і дослідження моделі парної лінійної регресії

Розрахунки проведемо для такої статистичної вибірки:

x	1,00	1,15	1,20	1,30	1,35	1,40	1,55	1,60	1,70	1,90	1,95	2,00	2,05	2,15	2,20
y	1,44	1,50	2,06	0,90	1,70	1,51	1,20	1,10	1,40	1,80	1,90	2,20	2,40	3,10	2,80
x	2,40	2,45	2,60	2,80	2,85	3,00	3,15	3,20	3,25	3,30	3,50	3,55	3,70	3,80	4,00
y	2,50	3,50	2,70	2,90	3,00	3,40	2,80	3,30	3,60	4,00	3,40	3,70	4,20	4,50	3,90

обсягом $n = 30$. Виконаємо такі кроки.

1. Побудова регресійного рівняння $\hat{Y} = \hat{a}_0 + \hat{a}_1 X$.

2. Обчислення модельних значень $\hat{y}_i = \hat{a}_0 + \hat{a}_1 x_i$ та залишків $u_i = y_i - \hat{y}_i$.

3. Обчислення відносної похибки $\delta_i = \frac{u_i}{y_i}$ і середнього значення

відносної похибки $\bar{\delta} = \frac{1}{n} \sum \delta_i$.

Результати розрахунків відобразимо в таблиці.

x	y	x^2	y^2	xy	\hat{y}	u	u^2	u/y	δ
1	2	3	4	5	6	7	8	9	10
1,00	1,44	1,00	2,07	1,44	1,15	0,29	0,08	0,20	0,20
1,15	1,50	1,32	2,25	1,73	1,31	0,19	0,04	0,13	0,13
1,20	2,06	1,44	4,24	2,47	1,36	0,70	0,49	0,34	0,34
1,30	0,90	1,69	0,81	1,17	1,47	-0,57	0,32	-0,63	-0,63
1,35	1,70	1,82	2,89	2,30	1,52	0,18	0,03	0,11	0,11
1,40	1,51	1,96	2,28	2,11	1,57	-0,06	0,00	-0,04	-0,04
1,55	1,20	2,40	1,44	1,86	1,73	-0,53	0,28	-0,44	-0,44
1,60	1,10	2,56	1,21	1,76	1,78	-0,68	0,46	-0,62	-0,62

1	2	3	4	5	6	7	8	9	10
1,70	1,40	2,89	1,96	2,38	1,88	-0,48	0,23	-0,34	-0,34
1,90	1,80	3,61	3,24	3,42	2,09	-0,29	0,08	-0,16	-0,16
1,95	1,90	3,80	3,61	3,71	2,14	-0,24	0,06	-0,13	-0,13
2,00	2,20	4,00	4,84	4,40	2,19	0,01	0,00	0,00	0,00
2,05	2,40	4,20	5,76	4,92	2,25	0,15	0,02	0,06	0,06
2,15	3,10	4,62	9,61	6,67	2,35	0,75	0,56	0,24	0,24
2,20	2,80	4,84	7,84	6,16	2,40	0,40	0,16	0,14	0,14
2,40	2,50	5,76	6,25	6,00	2,61	-0,11	0,01	-0,04	-0,04
2,45	3,50	6,00	12,25	8,58	2,66	0,84	0,70	0,24	0,24
2,60	2,70	6,76	7,29	7,02	2,82	-0,12	0,01	-0,04	-0,04
2,80	2,90	7,84	8,41	8,12	3,03	-0,13	0,02	-0,04	-0,04
2,85	3,00	8,12	9,00	8,55	3,08	-0,08	0,01	-0,03	-0,03
3,00	3,40	9,00	11,56	10,20	3,24	0,16	0,03	0,05	0,05
3,15	2,80	9,92	7,84	8,82	3,39	-0,59	0,35	-0,21	-0,21
3,20	3,30	10,24	10,89	10,56	3,45	-0,15	0,02	-0,04	-0,04
3,25	3,60	10,56	12,96	11,70	3,50	0,10	0,01	0,03	0,03
3,30	4,00	10,89	16,00	13,20	3,55	0,45	0,20	0,11	0,11
3,50	3,40	12,25	11,56	11,90	3,76	-0,36	0,13	-0,11	-0,11
3,55	3,70	12,60	13,69	13,14	3,81	-0,11	0,01	-0,03	-0,03
3,70	4,20	13,69	17,64	15,54	3,97	0,23	0,05	0,06	0,06
3,80	4,50	14,44	20,25	17,10	4,07	0,43	0,18	0,10	0,10
4,00	3,90	16,00	15,21	15,60	4,28	-0,38	0,14	-0,10	-0,10
\bar{x}	\bar{y}	$\overline{x^2}$	$\overline{y^2}$	\overline{xy}					$\bar{\delta}$
2,40	2,61	6,54	7,83	7,08					-0,04

В останньому рядку наведено середньостатистичні величини (середні значення).

4. Обчислення залишкової дисперсії $S_u^2 = \frac{\sum u_i^2}{n}$ і незміщеної дисперсії

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

5. Визначення коефіцієнта детермінації $R^2 = 1 - \frac{S_u^2}{S_y^2} = \overline{y^2} - (\bar{y})^2$.

Для цієї задачі отримано результати:

$$S_u^2 = 0,168, S_u = 0,41; R = 0,915.$$

6. Перевірка статистичних гіпотез.

6.1. Перевіряємо на значущість коефіцієнт детермінації R^2 :

гіпотеза $H_0 : R^2 = 0$ проти $H_1 : R^2 > 0$.

Це рівнозначно тому, що перевіряється значущість водночас усіх параметрів моделі (гіпотеза $H_0 : \hat{a}_0 = \hat{a}_1 = 0$ проти $H_1 : \hat{a}_j \neq 0, j = 0, 1$).

Для перевірки обчислюємо F -статистику:

$$F_{\text{стат}} = \frac{(n-2)R^2}{1-R^2}.$$

Знайдене значення F -критерію порівнюємо з табличним $F_{\text{табл}} = F(\alpha, k_1, k_2)$, де $\alpha = 0,05$ – рівень значущості, $k_1 = 1$, $k_2 = n - 2 = 30 - 2 = 28$ – степені вільності.

Зауваження. В Excel для визначення $F_{\text{табл}}$ застосовують статистичну функцію ФРАСПОБР(α, k_1, k_2).

Якщо $F_{\text{стат}} > F_{\text{табл}}$, нульову гіпотезу відхиляємо: змінна y істотно залежить від змінної x . Модель якісна.

Якщо $F_{\text{стат}} < F_{\text{табл}}$, нульову гіпотезу приймаємо: змінна y слабо залежить від змінної x . Модель неякісна.

Для цієї задачі:

$F_{\text{стат}} = 144,00; F_{\text{табл}} = 4,20; F_{\text{стат}} > F_{\text{табл}}$. Гіпотезу ($a_0 = a_1 = 0$) відхиляємо. Хоча б один коефіцієнт моделі є значущим.

6.2. Перевіряємо на значущість коефіцієнт кореляції (гіпотеза $H_0 : R = 0$ проти альтернативної $H_1 : R \neq 0$). Для цього обчислюємо t -статистику:

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

Знаходимо $t_{\text{табл}}(\alpha/2, k)$ – відповідне табличне значення t -розподілу з $k = n - 2$ степенями вільності та рівнем значущості $\alpha = 0,05$.

Зауваження. В Excel для визначення $t_{\text{табл}}$ застосовують статистичну функцію СТЬЮДРАСПОБР ($\alpha/2, n - 2$).

Якщо $|t_{\text{стат}}| > t_{\text{табл}}$, гіпотезу $R = 0$ відхиляємо. Коефіцієнт кореляції R має істотно ненульове значення. Він є значущим, тобто зв'язок між змінними суттєвий.

Якщо $|t_{\text{стат}}| < t_{\text{табл}}$, гіпотезу $R = 0$ приймаємо. Коефіцієнт кореляції R неістотний: зв'язок між змінними незначущий.

Для цієї задачі:

$$t_{\text{стат}} = 12,00; t_{\text{табл}} = 2,37; t_{\text{стат}} > t_{\text{табл}}. \text{ Гіпотезу } R = 0 \text{ відхиляємо.}$$

6.3. Перевіряємо на значущість кожний параметр моделі, тобто перевіряємо гіпотези $H_0^j: a_0 = 0, a_1 = 0$ проти відповідних альтернативних гіпотез $H_1^j: a_j \neq 0$ ($j = 0, 1$).

Для цього обчислюємо t -статистику для кожного коефіцієнта:

$$t_0 = \frac{a_0}{S_{a_0}}, \quad S_{a_0} = S_u \cdot \sqrt{c_{00}}, \quad c_{00} = \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}$$

$$t_i = \frac{a_i}{S_{a_i}}, \quad S_{a_i} = S_u \cdot \sqrt{c_{ii}}, \quad c_{11} = \frac{1}{\sum (x_i - \bar{x})^2}.$$

Маємо:

$t_0 = 0,64; t_0 < t_{\text{табл}}$. Гіпотезу $a_0 = 0$ приймаємо. Коефіцієнт a_0 незначущий.

$t_1 = 15,48; t_1 > t_{\text{табл}}$. Гіпотезу $a_1 = 0$ відхиляємо. Коефіцієнт a_1 значущий.

7. Обчислення довірчих інтервалів.

Для $R: (R - \Delta R, R + \Delta R), \Delta R = t_{\alpha/2, k} \cdot \frac{1 - R}{\sqrt{n}}$.

Маємо:

$$R = 0,915; \Delta R = 0,02; R - \Delta R = 0,90; R + \Delta R = 0,93.$$

Для $a_i: (a_i - \Delta a_i, a_i + \Delta a_i), \Delta a_i = t_{\alpha/2, k} \cdot \sigma_a$

Довірчі інтервали –

для $a_0, a_1: a_0 = 0,11; a_1 = 1,04; \Delta a_0 = 0,52; \Delta a_1 = 0,20;$

для регресії: $(y_i - \Delta y_i, y_i + \Delta y_i),$

$$\Delta y_i = t_{\alpha/2, n-2} \frac{S}{\sqrt{n}} \sqrt{1 + \frac{(x_i - \bar{x})^2}{\sigma_x^2}}, \sigma_x^2 = \overline{x^2} - (\bar{x})^2.$$

Обчислюємо прогнозні значення:

$$y_{pi} = a_0 + a_1 x_{pi},$$

де x_{pi} — значення, які перебувають за межами базового періоду (точковий прогноз).

Знаходимо межі довірчих інтервалів прогнозу індивідуального значення:

$$(y_{pi} - \Delta y_{pi}, y_{pi} + \Delta y_{pi}), \Delta y_{pi} = t_{\alpha/2, n-2} \cdot S_u \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_{pi} - \bar{x})^2}{\sum (x_i - \bar{x})^2}}.$$

Знаходимо межі довірчих інтервалів для математичних сподівань:

$$\Delta M(\hat{y}_{pi}) = t_{\alpha/2, n-2} \cdot S_u \cdot \sqrt{\frac{1}{n} + \frac{(x_{pi} - \bar{x})^2}{\sum (x_i - \bar{x})^2}}.$$

Результати розрахунків наведено на рис. 2.1.

Квадратні точки на графіку визначають полігон частот статистичної вибірки. Пряма лінія визначає лінійну регресію. Лінії вище та нижче прямої відображають довірчі інтервали для статистичних даних y_i . Жирні точки на графіку відповідають прогнозним значенням для факторів x , що не входять до статистичної вибірки. Лінії над точками та під ними визначають довірчі інтервали. За розрахунками маємо:

$$x_{p1} = 0,5, y_{p1} = 0,63, \Delta_{yp1} = 1,05, y_{p1} - \Delta_{yp1} = -0,43, y_{p1} + \Delta_{yp1} = 1,69;$$

$$x_{p2} = 0,8, y_{p2} = 0,8, \Delta_{yp2} = 0,94, y_{p2} - \Delta_{yp2} = 0,58, y_{p2} + \Delta_{yp2} = 1,31;$$

$$x_{p3} = 4,2, y_{p3} = 4,5, \Delta_{yp3} = 1,05, y_{p3} - \Delta_{yp3} = 3,44, y_{p3} + \Delta_{yp3} = 5,54;$$

$$x_{p4} = 4,5, y_{p4} = 4,8, \Delta_{yp4} = 1,07, y_{p4} - \Delta_{yp4} = 3,73, y_{p4} + \Delta_{yp4} = 5,87.$$

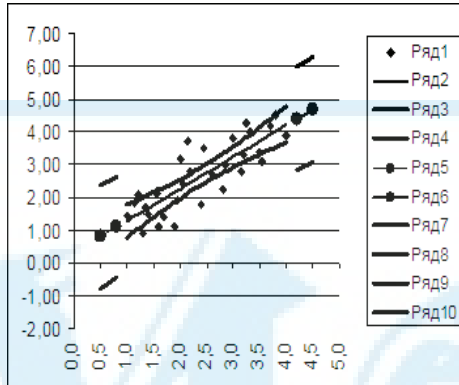


Рис. 2.1

Наведемо результати розрахунків для інших випадків (рис. 2.2–2.5).

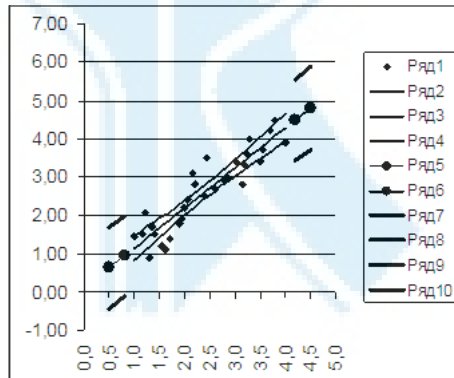


Рис. 2.2

Розрахунки парної регресії:

$$a_0 = 0,31; a_1 = 0,97, R = 0,82; t_{\text{табл}} = 2,37; F_{\text{табл}} = 4,20;$$

$$t_{\text{стат}} = 7,49; t_{\text{стат}} > t_{\text{табл}} \text{ Гіпотезу } R = 0 \text{ відхиляємо.}$$

$$F_{\text{стат}} = 56,04; F_{\text{стат}} > F_{\text{табл}} \text{ Гіпотезу } a_0 = a_1 = 0 \text{ відхиляємо.}$$

$t_0 = 1,21; t_0 < t_{\text{табл}}$ Гіпотезу $a_0 = 0$ приймаємо. Коефіцієнт a_0 незначущий.

$t_1 = 9,71$; $t_1 > t_{\text{табл}}$. Гіпотезу $a_1 = 0$ відхиляємо. Коефіцієнт a_1 значущий.

Довірчий інтервал для R : $\Delta R = 0,03$; $R - \Delta R = 0,78$; $R + \Delta R = 0,85$.

Довірчі інтервали для a_0, a_1 : $\Delta a_0 = 0,77$; $\Delta a_1 = 0,30$.

$x_{p1} = 0,5, y_{p1} = 0,80, \Delta_{yp1} = 1,58, y_{p1} - \Delta_{yp1} = -0,78, y_{p1} + \Delta_{yp1} = 2,38$;

$x_{p2} = 0,8, y_{p2} = 1,09, \Delta_{yp2} = 1,55, y_{p2} - \Delta_{yp2} = -0,46, y_{p2} + \Delta_{yp2} = 2,64$;

$x_{p3} = 4,2, y_{p3} = 4,41, \Delta_{yp3} = 1,57, y_{p3} - \Delta_{yp3} = 2,84, y_{p3} + \Delta_{yp3} = 5,97$;

$x_{p4} = 4,5, y_{p4} = 4,70, \Delta_{yp4} = 1,60, y_{p4} - \Delta_{yp4} = 3,10, y_{p4} + \Delta_{yp4} = 6,30$.

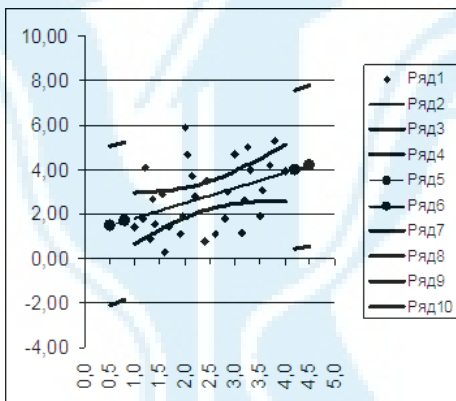


Рис. 2.3

Розрахунки парної регресії:

$a_0 = 0,78$; $a_1 = 0,80, R = 0,52$; $t_{\text{табл}} = 2,37$; $F_{\text{табл}} = 4,20$;

$t_{\text{стат}} = 3,24$; $t_{\text{стат}} > t_{\text{табл}}$. Гіпотезу $R = 0$ відхиляємо.

$F_{\text{стат}} = 10,50$; $F_{\text{стат}} > F_{\text{табл}}$. Гіпотезу $a_0 = a_1 = 0$ відхиляємо.

$t_0 = 1,30$; $t_0 < t_{\text{табл}}$. Гіпотезу $a_0 = 0$ приймаємо. Коефіцієнт a_0 незначущий.

$t_1 = 4,35$; $t_1 > t_{\text{табл}}$. Гіпотезу $a_1 = 0$ відхиляємо. Коефіцієнт a_1 значущий.

Довірчий інтервал для R : $\Delta R = 0,09$; $R - \Delta R = 0,43$; $R + \Delta R = 0,61$.

Довірчі інтервали для a_0, a_1 : $\Delta a_0 = 1,12$; $\Delta a_1 = 0,44$.

$x_{p1} = 0,5, y_{p1} = 1,18, \Delta_{yp1} = 2,90, y_{p1} - \Delta_{yp1} = -1,72, y_{p1} + \Delta_{yp1} = 4,08$;

$x_{p2} = 0,8, y_{p2} = 1,42, \Delta_{yp2} = 2,85, y_{p2} - \Delta_{yp2} = -1,43, y_{p2} + \Delta_{yp2} = 4,26$;

$x_{p3} = 4,2, y_{p3} = 4,15, \Delta_{yp3} = 12,88, y_{p3} - \Delta_{yp3} = 1,27, y_{p3} + \Delta_{yp3} = 7,03$;

$x_{p4} = 4,5, y_{p4} = 4,39, \Delta_{yp4} = 2,94, y_{p4} - \Delta_{yp4} = 1,45, y_{p4} + \Delta_{yp4} = 7,33$.

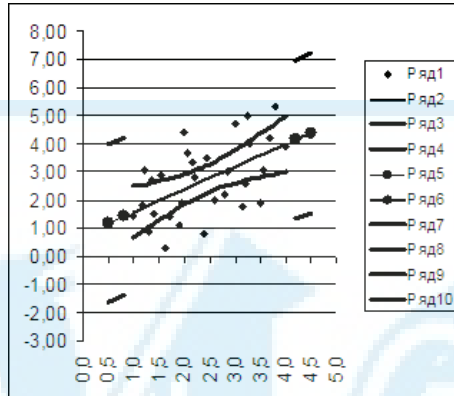


Рис. 2.4

Розрахунки парної регресії:

$$a_0 = 1,16; a_1 = 0,67, R = 0,36; t_{\text{табл}} = 2,37; F_{\text{табл}} = 4,20;$$

$t_{\text{стат}} = 2,02; t_{\text{стат}} < t_{\text{табл}}$. Гіпотезу $R = 0$ приймаємо.

$F_{\text{стат}} = 4,07; F_{\text{стат}} > F_{\text{табл}}$. Гіпотезу $a_0 = a_1 = 0$ відхиляємо.

$t_0 = 1,94; t_0 < t_{\text{табл}}$. Гіпотезу $a_0 = 0$ приймаємо. Коефіцієнт a_0 незначущий.

$t_1 = 2,89; t_1 > t_{\text{табл}}$. Гіпотезу $a_1 = 0$ відхиляємо. Коефіцієнт a_1 значущий.

Довірчий інтервал для R : $\Delta R = 0,12; R - \Delta R = 0,24; R + \Delta R = 0,47$.

Довірчі інтервали для a_0, a_1 : $\Delta a_0 = 1,78; \Delta a_1 = 0,70$.

$$x_{p1} = 0,5, y_{p1} = 1,49, \Delta_{yp1} = 3,66, y_{p1} - \Delta_{yp1} = -2,17, y_{p1} + \Delta_{yp1} = 5,16;$$

$$x_{p2} = 0,8, y_{p2} = 1,70, \Delta_{yp2} = 3,59, y_{p2} - \Delta_{yp2} = -1,90, y_{p2} + \Delta_{yp2} = 5,29;$$

$$x_{p3} = 4,2, y_{p3} = 3,98, \Delta_{yp3} = 3,64, y_{p3} - \Delta_{yp3} = 0,35, y_{p3} + \Delta_{yp3} = 7,62;$$

$$x_{p4} = 4,5, y_{p4} = 4,18, \Delta_{yp4} = 3,71, y_{p4} - \Delta_{yp4} = 0,47, y_{p4} + \Delta_{yp4} = 7,90.$$

Розрахунки парної регресії:

$$a_0 = 1,46; a_1 = 0,67, R = 0,37; t_{\text{табл}} = 2,37; F_{\text{табл}} = 4,20;$$

$t_{\text{стат}} = 2,27; t_{\text{стат}} < t_{\text{табл}}$. Гіпотезу $R = 0$ приймаємо.

$F_{\text{стат}} = 0,07; F_{\text{стат}} < F_{\text{табл}}$. Гіпотезу $a_0 = a_1 = 0$ приймаємо.

$t_0 = 2,02; t_0 < t_{\text{табл}}$. Гіпотезу $a_0 = 0$ приймаємо. Коефіцієнт a_0 незначущий.

$t_1 = 2,89; t_1 < t_{\text{табл}}$. Гіпотезу $a_1 = 0$ приймаємо. Коефіцієнт a_1 незначущий.

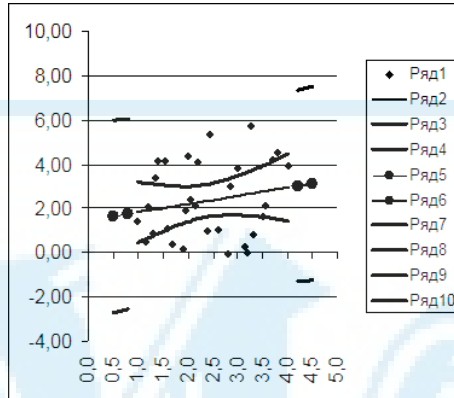


Рис. 2.5

Довірчий інтервал для R : $R = 0,10$; $\Delta R = 0,16$; $R - \Delta R = -0,07$; $R + \Delta R = 0,26$.

Довірчі інтервали для a_0 , a_1 : $\Delta a_0 = 2,12$; $\Delta a_1 = 0,83$.

$x_{p1} = 0,5$, $y_{p1} = 1,62$, $\Delta_{yp1} = 4,35$, $y_{p1} - \Delta_{yp1} = -2,73$, $y_{p1} + \Delta_{yp1} = 5,98$;

$x_{p2} = 0,8$, $y_{p2} = 1,74$, $\Delta_{yp2} = 4,27$, $y_{p2} - \Delta_{yp2} = -2,53$, $y_{p2} + \Delta_{yp2} = 6,01$;

$x_{p3} = 4,2$, $y_{p3} = 3,09$, $\Delta_{yp3} = 4,32$, $y_{p3} - \Delta_{yp3} = -1,23$, $y_{p3} + \Delta_{yp3} = 7,42$;

$x_{p4} = 4,5$, $y_{p4} = 3,21$, $\Delta_{yp4} = 4,41$, $y_{p4} - \Delta_{yp4} = -1,20$, $y_{p4} + \Delta_{yp4} = 7,63$.

Висновки

1. Рівняння парної регресії описує зв'язок між двома змінними (економічними показниками), що виявляється лише в середньому для сукупності спостережених даних.

2. Форму залежності вибирають за допомогою графічного, аналітичного або експериментального методу.

3. Залежність між змінними може бути лінійною або нелінійною як за параметрами, так і за змінними. Нелінійні за параметрами залежності перетворюються до лінійного вигляду.

4. Параметри лінійних моделей оцінюють методом найменших квадратів.

5. Після визначення форми рівняння регресії та оцінювання його параметрів перевіряють значущість рівняння загалом і окремих його параметрів.

6. Адекватність моделі перевіряють за допомогою критерію Фішера (F -критерію), а оцінку значущості параметрів — за допомогою критерію Стьюдента (t -критерію).

7. Побудована регресійна залежність дає змогу застосувати модель для аналізу зв'язку між змінними (показниками), а також для прогнозування значення результативної змінної для заданого значення факторної змінної.

8. Виокремлюють два види прогнозу: точковий та інтервальний. Точковий визначає найсподіваніше (єдине) значення результативної змінної, інтервальний — межі відхилення від цього значення, до яких з великою ймовірністю потраплять реальні значення показника.

Питання для самоперевірки

1. Що таке парна регресія?
2. Які види нелінійних функцій найчастіше застосовують в економічних дослідженнях? Як їх перетворити до лінійного вигляду?
3. Сутність методу найменших квадратів.
4. Наведіть формули для розрахунку коефіцієнтів емпіричного парного лінійного рівняння регресії за МНК.
5. Як пов'язані емпіричні коефіцієнти лінійної регресії з вибірковим коефіцієнтом кореляції між змінними рівняння регресії?
6. Які висновки можна зробити про оцінки коефіцієнтів регресії та випадкового відхилення, отриманих за МНК?
7. Проінтерпретуйте коефіцієнти емпіричного парного лінійного рівняння регресії.
8. Назвіть показники якості моделі, за якими обирають кращі моделі.
9. Від чого залежить якість прогнозу за лінійною моделлю?

Завдання

1. У наступній вибірці подано дані щодо ціни P і кількості Q блага, яке домогосподарство купує щомісяця упродовж року.

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

- а. Побудувати кореляційне поле і за розміщенням точок на графіку визначити формулу залежності між P та Q .
- б. Оцінити за МНК параметри рівняння лінійної регресії.
- в. Оцінити вибірковий коефіцієнт кореляції r_{pq} .
- г. Проінтерпретувати результати.

2. Дано таблицю тижневого прибутку X і тижневого споживання Y для 60 домашніх господарств:

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

- а. Для кожного рівня прибутку розрахувати середнє значення споживання, що є оцінкою умовного математичного сподівання $M(Y|X = x_i)$.
- б. Побудувати кореляційне поле для даної вибірки.
- в. Побудувати емпіричне лінійне рівняння регресії, використовуючи всі дані.
- г. Побудувати емпіричне лінійне рівняння регресії, використовуючи лише середні значення споживання для кожного рівня прибутку.
- д. Порівняти побудовані рівняння. Яке з них, на вашу думку, ближче до теоретичного?
- е. Розрахувати вибірковий коефіцієнт кореляції для пунктів в. і г. Чи буде лінійний зв'язок між цими змінними суттєвим? Відповідь обґрунтувати.

Вказівка. На підставі початкових даних утворити варіанти завдань, додаючи у першому наборі даних до кожного значення число N , а в другому — віднімаючи N , де N — остання цифра номера залікової книжки або номер за списком групи.

МЕТОДИ ПОБУДОВИ МНОЖИННОЇ РЕГРЕСІЇ

Незважаючи на те, що парна лінійна регресія легко інтерпретується, усе ж практичне застосування її обмежене. Найчастіше використовують моделі множинної регресії, які дають змогу врахувати вплив на результативну змінну не одного, а багатьох факторів.

Починаючи побудову багатофакторного регресійного рівняння, потрібно розв'язувати проблему не лише вибору форми залежності між змінними, а й добору факторів, що мають увійти до моделі.

Фактори мають бути кількісно вимірними. Якісним факторам, які не мають кількісної міри, потрібно надати кількісної визначеності (наприклад, задати якість у вигляді балів, проранжувати дані або встановити наявність чи відсутність якісної ознаки за допомогою бінарних змінних, тобто змінних, що набувають двох значень — 0 і 1).

3.1. Вибір факторів і форми рівняння

Незалежні змінні економетричної моделі повинні мати такі властивості:

- високу варіабельність;
- сильно корелювати з результативною змінною;
- слабо корелювати між собою;
- сильно корелювати зі змінними, що не використовуються в моделі як пояснювальні змінні, але пов'язані з результативною змінною.

Факторні (незалежні) змінні добирають за допомогою статистичних методів. Процедура вибору змінних складається з таких кроків:

- 1) на підставі нагромаджених знань складають множину так званих потенційних (первісних) змінних, до якої включають усі найважливіші показники, що впливають на результативну змінну. Позначимо їх X_1, X_2, \dots, X_m ;

- 2) збирають статистичну інформацію про реалізацію результативної й факторних змінних. Формують вектор спостережень залежної змінної Y і матрицю спостережень незалежних змінних X_1, X_2, \dots, X_m ;
- 3) змінні, що мають низький рівень варіабельності (змінюваності), виключають з розгляду як такі, що містять мінімальну інформацію;
- 4) обчислюють коефіцієнти кореляції між усіма змінними, що розглядаються.

3.1.1. Виключення квазінезмінних змінних

Мірою варіабельності незалежних змінних є коефіцієнт варіації

$$v_j = \frac{S_j}{\bar{x}_j}, \quad (j = 1, 2, \dots, m),$$

де \bar{x}_j – середнє арифметичне змінної X_j ; S_j – її стандартне середньоквадратичне відхилення,

$$S_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}.$$

Задають критичне значення коефіцієнта варіації $v_{\text{крит}}$, наприклад $v_{\text{крит}} = 0,1$. Змінні, для яких виконується нерівність $v_j \leq v_{\text{крит}}$, вважають квазінезмінними і виключають із множини факторних змінних, оскільки вони не містять значущої інформації.

Приклад 3.1 [26]. Обсяг виробництва деякого підприємства Y залежить від кількості працюючих (X_1), вартості машин і обладнання (X_2), тривалості простою обладнання (X_3) та інвестиційних витрат (X_4).

Значення конкретних змінних за десять років подано в таблиці.

Рік	Y	X_1	X_2	X_3	X_4
1	10	6	8	14	12
2	10	6	8	14	12
3	16	10	12	18	12
4	16	10	12	18	14
5	12	8	8	18	10
6	14	10	8	18	12
7	20	12	14	24	14
8	20	12	16	24	12
9	20	12	16	26	12
10	22	14	18	26	10

Перевіримо, чи мають факторні змінні достатньо високу варіабельність. Для цього обчислимо середні арифметичні значення змінних і їхні стандарти відхилення:

$$\bar{x}_1 = 10; \bar{x}_2 = 12; \bar{x}_3 = 20; \bar{x}_4 = 12;$$

$$S_1 = 2,51; S_2 = 3,69; S_3 = 4,38; S_4 = 1,27.$$

Коефіцієнти варіації змінних такі:

$$v_1 = \frac{2,51}{10} = 0,251; \quad v_2 = \frac{3,69}{12} = 0,307; \quad v_3 = \frac{4,38}{20} = 0,219;$$

$$v_4 = \frac{1,265}{12} = 0,105.$$

Останній коефіцієнт найменший; якщо критичним вибрати $v_{\text{крит}} = 0,15$, то змінну X_4 можна визнати квазінезмінною. Її потрібно вилучити із множини факторних змінних, оскільки змінна з низьким рівнем варіабельності непридатна для опису процесу формування результативної змінної.

3.1.2. Вектор і матриця коефіцієнтів кореляції

Для визначення сили (щільності) зв'язку обчислюємо коефіцієнти кореляції за формулою

$$r_j = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_{ij} - \bar{x}_j)}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}} \quad (j = 1, 2, \dots, m). \quad (3.1)$$

Ці коефіцієнти об'єднуємо у вектор кореляції $R = \begin{pmatrix} r_1 \\ r_2 \\ \dots \\ r_m \end{pmatrix}$.

Коефіцієнти кореляції між потенційними незалежними змінними X_1, X_2, \dots, X_m розраховані за формулою

$$r_{kj} = \frac{\sum_{i=1}^n (x_{ik} - \bar{x}_k)(x_{ij} - \bar{x}_j)}{\sqrt{\sum_{i=1}^n (x_{ik} - \bar{x}_k)^2 \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}} \quad (k, j = 1, 2, \dots, m), \quad (3.2)$$

об'єднуємо в матрицю кореляції:

$$R_0 = \begin{pmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{pmatrix}.$$

Матриця кореляції симетрична, тобто $r_{ij} = r_{ji}$.

Приклад 3.2 [26]. На основі п'яти спостережень залежної змінної Y і потенційних незалежних змінних X_1, X_2 розрахувати коефіцієнти кореляції змінної Y зі змінними X_1, X_2 , а також коефіцієнти кореляції між незалежними змінними. Скласти кореляційний вектор і кореляційну матрицю.

t	y_t	x_{t1}	x_{t2}
1	1,9	11	1,1
2	2,1	13	1,2
3	2,6	21	1,4
4	2,7	31	2,1
5	3,2	29	2,2

Спершу обчислюємо середні значення всіх змінних:

$$\bar{y} = 2,5; \quad \bar{x}_1 = 21; \quad \bar{x}_2 = 1,6.$$

Коефіцієнти кореляції між змінною Y і змінними X_1, X_2 такі: $r_1 = 0,7; r_2 = 0,6$. Коефіцієнт кореляції пари змінних X_1, X_2 становить $r_{12} = r_{21} = 0,5$.

Отже, вектор кореляції має вигляд $R = \begin{pmatrix} 0,7 \\ 0,6 \end{pmatrix}$, а кореляційна матриця $R_0 = \begin{pmatrix} 1 & 0,5 \\ 0,5 & 1 \end{pmatrix}$.

3.1.3. Метод аналізу матриці коефіцієнтів кореляції

Ідея методу зводиться до вибору факторних змінних, які сильно корелюють із залежною змінною і водночас слабко корелюють між собою.

Критичне значення коефіцієнта кореляції обчислюють за допомогою t -розподілу Стьюдента для заданого рівня значущості α і $n - 2$ степенів вільності за формулою

$$r^* = \sqrt{\frac{t_{\text{табл}}^2}{t_{\text{табл}}^2 + n - 2}}.$$

Критичне значення r^* можна також задавати апріорно.

Процедура вибору пояснювальних змінних складається з таких кроків:

- 1) із множини всіх потенційних незалежних змінних вилучають усі, що мають коефіцієнт кореляції, який менший за критичне значення r^* ;
- 2) серед змінних, що залишилися, вибирають таку, яка має найбільший коефіцієнт кореляції із залежною змінною, оскільки саме вона є носієм найбільшої кількості інформації про залежну змінну. Нехай це змінна X_i ;
- 3) із множини незалежних змінних вилучають такі, коефіцієнти кореляції яких із вибраною незалежною змінною перевищують критичне значення r^* , оскільки вони відтворюють ту саму інформацію, яку вже містить вибрана змінна.

Процес повторюють до повного спустошення множини незалежних змінних.

Приклад 3.3 [26]. З метою вивчення процесу формування в різних країнах рівня споживання м'яса Y попередньо було вибрано вісім потенційних пояснювальних змінних, що визначають рівні споживання інших продуктів харчування: X_1 – зернових продуктів; X_2 – картоплі; X_3 – овочів; X_4 – фруктів; X_5 – жирів; X_6 – риби; X_7 – молока; X_8 – яєць.

На основі статистичних даних про споживання цих продуктів у 28 країнах обчислено вектор коефіцієнтів кореляції R і матрицю коефіцієнтів кореляції R_0 :

$$R = \begin{pmatrix} -0,59 \\ -0,06 \\ 0,08 \\ 0,13 \\ 0,48 \\ -0,15 \\ -0,10 \\ 0,72 \end{pmatrix};$$

$$R_0 = \begin{pmatrix} 1 & -0,09 & 0,35 & -0,17 & -0,62 & -0,40 & -0,16 & -0,55 \\ -0,09 & 1 & -0,06 & -0,38 & 0,00 & 0,15 & 0,22 & 0,11 \\ 0,35 & -0,06 & 1 & 0,33 & -0,11 & -0,20 & -0,45 & -0,02 \\ -0,17 & -0,38 & 0,33 & 1 & 0,20 & -0,07 & -0,44 & 0,07 \\ -0,62 & 0,00 & -0,11 & 0,20 & 1 & 0,22 & 0,17 & -0,11 \\ -0,40 & 0,15 & -0,20 & -0,07 & 0,22 & 1 & -0,19 & 0,47 \\ -0,16 & 0,22 & -0,45 & -0,44 & 0,17 & -0,19 & 1 & 0,05 \\ -0,55 & 0,11 & -0,02 & 0,07 & -0,11 & 0,47 & 0,05 & 1 \end{pmatrix}.$$

Факторні змінні починаємо вибирати для рівня значущості $\alpha = 0,10$. Із таблиці розподілу Стюдента знаходимо $t_{\text{табл}} = t(0,1/2; 27) = 1,706$ і обчислюємо

$$r^* = \sqrt{\frac{1,706^2}{1,706^2 + 28 - 2}} = 0,318.$$

Насамперед із множини потенційних змінних виключаємо змінні, що мають слабку кореляцію із залежною змінною, тобто їхні коефіцієнти кореляції із залежною змінною менші за абсолютною величиною від критичного значення $r^* = 0,318$. Це змінні X_2, X_3, X_4, X_6 і X_7 , для яких $r_2 = -0,06$; $r_3 = 0,086$; $r_4 = 0,13$; $r_7 = -0,10$.

Серед змінних, що залишилися, вибираємо ту, що має найбільший коефіцієнт кореляції із залежною змінною. Це змінна X_8 , оскільки $r_8 = 0,72$. Саме її визнають як першу факторну змінну. Змінні, які мають з X_8 коефіцієнти кореляції більші, ніж r^* , виключають з розгляду як такі, що не дають додаткової інформації. Найбільші коефіцієнти кореляції з X_8 мають змінні X_1 і X_6 , для яких $r_{81} = r_{18} = 0,55$ і $r_{86} = r_{68} = 0,47$. Обидва значення більші за критичне, тому змінні потрібно вивести з розгляду (змінну X_6 вивели раніше). Отже, серед потенційних змінних залишилася лише одна – X_5 . Оскільки її коефіцієнт кореляції із залежною змінною перевищує критичне значення r^* ($r_5 = 0,48$), змінну розглядають як факторну. Отже, множина незалежних змінних моделі складається лише з двох змінних: X_5 – споживання жирів; X_8 – споживання яєць. Остаточна модель споживання м'яса відносно зазначених змінних можна записати у вигляді

$$Y = a_0 + a_1 X_5 + a_2 X_8 + u.$$

3.1.4. Метод показників інформаційної місткості

Ідея методу зводиться до вибору факторних змінних, які сильно корелюють із результативною змінною і водночас слабо корелюють між собою. Початковими даними для застосування методу є вектор коефіцієнтів кореляції R_0 і кореляційна матриця R .

Розглядають усі комбінації потенційних факторних змінних X_1, X_2, \dots, X_m . Загальна кількість таких комбінацій $L = 2^m - 1$.

Для кожної комбінації факторних змінних розраховують індивідуальні та інтегральні показники інформаційної місткості.

Індивідуальні показники інформаційної місткості для кожної конкретної ситуації визначають за формулою

$$h_{ij} = \frac{r_j^2}{1 + \sum_{\substack{i=1 \\ i \neq j}}^{m_l} |r_{ij}|}$$

($l = 1, 2, \dots, L$; m_l — кількість змінних у певній комбінації).

Інтегральні показники інформаційної місткості обчислюють за формулою

$$H_l = \sum_{j=1}^{m_l} h_{ij} \quad (l = 1, 2, \dots, L).$$

Індивідуальні й інтегральні показники інформаційної місткості нормуються в інтервалі $[0, 1]$. Їхні значення тим більші, чим сильніше факторні змінні корелюють із результативною змінною і чим слабкіше вони корелюють між собою.

Факторними змінними обирають ті, які мають максимальне значення інтегрального показника інформаційної місткості.

Приклад 3.4 [26]. Скласти лінійну модель, що описує виробництво худоби в живій вазі на деякому аграрному підприємстві в розрахунку на 1 га сільгоспугідь. Можливими факторними змінними можуть бути: X_1 — урожай кормової кукурудзи; X_2 — частка вартості рослинної агропродукції в загальній вартості продукції сільського господарства; X_3 — середні закупівельні ціни м'яса в живій вазі; X_4 — використання комбікормів.

На підставі статистичних даних за 15 років побудовано вектор коефіцієнтів кореляції R_0 і матрицю коефіцієнтів кореляції R :

$$R_0 = \begin{pmatrix} 0,43 \\ -0,8 \\ 0,18 \\ 0,63 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & -0,64 & 0,14 & 0,41 \\ -0,64 & 1 & -0,13 & -0,55 \\ 0,14 & -0,13 & 1 & -0,03 \\ 0,41 & -0,55 & -0,03 & 1 \end{pmatrix}.$$

Серед чотирьох потенційних змінних потрібно вибрати найвпливовіші. Скористаємось показниками інформаційної місткості. Для цього розглянемо $L = 2^4 - 1 = 15$ комбінацій змінних. Позначимо їх так:

$$\begin{array}{lll} C_1 = X_1; & C_5 = (X_1, X_2); & C_{11} = (X_1, X_2, X_3); \\ C_2 = X_2; & C_6 = (X_1, X_3); & C_{12} = (X_1, X_2, X_4); \\ C_3 = X_3; & C_7 = (X_1, X_4); & C_{13} = (X_1, X_3, X_4); \\ C_4 = X_4; & C_8 = (X_2, X_3); & C_{14} = (X_2, X_3, X_4); \\ & C_9 = (X_2, X_4); & C_{15} = (X_1, X_2, X_3, X_4). \\ & C_{10} = (X_3, X_4); & \end{array}$$

Для одноелементних комбінацій ($C_1 - C_4$) індивідуальні показники інформаційної місткості $h_{jj} = r_j^2$, тобто є квадратами коефіцієнтів кореляції відповідної факторної змінної із залежною змінною; інтегральні показники, які складаються лише з одного доданка, збігаються з індивідуальними показниками і мають такі значення:

$$\begin{aligned} H_1 &= h_{11} = r_1^2 = (0,43)^2 = 0,185; \\ H_2 &= h_{22} = r_2^2 = (-0,8)^2 = 0,640; \\ H_3 &= h_{33} = r_3^2 = (0,18)^2 = 0,032; \\ H_4 &= h_{44} = r_4^2 = (0,63)^2 = 0,370. \end{aligned}$$

Для двохелементної комбінації $C_5 = (X_1, X_2)$ обчислюємо два індивідуальних показники інформаційної місткості: h_{51} , пов'язаний зі змінною X_1 , і h_{52} , пов'язаний зі змінною X_2 :

$$h_{51} = \frac{r_1^2}{1 + |r_{12}|} = \frac{(0,43)^2}{1 + 0,64} = 0,113; \quad h_{52} = \frac{r_2^2}{1 + |r_{21}|} = \frac{(-0,8)^2}{1 + 0,64} = 0,390.$$

Інтегральний показник інформаційної місткості для цієї комбінації

$$H_5 = h_{51} + h_{52} = 0,113 + 0,390 = 0,503.$$

Після проведення аналогічних розрахунків для інших двохелементних комбінацій отримуємо

$$\begin{array}{lll}
h_{61} = 0,162; & h_{63} = 0,028; & H_6 = 0,190; \\
h_{71} = 0,131; & h_{74} = 0,281; & H_7 = 0,412; \\
h_{82} = 0,566; & h_{83} = 0,029; & H_8 = 0,595; \\
h_{92} = 0,413; & h_{94} = 0,255; & H_9 = 0,668; \\
h_{10,3} = 0,031; & h_{10,4} = 0,385; & H_{10} = 0,416.
\end{array}$$

Для трьохелементної комбінації $C_{11} = (X_1, X_2, X_3)$ індивідуальні показники інформаційної місткості мають такі значення:

$$\begin{aligned}
h_{11,1} &= \frac{r_1^2}{1 + |r_{12}| + |r_{13}|} = \frac{(0,43)^2}{1 + 0,64 + 0,14} = 0,104; \\
h_{11,2} &= \frac{r_2^2}{1 + |r_{21}| + |r_{23}|} = \frac{(-0,8)^2}{1 + 0,64 + 0,13} = 0,362; \\
h_{11,3} &= \frac{r_3^2}{1 + |r_{31}| + |r_{32}|} = \frac{(0,18)^2}{1 + 0,14 + 0,13} = 0,026.
\end{aligned}$$

Інтегральний показник інформаційної місткості для цієї комбінації

$$H_{11} = h_{11,1} + h_{11,2} + h_{11,3} = 0,492.$$

Після проведення аналогічних розрахунків для інших трьохелементних комбінації отримуємо

$$\begin{array}{llll}
h_{12,1} = 0,090; & h_{12,3} = 0,292; & h_{12,4} = 0,202; & H_{12} = 0,584; \\
h_{13,1} = 0,119; & h_{13,3} = 0,028; & h_{13,4} = 0,276; & H_{13} = 0,423; \\
h_{14,2} = 0,381; & h_{14,3} = 0,028; & h_{14,4} = 0,251; & H_{14} = 0,660.
\end{array}$$

Для останньої чотирьохелементної комбінації C_{15} індивідуальні показники інформаційної місткості мають такі значення:

$$\begin{aligned}
h_{15,1} &= \frac{r_1^2}{1 + |r_{12}| + |r_{13}| + |r_{14}|} = \frac{(0,43)^2}{1 + 0,64 + 0,14 + 0,41} = 0,084; \\
h_{15,2} &= \frac{r_2^2}{1 + |r_{21}| + |r_{23}| + |r_{24}|} = \frac{(-0,8)^2}{1 + 0,64 + 0,13 + 0,55} = 0,276; \\
h_{15,3} &= \frac{r_3^2}{1 + |r_{31}| + |r_{32}| + |r_{34}|} = \frac{(0,18)^2}{1 + 0,14 + 0,13 + 0,03} = 0,025; \\
h_{15,4} &= \frac{r_4^2}{1 + |r_{41}| + |r_{42}| + |r_{43}|} = \frac{(0,63)^2}{1 + 0,41 + 0,55 + 0,03} = 0,200.
\end{aligned}$$

Інтегральний показник інформаційної місткості

$$H_{15} = h_{15,1} + h_{15,2} + h_{15,3} + h_{15,4} = 0,585.$$

Максимальне значення інтегрального показника інформаційної місткості дорівнює 0,668 і відповідає комбінації C_9 , яка містить змінні X_2 і X_4 . Це означає, що в лінійній моделі, яка описує рівень виробництва худоби в живій вазі на 1 га сільгоспугідь на агропідприємстві, факторними змінними мають бути: X_2 – частка вартості рослинної продукції в загальній вартості продукції сільського господарства, X_4 – використання комбікормів. При цьому модель набирає вигляду

$$Y = a_0 + a_1X_2 + a_2X_4 + u.$$

3.1.5. Коефіцієнт множинної кореляції

Коефіцієнт множинної кореляції є мірою лінійного зв'язку результативної змінної Y з факторними змінними X_1, X_2, \dots, X_k . Його можна обчислити за допомогою елементів матриці R – матриці коефіцієнтів попарної кореляції незалежних змінних, а також вектора R_0 – коефіцієнтів кореляції залежної змінної з усіма незалежними змінними за формулою

$$R = \sqrt{1 - \frac{\det(W)}{\det(R)}},$$

де $\det(R)$ – детермінант (визначник) матриці R ; $\det(W)$ – визначник матриці W , яка складається з елементів вектора R_0 і матриці R за правилом

$$W = \begin{pmatrix} 1 & R_0^T \\ R_0 & R \end{pmatrix}$$

і в розгорненій формі набирає вигляду

$$W = \begin{pmatrix} 1 & r_1 & r_2 & \dots & r_k \\ r_1 & 1 & r_{12} & \dots & r_{1k} \\ r_2 & r_{21} & 1 & \dots & r_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ r_k & r_{k1} & r_{k2} & \dots & 1 \end{pmatrix}.$$

Коефіцієнт множинної кореляції нормується в інтервалі $[0, 1]$. Його значення тим більше, чим сильніший зв'язок результативної змінної з факторними змінними. Коефіцієнт множинної кореляції

може бути критерієм вибору найкращої комбінації факторних змінних серед комбінацій, що містять однакоvu кількість факторів.

Приклад 3.5 [26]. Побудувати лінійну модель, що відбиває зміну обсягу продажу (Y) деякого торговельного підприємства. Факторними змінними можуть бути: X_1 – зайнятість, X_2 – обсяг постачання товарів, X_3 – середні ціни товарів, що продаються, X_4 – торговельна площа магазинів.

Задано вектор коефіцієнтів кореляції R_0 і кореляційну матрицю R :

$$R_0 = \begin{pmatrix} 0,7 \\ 0,9 \\ 0,1 \\ 0,5 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0,8 & 0,2 & 0,4 \\ 0,8 & 1 & 0,1 & 0,6 \\ 0,2 & 0,1 & 1 & 0,3 \\ 0,4 & 0,6 & 0,3 & 1 \end{pmatrix}.$$

Розглянемо двохелементні комбінації потенційних змінних:

$$C_1 = (X_1, X_2), \quad C_2 = (X_1, X_3), \quad C_3 = (X_1, X_4),$$

$$C_4 = (X_2, X_3); \quad C_5 = (X_2, X_4); \quad C_6 = (X_3, X_4).$$

Для комбінації $C_1 = (X_1, X_2)$

$$R_{01} = \begin{pmatrix} 0,7 \\ 0,9 \end{pmatrix}, \quad R_1 = \begin{pmatrix} 1 & 0,8 \\ 0,8 & 1 \end{pmatrix}, \quad W = \begin{pmatrix} 1 & 0,7 & 0,9 \\ 0,7 & 1 & 0,8 \\ 0,9 & 0,8 & 1 \end{pmatrix}.$$

Коефіцієнт множинної кореляції для неї

$$R_1 = \sqrt{1 - \frac{\det(W_1)}{\det(R_1)}}.$$

Легко переконатися, що визначники матриць

$$\det(R_1) = 0,360, \quad \det(W_1) = 0,068.$$

$$\text{Отже, } R_1 = \sqrt{1 - \frac{0,068}{0,360}} = 0,901.$$

Для інших комбінацій отримаємо:

- для комбінації C_2

$$R_{02} = \begin{pmatrix} 0,7 \\ 0,1 \end{pmatrix}, \quad R_2 = \begin{pmatrix} 1 & 0,2 \\ 0,2 & 1 \end{pmatrix}, \quad W = \begin{pmatrix} 1 & 0,7 & 0,1 \\ 0,7 & 1 & 0,2 \\ 0,1 & 0,2 & 1 \end{pmatrix},$$

$$\det(R_2) = 0,960, \quad \det(W_2) = 0,508; \quad R_2 = \sqrt{1 - \frac{0,508}{0,960}} = 0,686;$$

- для комбінації C_3

$$R_{03} = \begin{pmatrix} 0,7 \\ 0,5 \end{pmatrix}, R_3 = \begin{pmatrix} 1 & 0,4 \\ 0,4 & 1 \end{pmatrix}, W = \begin{pmatrix} 1 & 0,7 & 0,5 \\ 0,7 & 1 & 0,4 \\ 0,5 & 0,4 & 1 \end{pmatrix},$$

$$\det(R_3) = 0,84, \det(W_3) = 0,38; R_2 = \sqrt{1 - \frac{0,38}{0,84}} = 0,740;$$

- для комбінації C_4

$$R_{04} = \begin{pmatrix} 0,9 \\ 0,1 \end{pmatrix}, R_2 = \begin{pmatrix} 1 & 0,1 \\ 0,1 & 1 \end{pmatrix}, W = \begin{pmatrix} 1 & 0,9 & 0,1 \\ 0,9 & 1 & 0,1 \\ 0,1 & 0,1 & 1 \end{pmatrix},$$

$$\det(R_4) = 0,99, \det(W_4) = 0,188; R_4 = \sqrt{1 - \frac{0,188}{0,99}} = 0,90;$$

- для комбінації C_5

$$R_{05} = \begin{pmatrix} 0,9 \\ 0,5 \end{pmatrix}, R_2 = \begin{pmatrix} 1 & 0,6 \\ 0,6 & 1 \end{pmatrix}, W = \begin{pmatrix} 1 & 0,9 & 0,5 \\ 0,9 & 1 & 0,6 \\ 0,5 & 0,6 & 1 \end{pmatrix},$$

$$\det(R_5) = 0,64, \det(W_5) = 0,12; R_4 = \sqrt{1 - \frac{0,12}{0,64}} = 0,902;$$

- для комбінації C_6

$$R_{06} = \begin{pmatrix} 0,1 \\ 0,5 \end{pmatrix}, R_2 = \begin{pmatrix} 1 & 0,3 \\ 0,3 & 1 \end{pmatrix}, W = \begin{pmatrix} 1 & 0,1 & 0,5 \\ 0,1 & 1 & 0,3 \\ 0,5 & 0,3 & 1 \end{pmatrix},$$

$$\det(R_6) = 0,91, \det(W_6) = 0,68; R_4 = \sqrt{1 - \frac{0,68}{0,91}} = 0,5.$$

Максимальне значення коефіцієнта кореляції дорівнює 0,902 і відповідає комбінації C_5 , яка об'єднує змінні X_2 і X_4 . Це означає, що в лінійній моделі з двома факторами найвпливовішими є X_2 — обсяг постачання товарів і X_4 — торговельна площа магазинів. Отже, модель має вигляд

$$Y = a_0 + a_1 X_2 + a_2 X_4 + u.$$

3.1.6. Ефект каталізу в економетричній моделі

Ефект каталізу в лінійній економетричній моделі означає сильну кореляцію результативної змінної з факторними, хоча високий рівень коефіцієнта множинної кореляції зумовлений сильною кореляцією факторних змінних між собою. Факторні змінні, що спричиняють ефект каталізу, вилучають з моделі.

Нехай у векторі R_0 коефіцієнти кореляції додатні й упорядковані за зростанням. Тоді величина (R, R_0) називається регулярною кореляційною парою. Для неї матриця

$$Q = \begin{pmatrix} 1 & q_{12} & \dots & q_{1k} \\ q_{21} & 1 & \dots & q_{2k} \\ \dots & \dots & \dots & \dots \\ q_{k1} & q_{k2} & \dots & 1 \end{pmatrix},$$

для якої $q_{ij} = q_{ji} = \frac{r_i}{r_j} (i < j)$, називається нейтральною. Величина q_{ij} є нейтральним значенням коефіцієнта кореляції $r_{ij} = r(X_i, X_j)$.

Якщо (R, R_0) — регулярна кореляційна пара і для довільної кількості факторів k виконується рівність $r_{ij} = q_{ij}$, то коефіцієнт множинної кореляції $R = r_k$, де $r_k = \max\{r_i\}$. У цьому разі всі факторні змінні X_1, X_2, \dots, X_{k-1} не надають додаткової інформації щодо результативної змінної порівняно зі змінною X_k .

Якщо для всіх елементів матриці Q виконується $r_{ij}^2 < q_{ij}^2$, множинний коефіцієнт кореляції перевищуватиме максимальний коефіцієнт r_k , тобто факторна змінна додає певну кількість інформації про результативну змінну.

Якщо виконується протилежна нерівність $r_{ij}^2 > q_{ij}^2$, квадрат коефіцієнта кореляції (коефіцієнт детермінації) також зростає порівняно з максимальним значенням простого коефіцієнта кореляції r_k , однак це зростання пов'язане з ефектом каталізу.

Отже, ефект каталізу виникає тоді, коли

або
$$r_{ij} < 0$$

$$r_{ij} > \frac{r_i}{r_j}.$$

Змінна X_i називається каталітичною змінною, або каталізатором.

Для виявлення в економетричній моделі ефекту каталізу застосовують такий показник:

$$U_l = R_l^2 - H_l,$$

де l – номер комбінації незалежних змінних; H_l – інтегральний показник інформаційної місткості l -ї комбінації змінних.

Якщо значення цього показника додатне, виникає підозра про наявність ефекту каталізу. Для її підтвердження чи спростування потрібно застосовувати статистичний критерій.

Приклад 3.6 [26]. Вивчаються коливання обсягу продажу готельних послуг. Потенційними незалежними змінними можуть бути X_1 – зайнятість, X_2 – середня ціна одного місця в готелі, X_3 – кількість місць у готелі. Задано вектор коефіцієнтів кореляції R_0 і кореляційну матрицю R :

$$R_0 = \begin{pmatrix} 0,43 \\ 0,50 \\ 0,79 \end{pmatrix}, \quad R = \begin{pmatrix} 1 & 0,25 & 0,74 \\ 0,25 & 1 & 0,40 \\ 0,74 & 0,40 & 1 \end{pmatrix}.$$

Величина (R, R_0) є регулярною кореляційною парою. Обчислюємо нейтральні значення коефіцієнтів кореляції $q_{ij} = q_{ji} = \frac{r_i}{r_j}$:

$$q_{12} = q_{21} = \frac{r_1}{r_2} = \frac{0,43}{0,50} = 0,86; \quad q_{13} = q_{31} = \frac{r_1}{r_3} = \frac{0,43}{0,79} = 0,54;$$

$$q_{23} = q_{32} = \frac{r_2}{r_3} = \frac{0,50}{0,79} = 0,63.$$

Отже, нейтральна матриця набирає вигляду

$$Q = \begin{pmatrix} 1 & 0,86 & 0,54 \\ 0,86 & 1 & 0,63 \\ 0,54 & 0,63 & 1 \end{pmatrix}.$$

Розглянемо лише дві комбінації потенційних пояснювальних змінних, а саме $C_1 = (X_1, X_3)$, і $C_2 = (X_2, X_3)$. Для першої отримаємо

$$R_{01} = \begin{pmatrix} 0,43 \\ 0,79 \end{pmatrix}, \quad R_1 = \begin{pmatrix} 1 & 0,74 \\ 0,74 & 1 \end{pmatrix}, \quad r_{13} = 0,74; \quad q_{13} = 0,54.$$

Оскільки $r_{13} > q_{13} = \frac{r_1}{r_3}$, то в разі поєднання змінних (X_1, X_3) спостерігається ефект каталізу. Змінна X_1 є каталітичною змінною.

Коефіцієнт множинної кореляції $R_1 = 0,856$, інтегральний показник інформаційної місткості $H_1 = 0,465$. Ступінь ефекту каталізу обчислюємо за формулою $U_1 = R_1^2 - H_1 = (0,856)^2 - 0,465 = 0,268$. Отримане значення суттєво відрізняється від нуля, а це означає, що велике значення множинного коефіцієнта кореляції за такої комбінації факторів зумовлене ефектом каталізу, а саме сильною кореляцією між змінними X_1 і X_3 .

Для другої комбінації факторів

$$R_{02} = \begin{pmatrix} 0,50 \\ 0,79 \end{pmatrix}, R_2 = \begin{pmatrix} 1 & 0,40 \\ 0,40 & 1 \end{pmatrix}, r_{23} = 0,40; q_{13} = 0,63.$$

Оскільки $r_{23} < q_{23} = \frac{r_2}{r_3}$, немає підстав припускати, що за такої комбінації факторів виникає ефект каталізу.

Коефіцієнт множинної кореляції $R_1 = 0,815$; інтегральний показник інформаційної місткості $H_2 = 0,624$. Ступінь ефекту каталізу $U_2 = R_2^2 - H_2 = (0,815)^2 - 0,624 = 0,040$. Досить мале значення показника U_2 підтверджує зроблений раніше висновок про те, що в моделі зі змінними X_2, X_3 відсутній ефект каталізу, а кожна незалежна змінна відображає певну кількість інформації щодо результативної змінної.

3.2. Процедури побудови моделей

У процедурах побудови економетричної моделі реалізовано ідею одночасного добору факторних змінних і оцінювання параметрів моделі. Під час виконання цих процедур остаточна форма моделі визначається поетапним "поліпшенням" її попередніх версій.

Послідовні процедури побудови моделі можна поділити на дві групи: процедури виключення і процедури селекції. У перших початковою точкою є модель, що містить найбільшу кількість потенційних змінних. У результаті певних перетворень кількість змінних поступово скорочується, допоки модель не відповідатиме заданим критеріям. Такими є метод побудови моделі з коінциденцією, а також процедура виключення *a posteriori*.

Порядок побудови моделей за процедурами селекції протилежний до попередніх методів. У них початковою є модель, яка містить лише одну змінну, а на наступних етапах до моделі поступово вводять нові змінні до отримання задовільної версії моделі. Зокрема, таким є метод покрокової регресії.

3.2.1. Побудова моделі з коінциденцією

Означення. Лінійна економетрична модель має властивість коінциденції, якщо для всіх її змінних параметри a_j , $j = 1, 2, \dots, m$, мають такі самі знаки, як і коефіцієнти кореляції між залежною та відповідною незалежною змінними.

Одним із способів побудови моделі, що має таку властивість, є метод виключення. Початковою точкою методу є модель, що містить усі можливі потенційні змінні, параметризована за методом найменших квадратів. Для всіх параметрів перевіряють умову коінциденції. Змінні, що не мають цієї властивості, виводять з моделі. Параметри скороченої моделі повторно оцінюють за методом найменших квадратів і знову перевіряють на коінцидентність. Перетворення повторюють до отримання рівняння, в якому всі змінні є коінцидентними.

Інший спосіб побудови моделі з коінциденцією полягає в тому, що модель містить усі потенційні змінні, жодна з яких не може бути виведена з моделі.

На основі вектора R_0 коефіцієнтів кореляції між залежною й незалежними змінними моделі будують універсальну матрицю G з елементами

$$g_{ij} = \begin{cases} 1, & i = j, \\ r_i r_j, & i \neq j. \end{cases}$$

На підставі цієї матриці обчислюють параметри моделі з усіма стандартизованими даними

$$Y = b_1 X_1 + b_2 X_2 + \dots + b_m X_m$$

за формулою

$$b = G^{-1} R_0,$$

причому всі параметри задовольняють нерівність $b_j > 0$, $j = 1, 2, \dots, m$.

Ваги змінних X_1, X_2, \dots, X_m розраховують за формулою

$$p_j = 2b_j r_j - \sum_{i=1}^m r_{ij} b_i b_j.$$

Ці ваги можна використати для вибору змінних, що мають виконувати роль факторних: змінні з “великою” вагою вводять до моделі, а з “малою” — виводять.

Приклад 3.7 [26]. Припустимо, що національний дохід країни залежить від таких величин: X_1 — споживання електроенергії, X_2 — споживання сталі, X_3 — споживання мінеральних добрив, X_4 — обсягу

експорту, X_5 – зайнятості у промисловості, X_6 – обсягу вантажоперевезень.

Вектор і матриця коефіцієнтів кореляції відповідно мають такі значення:

$$R_0 = \begin{pmatrix} 0,9007 \\ 0,8579 \\ 0,8393 \\ 0,7775 \\ 0,8187 \\ 0,8133 \end{pmatrix}, R = \begin{pmatrix} 1 & 0,9277 & 0,9126 & 0,8830 & 0,9190 & 0,9106 \\ 0,9277 & 1 & 0,8649 & 0,8483 & 0,8036 & 0,8735 \\ 0,9126 & 0,8649 & 1 & 0,8711 & 0,8000 & 0,8393 \\ 0,8830 & 0,8483 & 0,8711 & 1 & 0,7972 & 0,8529 \\ 0,9190 & 0,8036 & 0,8000 & 0,7972 & 1 & 0,8208 \\ 0,9106 & 0,8735 & 0,8393 & 0,8529 & 0,8208 & 1 \end{pmatrix}.$$

Економетрична модель, що містить усі перелічені потенційні змінні й побудована на основі стандартизованих змінних із застосуванням звичайного МНК, має вигляд

$$\hat{Y} = 0,740X_1 + 0,186X_2 + 0,147X_3 - 0,138X_4 + 0,200X_5 - 0,046X_6.$$

Змінні X_4, X_6 не відповідають принципу коінциденції, оскільки

$$r_4 = 0,7775 > 0, a_4 = -0,138 < 0,$$

$$r_6 = 0,8133 > 0, a_6 = -0,046 < 0.$$

Обидві виводяться з моделі. Після оцінювання параметрів скороченої моделі отримуємо

$$\hat{Y} = 0,6657X_1 + 0,1524X_2 + 0,0901X_3 + 0,0124X_5.$$

У новій моделі всі змінні коінцидентні й модель є остаточною.

Якщо з певних міркувань кількість незалежних змінних скорочувати не можна, модель потрібно будувати іншим способом, а саме із застосуванням універсальної матриці G .

Приклад 3.8 [26]. За даними попереднього прикладу побудуємо модель, що містить усі потенційні змінні, і при цьому всі змінні мають властивість коінциденції. Для цього складемо універсальну матрицю G на підставі коефіцієнтів кореляції, записаних у векторі R_0 . Отримаємо

$$G = \begin{pmatrix} 1 & 0,7727 & 0,7560 & 0,7003 & 0,7374 & 0,7326 \\ 0,7727 & 1 & 0,7200 & 0,6670 & 0,7024 & 0,6977 \\ 0,7560 & 0,7200 & 1 & 0,6526 & 0,6872 & 0,6826 \\ 0,7003 & 0,6670 & 0,6526 & 1 & 0,6366 & 0,7775 \\ 0,7374 & 0,7024 & 0,6872 & 0,6366 & 1 & 0,6639 \\ 0,7326 & 0,6977 & 0,6826 & 0,7775 & 0,6659 & 1 \end{pmatrix}.$$

Обернена до G матриця

$$G^{-1} = \begin{pmatrix} 3,874 & -0,970 & -0,848 & -0,587 & -0,742 & -0,718 \\ -0,970 & 3,128 & -0,578 & -0,400 & -0,505 & -0,489 \\ -0,848 & -0,578 & 2,879 & -0,349 & -0,469 & -0,427 \\ -0,587 & -0,400 & -0,349 & 2,287 & -0,305 & -0,296 \\ -0,742 & -0,505 & -0,469 & -0,305 & 2,647 & -0,373 \\ -0,718 & -0,488 & -0,427 & -0,296 & -0,373 & 2,593 \end{pmatrix}.$$

Тоді вектор значень структурних параметрів моделі, побудованої на основі стандартизованих даних,

$$b = G^{-1}R_0 = \begin{pmatrix} 0,2985 \\ 0,2032 \\ 0,1777 \\ 0,1230 \\ 0,1554 \\ 0,1503 \end{pmatrix},$$

а сама модель має вигляд

$$\hat{Y} = 0,2985X_1 + 0,2032X_2 + 0,1777X_3 + 0,123X_4 + 0,1554X_5 + 0,1503X_6.$$

3.2.2. Процедура виключення *a posteriori*

Процедура складається з таких етапів.

1. Будуємо модель, що містить усі потенційні незалежні змінні.
2. Для кожної потенційної змінної розраховуємо значення t -статистики:

$$t_{\text{експ}}^j = \frac{|a_j|}{S_{a_j}}, \quad j = 1, 2, \dots, m.$$

3. Найменше із цих значень $t_h = \min_j \{t_{\text{експ}}^j\}$ порівнюємо з критичним значенням розподілу Стьюдента $t_{\text{табл}}$ для заданого рівня значущості α двосторонньої критичної області і $(n - m - 1)$ степенів вільності, $t_{\text{табл}} = t(\alpha, n - m - 1)$. Якщо $t_h \leq t_{\text{табл}}$, змінну X_h потрібно вивести з моделі. Оцінюємо параметри моделі після виключення змінної і переходимо до етапу 2. Якщо $t_h > t_{\text{табл}}$, потрібно приймати модель з усіма змінними, що входять до моделі.

Приклад 3.9 [26]. Нехай за даними 13 спостережень побудовано модель, що містить чотири потенційні змінні й має вигляд

$$\hat{Y} = 62,405 + 1,551X_1 + 0,510X_2 + 0,102X_3 - 0,144X_4.$$

Стандартні похибки її параметрів мають значення $S_{a_1} = 0,745$; $S_{a_2} = 0,724$; $S_{a_3} = 0,755$; $S_{a_4} = 0,709$.

Тоді експериментальні значення t -статистик будуть такі:

$$t_1 = \frac{1,551}{0,745} = 2,08, \quad t_2 = \frac{0,510}{0,724} = 0,704,$$

$$t_3 = \frac{0,102}{0,755} = 0,135, \quad t_4 = \frac{0,144}{0,709} = 0,203.$$

Найменшою серед них є $t_3 = 0,135$. Для заданої кількості даних $n = 13$ і кількості оцінених параметрів $m + 1 = 5$, а також заданого рівня значущості $\alpha = 0,05$ $t_{\text{табл}} = t(0,05; 13 - 5) = 2,306$. Оскільки $t_3 < t_{\text{табл}}$, X_3 потрібно виключити з моделі.

Після повторного оцінювання параметрів моделі, яка містить лише змінні X_1, X_2, X_4 , отримаємо

$$\hat{Y} = 71,648 + 1,452X_1 + 0,416X_2 - 0,237X_4.$$

Стандартні похибки параметрів цієї моделі мають значення $S_{a_1} = 0,117$; $S_{a_2} = 0,186$; $S_{a_4} = 0,173$. При цьому експериментальні значення t -статистик такі: $t_1 = 12,410$; $t_2 = 2,237$; $t_4 = 1,370$. При зменшенні кількості змінних $t_{\text{табл}} = t(0,05; 13 - 4) = 2,262$. Оскільки лише $t_4 < t_{\text{табл}}$, з моделі потрібно вивести змінну X_4 .

Для моделі, що містить лише дві змінні X_1, X_2 , отримаємо

$$\hat{Y} = 52,577 + 1,468X_1 + 0,662X_2 \text{ і } S_{a_1} = 0,121; S_{a_2} = 0,046.$$

Тоді t -статистики дорівнюють: $t_1 = 12,132$; $t_2 = 14,391$; $t_{\text{табл}} = t(0,05; 13 - 3) = 2,228$. Оскільки обидва значення експериментальних статистик більші за відповідне табличне значення, це означає, що побудовано остаточний варіант моделі.

3.2.3. Процедура покрокової селекції

Процедуру покрокової селекції виконуємо у такому порядку:

1. Обчислюємо коефіцієнти кореляції між залежною змінною Y і всіма незалежними змінними X_1, X_2, \dots, X_m :

$$r_j = r(Y, X_j) \quad (j = 1, 2, \dots, m)$$

і вибираємо змінну, якій відповідає найбільше за абсолютною величиною значення коефіцієнта кореляції. Її визнають першою факторною змінною.

Зауваження. Якщо існує кілька змінних з однаковими за модулем значеннями коефіцієнтів кореляції, всі вони одночасно вважаються факторними змінними.

2. Будуємо модель у формі

$$Y = a_0 + a_1 X_k + \varepsilon_k,$$

а також допоміжні моделі

$$X_h = a_{0h} + a_{1h} X_k + \varepsilon_k, \quad (h = 1, 2, \dots, m, h \neq k).$$

Параметри всіх моделей оцінюємо за методом найменших квадратів і обчислюємо залишки всіх моделей.

3. Обчислюємо парні коефіцієнти кореляції між залишками першої моделі й залишками кожної з допоміжних моделей:

$$r_{kh} = r(E_k, E_h) \quad (h = 1, 2, \dots, m, h \neq k).$$

4. Вибираємо змінну, якій відповідає найбільше значення коефіцієнта r_{kh} , наприклад нехай це змінна X_l . Її вводимо до моделі, тобто оцінюємо параметри моделі

$$Y = a_0 + a_1 X_k + a_2 X_l + \varepsilon_{kl} \quad (*)$$

і перевіряємо гіпотезу про істотність її параметрів. Якщо параметри при незалежних змінних виявляються істотно відмінними від нуля, знову будуємо допоміжні регресії, які описують залежність кожної з потенційних пояснювальних змінних від змінних, що увійшли до моделі. Переходимо до кроку 3.

Якщо ці параметри істотно від нуля не відрізняються, відповідні змінні виводимо з моделі і шукаємо нові змінні, тобто першою вводимо до моделі змінну, що має найвищий коефіцієнт кореляції:

$$r_j = r(Y, X_j)$$

за винятком змінних, які вже випробувались. Розрахунки повторюємо з кроку 2.

Якщо лише один з параметрів моделі (*) істотно відмінний від нуля, то з моделі виводимо змінну, параметр якої виявився несуттєвим. Тоді допоміжні регресії будуємо для змінної, що залишилась в основній моделі, а також для інших потенційних змінних, крім тієї, яку щойно вивели з моделі.

Алгоритм повторюємо доти, доки жодну з потенційних змінних не можна буде ввести до моделі й жодну з тих, що увійшли до моделі, вивести з неї.

Приклад 3.10. Нехай за даними, що наведені в попередньому прикладі, отримано парні коефіцієнти кореляції

$$r_1 = 0,731; r_2 = 0,816; r_3 = -0,535; r_4 = -0,821.$$

Оскільки змінна X_4 найсильніше корелює із залежною змінною (має найвищий коефіцієнт кореляції), її насамперед визнають першою факторною змінною. Побудована модель парної залежності має вигляд

$$Y = 117,567 - 0,7382X_4,$$

а стандартна похибка коефіцієнта регресії дорівнює 0,1546.

Допоміжні моделі змінної X_4 відносно інших змінних мають вигляд

$$\hat{X}_1 = 10,0505 - 0,08626X_4,$$

$$\hat{X}_2 = 75,2894 - 0,90452X_4,$$

$$\hat{X}_3 = 11,4302 - 0,01130X_4.$$

Для всіх моделей обчислюємо вектори залишків:

$$E_4 = Y - (9117,567 - 0,7382X_4),$$

$$\hat{E}_1 = \hat{X}_1 - (10,0505 - 0,08626X_4),$$

$$\hat{E}_2 = \hat{X}_2 - (75,2894 - 0,90452X_4),$$

$$\hat{E}_3 = \hat{X}_3 - (11,4302 - 0,01130X_4)$$

і коефіцієнти кореляції між змінною E_4 та іншими змінними залишків. Отримуємо

$$\hat{r}_{41} = 0,9568, \quad \hat{r}_{42} = 0,1302, \quad \hat{r}_{43} = -0,8951.$$

Найбільшим серед цих коефіцієнтів є $\hat{r}_{41} = 0,9568$, отже, наступною факторною змінною має бути X_1 .

Модель, що описує змінювання результату і містить дві факторні змінні, має вигляд

$$Y = 103,097 + 1,440 X_1 - 0,614X_4.$$

Після перевірки значущості параметрів цієї моделі виявилось, що параметри при факторних змінних є значущими.

Наступним кроком є побудова допоміжних моделей залежності факторних змінних, які не увійшли до моделі, від уведених до моделі, тобто моделей

$$\hat{X}_2 = a_{02} + a_{12}X_1 + a_{42}X_4,$$

$$\hat{X}_3 = a_{03} + a_{13}X_1 + a_{43}X_4.$$

Після обчислення залишків останніх трьох моделей і коефіцієнтів кореляції між змінними залишків основної моделі й обох допоміжних отримуємо $\hat{r}_2 = 0,5986$, $\hat{r}_3 = 0,5657$. Отже, наступною факторною змінною визнаємо X_2 .

Модель, яка описує залежність результативної змінної від трьох факторних змінних, набирає вигляду

$$Y = 71,648 + 1,452X_1 + 0,416X_2 - 0,237X_4.$$

Після перевірки значущості її параметрів виявляється, що параметр, який відповідає X_4 , є незначущим. Отже, цю змінну виводимо з моделі.

Модель залежності Y від змінних X_1, X_2 має вигляд

$$Y = 52,577 + 1,468X_1 + 0,662X_2,$$

а допоміжні до неї

$$\hat{X}_3 = a_{03} + a_{13}X_1 + a_{23}X_2,$$

$$\hat{X}_4 = a_{04} + a_{14}X_1 + a_{24}X_2.$$

Квадрати коефіцієнтів кореляції змінних залишків основної і допоміжних моделей такі:

$$\hat{r}_3^2 = 0,1691, \quad \hat{r}_4^2 = 0,1715.$$

Оскільки $\hat{r}_4^2 = 0,1715$, це означає, що кандидатом на введення знову є змінна X_4 , яка щойно була вилучена з моделі. Отже, процес побудови моделі покроковим методом завершено: лише дві змінні з чотирьох потенційних факторних змінних мають бути введені до моделі.

3.3. Вибір аналітичної форми моделі

Одним з найпоширеніших методів вибору форми залежності між змінними є графічний метод, який детально розглядався на прикладі парних регресій. Саме парне дослідження впливу кожної незалежної змінної на результативну дає змогу вибрати форму залежності для моделей множинної регресії.

Форму рівняння для парної регресії вибирають на підставі візуальної оцінки розкиду емпіричних точок $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, що відповідають результатам спостережень двох змінних — незалежної X і залежної Y .

Пропонований метод можна застосувати для побудови моделей множинної регресії. У цьому разі будують n графіків, що відобража-

ють залежність Y від кожної незалежної змінної. Визначають аналітичні форми зв'язку між ними.

Загальну економічну модель будують як суму парних функцій регресії:

$$Y = f_1(X_1) + f_2(X_2) + \dots + f_k(X_r) + a_0.$$

Приклад 3.11 [26]. Визначити аналітичну форму залежності продуктивності праці Y від трудового стажу X_1 і віку працівника X_2 на підставі двадцяти спостережень за зазначеними показниками.

t	Y	X_1	X_2
1	84	1	19
2	92	2	23
3	80	0,5	21
4	85	2	23
5	94	2,5	25
6	89	3	21
7	113	9	35
8	118	9,5	31
9	111	3,5	25
10	102	4	25
11	110	6	47
12	102	5	49
13	108	7,5	48
14	112	8,5	46
15	113	7	28
16	115	8	27
17	105	7	45
18	116	11	43
19	121	11	40
20	122	12	35

Розміщення точок (x_{i1}, y_i) , що визначають залежність продуктивності праці від трудового стажу, свідчить про те, що можна застосувати функцію логарифмічної залежності

$$Y = a_{01} + a_{11} \log X_1,$$

а точки (x_{i2}, y_i) більше відповідають квадратичній залежності між продуктивністю праці та віком працюючих, тобто

$$Y = a_{02} + a_{12} X_2 + a_{32} X_2^2.$$

Сукупна залежність результативної змінної від обох факторних змінних має вигляд

$$Y = a_0 + a_{11} \log X_1 + a_{12} X_2 + a_{32} X_2^2.$$

Така модель є нелінійною відносно незалежних змінних і водночас лінійною відносно структурних параметрів рівняння. Після введення позначень

$$Z_1 = \log X_1, Z_2 = X_2, Z_3 = X_2^2$$

її можна розглядати як загальну лінійну модель

$$Y = a_0 + a_1 Z_1 + a_2 Z_2 + a_3 Z_3.$$

Добір незалежних змінних для моделі, нелінійної відносно цих змінних і водночас лінійної відносно параметрів, здійснюють на основі вектора коефіцієнтів кореляції результативної змінної з лінійними перетвореннями незалежних змінних і матриці коефіцієнтів кореляції між цими лінійними перетвореннями, причому будь-якій незалежній змінній може відповідати одне або кілька лінійних перетворень. Зокрема, степенева економетрична модель

$$Y = a_0 X_1^{a_1} X_2^{a_2} \dots X_m^{a_m},$$

після логарифмування набирає вигляду

$$Y = a_0 + a_1 Z_1 + a_2 Z_2 + \dots + a_m Z_m,$$

де $V = \log Y$, $a_0 = \log a_0$, $Z_i = \log X_i$, $i = 1, 2, \dots, m$.

У цьому разі незалежні змінні добирають на основі вектора коефіцієнтів кореляції між логарифмами результативної змінної й логарифмами факторних змінних, а також матриці коефіцієнтів кореляції між логарифмами незалежних змінних.

Висновки

1. Для побудови моделі множинної регресії фактори мають бути кількісно вимірними. Якщо до моделі потрібно включити якісний фактор, який не має кількісної міри, то йому потрібно надати кількісної визначеності (наприклад, задати якість у вигляді балів, проранжувати дані або визначити наявність чи відсутність якісної ознаки за допомогою бінарних змінних, тобто змінних, що набувають двох значень – 0 і 1).

2. Незалежні змінні економетричної моделі повинні мати такі властивості:

- високу варіабельність;
- сильно корелювати з результативною змінною;
- слабо корелювати між собою;

- сильно корелювати зі змінними, які не використовуються в моделі як пояснювальні змінні, але пов'язані з результативною змінною.

3. Факторні (незалежні) змінні добираються за допомогою статистичних методів. Процедура вибору складається з таких кроків:

- 1) на підставі нагромаджених знань складається множина так званих потенційних пояснювальних змінних (первісних змінних), до якої вносять усі найважливіші показники, що впливають на результативну змінну;
- 2) формується вектор спостережень залежної змінної і матриця спостережень усіх незалежних змінних;
- 3) змінні, що мають низький рівень варіабельності (змінюваності), виключаються з розгляду як такі, що надають мінімальну інформацію;
- 4) на основі коефіцієнтів кореляції між усіма змінними, що розглядаються, серед факторних змінних вибирають ті, які сильно корелюють із залежною змінною і водночас слабо корелюють між собою;
- 5) модель досліджується на наявність ефекту каталізу, який виявляється в тому, що високий рівень коефіцієнта множинної кореляції зумовлений сильною кореляцією факторних змінних між собою, а не сильною кореляцією результативної змінної з факторними. Факторні змінні, що спричинюють ефект каталізу, вилучають з моделі.

4. У процедурах побудови економетричної моделі реалізовано ідею одночасного добору пояснювальних змінних і оцінювання параметрів моделі. При виконанні цих процедур остаточна форма моделі визначається поетапним “поліпшенням” її попередніх версій.

5. Послідовні процедури побудови моделі можна поділити на дві групи: процедури виключення і процедури селекції.

6. У процедурах виключення початковою є модель, що містить найбільшу кількість потенційних змінних. У результаті певних перетворень кількість змінних поступово скорочується, аж поки модель не відповідатиме заданим критеріям. Одним із таких методів є метод побудови моделі з коінциденцією, а також процедура виключення *a posteriori*.

7. Порядок побудови моделей за процедурами селекції протилежний до попередніх методів. У них початковою є модель, яка містить лише одну змінну, а на наступних етапах до моделі поступово вводять

нові змінні до отримання задовільної версії моделі. Зокрема, таким є метод покрокової регресії.

8. Одним із найпоширеніших методів вибору форми залежності між змінними є графічний метод, який застосовують для парних регресій. Ідея методу полягає в тому, щоб за розташуванням точок кореляційного поля обирати ту чи іншу аналітичну форму залежності між показниками.

9. У разі множинної регресії будують n графіків, що відображають залежність результативного показника від кожної незалежної змінної. Після перетворення нелінійних залежностей до лінійного за параметрами вигляду будується загальна модель залежності результативного показника від усіх факторних змінних, що складається із суми парних моделей.

ПАРАМЕТРИЗАЦІЯ ТА ДОСЛІДЖЕННЯ МНОЖИННОЇ РЕГРЕСІЇ

4.1. Постановка задачі. Вибір інструментальних засобів для її розв'язання

У цьому розділі детально розглядається алгоритм параметризації та дослідження багатofакторної регресійної моделі на конкретному прикладі.

Задача. Розглянемо задачу дослідження впливу на економічний показник Y п'яти факторів X_1, X_2, X_3, X_4, X_5 , а саме досліджуватимемо залежність фінансового результату підприємства, яке реалізує продукцію, що належить до категорії продовольчих товарів (Y), від рівня вартості чистих активів (X_1), власного капіталу (X_2), запасів (X_3), готівки (X_4) і поточних зобов'язань (X_5).

Розв'язання. Необхідно виконати такі дії:

1) ідентифікувати вид моделі, яка описує залежність фінансового результату від рівня вартості чистих активів, власного капіталу, запасів, готівки і поточних зобов'язань (тобто визначити вигляд відповідної регресійної моделі); для цього необхідно перевірити фактори на мультіколінеарність;

2) оцінити параметри відповідної моделі;

3) провести повне її економетричне дослідження:

- обчислити відносну похибку розрахункових значень залежної змінної;
- розрахувати середньоквадратичну похибку дисперсії залишків;
- обчислити коефіцієнт детермінації;
- знайти значення вибіркового коефіцієнта кореляції і перевірити його значущість;
- перевірити адекватність моделі за критерієм Фішера;
- перевірити значущість окремих параметрів регресії та їх незміщеність;

- визначити граничний внесок кожного з факторів у значення загального коефіцієнта детермінації;
- обчислити скориговані коефіцієнти детермінації;
- для кожного із факторів обчислити коефіцієнти еластичності;
- дослідити залишки моделі на гетероскедастичність і автокореляцію і в разі потреби переоцінити параметри моделі за узагальненим МНК;

4) дослідити динамічні властивості факторів, тобто з'ясувати закономірність змінювання в часі кожної з факторних змінних;

5) розрахувати точковий прогноз індивідуального значення фінансового результату, обравши для прогнозного періоду вектор значень рівня вартості чистих активів, власного капіталу, запасів, готівки і поточних зобов'язань за наступний період: $X'_{\text{пр}} = (1, X1_{\text{пр}}, X2_{\text{пр}}, X3_{\text{пр}}, X4_{\text{пр}}, X5_{\text{пр}})$;

6) визначити інтервальний прогноз фінансового результату.

Статистичну інформацію, що використовувалась для дослідження, подамо у вигляді табл. 4.1.

Таблиця 4.1

Вихідні дані, тис. грн.

Дата	Y	X1	X2	X3	X4	X5
4 вересня	44,3	1690,8	241,3	122,1	74,4	336
4 жовтня	42,1	1730,2	241,3	123,5	142,6	784
4 листопада	46,8	1940,1	241,3	124,4	37,2	168
4 грудня	38,9	1560,3	269,4	138,9	80,6	470,4
5 січня	43,7	1640,1	269,1	136,4	161,2	694,4
5 лютого	45,9	1640,8	269,1	132,1	31	201,6
5 березня	48,4	1630,9	270,5	129,4	198,4	728
5 квітня	44,2	1610,2	270,9	111,5	173,6	739,2
5 травня	49,7	1610,2	272,3	118,7	99,2	224
5 червня	48,1	1600,3	274,5	122,1	130,2	392
5 липня	50,1	1440,5	274,5	124,5	49,6	257,6
5 серпня	50,7	1440,5	238,1	129,6	55,8	156,8
5 вересня	52,4	1430,2	236,2	144,4	62	336
5 жовтня	48,9	1470,2	241,1	146,2	68,2	280
5 листопада	56,3	1420,4	241,1	142,1	86,8	336
5 грудня	54,4	1420,1	249,7	136,5	117,8	392

Зауваження 1. Джерелом інформаційної бази даних є форма бухгалтерського обліку № 1 “Баланс” та форма бухгалтерського обліку № 2 “Звіт про фінансовий результат” за наведений період часу (з вересня 2004 року по грудень 2005).

Зауваження 2. Необхідні розрахунки проведено в середовищі пакету електронних таблиць Excel, оскільки він є стандартним і використовується практично в кожному навчальному закладі.

Після того як вибрано програмне середовище, перейдемо до дослідження поведінки факторів у часі.

4.2. Специфікація моделі

Задача полягає в дослідженні сукупного впливу всіх факторів (рівня вартості чистих активів, власного капіталу, запасів, готівки і поточних зобов'язань) на фінансовий результат, тобто необхідно визначити вигляд відповідної регресійної моделі або її специфікацію (побудова багатофакторної регресії).

Специфікація моделі — це її аналітична форма. Вона складається з певного виду функції чи функцій, а також має ймовірнісні характеристики, які притаманні стохастичним залишкам моделі.

Припустимо, що між економічним показником Y і кожною факторною змінною X_1, X_2, X_3, X_4 та X_5 існує лінійний зв'язок.

Запишемо рівняння регресії у вигляді

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + u, \quad (4.1)$$

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5, \quad (4.2)$$

де y, \hat{y} — відповідно фактичні та розрахункові значення фінансового результату; x_1, x_2, x_3, x_4, x_5 — рівень вартості відповідно чистих активів, власного капіталу, запасів, готівки та поточних зобов'язань; $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ — параметри моделі, які потрібно оцінити; $b_0, b_1, b_2, b_3, b_4, b_5$ — їхні оцінки; u — стохастична складова.

4.3. Перевірка факторів моделі на мультиколінеарність

В економетричних дослідженнях важливо з'ясувати, чи існують між факторами моделі взаємозв'язки, які називають мультиколінеарністю.

Адже від наявності чи, навпаки, від відсутності мультиколінеарності залежить специфікація моделі й вибір методу оцінювання параметрів моделі.

Мультиколінеарність означає існування щільної лінійної залежності (сильної кореляції) між двома або більше факторними змінними моделі. Наявність мультиколінеарності призводить до зміщення оцінок параметрів.

Перевіримо нашу модель на мультиколінеарність. Для цього застосуємо *алгоритм Феррара–Глобера* (див. п. 7.7.2). Обчислення за цим алгоритмом виконаємо по кроках.

Крок 1. Нормалізуємо змінні x_1, x_2, x_3, x_4, x_5 економетричної моделі:

$$x_{ij}^* = \frac{(x_{ij} - \bar{x}_j)}{\sqrt{n\sigma_{x_j}^2}}, \text{ або } x_{ij}^* = \frac{(x_{ij} - \bar{x}_j)}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}},$$

де $n = 16$ – кількість спостережень ($i = 1, 2, \dots, n$); $m = 5$ – кількість незалежних змінних ($j = 1, m$); \bar{x}_j – середня арифметична j -ї незалежної змінної:

$$\bar{x}_1 = 1579,738; \quad \bar{x}_2 = 256,275; \quad \bar{x}_3 = 130,15;$$

$$\bar{x}_4 = 98,0375; \quad \bar{x}_5 = 406;$$

$$\sigma_{x_j}^2 = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2} \text{ – дисперсія } j\text{-ї незалежної змінної:}$$

$$\sigma_{x_1}^2 = 18843; \quad \sigma_{x_2}^2 = 234,2781; \quad \sigma_{x_3}^2 = 92,6138;$$

$$\sigma_{x_4}^2 = 2417,966; \quad \sigma_{x_5}^2 = 43221,92.$$

Крок 2. Будемо нову матрицю X^* , елементами якої є нормалізовані незалежні змінні x_{ij}^* :

$$X^* = \begin{pmatrix} 0,20227 & -0,24459 & -0,20912 & -0,12018 & -0,08418 \\ 0,274027 & -0,24459 & -0,17275 & 0,226561 & 0,454548 \\ 0,656303 & -0,24459 & -0,14937 & -0,3093 & -0,2862 \\ -0,0354 & 0,214375 & 0,227306 & -0,08865 & 0,077442 \\ 0,109934 & 0,209475 & 0,162361 & 0,321125 & 0,346803 \\ 0,111209 & 0,209475 & 0,050657 & -0,34083 & -0,24579 \\ 0,093179 & 0,232341 & -0,01948 & 0,510254 & 0,387208 \\ 0,055479 & 0,238875 & -0,48449 & 0,384168 & 0,400676 \\ 0,055479 & 0,261741 & -0,29745 & 0,00591 & -0,21886 \\ 0,037449 & 0,297675 & -0,20912 & 0,163518 & -0,01684 \end{pmatrix}.$$

$$\begin{pmatrix} -0,25358 & 0,297675 & -0,14677 & -0,24626 & -0,17845 \\ -0,25358 & -0,29686 & -0,01429 & -0,21474 & -0,29966 \\ -0,27234 & -0,32789 & 0,370184 & -0,18322 & -0,08418 \\ -0,19949 & -0,24786 & 0,416944 & -0,1517 & -0,15152 \\ -0,29019 & -0,24786 & 0,310435 & -0,05713 & -0,08418 \\ -0,29074 & -0,10739 & 0,164959 & 0,100475 & -0,01684 \end{pmatrix}$$

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

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

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

$$R = \begin{pmatrix} 1 & 0,063967 & -0,50646 & 0,095057 & 0,177001 \\ 0,063967 & 1 & -0,3959 & 0,378531 & 0,264216 \\ -0,50646 & -0,3959 & 1 & -0,27813 & -0,18253 \\ 0,095057 & 0,378531 & -0,27813 & 1 & 0,88717 \\ 0,177001 & 0,264216 & -0,18253 & 0,88717 & 1 \end{pmatrix}.$$

Однак на основі цієї залежності не можна стверджувати, що отриманий зв'язок є явищем мультиколінеарності.

Зауваження. Якщо діагональні елементи матриці R не дорівнюють одиниці, то на діагоналі цієї матриці ми ставимо одиниці, а до інших елементів додаємо різницю між одиницею і значенням діагонального елемента.

Крок 3:

- визначаємо визначник кореляційної матриці R :

$$|R| = 0,09461;$$

- знаходимо значення критерію χ^2 :

$$\chi^2 = -[n-1 - \frac{1}{6}(2m+5)] \ln|R|; \quad \chi^2 = 29,47487;$$

- порівнюємо отримане значення χ^2 з табличним при $\frac{1}{2}m(m-1) = 10$ степенях вільності і рівні значущості $\alpha = 0,01$ (дод. 2):

$$\chi^2_{\text{табл}} = 23,20925.$$

Оскільки $\chi^2 < \chi^2_{\text{табл}}$, у масиві незалежних змінних не існує мультиколінеарності в сукупності даних.

Крок 4. Визначаємо матрицю похибок:

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

$$C = R^{-1} = \begin{pmatrix} 1,529783 & 0,18732 & 0,916082 & 0,818347 & -0,87907 \\ 0,18732 & 1,351828 & 0,507966 & -0,58331 & 0,21988 \\ 0,916082 & 0,507966 & 1,773538 & 0,890591 & -0,76275 \\ 0,818347 & -0,58331 & 0,890591 & 5,815387 & -4,98741 \\ -0,87907 & 0,21988 & -0,76275 & -4,98741 & 5,38296 \end{pmatrix}.$$

Крок 5:

- розраховуємо F -критерії:

$$F_k = \frac{(c_{kk} - 1) \cdot (n - m)}{(m - 1)}, \quad k = 1, 2, 3, 4, 5,$$

де c_{kk} — діагональні елементи матриці C ;

$$F_1 = 1,456903;$$

$$F_2 = 0,967526;$$

$$F_3 = 2,127229;$$

$$F_4 = 13,24232;$$

$$F_5 = 12,05314;$$

- значення критеріїв F_k порівнюємо з табличним при $(n - m) = 11$ і $(m - 1) = 4$ степенях вільності і рівні значущості $\alpha = 0,01$ (дод. 4):

$$F_{\text{табл}} = 14,45288.$$

Оскільки $F_1 < F_{\text{табл}}$, $F_2 < F_{\text{табл}}$, $F_3 < F_{\text{табл}}$, $F_4 < F_{\text{табл}}$, $F_5 < F_{\text{табл}}$, робимо **висновок**, що не існує залежності кожної незалежної змінної від сукупності інших незалежних змінних.

- розраховуємо коефіцієнти детермінації для кожної змінної:

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

$$R_1^2 = 0,346312; \quad R_2^2 = 0,260261; \quad R_3^2 = 0,436155;$$

$$R_4^2 = 0,828042; \quad R_5^2 = 0,814229.$$

Крок 6. Знаходимо часткові коефіцієнти кореляції, які характеризують щільність зв'язку між двома змінними за умови, що інші змінні $x_{l1}, x_{l2}, \dots, x_{lm}$ не впливають на цей зв'язок (існування парної мультиколінеарності):

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

де c_{kj} – елемент матриці C , розміщений на перетині k -го рядка та j -го стовпця, $k = 1, 2, \dots, m; j = 1, 2, \dots, m$, c_{kk} і c_{jj} – діагональні елементи матриці C ;

$$\begin{aligned} r_{12} &= -0,13026; & r_{13} &= -0,55616; & r_{14} &= -0,27437; & r_{15} &= 0,306336; \\ r_{23} &= -0,32806; & r_{24} &= 0,20804; & r_{25} &= -0,08151; & r_{34} &= -0,27731; \\ r_{35} &= 0,246859; & r_{45} &= 0,891405. \end{aligned}$$

Однак якщо порівняти абсолютні значення часткових і парних коефіцієнтів, то можна виявити, що перші значно менші за останні. Тому на основі знання парних коефіцієнтів кореляції висновок про мультиколінеарність робити неможливо. Для цього необхідно виконати сьомий крок.

Крок 7:

- розраховуємо t -критерії:

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

$$\begin{aligned} t_{12} &= 0,473691; & t_{13} &= 2,41285; & t_{14} &= 1,028726; & t_{15} &= 1,160292; \\ t_{23} &= 1,252137; & t_{24} &= 0,766878; & t_{25} &= 0,294872; & t_{34} &= 1,040678; \\ t_{35} &= 0,91849; & t_{45} &= 7,091697; \end{aligned}$$

- значення критеріїв t_{kj} порівнюємо з табличними при $(n - m) = 11$ степенях вільності і рівні значущості $\alpha = 0,01$ (дод. 3):

$$t_{\text{табл}} = 3,105807.$$

Оскільки $t_{12} < t_{\text{табл}}$, $t_{13} < t_{\text{табл}}$, $t_{14} < t_{\text{табл}}$, $t_{15} < t_{\text{табл}}$, $t_{23} < t_{\text{табл}}$, $t_{24} < t_{\text{табл}}$, $t_{25} < t_{\text{табл}}$, $t_{34} < t_{\text{табл}}$, $t_{35} < t_{\text{табл}}$ то між відповідними парами незалежних

змінних не існує мультиколінеарності. А оскільки $t_{45} > t_{\text{табл}}$, то між змінними x_4 і x_5 мультиколінеарність існує.

Якщо між незалежними змінними існує мультиколінеарність, то оцінки параметрів виявляються зміщеними, отже, необхідно вирішувати питання про її усунення. Наприклад, доцільно вивести одну із змінних мультиколінеарної пари з моделі (якщо це не суперечить логіці економічних зв'язків) і дослідити вплив на показник решти факторів. Позбутися мультиколінеарності можна також перетворенням певним чином пояснювальних змінних моделі (змінити спосіб їх вимірювання і від абсолютних значень перейти до відносних).

Якщо жоден із цих способів не дає змоги позбутися мультиколінеарності, параметри моделі слід оцінювати за методом головних компонент.

Зауваження. Усі перелічені обчислення можна виконати в середовищі пакету Excel, застосовуючи вбудовані функції:

- 1) нормалізація змінних при цьому складається з таких етапів:
 - обчислюють середні значені кожного з факторів \bar{x}_j , $j = 1, 2, 3, 4, 5$ (використовують статистичну функцію СРЗНАЧ(масив));
 - визначають відхилення кожного з факторів від його середнього значення (уводять формулу з клавіатури, фіксуючи адресу середнього значення (клавіша F4), і продовжують формулу у відповідний діапазон комірок);
 - обчислюють суму квадратів отриманих різниць (використовують математичну функцію СУММКВ (вектор-стовпець різниць));
 - з кожної суми добувають корінь квадратний (математична функція КОРЕНЬ(число));
 - ділять кожен елемент матриці, що містить відхилення факторів від своїх середніх значень, на відповідний корінь квадратний (формулу вводять з клавіатури і продовжують у діапазоні комірок); отримана матриця є матрицею нормалізованих змінних;
- 2) для обчислення визначника кореляційної матриці R потрібно:
 - транспонувати матрицю нормалізованих змінних (за допомогою функції ТРАНСП(масив) із категорії “ссылки и массивы”;
 - обчислити добуток $X^{*tr}X^* = R$ (математична функція МУМНОЖ (масив1, масив2));

- обчислити визначник $|R|$ (за допомогою математичної функції МОПРЕД(массив)).

Інші формули ввести з клавіатури.

4.4. Оцінювання параметрів множинної регресійної моделі

- записуємо рівняння регресії у вигляді:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + u,$$

де $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ – параметри моделі, які потрібно оцінити.

Оцінки параметрів моделі $b_0, b_1, b_2, b_3, b_4, b_5$ визначатимемо за методом найменших квадратів. Для цього складемо вектор-стовпець спостережень залежної змінної Y і матрицю спостережень незалежних змінних X :

$$Y = \begin{pmatrix} 44,3 \\ 42,1 \\ 46,8 \\ 38,9 \\ 43,7 \\ 45,9 \\ 48,4 \\ 44,2 \\ 49,7 \\ 48,1 \\ 50,1 \\ 50,7 \\ 52,4 \\ 48,9 \\ 56,3 \\ 54,4 \end{pmatrix}; \quad X = \begin{pmatrix} 1 & 1690,8 & 241,3 & 122,1 & 74,4 & 336 \\ 1 & 1730,2 & 241,3 & 123,5 & 142,6 & 784 \\ 1 & 1940,1 & 241,3 & 124,4 & 37,2 & 168 \\ 1 & 1560,3 & 269,4 & 138,9 & 80,6 & 470,4 \\ 1 & 1640,1 & 269,1 & 136,4 & 161,2 & 694,4 \\ 1 & 1640,8 & 269,1 & 132,1 & 31 & 201,6 \\ 1 & 1630,9 & 270,5 & 129,4 & 198,4 & 728 \\ 1 & 1610,2 & 270,9 & 111,5 & 173,6 & 739,2 \\ 1 & 1610,2 & 272,3 & 118,7 & 99,2 & 224 \\ 1 & 1600,3 & 274,5 & 122,1 & 130,2 & 392 \\ 1 & 1440,5 & 274,5 & 124,5 & 49,6 & 257,6 \\ 1 & 1440,5 & 238,1 & 129,6 & 55,8 & 156,8 \\ 1 & 1430,2 & 236,2 & 144,4 & 62 & 336 \\ 1 & 1470,2 & 241,1 & 146,2 & 68,2 & 280 \\ 1 & 1420,4 & 241,1 & 142,1 & 86,8 & 336 \\ 1 & 1420,1 & 249,7 & 136,5 & 117,8 & 392 \end{pmatrix};$$

- обчислюємо оцінки регресійних коефіцієнтів за формулою

$$B = (X'X)^{-1} X'Y, \quad (4.2)$$

де X' (або X^T) – транспонована матриця X ;

$$X = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1691 & 1730 & 1940 & 1560 & 1640,1 & 1641 & 1631 & 1610 & 1610 & 1600 & 1441 & 1440 & 1430 & 1470 & 1420,4 & 1420,1 \\ 241,3 & 241,3 & 241,3 & 269,4 & 269,1 & 269,1 & 1270,5 & 1270,9 & 1272,3 & 1274,5 & 1274,5 & 1238,1 & 1236,2 & 241,1 & 241,1 & 249,7 \\ 122,1 & 123,5 & 124,4 & 138,9 & 136,4 & 132,1 & 129,4 & 111,5 & 118,7 & 122,1 & 124,5 & 129,6 & 144,4 & 146,2 & 142,1 & 136,5 \\ 74,4 & 142,6 & 37,2 & 80,6 & 161,2 & 31 & 198,4 & 173,6 & 99,2 & 130,2 & 49,6 & 55,8 & 62 & 68,2 & 86,8 & 117,8 \\ 336 & 784 & 168 & 470,4 & 694,4 & 201,6 & 728 & 739,2 & 224 & 392 & 257,6 & 156,8 & 336 & 280 & 336 & 392 \end{pmatrix};$$

$$X^T X = \begin{pmatrix} 16 & 25275,8 & 4100,4 & 2082,4 \\ 25275,8 & 40230617,16 & 6479706 & 3278941 \\ 4100,4 & 6479706,05 & 1054578 & 532734 \\ 2082,4 & 3278940,6 & 532734 & 272506,2 \\ 1568,6 & 2488242,28 & 406551,4 & 202047,5 \\ 6496 & 10342795,68 & 1678215 & 839611,4 \end{pmatrix};$$

$$(X^T X)^{-1} = \begin{pmatrix} 16 & 183,11 & 134,35 & 264,06 \\ 183,11 & 2127,37 & 1550,42 & 3034,296 \\ 134,35 & 1550,42 & 1147,828 & 2224,617 \\ 264,06 & 3034,29 & 2224,617 & 4383,157 \end{pmatrix};$$

$$X^T Y = \begin{pmatrix} 764,9 \\ 1202589 \\ 195683,3 \\ 99770,98 \\ 74304,52 \\ 303942,2 \end{pmatrix} \quad B = \begin{pmatrix} 95,43762 \\ -0,014055 \\ -0,103316 \\ 0,016428 \\ 0,097328 \\ -0,026186 \end{pmatrix}.$$

Зауваження. Для виконання розрахунків за формулою (4.2) можна скористатися вбудованими функціями Excel:

1) транспоновану матрицю X^T сформувати за допомогою функції ТРАНСП(масив) із категорії “ссылки и массивы”;

2) добутки матриць $X^T X$ і $X^T y$ обчислити за допомогою математичної функції МУМНОЖ(масив1, масив2) (Матричное УМНОЖение);

3) обернену матрицю $(X^T X)^{-1}$ визначити за допомогою математичної функції МОБР(масив) (Матрица ОБРатная);

4) остаточний результат – вектор параметрів B – визначити також за допомогою функції МУМНОЖ.

• записуємо функцію регресії з урахуванням знайдених оцінок коефіцієнтів моделі:

$$\hat{y} = 95,438 - 0,014x_1 - 0,103x_2 + 0,016x_3 + 0,097x_4 - 0,026x_5; \quad (4.3)$$

• обчислюємо значення \hat{y}_i за заданими векторами спостережень $\bar{x}_1, \bar{x}_2, \bar{x}_3, \bar{x}_4, \bar{x}_5$ або за формулою $\hat{Y} = XB$;

$$\hat{Y} = \begin{pmatrix} 47,1926 \\ 41,5683 \\ 44,5052 \\ 43,4836 \\ 44,3309 \\ 44,4828 \\ 46,9413 \\ 44,1899 \\ 50,4133 \\ 48,9989 \\ 46,9591 \\ 54,0465 \\ 50,5416 \\ 51,5726 \\ 52,5490 \\ 53,1235 \end{pmatrix}.$$

Отже, ми побудували загальну лінійну модель (4.3) залежності фінансового результату від рівня вартості чистих активів, власного капіталу, запасів, готівки і поточних зобов'язань. Наступним кроком наших досліджень є проведення дисперсійно-кореляційного аналізу та аналізу залишків.

4.5. Дослідження моделі та перевірка її на адекватність

Виконаємо дослідження згідно із загальним алгоритмом побудови і аналізу регресійної моделі, наведеного на початку даного розділу:

• обчислюємо відносну похибку розрахункових значень регресії за формулою

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

$$\delta = \begin{pmatrix} -7,4360 \\ 1,2165 \\ 4,9994 \\ -9,4703 \\ -1,4275 \\ 2,8514 \\ 3,0324 \\ 0,0201 \\ -1,4070 \\ -1,7156 \\ 6,4230 \\ -5,9441 \\ 3,4160 \\ -5,5146 \\ 19,940 \\ 127,64 \end{pmatrix};$$

- обчислюємо середнє значення відносної похибки за формулою

$$\bar{\delta} = \frac{\sum_{i=1}^n \delta_i}{n}; \quad \bar{\delta} = 8,539268;$$

- знаходимо середньоквадратичну похибку дисперсії збурювань:

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

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

$$S = 2,968778;$$

- оцінюємо загальний вплив незалежних змінних на залежну змінну, а саме обчислюємо коефіцієнт детермінації:

$$R^2 = 1 - \frac{\sum_{i=1}^n u_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \text{ або } R^2 = 1 - \frac{Y'Y - B'X'Y}{Y'Y - ny^{-2}} = \frac{B'X'Y - ny^{-2}}{Y'Y - ny^{-2}};$$

маємо

$$R^2 = 0,719551.$$

Висновок. Оскільки коефіцієнт детермінації досить великий, варіація залежної змінної Y значною мірою визначається варіацією незалежних змінних;

- оцінюємо значущість вибіркового коефіцієнта кореляції, а саме обчислюємо коефіцієнт кореляції, який характеризує щільність лінійного зв'язку всіх незалежних факторів x_i із залежною змінною y :

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

$$R = 0,848264.$$

Коефіцієнт кореляції досить великий, тому *лінійний зв'язок усіх незалежних факторів x_1, x_2, x_3, x_4, x_5 із залежною змінною y є досить щільним;*

- обчислюємо t -статистику за формулою

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

$$t = 5,993329;$$

знаходимо табличне значення t -розподілу з $(n - m - 1) = 10$ степенями вільності й рівнем значущості $\alpha = 0,05$ (дод. 3) для двосторонньої критичної області:

$$t_{\text{табл}} = t(\alpha/2, n - m - 1);$$

$$t_{(0,025,10)} = 2,233769327.$$

Оскільки $|t| > t_{\text{табл}}$, можна зробити висновок про достовірність коефіцієнта кореляції, який характеризує щільність зв'язку між залежною і незалежними змінними моделі;

- для вибраного рівня значущості $\alpha = 0,05$ і відповідного степеня вільності $k = n - m - 1 = 10$ записуємо довірчі межі для множинного коефіцієнта кореляції R :

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

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

$$\Delta R = 2,63 \cdot (1 - 0,848264) / 4 = 0,099909614,$$

отже, довірчі межі для множинного коефіцієнта кореляції R :

$$(R - \Delta R; R + \Delta R) = (0,74835407; 0,9481733);$$

- обчислюємо F -статистику за формулою (спрощений варіант для перевірки нульової гіпотези $H_0 : \beta_1 = \beta_2 = \dots = \beta_m = 0$):

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

$$F_{\text{експ}} = 5,131428.$$

Знаходимо табличне значення F -статистики $F_{\text{табл}} = F(\alpha, m, n - m - 1)$ (дод. 4):

$$F_{\text{табл}} = F(0,05; 5; 10) = 3,325827$$

і порівнюємо його з обчисленою F -статистикою.

Оскільки $F_{\text{експ}} > F_{\text{табл}}$, нульову гіпотезу відхиляємо, коефіцієнти регресії є значущими.

Перевірку на значущість коефіцієнтів регресії здійснюємо також за t -тестом. Для цього:

- для кожного параметра b_j ($j = 0, 1, \dots, 5$) обчислюємо t -статистику за формулою

$$t_j = \frac{b_j}{\sqrt{\sigma_u^2 c_{jj}}} = \frac{b_j}{S_{b_j}} \quad (j = 0, 1, \dots, 5),$$

де c_{jj} — діагональний елемент матриці $(X'X)^{-1}$; S_{b_j} — стандартна помилка оцінки параметра моделі:

$$t_0 = 5,7585; \quad t_1 = -3,62118; \quad t_2 = -3,15749; \quad t_3 = 0,275598;$$

$$t_4 = 4,60727; \quad t_5 = -5,4472885.$$

Значення t_j -критерію порівнюємо з табличними при $k = n - m - 1 = 10$ степенях вільності й рівні значущості $\alpha = 0,05$:

$$t_{\text{табл}(0,025,10)} = 2,233769327.$$

Оскільки $|t_0| > t_{\text{табл}}$, $|t_1| > t_{\text{табл}}$, $|t_2| > t_{\text{табл}}$, $|t_3| < t_{\text{табл}}$, $|t_4| > t_{\text{табл}}$, $|t_5| > t_{\text{табл}}$, відповідно оцінки b_0, b_1, b_2, b_4, b_5 не є значущими, а оцінка b_3 є значущою;

- обчислюємо відношення

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

$$\delta_{b_0} = 17\%; \quad \delta_{b_1} = 28\%; \quad \delta_{b_2} = 32\%; \quad \delta_{b_3} = 363\%;$$

$$\delta_{b_4} = 22\%; \quad \delta_{b_5} = 18\%$$

(характеризує той факт, що оцінки $b_0, b_1, b_2, b_3, b_4, b_5$ зміщені);

- розраховуємо граничний вклад j -го регресора в коефіцієнт детермінації (тобто, на яку величину зменшиться частковий коефіцієнт детермінації, якщо j -й регресор буде вилучено з рівняння):

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

де $t_j = \frac{b_j}{\sqrt{\sigma_u^2 c_{jj}}} = \frac{b_j}{S_{b_j}}$; T – період чи обсяг вибірки (вважатимемо, що

$T = n = 16$);

$$\Delta R_0^2 = 0,929967; \quad \Delta R_1^2 = 0,367751; \quad \Delta R_2^2 = 0,2796; \quad \Delta R_3^2 = 0,00213;$$

$$\Delta R_4^2 = 0,595308; \quad \Delta R_5^2 = 0,832174;$$

- розраховуємо скоригований за Тейлом коефіцієнт:

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

$$\bar{R}_T^2 = 0,579327;$$

- визначаємо скоригований коефіцієнт детермінації за Амемією:

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

$$\bar{R}_A^2 = 0,383013.$$

Висновок. Оскільки з вилученням змінної з рівняння втрачається один степінь вільності, з двох варіантів рівнянь, які відрізняються величиною скоригованого коефіцієнта детермінації, але мають однаково задовільні інші критерії якості, перевагу віддають рівнянню з більшим значенням скоригованого коефіцієнта детермінації ($\bar{R}_T^2 = 0,579327$, який відображає втрату степеня вільності при введенні додаткового регресора чіткіше, ніж \bar{R}_A^2 , тобто в цьому разі $\bar{R}_T^2 > \bar{R}_A^2 = 0,383013$).

Обчислюємо коефіцієнти еластичності:

$$\alpha_i = \frac{\partial y}{\partial x_i} \cdot \frac{\bar{x}_i}{\bar{y}} = b_i \cdot \frac{\bar{x}_i}{\bar{y}}.$$

Коефіцієнт еластичності є показником впливу зміни питомої ваги x_i на y у припущенні, що вплив інших факторів відсутній: у нашому прикладі він показує, що *прибуток підприємства зменшиться на 0,45 %*, якщо *рівень вартості чистих активів збільшиться на 1 %*; *зменшиться на 0,55 %*, якщо *рівень власного капіталу зросте на 1 %*;

збільшиться на 0,044 %, якщо рівень запасів зросте на 1 %; збільшиться на 0,19 %, якщо рівень готівки зросте на 1 %; фінансовий зменшиться на 0,207 %, якщо рівень поточних зобов'язань зросте на 1 %.

Загальна еластичність Y від усіх факторів x_i :

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

$$\alpha = -0,972716 \%$$

Загальна еластичність показує, що зменшиться фінансовий результат на 0,97 %, якщо водночас збільшити на 1 % всі фактори: рівень вартості чистих активів, власного капіталу, запасів, готівки та поточних зобов'язань.

Такий висновок свідчить про те, що, можливо, потрібно повернутися до питання специфікації моделі або переоцінити її параметри з урахуванням особливих випадків поведінки стохастичної складової.

4.6. Особливі випадки поведінки залишків моделі: перевірка наявності гетероскедастичності та автокореляції

Почнемо з перевірки гетероскедастичності за допомогою параметричного тесту Гольдфелда–Квандта. Цей тест застосовують, коли сукупність спостережень невелика. У цьому разі розглядають випадок, коли $M(uu^{tr}) = \sigma_u^2 = k^2 x_{ij}^2$, тобто дисперсія залишків зростає пропорційно до квадрата однієї з незалежних змінних заданої моделі.

Тестування гетероскедастичності за допомогою параметричного тесту Гольдфелда–Квандта проводять за фактором, який найсильніше впливає на зміну дисперсії залишків моделі. Якщо важко апіорі визначити такий фактор, необхідно тестувати залишки за кожною незалежною змінною окремо і в кожному випадку застосовувати тест Гольдфелда–Квандта.

Параметричний тест Гольдфелда–Квандта

Припустимо, що зміну дисперсії залишків може спричинити зміна X_1 — рівень вартості чистих активів, і протестуємо за цим фактором.

Крок 1. Спостереження (вихідні дані) ранжуємо в порядку зростання (або спадання) значення відповідно до величини елементів вибраного нами вектора X_1 — рівня вартості чистих активів.

Крок 2. Відкидаємо c спостережень, які перебувають усередині векторів вихідних даних, де

$$c = \frac{4n}{15} = 4,$$

де $n = 16$ кількість елементів вектора X_1 .

Крок 3. Будуємо дві моделі парної регресії на основі звичайного МНК для двох створених сукупностей спостережень обсягом $n_1 = n_2 = \frac{n-c}{2} = 6$, адже $\frac{n-c}{2} \geq m$, де $m = 5$ – кількість змінних початкової моделі:

$$y_1 = 241,6984 - 0,13192x \text{ і } y_2 = 42,09889 + 0,001811x.$$

Крок 4. Знаходимо суму квадратів залишків S_1 і S_2 для першої і другої моделей:

$$S_1 = \sum_{i=1}^{n_1} u_{i1}^2, \quad S_2 = \sum_{i=1}^{n_2} u_{i2}^2,$$

де u_1 і u_2 – залишки відповідно для першої і другої моделей;

$$S_1 = 0,8033, \quad S_2 = 1,3557.$$

Крок 5:

- розраховуємо критерій $F_{\text{експ}} = \frac{S_2}{S_1}$, якщо $S_2 > S_1$, або $F_{\text{експ}} = \frac{S_1}{S_2}$, якщо $S_1 > S_2$ ($F_{\text{експ}} = 1,687625$), який у разі виконання гіпотези про гомоскедастичність відповідатиме F -розподілу з $\gamma_1 = \frac{(n-c-2m_1)}{2} = 4$ $\gamma_2 = \frac{(n-c-2m_2)}{2} = 4$ степенями вільності; $m_1 = m_2 = 2$ – кількість параметрів парних моделей;

- значення критерію $F_{\text{експ}} = 1,687625$ порівнюємо з табличним значенням F -критерію при вибраному довірчому рівні $\alpha = 0,05$ і відповідних степенях вільності (дод. 4):

$$F_{\text{табл}} = F_{(0,05; 4; 4)} = 6,388234.$$

Оскільки $F_{\text{експ}} \leq F_{\text{табл}}$, *гетероскедастичність відсутня.*

Отже, фактор X_1 не спричиняє гетероскедастичності залишків моделі.

Припустимо тепер, що зміну дисперсії залишків може зумовити вектор X_2 (рівень власного капіталу), і протестуємо за цим фактором.

Крок 1. Спостереження (вихідні дані) ранжуємо в порядку зростання (або спадання) значення відповідно до величини елементів вибраного нами вектора X_2 .

Крок 2. Відкидаємо c спостережень, які містяться всередині векторів вихідних даних, де

$$c = \frac{4n}{15} = 4,$$

$n = 16$ – кількість елементів вектора X_2 .

Крок 3. Будуємо дві моделі парної регресії на основі звичайного МНК для двох створених сукупностей спостережень обсягом $n_1 = n_2 = \frac{n-c}{2} = 6$ за умови, що $\frac{n-c}{2} \geq m$, де $m = 5$ – кількість змінних початкової моделі:

$$y_1 = 283,8086 - 0,97849x \text{ і } y_2 = -351,689 + 1,464085x.$$

Крок 4. Знаходимо суму квадратів залишків S_1 і S_2 для першої і другої моделей:

$$S_1 = \sum_{i=1}^{n_1} u_{i1}^2, \quad S_2 = \sum_{i=1}^{n_2} u_{i2}^2,$$

де u_1 і u_2 – залишки для першої і другої моделей відповідно;

$$S_1 = 2,8618, \quad S_2 = 1,7628.$$

Крок 5:

- розраховуємо критерій $F_{\text{експ}} = \frac{S_2}{S_1}$ ($F_{\text{експ}} = 0,615953$), який у разі виконання гіпотези про гомоскедастичність відповідатиме F -розподілу з $\gamma_1 = \frac{(n-c-2m_1)}{2} = 4$ і $\gamma_2 = \frac{(n-c-2m_2)}{2} = 4$ степенями вільності; $m_1 = m_2 = 2$ – кількість параметрів парних моделей;

- значення критерію $F_{\text{експ}} = 1,687625$ порівнюємо з табличним значенням F -критерію при вибраному довірчому рівні $\alpha = 0,05$ і відповідних степенях вільності (дод. 4):

$$F_{\text{табл}} = F_{(0,05; 4; 4)} = 6,388234.$$

Оскільки $F_{\text{експ}} \leq F_{\text{табл}}$, гетероскедастичність відсутня.

Отже, фактор X_2 не спричиняє гетероскедастичності залишків моделі.

Припустимо тепер, що зміну дисперсії залишків може зумовити вектор X_3 (рівень запасів), і протестуємо за цим фактором.

Крок 1. Спостереження (вихідні дані) ранжуємо в порядку зростання (або спадання) значення відповідно до величини елементів вибраного нами вектора X_3 .

Крок 2. Відкидаємо c спостережень, які перебувають у середині векторів вихідних даних, де

$$c = \frac{4n}{15} = 4,$$

$n = 16$ – кількість елементів вектора X_3 .

Крок 3. Будуємо дві моделі парної регресії на основі звичайного МНК для двох створених сукупностей спостережень обсягом

$$n_1 = n_2 = \frac{n-c}{2} = 6:$$

$$y_1 = 44,26846 + 0,013276x \text{ і } y_2 = -17,7792 + 0,475163x.$$

Крок 4. Знаходимо суму квадратів залишків S_1 і S_2 для першої і другої моделей:

$$S_1 = \sum_{i=1}^{n_1} u_{i1}^2, \quad S_2 = \sum_{i=1}^{n_2} u_{i2}^2,$$

де u_1 і u_2 – залишки для першої і другої моделей відповідно;

$$S_1 = 1,6893, \quad S_2 = 3,8248.$$

Крок 5:

- розраховуємо критерій $F_{\text{експ}} = \frac{S_2}{S_1}$ ($F_{\text{експ}} = 2,264125$), який у разі виконання гіпотези про гомоскедастичність відповідатиме F -розподілу з $\gamma_1 = \frac{(n-c-2m_1)}{2} = 4$ і $\gamma_2 = \frac{(n-c-2m_2)}{2} = 4$ степенями вільності; $m_1 = m_2 = 2$ – кількість параметрів парних моделей;

- значення критерію $F_{\text{експ}} = 2,264125$ порівнюємо з табличним значенням F -критерію при вибраному довірчому рівні $\alpha = 0,05$ і відповідних степенях вільності (дод. 4):

$$F_{\text{табл}} = F_{(0,05; 3; 3)} = 6,388234.$$

Оскільки $F_{\text{експ}} \leq F_{\text{табл}}$, гетероскедастичність відсутня.

Отже, фактор X_3 також не спричинює гетероскедастичності залишків моделі.

Припустимо тепер, що зміну дисперсії залишків може зумовити вектор X_4 (готівка), і протестуємо за цим фактором.

Крок 1. Спостереження (вихідні дані) ранжуємо в порядку зростання (або спадання) значення відповідно до величини елементів вибраного нами вектора X_4 .

Крок 2. Відкидаємо c спостережень, які перебувають усередині векторів вихідних даних, де

$$c = \frac{4n}{15} = 4,$$

$n = 16$ – кількість елементів вектора X_4 .

Крок 3. Будуємо дві моделі парної регресії на основі звичайного МНК для двох створених сукупностей спостережень обсягом

$$n_1 = n_2 = \frac{n-c}{2} = 6 :$$

$$y_1 = 42,5087 + 0,0130836x \text{ та } y_2 = 55,70116 - 0,0577x.$$

Крок 4. Знаходимо суму квадратів залишків S_1 і S_2 для першої і другої моделей:

$$S_1 = \sum_{i=1}^{n_1} u_{i1}^2, \quad S_2 = \sum_{i=1}^{n_2} u_{i2}^2,$$

де u_1 і u_2 – залишки для першої і другої моделей відповідно;

$$S_1 = 0,9395, \quad S_2 = 2,4752.$$

Крок 5:

- розраховуємо критерій $F_{\text{експ}} = \frac{S_2}{S_1}$ ($F_{\text{експ}} = 2,634619$), який у разі виконання гіпотези про гомоскедастичність відповідатиме F -розподілу з $\gamma_1 = \frac{(n-c-2m_1)}{2} = 4$ і $\gamma_2 = \frac{(n-c-2m_2)}{2} = 4$ степенями вільності;

$m_1 = m_2 = 2$ – кількість параметрів парних моделей;

- значення критерію $F_{\text{експ}} = 2,634619$ порівнюємо з табличним значенням F -критерію при вибраному довірчому рівні $\alpha = 0,05$ і відповідних степенях вільності (дод. 4):

$$F_{\text{табл}} = F_{(0,05; 3; 3)} = 6,388234.$$

Оскільки $F_{\text{експ}} \leq F_{\text{табл}}$, *гетероскедастичність відсутня.*

Отже, фактор X_4 також не спричиняє гетероскедастичності залишків моделі.

Припустимо тепер, що зміну дисперсії залишків може зумовити вектор X_5 (рівень власного капіталу), і протестуємо за цим фактором.

Крок 1. Спостереження (вихідні дані) ранжуємо в порядку зростання (або спадання) значення відповідно до величини елементів вибраного нами вектора X_5 .

Крок 2. Відкидаємо c спостережень, які містяться в середині векторів вихідних даних, де

$$c = \frac{4n}{15} = 4,$$

$n = 16$ – кількість елементів вектора X_3 .

Крок 3. Будуємо дві моделі парної регресії на основі звичайного МНК для двох створених сукупностей спостережень обсягом

$$n_1 = n_2 = \frac{n-c}{2} = 6:$$

$$y_1 = 47,02267 + 0,007736x \text{ і } y_2 = 52,80041 - 0,01184x.$$

Крок 4. Знаходимо суму квадратів залишків S_1 і S_2 для першої і другої моделей:

$$S_1 = \sum_{i=1}^{n_1} u_{i1}^2, \quad S_2 = \sum_{i=1}^{n_2} u_{i2}^2,$$

де u_1 і u_2 – залишки для першої і другої моделей відповідно;

$$S_1 = 1,1256, \quad S_2 = 3,035.$$

Крок 5:

- розраховуємо критерій $F_{\text{експ}} = \frac{S_2}{S_1}$ ($F_{\text{експ}} = 2,696445$), який у разі виконання гіпотези про гомоскедастичність відповідатиме F -розподілу з $\gamma_1 = \frac{(n-c-2m_1)}{2} = 4$ і $\gamma_2 = \frac{(n-c-2m_2)}{2} = 4$ степенями вільності; $m_1 = m_2 = 2$ – кількість параметрів парних моделей;

- значення критерію $F_{\text{експ}} = 2,696445$ порівнюємо з табличним значенням F -критерію при вибраному довірчому рівні $\alpha = 0,05$ і відповідних степенях вільності (дод. 4):

$$F_{\text{табл}} = F_{(0,05; 3; 3)} = 6,388234.$$

Оскільки $F_{\text{експ}} \leq F_{\text{табл}}$, *гетероскедастичність відсутня.*

Отже, фактор X_5 також не спричиняє гетероскедастичності залишків моделі.

Висновок. Залишки моделі гомоскедастичні за всіма факторами, отже, ефективність їх не порушиться через зміну дисперсії залишків.

Продовжимо аналіз залишків, дослідивши їх на наявність автокореляції. Застосуємо для цього тест Дарбіна–Уотсона (п. 7.7.6).

Крок 1. Обчислюємо d -статистику за формулою

$$d = \frac{\sum_{i=2}^T (\hat{u}_i - \hat{u}_{i-1})^2}{\sum_{i=1}^T \hat{u}_i^2}, \quad d = 2,984061$$

(d -статистика набуває значення з проміжку $[0, 4]$).

Крок 2. Вибираємо рівень значущості $\alpha = 0,05$ і за таблицею (дод. 5) знаходимо $DW1 = d_u(\alpha) = 0,73$ – нижнє значення і $DW2 = d_o(\alpha) = 1,94$ – верхнє значення статистики d .

Оскільки $4 - DW2 < d < 4 - DW1$, робимо висновок про *невизначеність автокореляції*. У цьому разі потрібно припустити, що, можливо, автокореляція існує, і переоцінити параметри моделі за узагальненим методом найменших квадратів (УМНК), який ще називають методом Ейткена (п. 7.7.7).

Оператор оцінювання УМНК має вигляд

$$\hat{B} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} Y$$

або

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

де Ω – дисперсійно-коваріаційна матриця залишків:

$$\Omega = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^9 \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^8 \\ \rho^2 & \rho & 1 & \rho & \dots & \rho^7 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \rho^9 & \rho^8 & \rho^7 & \rho^6 & \dots & 1 \end{pmatrix}; \quad S = \sigma_u^2 \cdot \Omega;$$

Ω^{-1} – матриця, обернена до матриці Ω ; S^{-1} – матриця, обернена до матриці S .

Оскільки в Ω коваріація залишків $\rho^s \rightarrow 0$ при $s > 2$, матриця Ω^{-1} матиме вигляд

$$\Omega^{-1} = \frac{1}{1-\rho^2} \cdot \begin{pmatrix} 1 & -\rho & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

На практиці для розрахунку ρ використовують співвідношення

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

або

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

Щоб сформувати Ω або S , необхідно визначити величину ρ , яка характеризує взаємозв'язок між послідовними членами ряду залишків. Припустимо, залишки описуються автокореляційною моделлю першого порядку

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

$$\text{де } \rho \approx r' \approx \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2} \cdot \frac{n}{n-1} + \frac{m+1}{n} = -0,10478.$$

Отже,

$$\Omega^{-1} = \begin{pmatrix} 1,0111 & 0,1059 & 0 & 0 & \dots & \dots & 0 \\ 0,1059 & 1,0222 & 0,1059 & 0 & \dots & \dots & 0 \\ 0 & 0,1059 & 1,0222 & 0 & \dots & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 0 & 1,0222 & 0,1059 & 0 \\ 0 & \dots & \dots & 0 & 0,1059 & 1,0222 & 0,1059 \\ 0 & \dots & \dots & \dots & 0 & 0,1059 & 1,0111 \end{pmatrix}.$$

Визначаємо:

$$X_t^T \Omega^{-1} = \begin{pmatrix} 1,117 & \dots & 1,117 \\ 1892,87 & \dots & 1586,349 \\ 269,543 & \dots & 278,0153 \\ 136,54 & \dots & 153,07 \\ 90,334 & \dots & 128,304 \\ 422,791 & \dots & 431,949 \end{pmatrix};$$

$$X_t^T \Omega^{-1} X_t = \begin{pmatrix} 19,5 & \dots & 7931,4 \\ 30828,5 & \dots & 12653663,5 \\ 5002,8 & \dots & 2048817,0 \\ 2539,6 & \dots & 1024836,2 \\ 1913,3 & \dots & 921092,6 \\ 7931,4 & \dots & 3906935,2 \end{pmatrix};$$

$$X_t^T \Omega^{-1} y_t = \begin{pmatrix} 932,4061 \\ 1465924 \\ 238679,5 \\ 121627,1 \\ 90659,64 \\ 371663,4 \end{pmatrix};$$

$$(X_t^T \Omega^{-1} X_t)^{-1} = \begin{pmatrix} 83,71 & \dots & 0,005 \\ -0,014 & \dots & 0 \\ -0,11 & \dots & 0 \\ -0,025 & \dots & 0 \\ -0,021 & \dots & 0 \\ 0,005 & \dots & 0 \end{pmatrix};$$

$$\hat{B} = \begin{pmatrix} 96,62511 \\ -0,014392 \\ -0,106536 \\ 0,017016 \\ 0,102629 \\ -0,02721 \end{pmatrix}$$

Отже,

$$\hat{y}_t = 96,62511 - 0,014392x_1 - 0,106536x_2 + 0,017016x_3 + 0,102629x_4 - 0,02721x_5.$$

Розраховуємо d -статистику Дарбіна–Уотсона:

$$d = \frac{\sum_{i=2}^T (\hat{u}_i - \hat{u}_{i-1})^2}{\sum_{i=1}^T \hat{u}_i^2},$$

$$d = 3,048909.$$

Порівнюємо d -статистику Дарбіна–Уотсона з критичними значеннями при $\alpha = 0,05$; $n = 16$ та $m = 5$ (дод. 5):

нижнє значення показника d

$$DW1 = d_u(\alpha) = 0,73,$$

верхнє значення

$$DW2 = d_o(\alpha) = 1,94.$$

Обчислюємо значення $4 - DW1 = 3,27$ і $4 - DW2 = 2,06$.

Оскільки $4 - DW2 < d < 4 - DW1$, то знову робимо висновок про *невизначеність автокореляції залишків*. А це, у свою чергу, означає, що гіпотеза про те, що залишки описуються авторегресійною схемою першого порядку, не справджується. Доцільно припустити, що залишки описуються авторегресійною схемою вищого порядку, та переоцінити параметри моделі методом Кочрена–Оркатта або Дарбіна.

4.7. Дослідження динамічних властивостей факторів

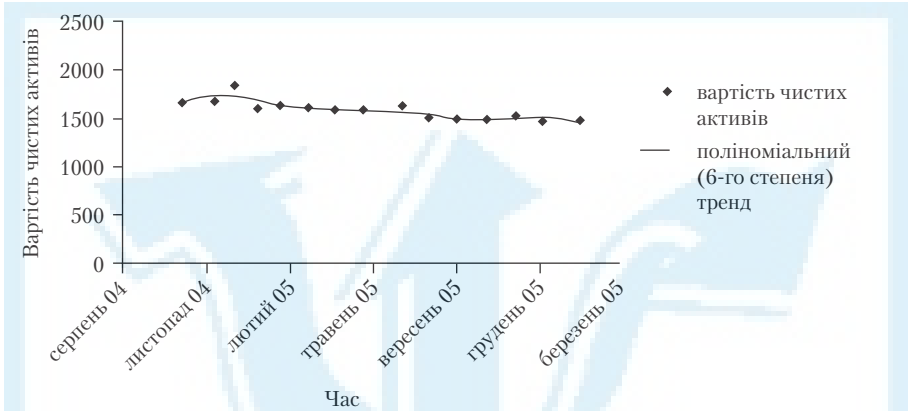
Для прогнозування значень залежної змінної потрібно дослідити поведінку кожної факторної змінної. Для цього проаналізуємо залежність кожного фактора моделі від дискретних моментів часу за допомогою графічних засобів Excel. Спочатку розглянемо динамічні властивості вартості чистих активів. За наведеними в табл. 4.1 даними будемо графіки залежності вартості чистих активів від дискретних моментів часу.

Апроксимуємо статистичну ламану вартості чистих активів трендами різних типів і для кожного тренда визначаємо коефіцієнти детермінації. Проаналізувавши статистичні дані, доходимо висновку, що достовірність апроксимації найбільша для поліноміального тренда 6-го степеня і становить $R^2 = 0,795$.

Динаміка вартості чистих активів

$$y = -1E - 11x^6 + 2E - 06x^5 - 0,2302x^4 + 11817x^3 - 3E + 08x^2 + 5E + 12x - 3E - 16$$

$$R^2 = 0,795$$



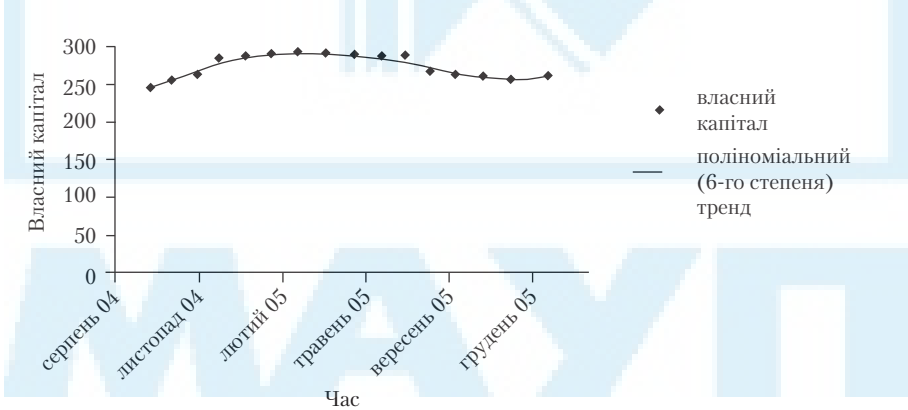
Отже, вартість чистих активів змінюється в часі як поліном 6-го степеня, а прогнозне значення вартості чистих активів в наступному місяці приблизно становить $X1_{пр} = 1368,26$ тис. грн.

Аналогічно аналізуємо динамічні властивості власного капіталу, запасів, готівки і поточних зобов'язань.

Динаміка власного капіталу

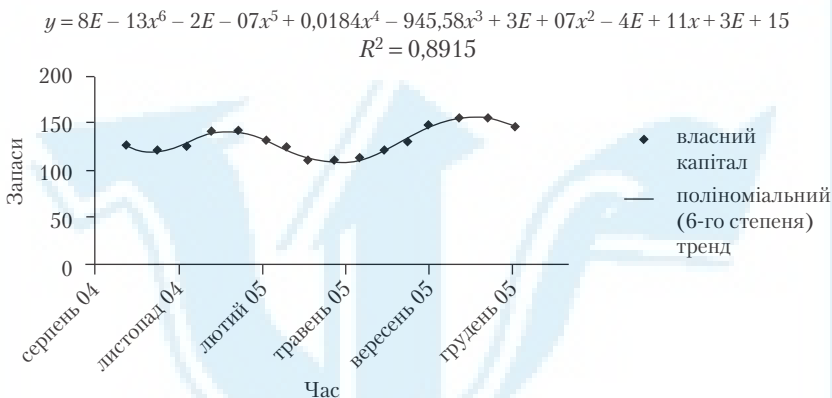
$$y = 2E - 13x^6 - 4E - 08x^5 - 0,0036x^4 - 187,1x^3 + 5E + 06x^2 + 8E + 10x + 5E + 14$$

$$R^2 = 0,8022$$



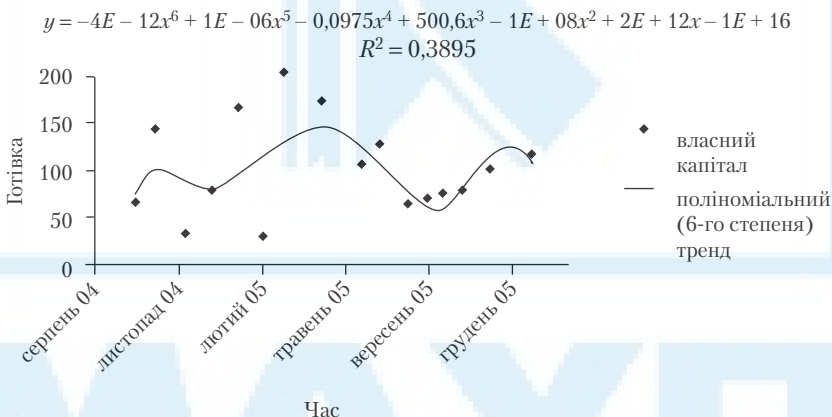
Власний капітал змінюється в часі як поліном 6-го степеня, а прогнозне значення власного капіталу в наступному місяці приблизно становить $X2_{пр} = 251,8475$ тис. грн.

Динаміка запасів



Запаси змінюються в часі як поліном 6-го степеня, а прогнозне значення запасів у наступному місяці приблизно становить $X3_{пр} = 138,235$ тис. грн.

Динаміка готівки

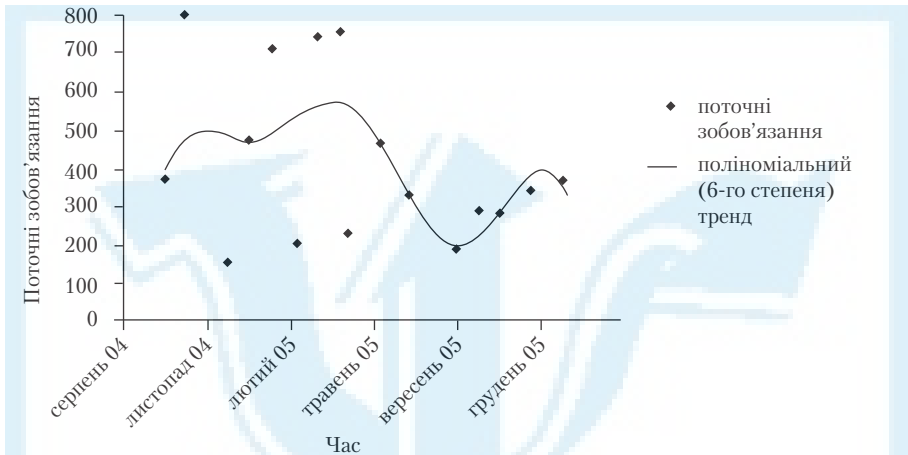


Готівка змінюється в часі як поліном 6-го степеня, а прогнозне значення готівки в наступному місяці приблизно становить $X4_{пр} = 107,74$ тис. грн.

Динаміка поточних зобов'язань

$$y = -2E - 11x^6 + 4E - 06x^5 - 0,3527x^4 + 18095x^3 - 5E + 08x^2 + 8E + 12x - 5E + 16$$

$$R^2 = 0,2932$$



Поточні зобов'язання змінюються в часі як поліном 6-го степеня, а прогнозне значення поточних зобов'язань у наступному місяці приблизно становить $X5_{\text{пр}} = 381,4$ тис. грн.

Отже, для прогнозного періоду ми отримали вектор прогнозних значень рівня вартості чистих активів, власного капіталу, запасів, готівки та поточних зобов'язань за наступний період: $X'_{\text{пр}} = (1, X1_{\text{пр}}, X2_{\text{пр}}, X3_{\text{пр}}, X4_{\text{пр}}, X5_{\text{пр}}) = (1, 1368,26; 251,8475; 138,235; 107,74; 381,4)$.

4.8. Прогнозування за побудованою моделлю

Використаємо модель (4.2) для знаходження точкового та інтервального прогнозу витрат на споживання, який відповідатиме вектору прогнозних значень факторів $X'_{\text{пр}} = (1, 1368,26; 251,8475; 138,235; 107,74; 381,4)$. Попередньо обчислимо довірчі інтервали регресії та її параметрів.

1) знаходження довірчого інтервалу для регресії:

- обчислюємо довірчі зони регресії (довірчий інтервал для математичного сподівання) як

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

$\hat{\sigma}_u^2$ — незміщена оцінка дисперсії залишків: $\hat{\sigma}_u = 2,968778$;

- знаходимо відповідне табличне значення $t_{\text{табл}} = t(\alpha/2, k)$ t -розподілу з $k = n - m - 1 = 10$ степенями вільності і при рівні значущості $\alpha = 0,05$ (дод. 3):

$$t_{\text{табл}} = 2,233769.$$

Модельні значення результативного показника обчислені за (4.3):

$$\hat{y}_i = \begin{pmatrix} 47,19264 \\ 41,56835 \\ 44,50525 \\ 43,48363 \\ 44,33098 \\ 44,48284 \\ 46,94139 \\ 44,18992 \\ 50,41335 \\ 48,99898 \\ 46,95910 \\ 54,04654 \\ 50,54166 \\ 51,57264 \\ 52,54909 \\ 53,12355 \end{pmatrix}.$$

Довірчі зони регресії такі:

$$\hat{y}_i - t_{\alpha/2, k} \hat{\sigma}_u \sqrt{X'_i (X'X)^{-1} X_i} = \begin{pmatrix} 46,50638 \\ 38,95742 \\ 41,57165 \\ 42,70105 \\ 43,87407 \\ 43,62867 \\ 45,19088 \\ 43,3381 \\ 47,96107 \\ 46,71737 \\ 44,5174 \\ 50,54187 \end{pmatrix},$$

$$\hat{y}_i - t_{\alpha/2, k} \hat{\sigma}_u \sqrt{X_i'(X'X)^{-1} X_i} = \begin{pmatrix} 47,44205 \\ 48,36893 \\ 49,05576 \\ 49,50933 \\ 47,87891 \\ 44,17929 \\ 47,43885 \\ 44,26623 \\ 44,7879 \\ 45,33701 \\ 48,69191 \\ 45,04175 \\ 52,86565 \\ 51,28061 \\ 49,40081 \\ 57,55123 \\ 53,64127 \\ 54,77636 \\ 56,04244 \end{pmatrix};$$

2) побудова довірчих інтервалів для параметрів регресії:

- обчислюємо довірчі інтервали для параметра β_i за формулами

$$(b_i - t_{\text{табл}} \sqrt{\hat{\sigma}_u^2 c_{ii}}; b_i + t_{\text{табл}} \sqrt{\hat{\sigma}_u^2 c_{ii}}),$$

де c_{ii} – діагональний елемент матриці $(X'X)^{-1}$; $\hat{\sigma}_u = 2,968778$;
 $t_{\text{табл}} = 2,633769$;

- виконуємо необхідні розрахунки:

$$\beta_0 \in (51,78701; 139,08822), \quad \beta_1 \in (-0,02428; -0,003832),$$

$$\beta_2 \in (-0,18949; -0,017136), \quad \beta_3 \in (-0,14057; 0,17342),$$

$$\beta_4 \in (0,04169; 0,15296), \quad \beta_5 \in (-0,03885; -0,013525);$$

3) прогнозні значення і довірчі інтервали прогнозу:

- для обчислення прогнозних значень $y_{\text{пр}}$ у рівняння (4.3)

$$\hat{y} = 95,438 - 0,014x_1 - 0,103x_2 + 0,016x_3 + 0,097x_4 - 0,026x_5$$

підставляємо задані значення $X1_{\text{пр}} = 1368,26$; $X2_{\text{пр}} = 251,8475$; $X3_{\text{пр}} = 138,235$; $X4_{\text{пр}} = 107,74$; $X5_{\text{пр}} = 381,4$, які лежать за межами базового періоду (точковий прогноз):

$$y_{\text{пр}} = \hat{y}_i = 52,95723;$$

4) обчислення інтервальних прогнозів:

- знаходимо межі інтервального прогнозу індивідуального значення за формулою

$$\begin{aligned} \hat{Y}_0 - t_{\text{табл}} \hat{\sigma}_u \sqrt{1 + X_0'(X'X)^{-1} X_0} &\leq Y_0 \leq \\ &\leq \hat{Y}_0 + t_{\text{табл}} \hat{\sigma}_u \sqrt{1 + X_0'(X'X)^{-1} X_0}. \end{aligned}$$

Зауважимо, що $\hat{Y}_0 = X_0 B$ можна розглядати як точкову оцінку математичного сподівання прогнозного значення Y_0 , а також як індивідуальне значення Y_0 для вектора незалежних змінних X_0 , що лежать за межами базового періоду:

$$\hat{Y}_0 = \hat{y}_i = 52,957; \quad \hat{\sigma}_u = 2,968778; \quad t_{\text{табл}} = t(0,025; 12) = 2,233769;$$

$$t_{\text{табл}} \hat{\sigma}_u \sqrt{1 + X_0'(X'X)^{-1} X_0} = 8,695897.$$

(44,261333; 61,65313) – *інтервальний прогноз індивідуального значення.*

Межі довірчих інтервалів для математичного сподівання значення $y_{\text{пр}}$ обчислюємо за формулою

$$\begin{aligned} \hat{Y}_0 - t_{\text{табл}} \hat{\sigma}_u \sqrt{X_0'(X'X)^{-1} X_0} &\leq M(Y_0) \leq \\ &\leq \hat{Y}_0 + t_{\text{табл}} \hat{\sigma}_u \sqrt{X_0'(X'X)^{-1} X_0}, \end{aligned}$$

$$t_{\text{табл}} \hat{\sigma}_u \sqrt{X_0'(X'X)^{-1} X_0} = 3,805346.$$

(49,5188456; 56,76258) – *інтервальний прогноз математичного сподівання індивідуального значення.*

Отже, ми отримали різні види теоретично обґрунтованих і статистично надійних прогнозів щодо фінансового результату, побудувавши і дослідивши лінійну регресійну модель залежності зазначеного показника від рівня вартості чистих активів, власного капіталу, запасів, готівки та поточних зобов'язань.

Висновки

Побудова і дослідження лінійної регресійної багатофакторної моделі передбачають:

- описування моделі (вибір впливових факторів і форми залежності);
- перевірку мультиколінеарності факторів;

- оцінювання параметрів моделі за МНК;
- повне економетричне дослідження (адекватність моделі, значущість параметрів, аналіз залишків на наявність автокореляції і гетероскедастичності);
- розрахунок прогнозних значень результативного показника.

Для виконання розрахунків застосовують спеціальні пакети прикладних програм або електронні таблиці *Excel*.

Питання для самоперевірки

1. Що означає мультиколінеарність змінних?
2. Які висновки можна зробити, дослідивши мультиколінеарність за алгоритмом Фаррара–Глобера?
3. Як позбутися мультиколінеарності? Наведіть методи її усунення.
4. До яких наслідків призводить порушення припущення про гомоскедастичність?
5. Як з'ясувати наявність гетероскедастичності?
6. Сутність узагальненого методу найменших квадратів.
7. Що означає автокореляція залишків?
8. До яких наслідків призводить автокореляції залишків?
9. Сутність критерію Дарбіна–Уотсона.
10. За допомогою яких методів можна оцінювати параметри моделі з автокорельованими залишками?
11. Як виконують обчислення за УМНК при автокореляції?

МАУП

ОДНОВИМІРНІ ЧАСОВІ РЯДИ ТА ЇХ МОДЕЛЮВАННЯ

5.1. Основні елементи часового ряду

Вивчення багатьох економічних явищ базується на дослідженні властивостей функції часу $f(t)$, яка визначає розвиток явища. Такі часові функції називають процесами. Прикладом часової функції можуть бути щорічна врожайність сільськогосподарської культури, щомісячний випуск продукції промислового підприємства, зміни валютного курсу, зміни цін товарів і послуг тощо.

Реєстрація значень досліджуваного процесу утворює сукупність величин y_1, y_2, \dots, y_N , які впорядковані за часом. Такий набір даних називають динамічним чи часовим рядом. Окремі його значення називають рівнями ряду і позначають $y_t, t = 1, 2, \dots, N$, де N — кількість рівнів ряду.

Часові ряди поділяють на моментні та інтервальні. Якщо рівень часового ряду фіксує значення показника на певний момент часу, такий ряд називається моментним, наприклад рівень цін на деякі види продуктів на початок кожного місяця. Якщо рівень часового ряду характеризує значення показника за певний період часу, ряд називається інтервальним, наприклад щомісячний обсяг випуску продукції.

Більшість часових рядів складається зі значень, вимірених у натуральному чи вартісному виразі за місяць, квартал, рік тощо. Інколи рівні ряду можна представляти не абсолютними значеннями, а деякими величинами, що походять від них, наприклад відносними чи середніми значеннями, відхиленнями від середнього тощо.

Якщо всі значення рівнів часового ряду можна точно визначити за відомою математичною функцією, ряд називається детермінованим (визначеним). Якщо рівні ряду можна описати функцією розподілу ймовірностей, ряд називається випадковим. Наприклад, чисельність працівників підприємства — детермінована величина, а

явочна чисельність працівників на певний момент часу — випадкова величина.

Процеси, що розвиваються з часом за законами теорії ймовірностей, називаються стохастичними чи випадковими.

Стохастичний процес називається стаціонарним, якщо його основні властивості не змінюються з часом, а саме:

1) математичне сподівання (середнє значення рівнів ряду) і дисперсія (квадрат відхилення від середнього значення) не змінюються з часом;

2) коваріація між рівнями ряду, зміщеними на кілька періодів, залежить лише від величини часового зміщення, а коефіцієнти кореляції між ними є сталими величинами.

Стохастичний процес, що не має таких властивостей, називається нестаціонарним.

За аналогією з просторовою вибіркою динамічний ряд можна вважати вибіркою з нескінченного ряду значень показника за часом. Однак на відміну від просторової вибірки елементи динамічного ряду виявляються статистично залежними і по-різному розподіленими величинами.

Значення рівнів часового ряду містять два види компонент — систематичну і випадкову. Систематична компонента є результатом впливу постійно діючих факторів.

Виокремлюють три основні систематичні компоненти часового ряду:

- тренд;
- сезонність;
- циклічність.

Тренд (тенденція) часового ряду (T_t) — це систематична лінійна чи нелінійна компонента, яка плавно змінюється з часом. Вона описує чистий вплив довготривалих факторів (наприклад, зростання чисельності населення, змінювання промислового виробництва тощо).

Сезонна компонента (S_t) — це періодичні коливання рівнів часового ряду протягом не дуже тривалого періоду (тижня, місяця, максимум — року). Сезонність відбиває повторюваність економічних процесів у межах одного року (наприклад, обсяг продажу певної групи товарів у різні пори року).

Циклічність (C_t) — це періодичні коливання, що виходять за межі одного року. Проміжок часу між двома сусідніми “вершинами” (найбільшими значеннями) чи “западинами” (найменшими значеннями) є довжиною циклу. Циклічність відбиває повторюваність економіч-

них процесів протягом тривалих періодів (наприклад, фазовий процес розвитку економіки).

Рівні часового ряду можуть одночасно містити всі систематичні компоненти або лише деякі з них.

Випадкова складова u_t відбиває різкі й несподівані впливи, що найсильніше позначаються на основній тенденції ряду, а також вплив поточних чинників, пов'язаних, наприклад, з похибками вимірювань.

Для опису і детального вивчення часових рядів застосовують різні математичні моделі, які дають змогу виявити їх основні компоненти.

При різних поєднаннях систематичних складових ряду залежність його рівнів від часу може набувати різних форм.

На рис. 5.1 показано компоненти гіпотетичного часового ряду, що ілюструє зростаючу тенденцію (а) і гіпотетичний часовий ряд, що містить лише сезонну компоненту (б).

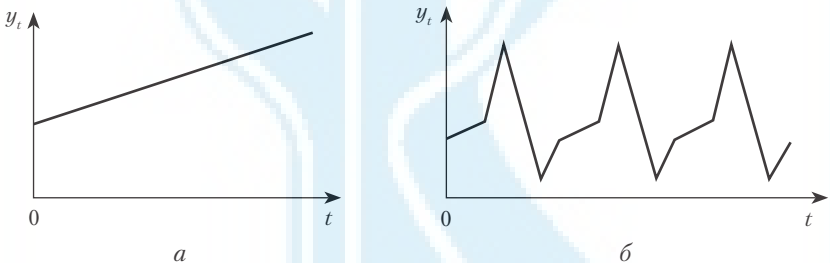


Рис. 5.1

Деякі часові ряди не містять тенденції й циклічної компоненти, а кожен наступний їх рівень утворюється як сума середнього рівня ряду та деякої (додатної чи від'ємної) випадкової компоненти. Приклад ряду, що містить лише випадкову компоненту, наведено на рис. 5.2.

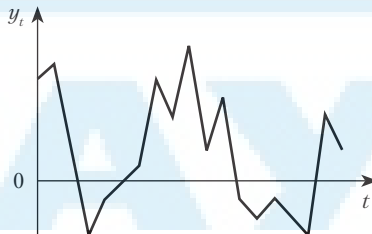


Рис. 5.2

Очевидно, що реальні дані не відповідають повною мірою жодній з описаних вище моделей. Найчастіше вони містять усі три компоненти. Кожен їхній рівень формується під впливом тенденції, сезонних коливань і випадкової компоненти.

Здебільшого фактичний рівень часового ряду можна представити як суму або добуток трендової, циклічної та випадкової компонент. Модель, в якій часовий ряд представлено як суму зазначених компонент, називається адитивною моделлю часового ряду:

$$y_t = T_t + S_t + C_t + U_t.$$

Модель, в якій часовий ряд представлено як добуток зазначених компонент, називається мультиплікативною моделлю часового ряду:

$$y_t = T_t S_t \cdot C_t U_t.$$

Основне завдання економетричного дослідження часових рядів – виявити та надати кількісного виразу кожній з цих компонент, з тим щоб використати отриману інформацію для прогнозування майбутніх значень ряду або для побудови моделей взаємозв'язку двох або більше часових рядів.

Більшість методів вивчення структури часового ряду (фільтрація рівнів ряду) спрямовані на виявлення та описування систематичних (регулярних) компонент ряду.

Дослідження часових рядів передбачає такі етапи аналізу:

- подання часового ряду в графічному вигляді;
- описування поведінки часового ряду;
- виявлення та вилучення систематичних (невипадкових) компонент часового ряду (тренда, сезонності, циклічності);
- згладжування і фільтрація часового ряду (вилучення випадкових викидів, які можуть бути пов'язані з похибками вимірювань чи з екстремальними подіями в економіці);
- дослідження випадкової складової часового ряду, перевірка адекватності побудованої моделі досліджуваному явищу;
- прогнозування майбутніх значень часового ряду на основі його поточних і попередніх значень;
- аналіз залежностей між різними часовими рядами.

5.2. Виявлення структури часового ряду з використанням автокореляції рівнів

У часових рядах значення кожного наступного рівня часто залежать від попередніх значень. 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 та рівнями цього ряду, зміщеними на кілька часових кроків (лагів).

Для розрахунку коефіцієнта кореляції застосуємо формулу

$$r_{xy} = \frac{\sum (x_j - \bar{x}) \cdot (y_j - \bar{y})}{\sqrt{\sum (x_j - \bar{x})^2 \cdot \sum (y_j - \bar{y})^2}}.$$

Якщо замість змінної x підставити значення рівнів ряду y_2, y_3, \dots, y_n , а замість змінної y – значення рівнів ряду y_1, y_2, \dots, y_{n-1} , то формула набуде вигляду

$$r_1 = \frac{\sum_{t=2}^n (y_t - \bar{y}_1) \cdot (y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^n (y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^n (y_{t-1} - \bar{y}_2)^2}}, \quad (5.1)$$

де

$$\bar{y}_1 = \frac{\sum_{t=2}^n y_t}{n-1}; \quad \bar{y}_2 = \frac{\sum_{t=2}^n y_{t-1}}{n-1}. \quad (5.2)$$

Цю величину називають коефіцієнтом автокореляції рівнів ряду першого порядку, тому що він вимірює залежність між сусідніми рівнями ряду t і $t-1$.

Аналогічно можна визначити коефіцієнти автокореляції другого і вищих порядків. Так, коефіцієнт автокореляції другого порядку характеризує щільність зв'язку між рівнями ряду y_t та y_{t-2} і визначається за формулою

$$r_2 = \frac{\sum_{t=3}^n (y_t - \bar{y}_3) \cdot (y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^n (y_t - \bar{y}_3)^2 \cdot \sum_{t=3}^n (y_{t-2} - \bar{y}_4)^2}}, \quad (5.3)$$

де

$$\bar{y}_3 = \frac{\sum_{t=3}^n y_t}{n-2}; \quad \bar{y}_4 = \frac{\sum_{t=3}^n y_{t-2}}{n-2}. \quad (5.4)$$

Приклад 5.1. Умовні дані про середні прибутки фірми за 8 років такі:

Рік	1	2	3	4	5	6	7	8
Рівень прибутку	7	8	8	10	11	12	14	16

Зрозуміло, що прибутки в поточному році залежать від прибутків попередніх років. Потрібно виявити автокореляцію рівнів ряду.

Обчислимо коефіцієнт кореляції між рядами y_t і y_{t-1} , який визначає щільність зв'язку між прибутками поточного та попереднього років.

За початковими даними і формулою (5.2) обчислимо:

$$\bar{y}_1 = \frac{8+8+10+11+12+14+16}{7} = \frac{79}{7} = 11,29;$$

$$\bar{y}_2 = \frac{7+8+8+10+11+12+14}{7} = \frac{70}{7} = 10.$$

Використовуючи формулу (5.1), отримуємо коефіцієнт автокореляції рівнів ряду першого порядку:

$$r_1 = \frac{44}{\sqrt{53,43 \cdot 38}} = 0,976.$$

Отримане значення свідчить про дуже щільний зв'язок між прибутками поточного та безпосередньо попереднього років, а отже, про наявність у часовому ряді прибутків сильної лінійної тенденції.

На підставі часових рядів y_t та y_{t-2} складемо табл. 5.1 і обчислимо

$$\bar{y}_3 = \frac{8+10+11+12+14+16}{6} = \frac{71}{6} = 11,83;$$

$$\bar{y}_4 = \frac{7+8+8+10+11+12}{6} = \frac{56}{6} = 9,33.$$

За формулою (5.3) розрахуємо коефіцієнт автокореляції другого порядку:

$$r_2 = \frac{27,3334}{\sqrt{40,8334 \cdot 19,3334}} = 0,973.$$

Таблиця 5.1

t	y_t	y_{t-2}	$y_t - \bar{y}_3$	$y_{t-2} - \bar{y}_4$	$(y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)$	$(y_t - \bar{y}_3)^2$	$(y_{t-2} - \bar{y}_4)^2$
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

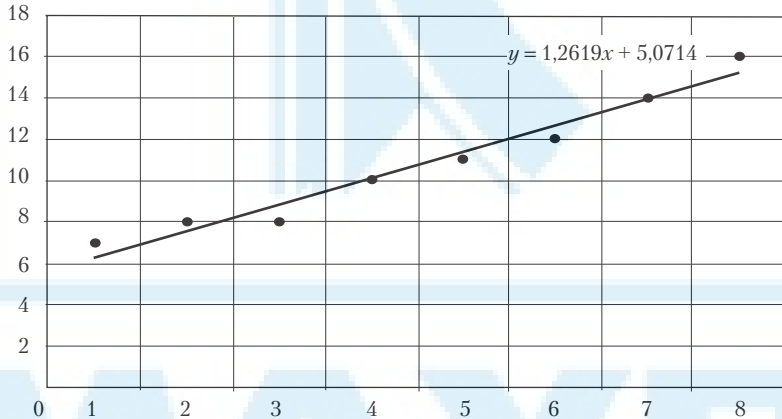


Рис. 5.3

Кількість періодів, за якими розраховують коефіцієнт автокореляції, називається лагом. Із збільшенням лага кількість пар значень,

за якими визначається коефіцієнт автокореляції, зменшується. Максимальний порядок коефіцієнта автокореляції не перевищує $n/4$, де n – обсяг вибірки.

Коефіцієнт автокореляції має дві основні властивості. По-перше, його визначають за аналогією з лінійним коефіцієнтом кореляції, отже, він характеризує щільність лінійного зв'язку поточного та попереднього рівнів ряду. Тому за коефіцієнтом автокореляції можна виявити лише наявність лінійної (або близької до лінійної) тенденції. Для деяких часових рядів, які мають сильну нелінійну тенденцію (наприклад, параболу другого порядку або експоненту), коефіцієнт автокореляції рівнів ряду може наближатися до нуля.

По-друге, за знаком коефіцієнта автокореляції не можна робити висновок про те, зростаючою чи спадною є тенденція в рівнях ряду. Більшість часових рядів економічних даних мають додатну автокореляцію рівнів, однак при цьому вони можуть мати спадну загальну тенденцію.

Послідовність коефіцієнтів автокореляції рівнів першого, другого і вищих порядків називають автокореляційною функцією часового ряду. Графік залежності її значень від величини лага (часового зміщення даних) називається корелограмою.

Аналіз автокореляційної функції та корелограми дає змогу визначити лаг, при якому автокореляція найвища, а саме лаг, при якому зв'язок між поточним і попередніми рівнями ряду найщільніший, що, у свою чергу, дає змогу виявити структуру ряду:

- якщо найбільшим виявився коефіцієнт автокореляції першого порядку, досліджуваний ряд містить лише тенденцію;
- якщо найбільшим виявився коефіцієнт автокореляції порядку τ , ряд містить циклічні коливання з періодичністю τ моментів часу;
- якщо жоден з коефіцієнтів автокореляції не виявився значущим, то, можливо, ряд не містить ані тенденції, ані циклічних коливань і має структуру, подібну до структури ряду, зображеної на рис. 5.2, або ряд містить сильну нелінійну тенденцію, для виявлення якої потрібен додатковий аналіз.

У розглянутому прикладі обидва обчислені коефіцієнти автокореляції рівнів високі, отже, між рівнями ряду існує щільний лінійний зв'язок. Більше значення першого з них свідчить, що часовий ряд прибутків містить лише тенденцію.

Приклад 5.2. Умовні дані про середні прибутки фірми за 16 місяців такі:

Місяць	1	2	3	4	5	6	7	8
Рівень прибутку	6,0	4,4	5,0	9,0	7,2	4,8	6,0	10,0
Місяць	9	10	11	12	13	14	15	16
Рівень прибутку	8,0	5,6	6,4	11,0	9,0	6,6	7,0	10,8

Потрібно визначити коефіцієнти автокореляції r_1, r_2, r_3, r_4 , зробити висновок про наявність або відсутність лінійної тенденції.

Нанесемо значення y_t на графік (рис. 5.3), а зміщені часові ряди $y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}$ запишемо у табл. 5.2.

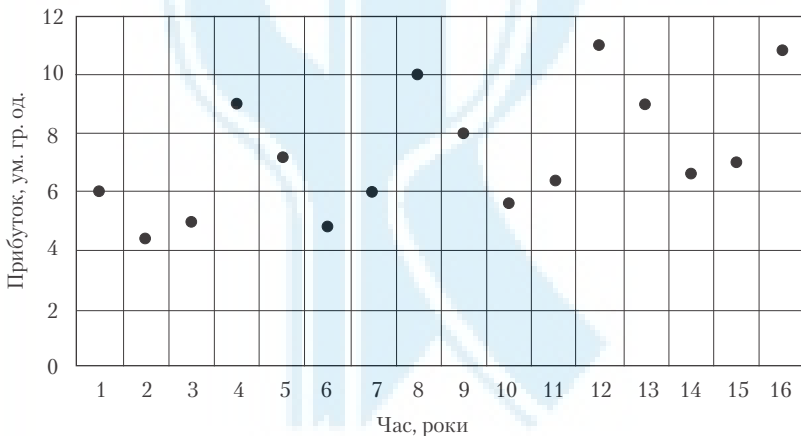


Рис. 5.4. Середні прибутки фірми

Таблиця 5.2

T	y_t	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}
1	2	3	4	5	6
1	6,0	–	–	–	–
2	4,4	6,0	–	–	–
3	5,0	4,4	6,0	–	–
4	9,0	5,0	4,4	6,0	–
5	7,2	9,0	5,0	4,4	6,0

Закінчення табл. 5.2

1	2	3	4	5	6
6	4,8	7,2	9,0	5,0	4,4
7	6,0	4,8	7,2	9,0	5,0
8	10,0	6,0	4,8	7,2	9,0
9	8,0	10,0	6,0	4,8	7,2
10	5,6	8,0	10,0	6,0	4,8
11	6,4	5,6	8,0	10,0	6,0
12	11,0	6,4	5,6	8,0	10,0
13	9,0	11,0	6,4	5,6	8,0
14	6,6	9,0	11,0	6,4	5,6
15	7,0	6,6	9,0	11,0	6,4
16	10,8	7,0	6,6	9,0	11,0

Визначимо коефіцієнт автокореляції першого порядку. За даними таблиці $r_1 = 0,165$. Таке мале його значення свідчить про слабку залежність поточних рівнів ряду від рівнів, що безпосередньо передують їм.

Після розрахунку коефіцієнтів автокореляції другого, третього і четвертого порядків отримаємо кількісну характеристику кореляційного зв'язку рівнів ряду. Маємо $r_2 = 0,567$, $r_3 = 0,156$, $r_4 = 0,937$.

Це означає, що структура ряду така, що кожен наступний рівень y_t залежить від рівнів y_{t-4} і y_{t-2} набагато більшою мірою, ніж від рівня y_{t-1} .

Продовжимо розрахунки і за отриманими значеннями побудуємо автокореляційну функцію цього ряду (рис. 5.5).

Аналіз значень автокореляційної функції дає змогу зробити висновки про наявність в досліджуваному часовому ряді сезонних коливань періодичністю в 4 місяці.

Цей висновок підтверджується і графічним аналізом структури ряду (рис. 5.6).

Якби серед коефіцієнтів автокореляції часового ряду найбільшим виявився коефіцієнт другого порядку, то це б свідчило про наявність циклічних коливань з циклом, що дорівнює двом періодам часу, і про пілкоподібну структуру ряду.

Але такий метод аналізу дає досить наближені результати, тому для дослідження періодичності сезонних коливань застосовують точніші методи, один з яких має назву спектрального аналізу і розглядається далі.

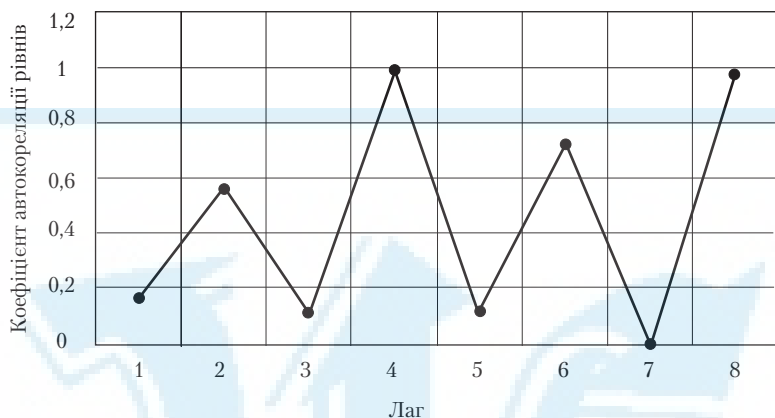


Рис. 5.5

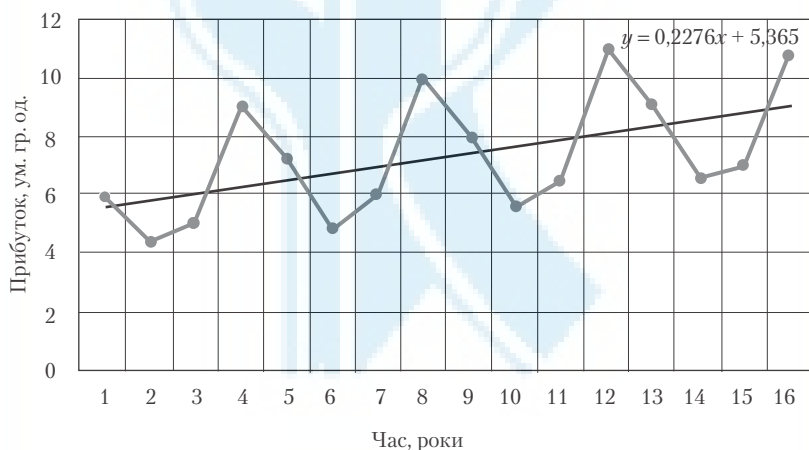


Рис. 5.6

5.3. Перевірка гіпотези про існування тенденції

Під *тенденцією* розуміють деякий загальний напрям розвитку, довготривалу еволюцію. Тенденцію ряду динаміки зображують у вигляді гладкої кривої (траєкторії) як функцію часу і називають трендом. *Тренд* характеризує основну закономірність розвитку економічного

явища в часі, вільну в основному (але не цілковито) від випадкових впливів.

Здебільшого отриману траєкторію пов'язують виключно з часом. Припускають, що, розглядаючи будь-яке явище як функцію часу, можна виявити спільний вплив усіх основних чинників, не визначаючи впливу кожного з них в явному вигляді. У зв'язку з цим під трендом зазвичай розуміють регресію на час. На практиці зручніше користуватися іншим загальним поняттям, згідно з яким тренд — це детермінована складова динаміки розвитку, зумовлена впливом постійно діючих факторів. При цьому окремі рівні часового ряду не збігаються із загальною тенденцією, а мають певні випадкові відхилення від неї, які характеризують випадкові впливи. Отже, рівняння, що описує процес у часі, має випадкову складову — відхилення від тренда. Тому рівні часового ряду описують рівнянням

$$y_t = f(t) + u,$$

де $f(t)$ — систематична складова, яка характеризує основну тенденцію явища в часі; u — випадкова складова.

У часових рядах можна спостерігати тенденції трьох видів: середнього рівня; дисперсії; автокореляції.

Тенденцію середнього рівня наочно можна представити графіком часового ряду. Він має вигляд функції $f(t)$, навколо якої варіюють фактичні значення явища, що вивчається.

Тенденція дисперсії — це зміни відхилень емпіричних значень часового ряду від значень, обчислених за рівнянням тренда.

Тенденція автокореляції — це тенденція зміни зв'язку між окремими рівнями часового ряду.

5.3.1. Перевірка наявності тенденції середнього рівня

Один із способів перевірки наявності тенденції заснований на порівнянні середніх рівнів ряду: часовий ряд розбивають на дві приблизно рівні частини, кожна з яких розглядають як деяку самостійну вибірку сукупність, що має нормальний розподіл. Якщо часовий ряд має тенденцію до змінювання, то середні значення, обчислені для кожної сукупності, мають істотно (значно) різнитися між собою. Якщо розбіжність буде незначною (неістотною, випадковою), це означатиме, що часовий ряд не має тенденції. Отже, перевірка наявності тренда в досліджуваному ряді зводиться до перевірки гіпотези про рівність середніх двох нормально розподілених сукупностей.

Обчислення за цим методом складається з чотирьох етапів:

1) вихідний часовий ряд $y_1, y_2, y_3, \dots, y_n$ розбивають на дві приблизно рівні частини обсягом $n_1 \approx n_2$, де $(n_1 + n_2 = n)$;

2) для кожної з частин обчислюють середні значення та дисперсії:

$$\bar{y}_1 = \frac{\sum_{t=1}^{n_1} y_t}{n_1}; \quad \sigma_1^2 = \frac{\sum_{t=1}^{n_1} (y_t - \bar{y}_1)^2}{n_1 - 1};$$
$$\bar{y}_2 = \frac{\sum_{t=1}^{n_2} y_t}{n_2}; \quad \sigma_2^2 = \frac{\sum_{t=1}^{n_2} (y_t - \bar{y}_2)^2}{n_2 - 1};$$

3) висувають основну гіпотезу про рівність середніх значень: $H_0: \bar{y}_1 = \bar{y}_2$ проти альтернативної $H_1: \bar{y}_1 \neq \bar{y}_2$; і допоміжну гіпотезу про рівність дисперсій $H_1^{\text{доп}}: \sigma_1^2 = \sigma_2^2$ проти $H_1^{\text{доп}}: \sigma_1^2 \neq \sigma_2^2$;

4) перевіряють допоміжну гіпотезу за допомогою F -критерію Фішера. Для цього порівнюють розрахункове (експериментальне) значення критерію:

$$F_{\text{експ}} = \begin{cases} \frac{\sigma_1^2}{\sigma_2^2}, & \text{якщо } \sigma_1^2 > \sigma_2^2, \\ \frac{\sigma_2^2}{\sigma_1^2}, & \text{якщо } \sigma_1^2 < \sigma_2^2, \end{cases}$$

з табличним (критичним) значенням розподілу Фішера $F_{\text{табл}} = F(\alpha, k_1, k_2)$, де α – заданий рівень значущості, $k_i = n_i - 1$ – степені вільності, $i = 1, 2$.

Якщо за критерієм Фішера дисперсії виявляться нерівними ($F_{\text{експ}} > F_{\text{табл}}$), то основну гіпотезу не перевіряють. Інакше переходять до наступного пункту;

5) основну гіпотезу про відсутність тренда перевіряють за допомогою t -критерію Стьюдента. Для цього обчислюють вибірккову статистику – розрахункове значення критерію Стьюдента за формулою

$$t_{\text{експ}} = \frac{|\bar{y}_1 - \bar{y}_2|}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

де σ – середньоквадратичне відхилення різниці середніх;

$$\sigma = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}}.$$

Якщо розрахункове значення $t_{\text{експ}}$ менше від табличного значення розподілу Стьюдента ($t_{\text{експ}} < t_{\text{табл}}$, де $t_{\text{табл}} = t(\alpha/2, n - 2)$), то основна гіпотеза H_0 приймається, тобто середні значення рівні, отже, ряд не має тренда. Якщо H_0 відхиляється, то ряд має тенденцію до змінювання (тренд ϵ).

5.3.2. Метод Фостера–Стюарта

Цей метод базується на послідовному порівнянні рівнів ряду. Він має більші можливості й дає надійніші результати порівняно з попереднім. Реалізація методу складається з чотирьох етапів.

1. Будують допоміжні послідовності: при порівнянні кожного рівня часового ряду, починаючи з другого, з усіма попередніми визначають значення:

$$k_t = \begin{cases} 1, & \text{якщо } y_t > y_{t-1}, y_{t-2}, \dots, y_1, \\ 0 & \text{в інших випадках,} \end{cases}$$

$$l_t = \begin{cases} 1, & \text{якщо } y_t < y_{t-1}, y_{t-2}, \dots, y_1, \\ 0 & \text{в інших випадках,} \end{cases}$$

$t = 2, 3, \dots, n$.

2. За утвореними послідовностями обчислюють величини s і d :

$$s = \sum_{t=2}^n (k_t + l_t), \quad d = \sum_{t=2}^n (k_t - l_t).$$

Обидві величини асимптотично нормальні й мають незалежні розподіли. За допомогою s можна перевірити, чи існує тенденція в середній, а d дає змогу виявити тенденцію в дисперсіях. Величина s , що характеризує зміну часового ряду, може набувати значень від 0 (усі рівні ряду рівні між собою) до $n - 1$ (ряд монотонний). Величина d характеризує зміну дисперсії рівнів часового ряду і змінюється від $-(n - 1)$ (ряд монотонно спадає) до $(n - 1)$ (ряд монотонно зростає).

3. Перевіряють гіпотези:

1) про рівність s і μ ($H_0 : s = \mu$), де μ – математичне сподівання величини s для ряду, в якому всі рівні випадкові;

2) про рівність $d = 0$.

Якщо на підставі спостережених значень гіпотези приймають, це означає, що відповідних тенденцій не існує.

Для перевірки гіпотез обчислюють вибіркові значення статистик Стьюдента для середніх і дисперсії:

$$t_s = \frac{|s - \mu|}{\sigma_1}; \quad \sigma_1 = \sqrt{2 \ln n - 3,4253};$$

$$t_d = \frac{|d - 0|}{\sigma_2}; \quad \sigma_2 = \sqrt{2 \ln n - 0,8456},$$

де σ_1, σ_2 — середньоквадратичні відхилення відповідно для величин s і d .

Для розрахунків використовують табельовані значення величин μ, σ_1, σ_2 ; фрагмент цих значень представлено в табл. 5.3.

Таблиця 5.3

n	10	15	20	30	40
μ	3,858	4,636	5,195	5,990	6,557
σ_1	1,288	1,41	1,677	1,882	2,019
σ_2	1,964	2,14	2,279	2,447	2,561

4. Розрахункові значення t_s і t_d порівнюють з табличним значенням t -критерію Стьюдента $t_{\text{табл}} = t(\alpha, k)$, де α — заданий рівень значущості, $k = n - 1$ — число степенів вільності.

Якщо розрахункове значення менше табличного $t_{\text{експ}}^s < t_{\text{табл}}$, гіпотезу про рівність s і μ приймають. Це означає, що відхилення між s і μ є випадковими, отже, тренда немає. Якщо $t_{\text{експ}}^d > t_{\text{табл}}$, то розбіжності між s і μ є не випадковими (суттєвими), отже, тренд є.

Аналогічно перевіряють і другу гіпотезу. Якщо $t_{\text{експ}}^d > t_{\text{табл}}$, то існує тренд дисперсії.

Зокрема, якщо $t_{\text{експ}}^s > t_{\text{табл}}$, а $t_{\text{експ}}^d < t_{\text{табл}}$, то часовий ряд має тренд у середньому, а тренда дисперсії рівнів ряду немає.

5.3.3. Кумулятивний критерій

Згідно з цим критерієм вибірккову статистику обчислюють як відношення суми відхилень рівнів ряду y_t від середнього рівня \bar{y}_t і самих відхилень:

$$T = \frac{\sum_{i=1}^n z_i^2}{s_y^2},$$

де s_y^2 — загальна сума квадратів відхилень:

$$s_y^2 = \sum_{i=1}^n (y_t - \bar{y}_t)^2,$$

z_i — сума відхилень, що відображає тенденцію їх зміни (накопичені відхилення):

$$z_i = \sum_{t=1}^i (y_t - \bar{y}_t).$$

Якщо довжина ряду велика, можна використати нормоване відхилення:

$$t = \frac{T - \left(\frac{n+1}{6}\right)}{\sqrt{(n-2) \frac{2n-1}{90}}}.$$

Фактичне значення T або t порівнюють з критичним при заданому рівні значущості α . Якщо фактичне значення перевищує критичне, гіпотезу про відсутність тренда відкидають, інакше наявність тренда вважають істотною.

Критичні значення кумулятивного критерію T і t наведено в табл. 5.4.

Таблиця 5.4

Обсяг вибірки, N	Критичні значення	
	T	t
6	2,62	2,08
7	3,11	2,10
8	3,59	2,09
9	4,07	2,09
10	4,55	2,09
11	5,02	2,08
12	5,49	2,08
13	5,96	2,07
14	6,42	2,04
15	6,89	2,06
16	7,36	2,06
17	7,82	2,06
18	8,29	2,05
19	8,76	2,05
20	9,22	2,04

Приклад 5.3. Фірма досліджує наявність тенденції прибутковості за 15 років свого існування. Після первинних статистичних досліджень отримано такі результати:

1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
14,1	9,3	19,4	19,7	5,4	24,2	13,8	24,5	14,7	16,6	5,6	16,2	25,3	11,9	18,5

Зобразимо графічно зазначений часовий ряд (рис. 5.8).

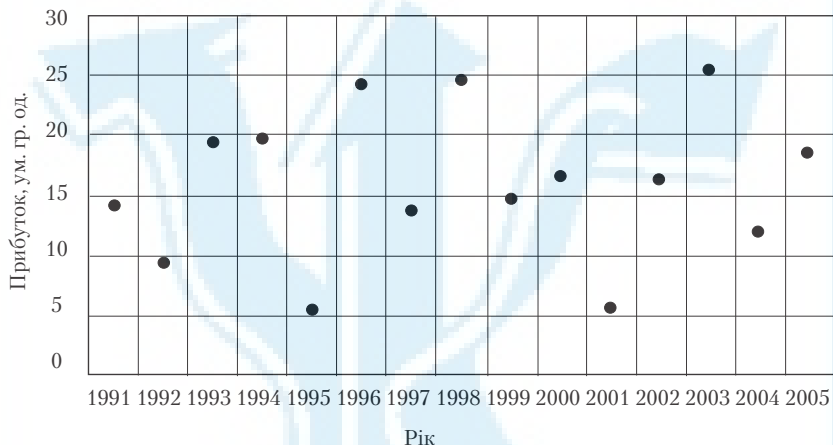


Рис. 5.8

Визначимо наявність основної тенденції (тренда), порівнюючи середні рівні частин ряду.

Ділимо ряд на дві частини обсягом відповідно $n_1 = 7$ і $n_2 = 8$. Для кожної обчислимо середні й дисперсії:

	Рік	Y	$(Y - Y_{\text{сеп}})^2$
	1991	14,10	1,06
	1992	9,30	33,97
	1993	19,40	18,25
	1994	19,70	20,90
	1995	5,40	94,65
	1996	24,20	82,29
	1997	13,80	8,19
Середнє		15,13	
Сума			259,30
Sigma^21 =			43,22

Рік	Y	(Y - Yсеп)^2
1998	24,50	61,43
1999	14,70	3,85
2000	16,60	0,00
2001	5,60	122,38
2002	16,20	0,21
2003	25,30	74,61
2004	11,90	22,68
2005	18,50	3,38

Середнє	16,66	
Сума		288,54
Sigma^2 =		41,22

Отримуємо $\bar{y}_1 = 15,13$; $\bar{y}_2 = 16,66$. $s_1^2 = 43,22$; $s_2^2 = 41,22$;

Перевіряємо допоміжну гіпотезу про рівність дисперсій при рівні значущості $\alpha = 0,05$:

$$H_1^{\text{доп}} : \sigma_1^2 = \sigma_2^2; H_1^{\text{доп}} : \sigma_1^2 \neq \sigma_2^2.$$

Розрахункове значення F -критерію $F_{\text{експ}} = \frac{s_1^2}{s_2^2} = 1,05$ порівняємо з табличним $F_{\text{табл}} = F(0,05; 6, 7) = 3,87$.

Оскільки $F_{\text{експ}} < F_{\text{табл}}$, немає підстав відкидати допоміжну гіпотезу. За даними спостереження, дисперсії генеральних сукупностей рівні: $\sigma_1^2 = \sigma_2^2$, виправлені вибіркові дисперсії (s_1^2 і s_2^2) різняться несуттєво (розбіжність між ними випадкова). Отже, можна перевіряти основну гіпотезу:

$$H_0 : \bar{y}_1 = \bar{y}_2 \quad H_1 : \bar{y}_1 \neq \bar{y}_2;$$

Для цього обчислюємо:

$$t_{\text{експ}} = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}} \sqrt{\frac{n_1 n_2 (n_1 + n_2 - 2)}{n_1 + n_2}} \approx -0,46.$$

Оскільки $|t_{\text{експ}}| < t_{\text{табл}}$, де $t_{\text{табл}} = t(0,05; 13) = 2,16$, немає підстав відкидати нульову гіпотезу про рівність середніх, розбіжність між ними незначна. Отже, тренд відсутній.

Метод Фостера–Стюарта.

Рік	Y	kt	lt	st = kt + lt	dt = kt - lt
1991	14,1	0	0	0	0
1992	9,3	0	1	1	-1
1993	19,4	1	0	1	1
1994	19,7	1	0	1	1
1995	5,4	0	1	1	-1
1996	24,2	1	0	1	1
1997	13,8	0	0	0	0
1998	24,5	1	0	1	1
1999	14,7	0	0	0	0
2000	16,6	0	0	0	0
2001	5,6	0	0	0	0
2002	16,2	0	0	0	0
2003	25,3	1	0	1	1
2004	11,9	0	0	0	0
2005	18,5	0	0	0	0
Суми				S	d
				7	3

З табл. 5.3 вибираємо: $\mu = 4,636$, $\sigma_1 = 1,41$, $\sigma_2 = 2,14$ при $n = 15$. Тоді при $s = 5 + 2 = 7$, $d = 5 - 2 = 3$, дорівнюють експериментальні значення відповідних статистик $t_{\text{експ}}^d = \frac{3-0}{2,14} = 1,68$; $t_{\text{експ}}^s = \frac{7-4,636}{1,41} = 1,40$. За таблицями розподілу Стьюдента $t_{\text{табл}} = t(0,05; 14) = 2,14$.

Оскільки $t_{\text{експ}}^s < t_{\text{табл}}$, немає підстав відкидати гіпотезу про відсутність тенденції в середній. Отже, метод Фостера–Стюарта ще раз підтвердив, що тренд у динаміці відсутній.

Оскільки $t_{\text{експ}}^d < t_{\text{табл}}$, гіпотезу про відсутність тенденції зміни в дисперсіях не відкидаємо, тенденції немає.

За кумулятивним критерієм отримаємо:

Рік	Y	Y - Yсер	(Y - Yсер)^2	Zi	Zi^2
1	2	3	4	5	6
1991	14,10	-1,85	3,41	-1,85	3,41
1992	9,30	-6,65	44,18	-8,49	72,14
1993	19,40	3,45	11,93	-5,04	25,40

1	2	3	4	5	6
1994	19,70	3,75	14,09	-1,29	1,66
1995	5,40	-10,55	111,23	-11,83	140,03
1996	24,20	8,25	68,12	-3,58	12,82
1997	13,80	-2,15	4,61	-5,73	32,79
1998	24,50	8,55	73,16	2,83	7,99
1999	14,70	-1,25	1,55	1,58	2,50
2000	16,60	0,65	0,43	2,23	4,99
2001	5,60	-10,35	107,05	-8,11	65,83
2002	16,20	0,25	0,06	-7,86	61,78
2003	25,30	9,35	87,48	1,49	2,23
2004	11,90	-4,05	16,38	-2,55	6,52
2005	18,50	2,55	6,52	0,00	0,00

Середнє

15,95

Суми

550,20

440,07

$T = 0,80$

$t = -0,91$

Порівнюємо обчислені значення $t = -0,91$; $T = 0,80$ з відповідними критичними значеннями табл. 5.5:

$$T_{\text{крит}} = 6,89; \quad t_{\text{крит}} = 2,06.$$

Оскільки $T < T_{\text{крит}}$, $t < t_{\text{крит}}$, доходимо висновку, що тренда немає.

5.4. Моделювання тенденції часового ряду

Після встановлення наявності тенденції динамічного ряду переходять до її моделювання.

Найпростішим способом моделювання тенденції соціально-економічного явища є згладжування (аналітичне вирівнювання) часового ряду. Існують різні прийоми згладжування, але суть їх одна – заміна фактичних рівнів ряду розрахунковими, які мають меншу варіативність, ніж початкові дані.

Найпоширенішими є лінійні тренди, загальна формула яких має вигляд

$$\bar{y}_t = \sum_{\tau=-q}^s \alpha_{\tau} y_{t+\tau}, \quad (5.5)$$

де \bar{y}_t – згладжене (вирівняне) значення рівня на момент t ; α_τ – вага рівня ряду, який знаходиться на відстані τ від моменту t ; s – кількість рівнів після моменту t ; q – кількість рівнів до моменту t .

Залежно від того, якого значення набуває вага α_τ , згладжування за формулою (5.5) виконують за допомогою або ковзних середніх, або експоненціальних середніх.

Процес вирівнювання складається з двох основних етапів: вибору типу кривої та оцінювання параметрів кривої.

Існують різні прийоми, які дають змогу так вибрати форму кривої, щоб вона достатньо добре описувала дійсний розвиток процесу. Найпростіший шлях, як зазначалося раніше, є *візуальний*, на основі графічного зображення часового ряду – кореляційного поля. За розташуванням точок на графіку добирають рівняння кривої, яка найближче підходить до емпіричного тренда (траєкторії). Це можуть бути:

1) поліноми:

$$\bar{y}_t = a_0 + a_1 t \text{ – першого степеня;}$$

$$\bar{y}_t = a_0 + a_1 + a_2 t^2 \text{ – другого степеня;}$$

$$\bar{y}_t = a_0 + a_1 + a_2 t^2 + a_3 t^3 \text{ – третього степеня;}$$

$$\bar{y}_t = a_0 + a_1 + a_2 t^2 \dots + a_k t^k \text{ – } \kappa\text{-го степеня;}$$

2) різні експоненти:

$$\bar{y}_t = a_0 + a_1^t;$$

$$\bar{y}_t = a_0 + a_1^{b_1 t + b_2 t^2};$$

$$y_t = a_0 a_1^t \text{ – модифікована експонента;}$$

3) логістичні криві:

$$\bar{y}_t = \frac{k}{1 + a_0 e^{-a_1 t}},$$

де e – основа натурального логарифма,

$$\bar{y}_t = \frac{k}{1 + 10^{a_0 + a_1 t}};$$

4) крива Гомперца

$$y = e^{ab^t + c}.$$

Форму кривої можна виявити за допомогою *методу послідовних різниць*. Сутність цього методу полягає в обчисленні перших, других і т. д. різниць рівнів, тобто

$$\Delta_t^1 = y_t - y_{t-1}; \Delta_t^2 = \Delta_t^1 - \Delta_{t-1}^1; \Delta_t^3 = \Delta_t^2 - \Delta_{t-1}^2.$$

Обчислення виконують доти, поки різниці не стануть приблизно рівними. Порядок цих різниць i є порядком шуканого полінома.

Приклад 5.4. Фірма досліджує рівень прибутковості за 15 років свого існування. Отримано такі дані:

Рік	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Рівень прибутку, ум. гр. од.	0,81	0,85	0,9	0,94	0,98	1,03	1,07	1,12	1,16	1,2	1,26	1,31	1,35	1,39	1,42

1. Досліджуємо наявність тенденції.

Рік	Рівень прибутку, ум. гр. од.	k_t	l_t	s_t	d_t	Δ_t^1	Δ_t^2	Δ_t^3	Δ_t^4
1991	0,81	0	0	0	0	–	–	–	–
1992	0,85	1	0	1	1	0,04	–	–	–
1993	0,90	1	0	1	1	0,05	0,01	–	–
1994	0,94	1	0	1	1	0,04	-0,01	-0,02	–
1995	0,98	1	0	1	1	0,04	0,00	0,01	0,03
1996	1,03	1	0	1	1	0,05	0,01	0,01	0,00
1997	1,07	1	0	1	1	0,04	-0,01	-0,02	-0,03
1998	1,12	1	0	1	1	0,05	0,01	0,02	0,04
1999	1,16	1	0	1	1	0,04	-0,01	-0,02	-0,04
2000	1,20	1	0	1	1	0,04	0,00	0,01	0,03
2001	1,26	1	0	1	1	0,06	0,02	0,02	0,01
2002	1,31	1	0	1	1	0,05	-0,01	-0,03	-0,05
2003	1,35	1	0	1	1	0,04	-0,01	0,00	0,03
2004	1,39	1	0	1	1	0,04	0,00	0,01	0,01
2005	1,42	1	0	1	1	0,03	-0,01	-0,01	-0,02
Суми						14	14		

Обчислимо k_t, l_t, Δ_t і перевіримо наявність тенденції середнього рівня прибутку фірми за методом Форстера–Стюарта.

$$\text{Маємо } s = 14; \mu = 4,6; \sigma_1 = 1,521. \text{ Тоді } T_{\text{іексп}}^s = \frac{14 - 4,636}{1,521} \approx 5,9.$$

Оскільки $T_{\text{іексп}}^S > t_{\text{табл}}$, де $t_{\text{табл}}(0,01;14) = 2,14$, гіпотезу про відсутність тенденції в досліджуваному ряді відкидаємо. Тенденція наявна, і її потрібно представити у вигляді деякої функції.

За допомогою послідовних різниць визначимо можливий порядок полінома, який можна прийняти за рівняння тренда.

З таблиці видно, що перші різниці можна вважати практично рівними, а середнє арифметичне різниць другого порядку близьке до нуля, ними можна знехтувати. Отже, тенденцію досліджуваного ряду можна описати поліномом першого степеня, тобто

$$\bar{y}_t = a_0 + a_1 t + \varepsilon_t.$$

У нашому прикладі засобами *Excel* отримали таке рівняння:

$$\bar{y} = 0,7619 + 0,0447t + \varepsilon.$$

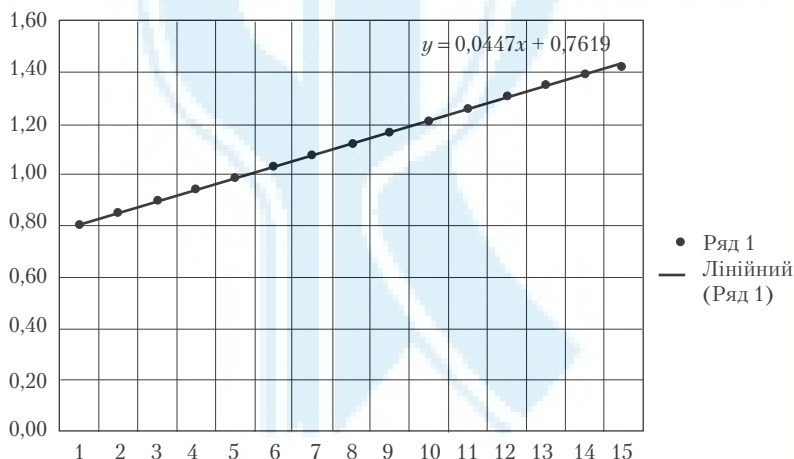


Рис. 5.9

Недолік цього методу в тому, що він прийнятний лише в разі добору кривих, описуваних поліномами.

Форму кривої можна вибрати, виходячи з *теоретичного аналізу* сутності досліджуваного явища і спираючись на досвід і знання дослідника. Якщо рівні ряду зростають в арифметичній прогресії, згладжування проводять за прямою, а якщо зростання рівнів близьке до геометричної прогресії, згладжування треба проводити за показниковою функцією. Для вирівнювання часових рядів економічних явищ,

що наближаються до деякої граничної величини (характеризуються насиченням), застосовують логістичні функції.

Добираючи функцію тренда, разом з теоретичним аналізом закономірностей розвитку явища використовують і емпіричні методи.

Після вибору форми кривої необхідно оцінити параметри відповідної моделі. Цю задачу розв'язують в основному методом найменших квадратів.

5.5. Моделювання сезонних і циклічних коливань

Моделювання циклічних коливань загалом здійснюється так само, як і сезонних коливань, тому розглянемо лише методи моделювання останніх.

Відомо кілька підходів до аналізу структури часових рядів, які містять сезонні або циклічні коливання:

- аналіз сезонності за допомогою автокореляційної функції;
- розрахунок сезонної компоненти і побудова адитивної чи мультиплікативної моделі часового ряду;
- застосування сезонних фіктивних змінних;
- застосування рядів Фур'є.

Найпростіший підхід – розрахунок значень сезонної компоненти методом ковзного середнього та побудова адитивної та мультиплікативної моделі часового ряду.

Якщо амплітуда сезонних коливань не змінюється з часом, застосовують адитивну модель часового ряду, загальний вигляд якої

$$Y = T + S + E.$$

Ця модель передбачає, що кожен рівень часового ряду може бути представлений як сума трендової T , сезонної S і випадкової E компонент.

Якщо амплітуда сезонних коливань змінюється з часом, застосовують мультиплікативну модель, загальний вигляд якої

$$Y = T \cdot S \cdot E.$$

Ця модель передбачає, що кожен рівень часового ряду може бути представлений як добуток трендової T , сезонної S і випадкової E компонент. Вибір однієї з двох моделей проводять на основі аналізу структури сезонних коливань.

Побудова адитивної та мультиплікативної моделей зводиться до розрахунку значень T , S і E для кожного рівня.

Процес побудови моделі складається з таких кроків.

Крок 1. Вирівнювання вихідного ряду методом ковзного середнього.

Крок 2. Обчислення значень сезонної компоненти S .

Крок 3. Усунення сезонної компоненти з вихідних рівнів ряду та отримання вирівняних даних $(T + E)$ в адитивній моделі або $(T \cdot E)$ в мультиплікативній моделі.

Крок 4. Аналітичне вирівнювання рівнів $(T + E)$ або $(T \cdot E)$ і розрахунок значень T з використанням отриманого рівняння тренда.

Крок 5. Розрахунок отриманих за моделлю значень $(T + S)$ або $(T \cdot S)$.

Крок 6. Обчислення абсолютних і/або відносних похибок.

Якщо отримані значення похибок не містять автокореляції, ними можна замінити вихідні рівні ряду і в подальшому використовувати часовий ряд похибок E для аналізу взаємозв'язку вихідного ряду та інших часових рядів.

5.5.1. Адитивна модель часового ряду

Розглянемо приклад побудови адитивної моделі часового ряду. Звернемося до даних про прибутки фірми за останні чотири роки з прикладу 5.2.

Раніше було показано, що цей часовий ряд містить сезонні коливання з періодичністю 4. Обсяги прибутків в осінньо-зимовий період (I і IV квартали) вищі, ніж навесні та влітку (II і III квартали). За графіком цього ряду можна встановити наявність приблизно рівної амплітуди коливань. Це свідчить про відповідність цього ряду адитивній моделі. Розрахуємо її компоненти.

Крок 1. Проведемо вирівнювання вихідних рівнів ряду методом ковзного середнього:

- складемо рівні ряду послідовно за кожні чотири квартали із зсувом на один момент часу та визначимо умовні річні обсяги прибутків;
- розділивши отримані суми на 4, знайдемо ковзні середні. Зазначимо, що отримані вирівняні значення вже не містять сезонної компоненти;
- приведемо ці значення у відповідність з фактичними моментами часу, для цього знайдемо середні значення з двох послідовних ковзних середніх — центровані ковзні середні.

Таблиця 5.5

Номер кварталу	Середній прибуток, ум. гр. од.	Усього за чотири квартали	Ковзна середня за чотири квартали	Центрована ковзна середня	Оцінка сезонної компоненти
1	6,0				
2	4,4				
3	5,0	–	–	–	–
4	9,0	24,4	6,10	6,250	–1,250
5	7,2	25,6	6,40	6,450	2,550
6	4,8	26,0	6,50	6,625	0,575
7	6,0	27,0	6,75	6,875	–2,075
8	10,0	28,0	7,00	7,100	–1,100
9	8,0	28,8	7,20	7,300	2,700
10	5,6	29,6	7,40	7,450	0,550
11	6,4	30,0	7,50	7,625	–2,025
12	11,0	31,0	7,75	7,875	–1,475
13	9,0	32,0	8,00	8,125	2,875
14	6,6	33,0	8,25	8,325	0,675
15	7,0	33,6	8,40	8,375	–1,775
16	10,8	33,4	8,35	–	–

Крок 2. Знайдемо оцінки сезонної компоненти як різницю між фактичними рівнями ряду та центрованими ковзними середніми. Використаємо ці оцінки для розрахунку значень сезонної компоненти S_t . Для цього визначимо середні за кожен квартал (по всіх роках) оцінки сезонної компоненти S_t . У моделях із сезонною компонентою зазвичай припускають, що сезонні впливи за період взаємознищуються. В адитивній моделі це виявляється в тому, що сума значень сезонної компоненти по всіх кварталах має дорівнювати нулю.

$$\bar{P}_1 = \frac{1}{4}(6,0 + 7,2 + 8,0 + 9,0) = 7,55;$$

$$\bar{P}_2 = \frac{1}{4}(4,4 + 4,8 + 5,6 + 6,6) = 5,35;$$

$$\bar{P}_3 = \frac{1}{4}(5,0 + 6,0 + 6,4 + 7,0) = 6,1;$$

$$\bar{P}_4 = \frac{1}{4}(9,0 + 10,0 + 11,0 + 10,8) = 10,2.$$

Таблиця 5.6

Показник	Рік	Номер кварталу 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	–	–
Усього за i -й квартал (за всі роки)	×	1,800	–5,875	–3,825	8,125
Середня оцінка сезонної компоненти для i -го кварталу S_i	×	0,600	–1,958	–1,275	2,708
Скоригована сезонна компонента S_i	×	0,581	–1,977	–1,294	2,690

Для даної моделі сума середніх оцінок сезонної компоненти дорівнює

$$0,6 - 1,958 - 1,275 + 2,708 = 0,075.$$

Отже, коригуючий коефіцієнт

$$k = 0,075 / 4 = 0,01875.$$

Розрахуємо скориговані значення сезонної компоненти (останній рядок таблиці) як різницю між її середньою оцінкою та коригуючим коефіцієнтом k :

$$S_i = \bar{S}_i - k,$$

де $i = 1, 2, 3, 4$.

Перевіримо умову рівності нулю суми значень сезонної компоненти:

$$0,581 - 1,977 - 1,294 + 2,690 = 0.$$

Отже, отримано такі значення сезонної компоненти:

I квартал: $S_1 = 0,581$;

II квартал: $S_2 = -1,979$;

III квартал: $S_3 = -1,294$;

IV квартал: $S_4 = 2,690$.

Занесемо ці значення в таблицю для відповідних кварталів кожного року.

Крок 3. Вилучимо (елімінуємо) вплив сезонної компоненти відніманням її значення від кожного рівня вихідного часового ряду: $T + E = Y - S$. Ці значення, розраховані для кожного моменту часу, містять лише тенденцію та випадкову компоненту.

Таблиця 5.7

t	y_t	S_i	$T + E = y_t - S_i$	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	0,1282
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

Крок 4. Визначимо трендову компоненту T моделі, провівши аналітичне вирівнювання ряду $(T + E)$ за допомогою лінійного тренда. Отримаємо лінійний тренд

$$T = 5,365 + 0,2276 t.$$

Підставивши в це рівняння значення $t = 1, \dots, 16$, знайдемо рівні T для кожного моменту часу. Графік рівняння тренда зображено на рис. 5.9.

Крок 5. Знайдемо значення рівнів ряду, отримані за адитивною моделлю. Для цього додамо до рівнів T значення сезонної компоненти для відповідних кварталів. Графічно значення $(T + S)$ подано на рис. 5.9.

Крок 6. Відповідно до методики побудови адитивної моделі похибку обчислюють за формулою

$$E = Y - (T + S).$$

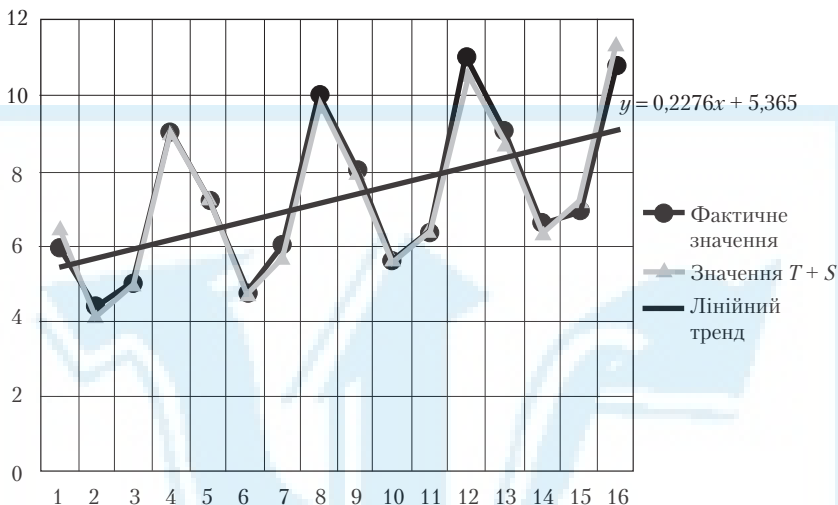


Рис. 5.9

Це абсолютна похибка. Числові значення абсолютних похибок наведено в табл. 5.7 (передостанній стовпець).

За аналогією з моделлю регресії для оцінки якості побудови моделі і вибору найкращої моделі визначають коефіцієнт детермінації R^2 :

$$R^2 = 1 - \frac{1,10}{71,59} = 0,985.$$

Це означає, що адитивна модель пояснює 98,5 % загальної варіації рівнів часового ряду середніх прибутків за останні 16 кварталів.

5.5.2. Мультиплікативна модель

Методику побудови мультиплікативної моделі можна проілюструвати на такому прикладі.

Приклад 5.5. Нехай є поквартальні дані про прибутки компанії за чотири роки (табл. 5.8):

Таблиця 5.8

Рік	Квартал			
	I	II	III	IV
1	72	100	90	64
2	70	92	80	58
3	62	80	68	48
4	52	60	50	30

Графік часового ряду (рис. 5.10) свідчить про наявність сезонних коливань (період коливань дорівнює чотирьом) і загальної спадної тенденції рівнів ряду. Прибуток компанії у весняно-літній період вищий, ніж в осінньо-зимовий. Оскільки амплітуда сезонних коливань зменшується, модель можна вважати мультиплікативною. Визначимо її основні компоненти.

Крок 1. Проведемо вирівнювання вихідних рівнів ряду методом ковзного середнього. Методика, яка застосовується на цьому кроці, повністю збігається з методикою адитивної моделі. Результати розрахунків оцінок сезонної компоненти наведено в табл. 5.9.

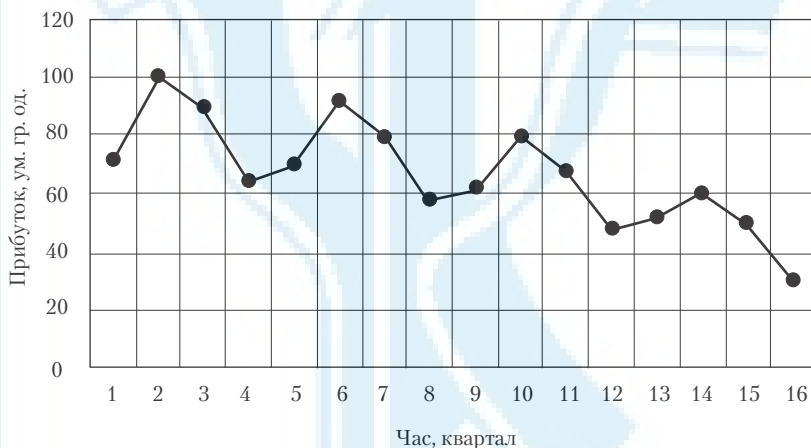


Рис. 5.10

Крок 2. Знайдемо оцінки сезонної компоненти як частку від ділення фактичних рівнів ряду на центровані ковзні середні (табл. 5.9). Використаємо ці оцінки для розрахунку значень сезонної компоненти S (табл. 5.10). Для цього знайдемо середні за кожен квартал оцінки сезонної компоненти S_t . Взаємокомпенсація сезонних впливів у мультиплікативній моделі виявляється в тому, що сума значень сезонної компоненти по всіх кварталах має дорівнювати кількості періодів у циклі, тобто чотирьом, оскільки в нашому прикладі кількість періодів одного циклу (рік) дорівнює чотирьом кварталам.

Таблиця 5.9

Номер кварталу t	Прибуток компанії y_t	Усього за чотири квартали	Ковзна середня за чотири квартали	Центрована ковзна середня	Оцінка сезонної компоненти
1	72				
2	100				
3	90	–	–	–	–
4	64	326	81,5	81,25	1,108
5	70	324	81,0	80,00	0,800
6	92	316	79,0	77,75	0,900
7	80	306	76,5	75,75	1,215
8	58	300	75,0	74,00	1,081
9	62	292	73,0	71,50	0,811
10	80	280	70,0	68,50	0,905
11	68	268	67,0	65,75	1,217
12	48	258	64,5	63,25	1,075
13	52	248	62,0	59,50	0,807
14	60	228	57,0	54,75	0,950
15	50	210	52,5	50,25	1,194
16	30	192	48,0	–	–

Таблиця 5.10

Показник	Рік	Квартал			
		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 за i -й квартал (за всі роки)	×	2,755	3,626	3,264	2,424
Середня оцінка сезонної компоненти для i -го кварталу \bar{S}_i	×	0,918	1,209	1,088	0,808
Скоригована сезонна компонента S_i	×	0,913	1,202	1,082	0,803

Маємо суму середніх значень

$$0,918 + 1,209 + 1,088 + 0,808 = 4,203.$$

Коригуючий коефіцієнт

$$k = 4 / 4,023 = 0,9943.$$

Визначимо скориговані значення сезонної компоненти множенням її середньої оцінки на коригуючий коефіцієнт k :

$$S_i = \bar{S}_i k,$$

де $i = 1, 2, 3, 4$.

Отримаємо такі сезонні компоненти по кварталах:

I квартал: $S_1 = 0,913$;

II квартал: $S_2 = 1,202$;

III квартал: $S_3 = 1,082$;

IV квартал: $S_4 = 0,803$.

Результати розрахунків внесемо до табл. 5.10 (останній рядок).

Крок 3. Розділимо кожен рівень вихідного ряду на відповідні значення сезонної компоненти. Отримаємо $T \cdot E = Y / S$, які містять лише тенденцію та випадкову компоненту.

Таблиця 5.11

t	y_t	S_t	$T \cdot E = Y / S$	T	$T \cdot S$	$E = y_t : (T \cdot S)$	$V = y_t - (T \cdot S)$	$(V)^2$
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

Крок 4. Визначимо компоненту T у мультиплікативній моделі. Для цього визначимо параметри лінійного тренда, використовуючи рівні $(T \cdot S)$.

Отримаємо рівняння тренда:

$$T = 92,1 - 2,9235t.$$

Підставимо в це рівняння значення $t = 1, \dots, 16$ і знайдемо рівні T для кожного моменту часу. Графік тренда зображено на рис. 5.11.

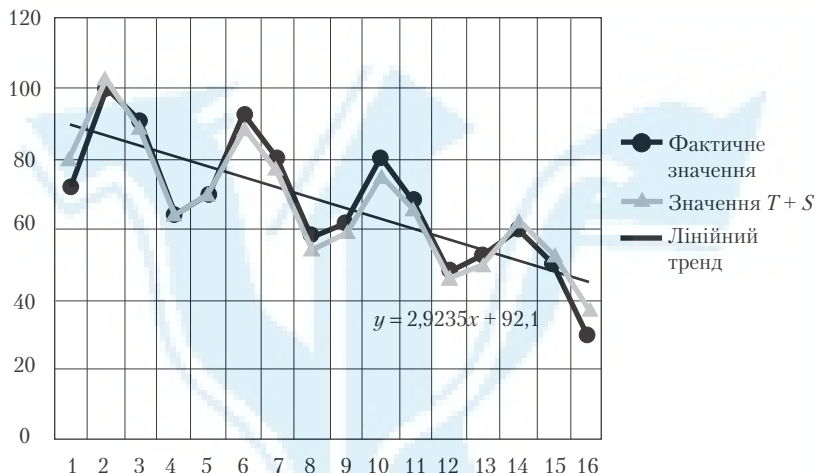


Рис. 5.11

Крок 5. Знайдемо рівні ряду за мультиплікативною моделлю множенням рівнів T на значення сезонної компоненти для відповідних кварталів (рис. 5.11).

Крок 6. Розрахунок похибки в мультиплікативній моделі виконуємо за формулою

$$E = Y : (T \cdot S).$$

Числові значення похибки наведено в табл. 5.11. Якщо часовий ряд похибок не містить автокореляції, його можна використовувати замість вихідного ряду для вивчення взаємозв'язку з іншими часовими рядами.

Для того щоб порівняти мультиплікативну модель з іншими моделями часового ряду, за аналогією з адитивною моделлю використовують коефіцієнт детермінації, але при цьому абсолютні похибки мультиплікативної моделі обчислюють за формулою

$$V = y_t - (T \cdot S).$$

У цій моделі сума квадратів абсолютних похибок становить 207,40, загальна сума квадратів відхилень фактичних рівнів цього ряду від середнього значення дорівнює 5023, при цьому коефіцієнт детермінації

$$R^2 = 1 - \frac{207,4}{5023} = 0,959.$$

Отже, частина пояснювальної дисперсії рівнів ряду становить 95,9 %.

Виявлення та усунення сезонного ефекту (у деяких джерелах вживають термін “десезоналізація рівнів ряду”) потрібні насамперед для аналізу структури одновимірних часових рядів і прогнозування рівнів ряду в наступні моменти часу, а також при вивченні взаємозв’язку кількох часових рядів. Вплив сезонних коливань слід усувати на етапі попередньої обробки вихідних даних.

Задача 5.1. За даними прикладу 5.2 спрогнозувати прибуток фірми на наступне півріччя за адитивною моделлю часового ряду.

Прогнозоване значення F_t рівня часового ряду в адитивній моделі згідно із співвідношенням

$$Y = T + S + E$$

визначають як суму трендової та сезонної компонент.

Обсяг прибутку, отриманого упродовж першого півріччя найближчого наступного, тобто п’ятого, року, розраховують як суму обсягів прибутку в I і в II кварталах п’ятого року — відповідно F_{17} і F_{18} . Для визначення трендової компоненти скористаємось рівнянням тренда

$$T = 5,365 + 0,2276t.$$

Отримаємо:

$$T_{17} = 5,365 + 0,2276 \cdot 17 = 9,2342;$$

$$T_{18} = 5,365 + 0,2276 \cdot 18 = 9,4618.$$

Значення сезонної компоненти:

$$S_1 = 0,581 \text{ (I квартал);}$$

$$S_2 = -1,977 \text{ (II квартал).}$$

Отже,

$$F_{17} = T_{17} + S_1 = 9,2342 + 0,581 = 9,8152;$$

$$F_{18} = T_{18} + S_2 = 9,4618 + 1,977 = 11,4388.$$

Прогноз обсягу прибутку на перше півріччя наступного (п’ятого) року такий:

$$9,8152 + 11,4388 = 21,254 \text{ (ум. гр. од.).}$$

Задача 5.2. За даними прикладу 5.5 спрогнозувати прибуток фірми на наступне півріччя за мультиплікативною моделлю часового ряду.

Прогнозоване значення F_t рівня часового ряду в мультиплікативній моделі згідно із співвідношенням

$$Y = T \cdot S \cdot E$$

розраховують як добуток трендової та сезонної компонент. Для визначення трендової компоненти за кожен квартал скористаємось рівнянням тренда

$$T = 92,1 - 2,9235t.$$

Отримаємо:

$$T_{17} = 92,1 - 2,9235 \cdot 17 = 42,401;$$

$$T_{18} = 92,1 - 2,9235 \cdot 18 = 39,477.$$

Значення сезонної компоненти такі:

$$S_1 = 0,913 \text{ (I квартал);}$$

$$S_2 = 1,202 \text{ (II квартал).}$$

Отже,

$$F_{17} = T_{17} \cdot S_1 = 42,401 \cdot 0,913 = 38,712;$$

$$F_{18} = T_{18} \cdot S_2 = 39,477 \cdot 1,202 = 47,451.$$

Прогноз обсягу прибутку на перше півріччя наступного (п'ятого) року такий:

$$38,712 + 47,451 = 86,163 \text{ (ум. гр. од.).}$$

5.6. Використання бінарних змінних у сезонному аналізі

Зазвичай незалежні змінні в регресійних моделях мають “неперервні” області змінювання (національний дохід, обсяги виробництва, розмір заробітної плати тощо), тобто є метрично (кількісно) вимірними величинами. У реальних ситуаціях на залежну змінну крім кількісних факторів впливають також і якісні: якість продукції, рівень професійної підготовки працівників, їхня стать, страйки, зміни в економічній політиці тощо. Часто змінні, що відображають якісні характеристики об'єкта, набувають лише двох значень: “1” — якщо певна ознака присутня і “0” — за її відсутності. Такі змінні називають бінарними, дихотомними або *dummy*-змінними.

Dummy variables у перекладі з англійської мови означає “фіктивні змінні”, хоча насправді їх “фіктивність” полягає лише в тому, що вони кількісно описують деяку якісну ознаку.

Dummy-змінні використовують у регресійних моделях разом з кількісними змінними або утворюють регресійні моделі, в яких всі фактори є якісними (бінарними) змінними. Регресійна модель, факторними змінними якої є бінарні змінні, називається регресійною моделлю зі змінною структурою.

Застосуємо таку модель до аналізу сезонних коливань.

Приклад 5.6. Нехай y — відомий обсяг споживання певного продукту. За допомогою бінарних змінних потрібно з'ясувати, чи залежить обсяг споживання від пори року.

Для виявлення сезонності вводять бінарні змінні d_1, d_2, d_3 :

$d_1 = 1$, якщо місяць року є зимовим, $d_1 = 0$ — в інших випадках;

$d_2 = 1$, якщо місяць року є весняним, $d_2 = 0$ — в інших випадках;

$d_3 = 1$, якщо місяць року є літнім, $d_3 = 0$ — в інших випадках.

Лінійна регресійна модель залежності має вигляд

$$y = a_0 + a_1 d_1 + a_2 d_2 + a_3 d_3 + u,$$

де коефіцієнт a_0 визначає середньомісячний обсяг споживання досліджуваного продукту; параметри a_1, a_2, a_3 вказують на сезонні відхилення в обсягах споживання продукту відносно осінніх місяців. Суми коефіцієнтів $a_0 + a_1, a_0 + a_2, a_0 + a_3$ визначають обсяг споживання відповідно взимку, навесні та влітку.

Параметри a_0, a_1, a_2, a_3 оцінюють методом найменших квадратів на базі відповідних статистичних даних про обсяг споживання y .

Перевірку статистичної значущості кожного з коефіцієнтів регресії виконують за допомогою t -тесту Стьюдента. Прийняття гіпотези про рівність нулю кожного з параметрів означає несуттєву різницю між осіннім споживанням та споживанням іншого сезону. Комплексну гіпотезу $a_1 = a_2 = a_3$ перевіряють за допомогою F -тесту. Зокрема, якщо приймають припущення $a_1 = a_2$, це означає, що показники споживання взимку та навесні не різняться між собою.

Розглянемо інші приклади застосування фіктивних змінних.

Як зазначалося, у лінійній регресійній моделі залежності споживання від доходу

$$Y = a_0 + a_1 I + u,$$

де y — середньомісячний обсяг споживання деякого індивіда, а I — його середньомісячний дохід, коефіцієнт a_1 називається “схильністю до споживання”. Щоб визначити вплив сезону на схильність до споживання, також застосовують бінарні змінні d_1, d_2, d_3 . Модель при цьому набирає вигляду

$$y = a_0 + a_1 d_1 + a_2 d_2 + a_3 d_3 + a_4 d_1 I + a_5 d_2 I + a_6 d_3 I + a_7 I + u.$$

Коефіцієнти цієї моделі $a_7, a_4 + a_7, a_5 + a_7, a_6 + a_7$ визначають схильність до споживання відповідно восени, взимку, весною та влітку. Як і в попередній моделі, гіпотези про відсутність сезонних впливів на схильність до споживання перевіряють за критерієм Стьюдента.

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

Приклад 5.7. Нехай досліджується залежність обсягу випущеної підприємством продукції y від обсягу його основного фонду x . Припускається, що при досягненні основним фондом підприємства розміру \bar{x} відбувається певна структурна перебудова підприємства. Залежність випуску продукції від основного фонду в результаті перебудови змінюється, але загалом залишається неперервною. У такому разі функція залежності матиме кусково-лінійний графік, що відображає така регресійна модель:

$$y = a_0 + a_1x + a_2(x - \bar{x})d + u,$$

де бінарна змінна $d = 0$, якщо $x \leq \bar{x}$, і $d = 1$, якщо $x > \bar{x}$.

Якщо в результаті тестування значущості параметрів моделі приймають гіпотезу $a_2 = 0$, це означає, що структурна зміна на підприємстві не відбулася.

Узагалі спосіб залучення до моделі *dummy*-змінних залежить від апріорної інформації відносно впливу якісних факторів на залежну змінну і від гіпотез, які необхідно перевірити на підставі цієї інформації. У свою чергу, цей самий спосіб визначає, як саме будуть інтерпретовані отримані оцінки параметрів моделі.

5.7. Спектральний аналіз часових рядів

Стационарний випадковий процес може бути представлений у вигляді суми гармонійних коливань різних частот, які називаються *гармоніками*. Функція, що описує розподіл амплітуд цього процесу за різними частотами, називається *спектром*. Спектр показує, якого роду коливання переважають у цьому процесі, яка його внутрішня структура.

Стационарна випадкова функція $X(t)$ може бути зображена у вигляді канонічного розкладання:

$$X(t) = \sum_{k=0}^{\infty} (u_k \cos \omega_k t + v_k \sin \omega_k t),$$

де u_k, v_k — некорельовані випадкові величини з нульовими математичними сподіваннями й однаковими дисперсіями, тобто $M(u_k) = M(v_k) = 0$, $D(u_k) = D(v_k) = D_k$.

Подібне зображення стаціонарного випадкового процесу $X = X(t)$ називають його спектральним розкладанням або спектром.

Дисперсія стаціонарної випадкової функції $X(t)$ дорівнює сумі дисперсій усіх гармонік її спектрального розкладання:

$$D[X(t)] = D\left[\sum_{k=0}^{\infty} (u_k \cos \omega_k t + v_k \sin \omega_k t)\right] = \\ = \sum_{k=0}^{\infty} (\cos^2 \omega_k t + \sin^2 \omega_k t) D_k = \sum_{k=0}^{\infty} D_k.$$

Виходячи з цього можна зробити висновок, що дисперсія величини $X(t)$ певним чином розподілена за різними частотами: одним частотам відповідає більша дисперсія, іншим — менша.

Покажемо розподіл дисперсій за частотами на графіку. На осі абсцис відкладаємо частоти $\omega_0 = 0, \omega_1, \omega_2, \dots, \omega_k$, а на осі ординат — відповідні дисперсії.

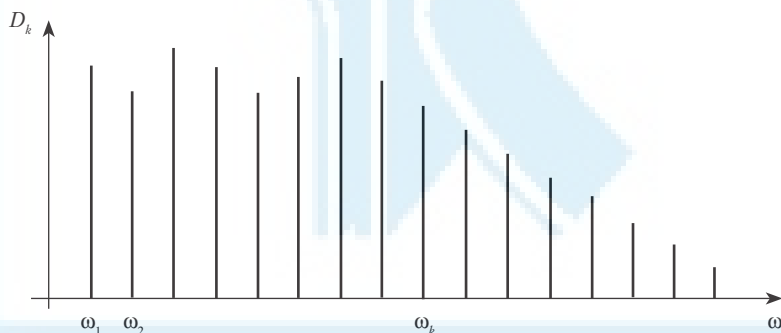


Рис. 5.12

На графіку спектр дисперсій випадкової величини на кінцевій ділянці часу $(0, T)$ має вигляд окремих дискретних ліній, розділених рівними проміжками (рис. 5.12). Такий спектр називається *дискретним*. Сума всіх побудованих ординат дорівнює дисперсії випадкового процесу $X = X(t)$. Чим більше ділянка часу $(0, T)$, тим повнішу інформацію можна одержати про випадкову функцію $X(t)$.

Якщо $T \rightarrow \infty$, $\omega_k \rightarrow \frac{2\pi}{T} \rightarrow 0$, відстані між частотами необмежено зменшуються. У цьому разі дискретний спектр наближається до *неперервного*, в якому кожному як завгодно малому інтервалові частот $\Delta\omega = \frac{2\pi}{T}$ відповідатиме елементарна дисперсія $\Delta D(\omega)$. Графік неперервного спектра вже інший. На осі ординат відкладаємо не дисперсію D_k , а середню щільність дисперсії, що припадає на одиницю довжини певного інтервалу частот. На кожному відрізку $\Delta\omega$ як на основі будуємо прямокутники площею D_k . Висота кожного прямокутника $\rho_x(\omega) = D_k / \Delta\omega$ — середня щільність дисперсії на цій ділянці. Сукупність усіх прямокутників утворить східчасту фігуру, що нагадує гістограму статистичного розподілу. Площа східчастої фігури дорівнює дисперсії випадкового процесу $X = X(t)$.

За необмеженого збільшення $T \rightarrow \infty$, $\Delta\omega \rightarrow 0$ східчаста крива необмежено наближається до плавної кривої $\rho_x(\omega)$. Ця крива і є графічним зображенням щільності розподілу дисперсій за частотами неперервного спектра (рис. 5.13).

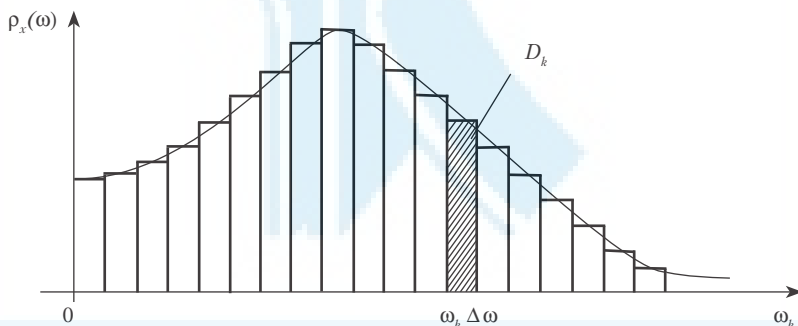


Рис. 5.13

Функція $\rho_x(\omega)$ називається *спектральною щільністю дисперсії* або *спектральною площиною* стаціонарної випадкової функції $X(t)$. Мета спектрального аналізу часових рядів — дати оцінку спектра ряду. Під *спектром часового ряду* розуміють розкладання дисперсії ряду за частотами для визначення істотних гармонічних складових.

Значення спектра оцінюють за формулою

$$f(\omega_j) = \frac{1}{2\pi} \left[\lambda_0 c_0 + 2 \sum_{k=1}^m \lambda_k c_k \cos \omega_j k \right], \quad (5.6)$$

де ω_j — частоти, для яких оцінюють спектри $\omega_j = \frac{\pi j}{m}$, $j = 1, 2, \dots, m$;

c_k — автоковаріаційна функція; λ_k — спеціально дібрані ваги значень коваріаційної функції, що залежать від частоти m ; λ_k ще називаються кореляційним вікном; m — ціле число, яке називають точкою відтину або числом використовуваних зрушень, являє собою кількість частотних смуг, для яких оцінюють спектр. Чим більше m , тим більше точок оцінюваного спектра, а отже, і більша дисперсія оцінки в кожній точці. Чим менше m , тим краща оцінка. Величина m залежить від довжини часового ряду.

Оцінка коваріації (значення автоковаріаційної функції)

$$c_k = \frac{1}{n-k} \left[\sum_{t=1}^{n-k} z_t z_{t+k} - \frac{1}{n-k} \sum_{t=k+1}^n z_t + \sum_{t=1}^{n-k} z_t \right],$$

де z_t — відхилення від тренда (\hat{y}_t).

Кореляційні вікна (ваги автоковаріаційної функції) λ_k можна розраховувати різними методами.

Існують оцінки Тьюкі–Хеннінга і Парзена. Оцінки Тьюкі–Хеннінга розраховують за формулою

$$\lambda_k = \frac{1}{2} \left(1 + \cos \frac{\pi k}{m} \right). \quad (5.7)$$

Підставляючи (5.7) у формулу (5.6), отримуємо

$$L_j = \frac{1}{2\pi} \left[c_0 + 2 \sum_{k=1}^{n-1} c_k \cos \left(\frac{\pi k j}{m} \right) + c_m \cos(\pi j) \right], \quad (5.8)$$

$$u_j = 0,25L_{j-1} + 0,5L_j + 0,25L_{j+1},$$

де $L_{-1} = L_{+1}$, $L_{m+1} = L_{m-1}$.

Значення L_j є наближеними оцінками спектра, а u_j — згладженіми.

Але оцінки Тьюкі–Хеннінга іноді дають від'ємні значення спектра, що суперечить його змістові (розкладання дисперсії ряду). У цьому разі від'ємні оцінки можна замінити нулями. Кращі результати дають оцінки Парзена, вони ніколи не набувають від'ємних значень. Ваги для них знаходять за формулою

$$\lambda_k = \begin{cases} 1 - \frac{6k^2}{m^2} \left(1 - \frac{k}{m}\right), & 0 \leq k \leq \frac{m}{2}; \\ 2 \left(1 - \frac{k}{m}\right)^3 - \frac{m}{2} + 1 \leq k \leq m. \end{cases} \quad (5.9)$$

Тоді формула (5.8) для розрахунку оцінок спектра набере вигляду

$$u_j = \frac{c_0}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{m/2} \left[1 - \frac{6k^2}{m^2} \left(1 - \frac{k}{m}\right) \right] c_k \cos\left(\frac{\pi k}{m} j\right) + \frac{2}{\pi} \sum_{k=\frac{m}{2}+1}^m \left(1 - \frac{k}{m}\right)^3 c_k \cos\left(\frac{\pi k}{m} j\right),$$

або

$$u_j = \frac{c_0}{2\pi} + \frac{1}{\pi} \sum_{k=1}^{m/2} \lambda_k c_k \cos\left(\frac{\pi k}{m} j\right) + \sum_{k=\frac{m}{2}+1}^m \lambda_k c_k \cos\left(\frac{\pi k}{m} j\right), \quad (5.10)$$

де λ_k визначають за формулою (5.16). У цьому разі значення автоковаріаційної функції

$$c_k = \frac{1}{n} \sum_{t=1}^{n-k} (z_t - \bar{z})(z_{t+k} - \bar{z}).$$

За максимальною оцінкою u_j знаходять пік спектра; гармонічна складова має період $t = \frac{2m}{j}$.

Спектральний аналіз, використаний у вивченні сезонних коливань, дає змогу виявити періодичні складові досліджуваного ряду з метою підвищення точності прогнозування.

Приклад 5.8. Розглянемо застосування методики спектрального аналізу динаміки отримання щомісячного прибутку фірмою за період 2001–2003 рр. (в умовних грошових одиницях). Нехай дані задано такою таблицею:

Умовний номер місяця	1	2	3	4	5	6	7	8	9	10	11	12
Прибуток, ум. гр. од.	13	43	76	92	131	186	186	152	110	69	10	4
Умовний номер місяця	13	14	15	16	17	18	19	20	21	22	23	24
Прибуток, ум. гр. од.	21	37	58	78	133	209	186	167	122	62	27	22

Умовний номер місяця	25	26	27	28	29	30	31	32	33	34	35	36
Прибуток, ум. гр. од.	39	52	96	111	191	219	198	170	111	42	20	34

Зобразимо графічно дану інформацію (рис. 5.14).

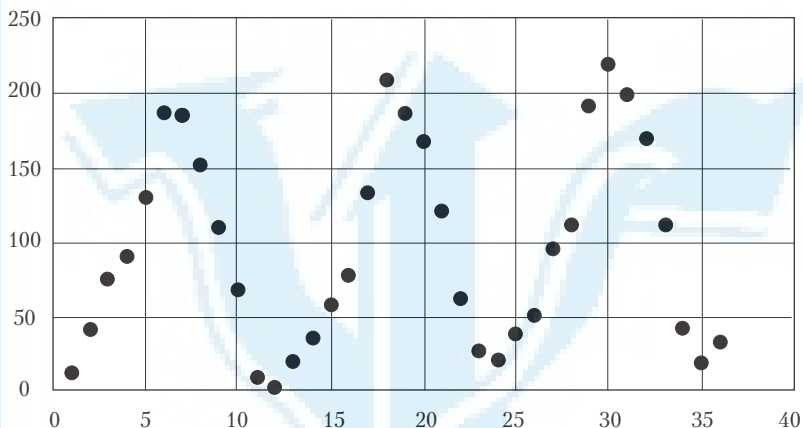


Рис. 5.14

Ряд має яскраво виражену сезонну періодичність. Відзначається весняно-літній пік прибутковості й різке зниження прибутковості в осінні місяці.

Узагалі можна сказати, що ряд має періодичність з періодом s , коли подібні особливості ряду повторюються після s опорних часових інтервалів. У наведеному прикладі опорний часовий інтервал дорівнює одному місяцю, а період s — 12 місяцям.

Попередній аналіз структури ряду показує, що він має незначну тенденцію. Щоб уникнути впливу основної тенденції як однієї з основних передумов для спектрального аналізу, використаємо метод найменших квадратів для оцінювання параметрів тренда. Отримаємо

$$\hat{y}_t = 86,54 + 0,54t.$$

З рівняння тренда бачимо, що щомісячні прибутки мають тенденцію до зростання, але в середньому щомісячні вони збільшуються лише на 0,54 ум. гр. од., отже, досліджуваний ряд має слабку тенденцію. За рівнянням тренда розраховуємо вирівняні значення прибутку, а також відхилення фактичних рівнів від вирівняних. Графічно вирівняний ряд прибутків — це пряма лінія (рис. 5.15).

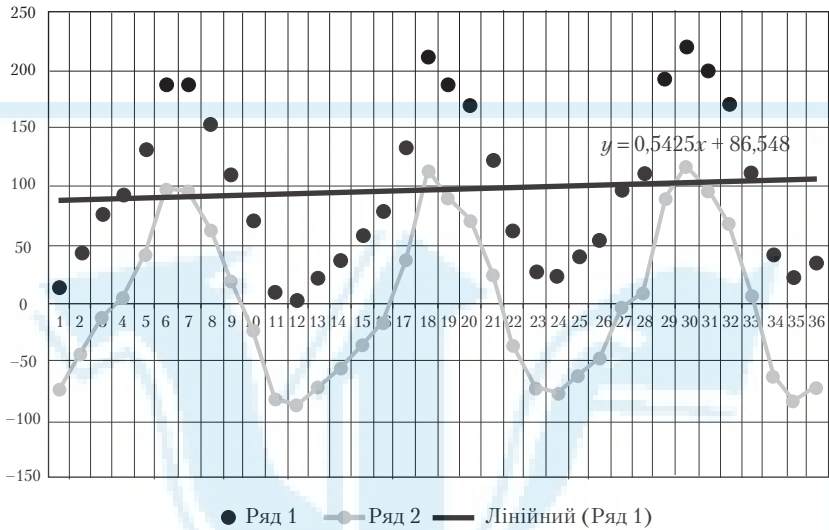


Рис. 5.15

Щоб оцінити спектр ряду прибутків, скористаємося оцінками Парзена. Попередньо знайдемо значення автоковаріаційної функції, узявши

$$m = \frac{n}{3} = 12.$$

Оскільки $\sum_{t=1}^n z = 0$, то $c_k = \frac{1}{n} \sum_{t=1}^{n-k} z_t z_{t+k}$ $k = \overline{0, 12}$

k	c_k	k	c_k	k	c_k
0	4298,10	5	-3216,48	10	1721,251
1	3416,39	6	-3287,78	11	2635,527
2	1660,39	7	-2511,96	12	2736,572
3	-366,115	8	-1195,67		
4	-2161,93	9	300,8098		

Ваги автоковаріаційної функції (оцінки Парзена) розрахуємо за формулою (5.9).

k	λk	k	λk	k	λk
0	1,000	5	0,392	10	0,009
1	0,962	6	0,250	11	0,001
2	0,861	7	0,145	12	0,000
3	0,719	8	0,074		
4	0,556	9	0,031		

Використовуючи значення автоковаріаційної функції та її ваги, обчислюємо оцінки спектра за формулою (5.10):

$$\begin{aligned}
 u_0 &= 918,75, & u_5 &= 169,48, & u_{10} &= 26,50, \\
 u_1 &= 1772,55, & u_6 &= 76,39, & u_{11} &= 20,51, \\
 u_2 &= 2733,39, & u_7 &= 64,43, & u_{12} &= 26,94. \\
 u_3 &= 2038,42, & u_8 &= 49,87, & & \\
 u_4 &= 750,38, & u_9 &= 39,64, & &
 \end{aligned}$$

Результати розрахунків занесемо до табл. 5.10 (с. 150–151).

Порівнюючи оцінки спектра u_p , бачимо, що пік у спектрі часового ряду залишків прибутків перебуває в точці $u_2 = 2733,39$. Отже, досліджуваний ряд має суттєву гармонічну складову з періодом $t = \frac{2 \cdot 12}{2} = 12$ місяців.

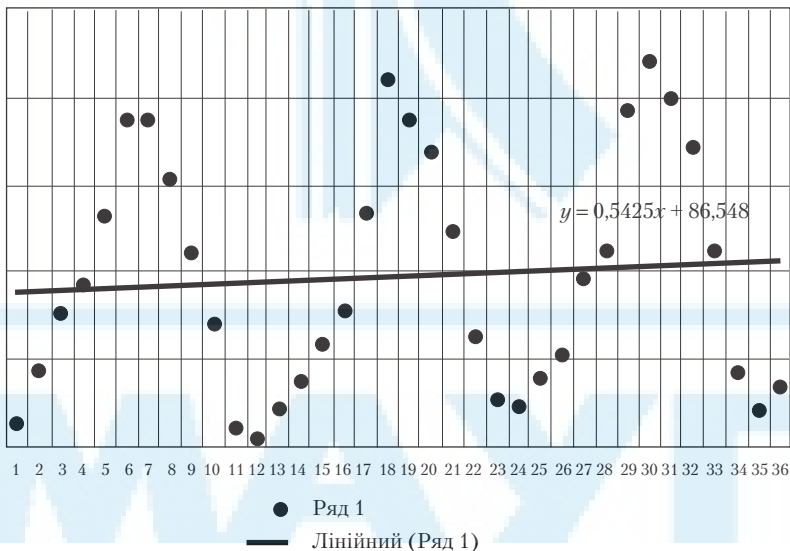


Рис. 5.16

Зауваження. Оскільки спектральний аналіз часових рядів супроводжується громіздкими обчисленнями, тому подібні задачі, як правило, розв'язують на ПЕОМ.

При моделюванні й прогнозуванні часових рядів необхідно досліджувати питання про наявність і тривалість періодичних змін. Це можна зробити за допомогою спектрального аналізу. Визначивши в спектрі часового ряду пік відхилень від тренда, рівні досліджуваного емпіричного часового ряду конструюють як суму рівнів, отриманих за трендом і відповідною гармонікою. У цьому полягає велике практичне значення спектрального аналізу часових рядів (рис. 5.16).

5.8. Прогнозування на адаптивній моделі Хольга–Брауна

У практиці статистичного прогнозування найчастіше використовують дві базові моделі ковзного середнього — Брауна і Хольга, перша з яких є частковим випадком другої. Ці моделі представляють процес розвитку як лінійну тенденцію з параметрами, що постійно змінюються.

Модель Брауна (модель експоненціального згладжування) може відбивати розвиток не лише у вигляді випадкового процесу, що має тенденцію, а й у вигляді випадкового процесу, який не має тенденції, а також у вигляді параболічної змінюваної тенденції. Відповідно розрізняють моделі Брауна:

- нульового порядку, яка описує процеси, що не мають тенденції розвитку. Вона має один параметр A_0 (оцінка поточного рівня). Прогноз розвитку на k кроків уперед виконують за формулою $Y(t+k) = A_0$. Така модель також називається “наївною”;
- першого порядку — $Y(t+k) = A_0 + A_1k$. Коефіцієнт A_0 — значення, близьке до останнього рівня, є немов би закономірною складовою цього рівня. Коефіцієнт A_1 визначає приріст, що сформувався в основному до кінця періоду спостережень, але визначає також (щоправда меншою мірою) швидкість зростання на більш ранніх етапах;
- другого порядку, що відображає розвиток у вигляді параболічної тенденції з параметрами “швидкість” і “прискорення”. Вона має три параметри A_0, A_1, A_2 (A_0, A_1 мають раніше визначений зміст, A_2 — оцінка поточного приросту або “прискорення”). Прогноз виконують за формулою $Y(t+k) = A_0 + A_1k + A_2k^2$.

	x	y_i	xy_i	\hat{y}_i	$z_i = y_i - \hat{y}_i$	z_i^2	$z_i^* z_{i+1}$	$z_i^* z_{i+2}$	$z_i^* z_{i+3}$
	1	13	13	87,09	-74,09	5489,34	3306,83	902,05	-243,20
	2	43	86	87,63	-44,63	1992,07	543,40	-146,51	-1862,96
	3	76	228	88,18	-12,18	148,23	-39,96	-508,19	-1171,21
	4	92	368	88,72	3,28	10,77	137,01	315,77	313,99
	5	131	655	89,26	41,74	1742,23	4015,29	3992,65	2550,84
	6	186	1116	89,80	96,20	9253,97	9201,79	5878,88	1786,40
	7	186	1302	90,34	95,66	9149,90	5845,73	1776,33	-2197,42
	8	152	1216	90,89	61,11	3734,75	1134,87	-1403,90	-5042,70
	9	110	990	91,43	18,57	344,85	-426,60	-1532,31	-1653,81
	10	69	690	91,97	-22,97	527,73	1895,56	2045,85	1667,79
	11	10	110	92,51	-82,51	6808,69	7348,54	5990,55	4715,08
	12	4	48	93,06	-89,06	7931,20	6465,53	5088,93	3267,04
	13	21	273	93,60	-72,60	5270,72	4148,51	2663,30	1250,69
	14	37	518	94,14	-57,14	3265,23	2096,24	984,40	-2127,43
	15	58	870	94,68	-36,68	1345,77	631,97	-1365,78	-4133,92
	16	78	1248	95,23	-17,23	296,77	-641,37	-1941,29	-1535,72
	17	133	2261	95,77	37,23	1386,10	4195,41	3318,92	2591,34
	18	209	3762	96,31	112,69	12698,56	10045,61	7843,41	2711,33
	19	186	3534	96,85	89,15	7946,91	6204,79	2144,88	-3252,20
	20	167	3340	97,40	69,60	4844,57	1674,68	-2539,25	-5013,12
	21	122	2562	97,94	24,06	578,91	-877,77	-1732,94	-1866,30
	22	62	1364	98,48	-36,48	1330,94	2627,59	2829,80	2229,39
	23	27	621	99,02	-72,02	5187,52	5586,72	4401,37	3504,12
	24	22	528	99,57	-77,57	6016,63	4740,07	3773,78	402,91
	25	39	975	100,11	-61,11	3734,36	2973,09	317,42	-566,07
	26	52	1352	100,65	-48,65	2367,00	252,71	-450,67	-4316,43
	27	96	2592	101,19	-5,19	26,98	-48,12	-460,85	-603,47
	28	111	3108	101,74	9,26	85,81	821,84	1076,18	876,63
	29	191	5539	102,28	88,72	7871,37	10307,42	8396,15	5863,85
	30	219	6570	102,82	116,18	13497,39	10994,62	7678,60	761,06
	31	198	6138	103,36	94,64	8955,93	6254,79	619,94	-5961,26
	32	170	5440	103,91	66,09	4368,33	432,97	-4163,33	-5653,23
	33	111	3663	104,45	6,55	42,91	-412,65	-560,32	-472,16
	34	42	1428	104,99	-62,99	3967,95	5387,93	4540,22	
	35	20	700	105,53	-85,53	7316,08	6165,01		
	36	34	1224	106,08	-72,08	5195,03			
Середні	18,50	96,58	1845,33		0,00				
Суми			66432			154731,49	122990,05	59774,05	-13180,14

Таблица 5.10

$z_t^* z_{t+4}$	$z_t^* z_{t+5}$	$z_t^* z_{t+6}$	$z_t^* z_{t+7}$	$z_t^* z_{t+8}$	$z_t^* z_{t+9}$	$z_t^* z_{t+10}$	$z_t^* z_{t+11}$	$z_t^* z_{t+12}$
-3092.52	-7127.29	-7087.09	-4527.84	-1375.86	1702.02	6113.53	6598.26	5378.92
-4293.54	-4269.33	-2727.61	-828.83	1025.31	3682.85	3974.85	3240.31	2550.40
-1164.60	-744.05	-226.09	279.69	1004.62	1084.28	883.90	695.71	446.64
200.60	60.96	-75.41	-270.85	-292.33	-238.31	-187.57	-120.42	-56.55
775.12	-958.87	-3444.17	-3717.25	-3030.32	-2385.12	-1531.22	-719.06	1554.00
-2209.88	-7937.72	-8567.09	-6983.92	-5496.94	-3528.98	-1657.21	3581.47	10840.30
-7892.96	-8518.78	-6944.53	-5465.94	-3509.08	-1647.87	3561.27	10779.17	8527.21
-5442.52	-4436.76	-3492.11	-2241.90	-1052.80	2275.25	6886.65	5447.91	4253.62
-1348.19	-1061.14	-681.24	-319.91	691.37	2092.63	1635.44	1292.54	446.81
1312.69	842.73	395.75	-855.27	-2588.70	-2047.88	-1598.94	-552.73	838.08
3027.03	1421.50	-3072.06	-9298.42	-7355.82	-5743.27	-1985.35	3010.30	5943.08
1534.20	-3315.64	-10035.68	-7939.05	-6198.65	-2142.76	3248.99	6414.30	6907.90
-2702.92	-8181.11	-6471.94	-5053.16	-1746.79	2648.58	5228.96	5631.34	4436.53
-6439.24	-5093.97	-3977.27	-1374.87	2084.66	4115.64	4432.35	3491.93	2780.08
-3270.27	-2553.36	-882.65	1338.33	2642.19	2845.52	2241.78	1784.78	190.55
-1199.06	-414.49	628.48	1240.78	1336.26	1052.74	838.13	89.48	-159.58
895.78	-1358.24	-2681.50	-2887.85	-2275.13	-1811.33	-193.39	344.87	3303.11
-4111.08	-8116.28	-8740.85	-6886.29	-5482.48	-585.34	1043.85	9997.75	13091.88
-6420.65	-6914.74	-5447.62	-4337.09	-463.05	825.77	7909.05	10356.76	8436.35
-5398.89	-4253.39	-3386.31	-361.54	644.75	6175.22	8086.35	6586.93	4600.29
-1470.32	-1170.59	-124.98	222.88	2134.66	2795.31	2276.98	1590.24	157.62
1774.92	189.50	-337.94	-3236.71	-4238.41	-3452.50	-2411.21	-238.99	2298.06
374.12	-667.18	-6390.06	-8367.67	-6816.09	-4760.33	-471.82	4536.94	6160.55
-718.52	-6881.79	-9011.59	-7340.61	-5126.65	-508.13	4886.07	6634.62	5590.76
-5421.67	-7099.58	-5783.14	-4038.92	-400.32	3849.38	5226.94	4404.56	
-5652.29	-4604.21	-3215.56	-318.71	3064.66	4161.39	3506.66		
-491.57	-343.31	-34.03	327.20	444.29	374.39			
612.23	60.68	-583.50	-792.32	-667.66				
581.19	-5588.66	-7588.65	-6394.69					
-7318.26	-9937.20	-8373.73						
-8094.59	-6821.02							
-4763.78								
-77829.42	-115793.34	-118360.17	-90430.73	-43044.28	10829.15	61965.04	94878.97	98516.59

Порядок моделі зазвичай визначають або апріорно на основі візуального аналізу графіка процесу й знання законів розвитку характеру зміни досліджуваного явища, або методом спроб, порівнюючи характеристики моделей різного порядку на ділянці ретроспективного прогнозування.

Розглянемо етапи побудови лінійної адаптивної моделі Брауна.

1. За першими n точками часового ряду ($n > N/2$) оцінюють початкові значення параметрів моделі A_0 і A_1 за допомогою методу найменших квадратів для лінійної апроксимації:

$$Y_p(t) = A_0 + A_1 t \quad (t = 1, 2, \dots, n). \quad (5.11)$$

2. З використанням параметрів A_0 і A_1 за моделлю Брауна знаходять прогноз на один крок ($k = 1$):

$$Y_p(t, k) = A_0(t) + A_1(t) k, \quad t = n + 1, n + 1, \dots, N. \quad (5.12)$$

3. Розрахункове значення $Y_p(t, k)$ економічного показника порівнюють з фактичним $Y(t)$ і обчислюють розбіжності між ними (похибки). При $k = 1$ маємо

$$e(t + 1) = Y(t + 1) - Y_p(t, 1). \quad (5.13)$$

4. Відповідно до цієї величини коригують параметри моделі. У моделі Брауна модифікацію виконують за формулами

$$\begin{aligned} A_0(t) &= A_0(t-1) + A_1(t-1) + (1 - \beta)^2 e(t); \\ A_1(t) &= A_1(t-1) + (1 - \beta)^2 e(t), \end{aligned}$$

де β — коефіцієнт дисконтування даних, що змінюється в межах від 0 до 1 ($\alpha + \beta = 1$) і характеризує знецінювання даних за одиницю часу (оптимальне значення $\beta = (N - 3) / (N - 1)$, хоча його можна добирати й евристично); N — довжина часового ряду, α — параметр згладжування ($\alpha = 1 - \beta$); $e(t)$ — похибка прогнозування рівня $Y(t)$, обчислена в момент часу $(t - 1)$ на один крок уперед.

Приклад 5.9. За статистичними даними про прибутки фірми за 8 років побудувати прогноз на 2 роки за моделлю Брауна і оцінити межі прогнозу

t , роки	1	2	3	4	5	6	7	8
$Y(t)$, ум. гр. од.	178,3	185,5	177,5	143,7	212,6	239,5	305,2	244,04

За першими $n = 8/2 = 4$ точками спостережень отримаємо початкові значення параметрів лінійної регресії:

$A_1(0) =$	2,68
$A_0(0) =$	171,48

Кількість кроків прогнозування				2	
Коефіцієнт дисконтування даних				0,714	
t	$Y(t)$	$A_0(t)$	$A_1(t)$	$Y_p(t)$	$E(t)$
1	178,30			174,16	
2	185,50			176,84	
3	177,50			179,52	
4	143,70			182,20	
5	212,60	171,48	2,68	184,88	
6	239,50	178,63	7,15	221,51	54,62
7	305,20	192,62	13,99	290,57	83,69
8	244,04	202,81	10,19	284,30	-46,53
9		212,99	10,19	304,68	
10				484,12	

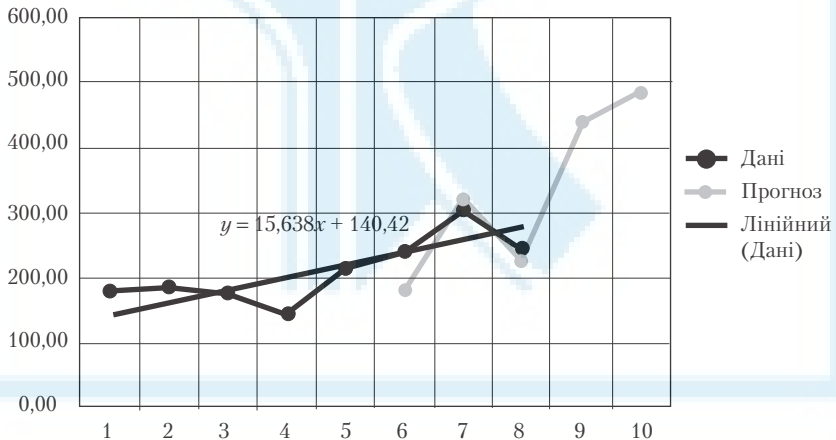


Рис. 5.17

Верхню і нижню межі прогнозу в моделі Брауна визначають за формулою

$$(Y_p(N+k) - \Delta Y_p, Y_p(N+k) + \Delta Y_p),$$

де $Y_p(N+k)$ – прогнозоване значення показника Y у момент $(N+k)$; ΔY_p – похибка прогнозу,

$$\Delta Y_p = St_{\text{табл}} \sqrt{1 + 1/N + (N + k - \bar{t})^2 / \sum_{t=1}^N (t - \bar{t})^2},$$

S – середнє квадратичне відхилення залишків; $t_{\text{табл}}$ – табличне значення t -статистики Стьюдента, $t_{\text{табл}} = 2,5$; N – кількість значень у прогнозованому ряді; k – порядковий номер прогнозованого значення (відносно N); \bar{t} – середнє арифметичне значень незалежної змінної.

Для двох наступних (прогнозованих) періодів отримаємо $\Delta Y_p(1) = 7,22$; $\Delta Y_p(2) = 7,62$; межі прогнозу визначаються відповідно інтервалами: (429,37; 443,82) і (476,50; 491,75), що відображено на графіку (рис. 5.18).

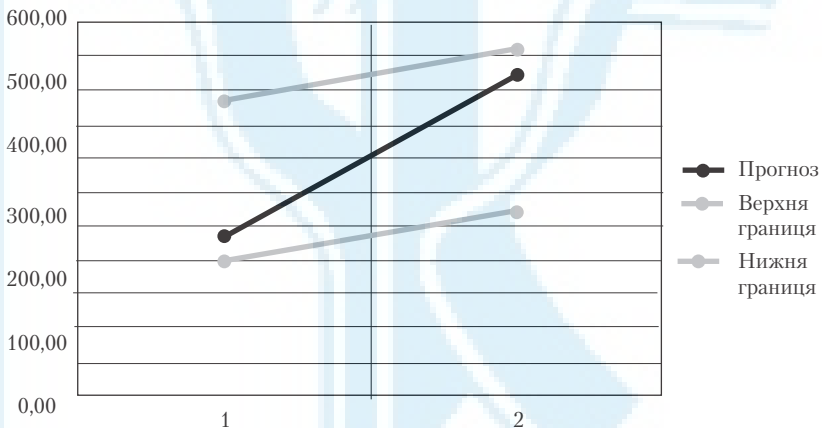


Рис. 5.18

Отже, на дев'ятий рік очікують прибуток у розмірі 436,59 ум. гр. од., а на десятий – 484,12 ум. гр. од.

Висновки

1. Часовий або динамічний ряд – це ряд упорядкованих за часом спостережень деякого показника. На відміну від спостережень просторового ряду, елементи часового ряду хоча й розподілені за однаковою законом імовірностей, але часто є статистично залежними.

2. Часові дані містять зазвичай два види компонент – систематичну і випадкову.

3. Виокремлюють три складові систематичної компоненти: тренд, сезонність, циклічність.

4. Тренд — це компонента, яка, плавно змінюючись, описує чистий вплив довготривалих факторів, загальний закон змінювання явища.

5. Сезонність — це періодичні коливання рівнів часового ряду протягом одного року. Циклічність — це періодичні коливання, які виходять за межі одного року і відображають економічні цикли.

6. Більшість методів дослідження часового ряду спрямовані на виявлення і опис систематичної та випадкової компонент ряду — фільтрацію рівнів ряду.

7. Одним з найпоширеніших методів визначення структури ряду є автокореляційна функція, тобто така, що складається із значень коефіцієнтів кореляції для рівнів одного ряду, зміщених на різні часові проміжки (часові лаги). Значущість коефіцієнтів кореляції різного порядку дає змогу виявити наявність трендової та сезонної компонент ряду.

8. Для визначення основної тенденції (тренда) в рівнях часового ряду зазвичай застосовують методи механічного або аналітичного вирівнювання. Однак перш ніж будувати трендове рівняння, необхідно перевірити гіпотезу про існування тенденції часового ряду.

9. Найчастіше для перевірки такої гіпотези застосовують:

- 1) порівняння середніх рівнів фрагментів ряду;
- 2) критерій “зростаючих і спадних” серій;
- 3) метод Фостера–Стюарта тощо.

10. Після перевірки гіпотези про наявність тенденції переходять до її моделювання. Воно може бути графічним, механічним або аналітичним.

11. На підставі графічного зображення даних — діаграми розсіювання — висувають початкові припущення щодо тенденції розвитку процесу.

12. Якщо графічно форму тренда виявити неможливо, застосовують механічне вирівнювання — згладжування ковзними середніми, експоненційне згладжування тощо. При цьому початкові значення рівнів ряду заміняють розрахунковими середніми значеннями з меншими коливаннями.

13. Метод аналітичного вирівнювання зводиться до моделювання процесу у вигляді функцій часу або кривих зростання.

14. Сезонні й циклічні коливання вивчають за допомогою:

- 1) розрахунку сезонної компоненти та побудови адитивної чи мультиплікативної моделі часового ряду;
- 2) застосування фіктивних сезонних змінних;
- 3) аналізу сезонності за допомогою автокореляційної функції;
- 4) застосування рядів Фур'є.

15. Спектральний аналіз часового ряду дає змогу визначати пік відхилень від тренда і тим самим розраховувати тривалість періодичної компоненти ряду.

Питання для самоперевірки

1. Що таке часовий ряд?
2. З яких компонент складається часовий ряд?
3. Що таке тенденція часового ряду?
4. Що таке сезонна і циклічна компоненти ряду?
5. Які характеристики має випадкова складова часового ряду?
6. Вкажіть основні етапи дослідження часових рядів.
7. Якими методами виявляють структуру часового ряду?
8. Що таке автокореляційна функція і як вона застосовується при аналізі часових рядів?
9. Вкажіть методи перевірки гіпотези про наявність тенденції часового ряду.
10. У чому полягає ідея методу порівняння середніх рівнів ряду?
11. Вкажіть переваги методу Фостера–Стьюарта.
12. Назвіть методи моделювання сезонних і циклічних коливань рівнів часового ряду.
13. Чим різняться адитивна та мультиплікативна моделі ряду? Що між ними спільного?
14. Як будують адитивну і мультиплікативну моделі часового ряду?
15. Що таке фіктивні змінні та яка їх роль у сезонному аналізі?
16. Які результати можна отримати за допомогою спектрального аналізу?

Завдання

1. Перевірити наявність тенденції в часовому ряді*.

1991	1992	1993	1994	1995	1996	1997	1998
14,1+N/10	9,3-N/10	19,4+N/10	19,7-N/10	5,4+N/10	24,2-N/10	13,8+N/10	24,5-N/10
1999	2000	2001	2002	2003	2004	2005	
14,7+N/10	16,6-N/10	5,6+N/10	16,2-N/10	25,3+N/10	11,9-N/10	18,5-N/10	

2. Перевірити наявність тенденції середнього рівня прибутку фірми і визначити порядок полінома, який описує цю тенденцію*.

Роки	1991	1992	1993	1994	1995	1996	1997	
Рівень прибутку y_t								
ум. гр. од.	0,81+ +N/10	0,85+ +N/10	0,9+ +N/10	0,94+ +N/10	0,98+ +N/10	1,03+ +N/10	1,07+ +N/10	
Роки	1998	1999	2000	2001	2002	2003	2004	2005
Рівень прибутку y_t								
ум. гр. од.	1,12+ +N/10	1,16+ +N/10	1,2+ +N/10	1,26+ +N/10	1,31+ +N/10	1,35+ +N/10	1,39+ +N/10	1,42+ +N/10

3. Визначити структуру обох часових рядів.

4. Побудувати прогноз на 2 роки за наведеними статистичними даними, використовуючи адаптивну модель Хольта–Брауна*.

t, роки	1	2	3	4	5	6	7	8
Y(t), ум. гр. од.	178,3+N	185,5-N	177,5+N	143,7-N	212,6+N	239,5-N	305,2+N	244,04-N

* N – остання цифра у номері залікової книжки або порядковий номер у списку групи.

СИСТЕМИ ОДНОЧАСНИХ РІВНЯНЬ

6.1. Поняття про системи одночасних рівнянь

При одночасному дослідженні не одного, а кількох показників застосовують модель, що складається із системи регресійних рівнянь.

Кожне рівняння такої економетричної моделі описує одно- та багатосторонні стохастичні (кореляційні) причинні співвідношення між економічними величинами в їх безпосередньому вигляді. Ці рівняння містять всю суттєву інформацію про залежності між економічними показниками і називаються структурними.

Наприклад, одночасне дослідження доходів і витрат потребує двох співвідношень, які описують змінювання обох показників. Споживання C є функцією від доходу Y , яку можна представити регресійним рівнянням

$$C = a_0 + a_1 Y.$$

У свою чергу дохід складається із витрат (споживання) C і збережень (інвестицій) I . Тоді одночасне змінювання доходу і витрат описують системою

$$\begin{aligned} C &= a_0 + a_1 Y, \\ Y &= C + S, \end{aligned}$$

де перше співвідношення — регресійне рівняння, друге — тотожність. Тотожності не містять випадкових складових, а параметри їх заздалегідь відомі (найчастіше вони дорівнюють одиниці), тому оцінювати їх не потрібно.

Означення 6.1. Для систем одночасних рівнянь усі змінні, що можуть бути визначені із системи рівнянь, називаються ендогенними (залежними), причому їх кількість не перевищує загальної кількості рівнянь системи.

На значення залежних змінних можуть впливати показники, виміряні в той самий момент часу, і відомі заздалегідь показники, тобто визначені в попередні моменти часу.

Означення 6.2. Для систем одночасних рівнянь усі змінні, що їх задають за межами моделі або вони заздалегідь відомі, називаються відповідно екзогенними або предетермінованими.

Класифікація на ендогенні та екзогенні змінні залежить від теоретичної концепції моделі. Доцільно екзогенними змінними вибирати ті, які можна регулювати і тим самим впливати на ендогенні змінні.

6.2. Структурна та зведена (прогнозна) форми системи рівнянь

1. Структурна форма моделі.

Означення 6.3. Економетрична модель, що складається зі структурних рівнянь і тотожностей і відображає структуру зв'язків між змінними, називається структурною формою моделі.

У загальному випадку структурна форма моделі має вигляд

$$Ay_t = Bx_t + u_t, \quad (6.1)$$

де x_t — вектор незалежних (екзогенних) змінних; y_t — вектор залежних змінних; u_t — вектор залишків, $t = 1, 2, \dots, T$.

2. Повна економетрична модель.

Економетрична модель називається *повною*, якщо вона:

- охоплює змінні, що суттєво впливають на спільно залежні (ендогенні) змінні, тобто на змінні, що є залежними для системи загалом, а вектор залишків для кожного рівняння має випадковий характер;
- містить стільки рівнянь, скільки в ній є спільно залежних змінних (кожна залежна змінна пояснюється окремим рівнянням);
- окрім того, система рівнянь має однозначний розв'язок відносно спільно залежних змінних, тобто матриця A в моделі (6.1) невіроджена (має відмінний від нуля визначник):

$$\det A \neq 0.$$

3. Зведена форма економетричної моделі.

Якщо економетричну модель застосовують не для аналізу системи, а для передбачення її розвитку чи оцінювання параметрів, структурна форма моделі стає неприйнятною. Засобами алгебраїчних перетворень систему структурних рівнянь зводять до такої форми, де кожне рівняння містить лише одну ендогенну змінну, яка є функцією від екзогенних змінних. Таку форму рівнянь називають *зведеною*.

Зведену форму рівнянь можна назвати *скороченою*. Це пов'язано з тим, що при певних перетвореннях багато окремих економічних

залежностей можна виключити з розгляду, тобто загальна кількість рівнянь може скоротитися.

Унаслідок таких перетворень зведена форма рівнянь, на відміну від структурної форми, не має ані безпосередньої, ані будь-якої іншої економічної інтерпретації. Рівняння у зведеній формі дають змогу передбачити, як зміниться значення тієї чи іншої ендогенної змінної, якщо змінюватимуться значення екзогенних змінних, але на підставі цих рівнянь неможливо пояснити, як і чому це відбувається. Саме через це зведену форму рівнянь називають також *прогнозною* формою.

Отже, коли виникає питання про консультації чи практичні поради, системи рівнянь у зведеній формі особливо корисні, тому що дають змогу звести формальну модель до мінімальної кількості співвідношень. Звичайно, зведена модель матиме цінність, якщо правильно є початкова структурна модель.

Зокрема, якщо економетрична модель є повною, залежні змінні такої моделі можна представити в явному вигляді як функції від спільно незалежних змінних, розв'язавши її відносно вектора залежних змінних y_t . Це можливо, оскільки за означенням матриця A такої моделі є невіродженою, і після множення системи (6.1) на її обернену матрицю отримаємо

$$y_t = A^{-1}Bx_t + A^{-1}u_t. \quad (6.2)$$

При таких перетвореннях параметри зведеної форми стають функціями від параметрів вихідних структурних рівнянь, і залишки такої моделі, очевидно, є лінійною комбінацією залишків структурної моделі.

Увівши позначення $v_t = A^{-1}u_t$, $R = A^{-1}B$, отримаємо спрощений вигляд моделі:

$$y_t = Rx_t + v_t. \quad (6.3)$$

У цій системі кожна залежна змінна визначається через незалежні змінні моделі, тобто система (6.3) є зведеною формою економетричної моделі.

6.3. Проблеми оцінювання параметрів систем рівнянь

Залежно від зв'язків між змінними в моделі системи рівнянь поділяють на класи:

- 1) системи незалежних рівнянь, в яких кожна залежна змінна є функцією від того самого набору факторів (незалежних змінних) або різної їх комбінації;

- 2) рекурсивні системи, в яких залежна змінна є функцією від усіх залежних змінних попередніх рівнянь, а також певного набору незалежних факторів;
- 3) системи одночасних рівнянь — системи взаємопов'язаних сумісних рівнянь, в яких ті самі змінні в одних рівняннях є залежними, а в інших незалежними змінними.

Для отримання незміщених, ефективних і обґрунтованих оцінок параметрів регресійних рівнянь за звичайним МНК необхідно виконання ряду передумов, зокрема залишки моделі мають бути випадковими величинами з нульовим математичним сподіванням, сталими дисперсіями, некорельованими між собою та незалежними відносно ендогенних змінних моделі.

У системах незалежних рівнянь, а також у рекурсивних системах зазначені умови зазвичай виконуються і тому після відповідної перевірки параметри окремих рівнянь таких систем оцінюють методом найменших квадратів. Зокрема, у рекурсивних системах рівняння оцінюють поетапно одне за одним.

У системах одночасних рівнянь, а саме в рівняннях структурної форми моделі, залишки корелюють із залежною змінною, тому застосування звичайного МНК призводить до зміщених і необґрунтованих оцінок параметрів моделі. При переході до зведеної форми моделі кореляційна залежність між залишками й змінними моделі усувається. При цьому параметри зведеної моделі є функціями від параметрів структурної форми. Тому цілком природно постає питання визначення параметрів структурної форми через параметри зведеної.

Перехід від структурної до зведеної форми системи рівнянь хоча і дає змогу усунути проблему корельованості залежної змінної та випадкового відхилення, але призводить до іншої, не менш серйозної проблеми — *проблеми ідентифікації*.

6.4. Ідентифікація (ототожнення) системи рівнянь

Щоб правильно визначити метод оцінювання параметрів системи одночасних рівнянь, необхідно провести попереднє дослідження моделі, а саме перевірити ідентифікованість системи.

Під проблемою ідентифікації розуміють можливість чисельного оцінювання параметрів структурних рівнянь за оцінками коефіцієнтів зведених рівнянь.

Означення 6.4. Економетрична модель, задана системою одночасних рівнянь, називається точно (строго) ідентифікованою (ототоженою), якщо оцінки її параметрів можна однозначно отримати на основі оцінок параметрів зведеної моделі.

Означення 6.5. Надідентифікованою (переототоженою) називається така модель, що для деяких її параметрів можна отримати кілька кількісних оцінок, обчислених на підставі оцінок параметрів зведеної форми.

Окрім того, модель може бути неідентифікованою (неототоженою). Це трапляється в тому разі, коли кількість невідомих параметрів значно перевищує кількість рівнянь, через які їх треба оцінити.

Умову ідентифікованості перевіряють для кожного рівняння структурної моделі. Можливість ідентифікації рівнянь системи з'ясовують, користуючись таким спрощеним правилом: для того щоб структурне рівняння було точно ідентифікованим, кількість змінних (ендогенних чи екзогенних), відсутніх у цьому рівнянні, має дорівнювати кількості ендогенних змінних у системі мінус одиниця.

Нехай N — кількість рівнянь і кількість ендогенних змінних у структурній моделі, M — кількість екзогенних змінних моделі, n — кількість ендогенних змінних, що входять до рівняння, m — кількість екзогенних змінних, що входять до цього самого рівняння. Змінні, що не увійшли до рівняння, називаються виключеними і їх кількість відповідно становить $N - n$ (ендогенних) і $M - m$ (екзогенних). Тоді необхідна умова ідентифікованості має вигляд

$$(N - n) + (M - m) \geq N - 1$$

або

$$M - m \geq n - 1.$$

Якщо співвідношення виконуються як рівність ($M - m = n - 1$), відповідне рівняння точно (строго) ідентифіковане, якщо як нерівність, тобто $M - m > n - 1$, то рівняння надідентифіковане. Якщо співвідношення не виконується, тобто $M - m < n - 1$, то рівняння неідентифіковане.

Система загалом ідентифікована, якщо ідентифіковане кожне її рівняння. Якщо хоча б одне рівняння неідентифіковане, то й уся система неідентифікована. Якщо хоча б одне з рівнянь надідентифіковане і немає жодного неідентифікованого, то система надідентифікована.

6.5. Методи оцінювання параметрів систем одночасних рівнянь

Загальна характеристика методів оцінювання параметрів систем рівнянь

Головною проблемою, що виникає при оцінюванні параметрів систем рівнянь, є порушення передумови про незалежність змінних і залишків моделі. Одним із способів усунення цієї залежності є перетворення початкової структурної системи рівнянь до зведеного вигляду. Для оцінювання параметрів окремих рівнянь зведеної моделі можна застосовувати звичайний МНК, однак визначати за їх оцінками оцінки параметрів структурної моделі вдається не завжди, а лише за умови точної (строгої) ідентифікованості кожного рівняння системи. Якщо серед рівнянь системи виявляться надідентифіковані, потрібно застосовувати інші методи оцінювання. Такими є двокроковий МНК, метод інструментальних змінних, головних компонентів тощо. Кожен з них полягає в заміні змінних, які корелюють із залишками, іншими — інструментальними, які тісно пов'язані з незалежними змінними моделі, але не корелюють з її залишками.

Отже, залежно від ідентифікованості систем для оцінювання їх параметрів застосовують такі методи:

- 1) непрямий МНК (НМНК) (для оцінювання параметрів строго ідентифікованих систем);
- 2) двокроковий МНК (2МНК) і його модифікації для оцінювання параметрів надідентифікованих систем;
- 3) метод інструментальних змінних;
- 4) метод головних компонентів;
- 5) трикроковий МНК (3МНК) для одночасного оцінювання параметрів усіх рівнянь системи.

6.5.1. Непрямий метод найменших квадратів оцінювання параметрів точно ідентифікованих систем

Ідея методу полягає в тому, щоб від структурної форми моделі перейти до зведеної, оцінити її параметри звичайним МНК і оберненим перетворенням отримати оцінки параметрів структурної форми.

Непрямий МНК складається з таких кроків:

- 1) перевіряють умову ідентифікованості для кожного рівняння структурної форми. Якщо всі рівняння точно ідентифіковані, виконують наступний крок, інакше застосовують інший метод;

- 2) початкову структурну систему рівнянь перетворюють до зведеної форми;
- 3) до кожного рівняння зведеної форми застосовують звичайний МНК;
- 4) на основі оцінок, знайдених для зведеної форми, обчислюють оцінки параметрів структурних рівнянь шляхом обернених перетворень.

Приклад 6.1. Для ілюстрації НМНК розглянемо кейнсіанську модель формування прибутків:

$$C_t = a_0 + a_1 Y_t + u_t,$$

$$Y_t = C_t + Z_t.$$

У зведеній формі ця модель має вигляд

$$Y_t = \frac{a_0}{1-a_1} + \frac{1}{1-a_1} Z_t + \frac{u_t}{1-a_1}, \quad (6.4)$$

$$C_t = \frac{a_0}{1-a_1} + \frac{a_1}{1-a_1} Z_t + \frac{u_t}{1-a_1}. \quad (6.5)$$

Позначимо

$$\frac{a_0}{1-a} = \lambda_{10}, \quad \frac{1}{1-a_1} = \lambda_{11},$$

$$\frac{a_0}{1-a_1} = \lambda_{20}, \quad \frac{a_1}{1-a_1} = \lambda_{21},$$

$$\frac{u_t}{1-a_1} = v_t$$

залишки такої моделі розподілені нормально за законом

$$N\left(0, \frac{\sigma^2}{1-a_1}\right).$$

Тоді замість останніх співвідношень отримаємо

$$Y_t = \lambda_{10} + \lambda_{11} Z_t + v_t,$$

$$C_t = \lambda_{20} + \lambda_{21} Z_t + v_t.$$

Через те, що обсяг інвестицій Z_t є екзогенною змінною моделі, ця змінна не корелює з випадковими залишками u_t у зведеній формі системи рівнянь (6.4), (6.5), а отже, і з залишками v_t останньої системи. Це означає, що для випадкового члена v_t виконуються передумови

МНК. Тому оцінки $\hat{\lambda}_{10}, \hat{\lambda}_{11}, \hat{\lambda}_{20}, \hat{\lambda}_{21}$, отримані за МНК, будуть незміщеними й ефективними оцінками параметрів $\lambda_{10}, \lambda_{11}, \lambda_{20}, \lambda_{21}$. Маючи ці оцінки, нескладно визначити оцінки \hat{a}_0 і \hat{a}_1 коефіцієнтів a_0 і a_1 рівняння початкової структурної системи:

$$\hat{a}_1 = \frac{\hat{\lambda}_{21}}{\hat{\lambda}_{11}}; \quad \hat{a}_0 = \frac{\hat{\lambda}_{20}}{\hat{\lambda}_{11}}.$$

Зміст такої назви методу очевидний: перш ніж застосувати звичайний МНК, початкову систему перетворюємо до зведеної форми, а потім на підставі МНК-оцінок зведеної форми визначаємо оцінки параметрів початкової системи.

Оцінки параметрів a_0 та a_1 , отримані за НМНК, є обґрунтованими, а тому при великих вибірках існує велика ймовірність, що вони будуть близькими до істинних значень параметрів.

Зауважимо, що в цьому разі оцінки визначаються точно, а регресійне рівняння в розглянутій моделі доходу є ідентифікованим (однозначно визначеним).

Приклад 6.2. Розглядається модель “попит — пропозиція” сільськогосподарської продукції:

Пропозиція: $q_t = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + \varepsilon_{1t}$

Попит: $q_t = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \varepsilon_{2t}$.

Умова рівноваги: $q_t^s = q_t^d$.

Функція пропозиції залежить від ринкової ціни p_t і ціни минулого періоду p_{t-1} . Функція попиту залежить від ринкової ціни p_t і доходу споживачів y_t . В умовах ринкової рівноваги запропонована кількість продукції дорівнює кількості, що споживається: $q_t^s = q_t^d = q_t$. У цій системі рівнянь залежними є змінні q_t і p_t , а незалежними — p_{t-1} і y_t , тобто $N = 2, M = 2$. В обох рівняннях $n = 2, m = 1$, тобто умова ідентифікованості виконується як рівність:

$$(N - n) + (M - m) = (2 - 2) + (2 - 1) = 1 = N - 1.$$

Отже, система рівнянь є строго ідентифікованою і оцінювати її параметри можна непрямим методом найменших квадратів.

Передусім розв'яжемо систему відносно спільно залежних змінних. Скористаємось умовою рівноваги і підставимо значення q_t із другого рівняння у перше:

$$\alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \varepsilon_{2t} = \beta_0 + \beta_1 p_t + \beta_2 p_{t-1} + \varepsilon_{1t};$$

отриманий результат

$$p_t = \gamma_{10} + \gamma_{11}y_t + \gamma_{12}p_{t-1} + \xi_{1t}$$

підставимо в одне з початкових рівнянь. Дістанемо

$$q_t = \gamma_{20} + \gamma_{21}y_t + \gamma_{22}p_{t-1} + \xi_{2t},$$

де

$$\begin{cases} \gamma_{10} = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}; & \gamma_{11} = -\frac{\alpha_2}{\alpha_1 - \beta_1}; & \gamma_{12} = -\frac{\beta_2}{\alpha_1 - \beta_1}; \\ \gamma_{20} = \frac{\alpha_1\beta_0 - \alpha_0\beta_1}{\alpha_1 - \beta_1}; \\ \gamma_{21} = -\frac{\alpha_2\beta_1}{\alpha_1 - \beta_1}; & \gamma_{22} = -\frac{\alpha_1\beta_2}{\alpha_1 - \beta_1}; \\ \xi_{1t} = \frac{\alpha_1\varepsilon_{t2} - \beta_1\varepsilon_{t1}}{\alpha_1 - \beta_1}; & \xi_{2t} = \frac{\varepsilon_{t2} - \varepsilon_{t1}}{\alpha_1 - \beta_1}. \end{cases}$$

Маємо систему шести рівнянь відносно шести невідомих. Очевидно, за відомими значеннями γ_{12} і γ_{22} , γ_{21} і γ_{11} можна обчислити $\alpha_1 = \frac{\gamma_{22}}{\gamma_{12}}$ і

$\beta_1 = \frac{\gamma_{22}}{\gamma_{12}}$, а потім за відомими α_1 , β_1 знайти $\alpha_2 = \gamma_{11}(\beta_1 - \alpha_1)$ і

$\beta_2 = \gamma_{12}(\beta_1 - \alpha_1)$.

Відомі значення γ_{10} і γ_{20} дають змогу обчислити α_0 , β_0 . Так, із співвідношення для γ_{10} , визначивши $\alpha_0 = \gamma_{10}(\beta_1 - \alpha_1) + \beta_0$ і підставивши його у співвідношення для γ_{20} , матимемо $\beta_0 = \gamma_{20} - \beta_1\gamma_{10}$.

Приклад 6.3. Розглядається модель “попит – пропозиція”.

Пропозиція: $q_t = \beta_0 + \beta_1 p_t + \varepsilon_{1t}$.

Попит: $q_t = \alpha_0 + \alpha_1 p_t + \alpha_2 y_t + \varepsilon_{2t}$,

де q_t , p_t – ендogenous змінні (кількість товару і ціна в році t); y_t – екogenous (прибуток споживачів); ε_{1t} , ε_{2t} – випадкові відхилення.

Для цієї моделі $N = 2$, $M = 1$. У першому рівнянні $n = 1$, $m = 0$, тоді $M - m = n - 1$, тобто рівняння ідентифіковане. У другому рівнянні $n = 1$, $m = 1$, тоді $M - m = 0$, $n - 1 = 2 - 1 = 1$, умова не виконується, тобто рівняння неідентифіковане.

На підставі наступних статистичних даних необхідно оцінити коефіцієнти функції пропозиції, використовуючи для цього МНК і НМНК. Результати порівняти.

p_t	q_t	y_t	p_t^2	y_t^2	$p_t q_t$	$p_t y_t$	$q_t y_t$
1	8	2	1	4	8	2	16
2	10	4	4	16	20	8	40
3	7	3	9	9	21	9	21
4	5	5	16	25	20	20	25
5	1	2	25	4	5	10	2
15	31	16	55	58	74	49	104
3	6,2	3,2	11	11,6	14,8	9,8	20,8

Сума
Середнє

Побудуємо зведені рівняння системи. Для цього віднімо від функції пропозиції функцію попиту:

$$(\beta_0 - \alpha_0) + (\beta_1 - \alpha_1) p_t - \alpha_2 y_t + (\varepsilon_{1t} - \varepsilon_{2t}) = 0.$$

Звідки

$$p_t = \pi_{10} + \pi_{11} y_t + v_{2t},$$

$$q_t = \pi_{20} + \pi_{21} y_t + v_{2t},$$

де

$$\pi_{10} = \frac{\alpha_0 - \beta_0}{\beta_1 - \alpha_1}, \quad \pi_{11} = \frac{\alpha_2}{\beta_1 - \alpha_1}, \quad v_{1t} = \frac{\varepsilon_{2t} - \varepsilon_{1t}}{\beta_1 - \alpha_1};$$

$$\pi_{20} = \beta_0 + \beta_1 \pi_{10}, \quad \pi_{21} = \beta_1 \pi_{11}, \quad v_{2t} = \beta_1 v_{1t} + \varepsilon_{1t}.$$

Неважко помітити, що функція пропозиції точно ідентифікована. Оцінки b_1 і b_0 параметрів β_1 і β_0 можуть бути визначені на основі оцінок коефіцієнтів таких рівнянь:

$$\beta_1 = \frac{\pi_{21}}{\pi_{11}}, \quad \beta_0 = \pi_{20} - \beta_1 \pi_{10} \Rightarrow \bar{\beta}_1 = \frac{\bar{\pi}_{21}}{\bar{\pi}_{11}}, \quad \bar{b}_0 = \bar{\pi}_{20} - \bar{b}_1 \bar{\pi}_{10}.$$

За наявними статистичними даними оцінимо коефіцієнти зведених рівнянь:

$$\bar{\pi}_{11} = \frac{\overline{yp - y \cdot p}}{\overline{y^2 - y}} = \frac{0,2}{1,36} = 0,1471,$$

$$\bar{\pi}_{10} = \bar{p} - \bar{\pi}_{11} \bar{y} = 3 - 0,147 \cdot 3,2 = 2,5293,$$

$$\bar{\pi}_{21} = \frac{\overline{yq - y \cdot q}}{\overline{y^2 - y}} = \frac{0,96}{1,36} = 0,9411,$$

$$\bar{\pi}_{20} = \bar{q} - \bar{\pi}_{21} \bar{y} = 3,9411.$$

Отже, оцінки коефіцієнтів функції пропозиції за МНК такі:

$b_1 = 0,7059/0,1471 = 4,7988$, $b_0 = 3,9411 - 4,7988 \cdot 2,5293 = -8,1965$,
функція пропозиції має вигляд

$$q_t = -8,1965 + 4,7988p_t.$$

Водночас розраховані безпосередньо за МНК оцінки рівняння становитимуть:

$$b_1 = \frac{\overline{pq} - \bar{p} \cdot \bar{q}}{\overline{p^2} - \bar{p}^2} = \frac{-3,8}{2} = -1,9, b_0 = \bar{q} - b_1 \bar{p} = 11,9,$$

тобто функція пропозиції має вигляд

$$\bar{q}_t = 11,9 - 1,9p_t.$$

Від'ємне значення параметра при p_t вказує на спадну залежність обсягу пропозиції від ціни продукції, що суперечить економічній теорії.

Отримані результати дають змогу зробити висновок про те, що застосування МНК у невідповідних ситуаціях може істотно спотворити картину залежності.

6.5.2. Двокроковий МНК (2МНК)

Цей метод застосовують у випадку, коли всі рівняння системи є надідентифікованими. Якщо серед рівнянь є також строго ідентифіковані, то 2МНК застосовують у поєднанні з непрямим МНК. Оцінювання параметрів за цим методом виконують у два етапи: на першому — ендогенні змінні “звільняють” від стохастичних залишків; на другому — оцінені рівняння підставляють у структурну систему рівнянь, до яких потім застосовують звичайний МНК.

Алгоритм 2МНК.

1. Усі рівняння структурної форми моделі перевіряють на ідентифікованість. Якщо умова виконується для всіх рівнянь як рівність чи строга нерівність, застосовують 2МНК.
2. Початкову систему розв'язують відносно всіх спільно залежних змінних, тобто переходять до зведеної форми моделі.
3. До кожного рівняння зведеної моделі застосовують звичайний МНК.
4. За обчисленими оцінками параметрів зведених рівнянь і відомими спостереженнями незалежних змінних обчислюють модельні значення залежних змінних.

5. Звичайним МНК оцінюють параметри рівнянь структурної форми моделі за умови, що значення залежних змінних, які входять до правої частини структурних рівнянь, є предетермінованими і такими, що дорівнюють модельним значенням.

2МНК має властивості, які роблять його дуже привабливим для практичного застосування.

1. У цьому методі перший етап (побудову зведених рівнянь) виконують для надідентифікованих рівнянь системи. Інші рівняння моделі не перетворюються. Це дає змогу мінімізувати обсяг обчислень.

2. За наявності перевизначених (надідентифікованих) рівнянь 2МНК, на відміну від непрямого МНК, визначає оцінки параметрів моделі однозначно.

3. При застосуванні цього методу достатньо використовувати лише екзогенні та предетерміновані змінні моделі.

4. Застосування 2МНК буде ефективним лише в тому разі, коли коефіцієнт детермінації R^2 для зведених рівнянь, складених на першому етапі, буде досить високим. За низького значення R^2 використання 2МНК малопродуктивне, тому що в цьому разі зв'язок між заданими та модельними значеннями ендогенних змінних надто слабкий і не відтворює істинних значень заміненої змінної.

Одним з недоліків 2МНК є те, що його застосування хоча й гарантує обґрунтованість оцінок параметрів, однак для вибірок малого обсягу вони можуть виявитися суттєво зміщеними. Крім того, цей метод надзвичайно чутливий до мультиколінеарності незалежних змінних.

Зазначених недоліків можна позбутися, застосовуючи 2МНК у комбінації з методом інструментальних змінних чи з методом головних компонентів.

6.5.3. Метод інструментальних змінних

Цей метод, як і 2МНК, застосовують у тих випадках, коли порушується умова незалежності змінних і залишків. В основі методу лежить перетворення моделі за допомогою інструментальних змінних, тобто змінних, які мають найменшу кореляцію з похибками моделі й досить високу кореляцію з ендогенними змінними. Очевидно, інструментальними змінними насамперед можуть бути незалежні чи предетерміновані змінні моделі або їхні лінійні комбінації, причому не всі одразу, а лише та їх частина, що досить повно відображає поведінку певної ендогенної (залежної) змінної і до того ж має найменшу кореляцію із залишками.

Одним із способів вибору інструментальних змінних є метод послідовних регресій. Згідно з цим методом з матриці предетермінованих і незалежних змінних вибирають вектори, які мають максимальну кореляцію з відповідною залежною змінною. Параметри допоміжних регресій оцінюють за звичайним МНК. Розрахункові значення ендогенної змінної, обчислені за цими регресійними рівняннями, використовують замість спостережених значень при оцінюванні параметрів рівнянь структурної форми моделі.

Окремим випадком методу інструментальних змінних є метод головних компонентів. Головні компоненти — це лінійні комбінації незалежних і предетермінованих змінних моделі, які взаємно незалежні і пояснюють більшу (головну) частину дисперсії залежної змінної. За допомогою цього методу розв'язують проблеми, пов'язані з мультиколінеарністю і великою кількістю незалежних змінних. Метод потребує великої кількості обчислень, але дає змогу отримати надійні оцінки параметрів рівнянь навіть у разі мультиколінеарності незалежних змінних.

Серед недоліків методу можна виокремити неоднозначність у виборі головних компонентів і їх кількості. У будь-якому разі ця кількість має бути суттєво меншою, ніж загальна кількість спостережених змінних, і все ж достатньою для відображення основної частини інформації, що міститься в матриці незалежних змінних моделі.

6.5.4. Модифікований 2МНК

У результаті застосування 2МНК виявляється, що предетерміновані змінні, які входять до структурної моделі й визначаються як модельні значення зведеної системи, мають відхилення, що є перетвореними відхиленнями структурних рівнянь. Саме цю особливість системи рівнянь використовують у модифікованому 2МНК. Згідно з цим методом проблема оцінювання параметрів надідентифікованої структурної системи регресійних рівнянь зводиться до оцінювання параметрів системи рівнянь, які характеризують зв'язок між відхиленнями (залишками) моделі у структурній і зведеній (прогнозній) формах. Такий спосіб оцінювання параметрів також розв'язує проблему залежності змінних і залишків і неоднозначності оцінок параметрів надідентифікованих рівнянь.

Алгоритм модифікованого 2МНК

1. Оцінюють звичайним МНК параметри зведених рівнянь.
2. Обчислюють модельні значення ендогенних змінних.

3. Вважаючи, що ендогенними змінними є відхилення між заданими та модельними значеннями початкових ендогенних змінних, оцінюють за звичайним МНК параметри при ендогенних змінних у кожному рівнянні зведеної системи.
4. На основі оцінених параметрів двох матриць визначають елементи (параметри) третьої матриці, яка описує залежність між залишками обох форм моделі. Останню модель можна подати у вигляді системи регресій, в якій відсутні екзогенні величини, а змінними є залишки обох форм моделі.
5. До кожного рівняння нової системи рівнянь застосовують звичайний МНК.

6.5.5. Трикроковий МНК

Розглянуті методи орієнтовані на оцінювання параметрів окремих рівнянь. Кожен із цих методів має певні недоліки й переваги; незважаючи на відмінності між ними можна виокремити спільну рису, яка їх поєднує, — значний обсяг розрахунків у роботі із системами великої розмірності (з великою кількістю рівнянь і змінних). Скорочення обсягу розрахунків особливо актуальне під час дослідження швидкоплинних процесів, а також у тому разі, коли змінюється пріоритетність окремих незалежних змінних. У таких випадках доцільно застосовувати трикроковий МНК (ЗМНК) або інший метод, який дає змогу одночасно оцінювати параметри всіх рівнянь системи.

Особливістю ЗМНК є те, що навіть у разі оцінювання параметрів усієї системи необхідно враховувати залежності між її окремими рівняннями. Це пов'язано з тим, що залишки деяких рівнянь системи корелюють між собою, отже, загальна матриця коваріацій системи є недиагональною (подібно до кореляційної матриці при автокореляції залишків одного рівняння). У таких випадках для оцінювання параметрів зазвичай застосовують узагальнений МНК (метод Ейткена), але заздалегідь має бути відомою коригуюча матриця (обернена до матриці коваріацій). При визначенні цієї матриці в разі одного рівняння використовують залишки моделі, параметри якої оцінено звичайним МНК. У разі систем рівнянь (особливо за наявності надієнтіфікованих рівнянь) початкове оцінювання виконують за 2МНК, а потім на підставі залишків оцінених співвідношень будують матрицю коваріацій для системи загалом і оцінюють параметри за методом Ейткена. Саме поєднання 2МНК і узагальненого МНК дало назву цього методу.

Алгоритм ЗМНК

1. Виключають із розгляду всі тотожності, параметри яких відомі.
2. Виключають із розгляду всі неідентифіковані рівняння, оскільки їхні параметри оцінити в принципі неможливо.
3. Точно ідентифіковані та надідентифіковані рівняння об'єднують в окремі групи.

Зауваження. Якщо в утворених групах є рівняння, не пов'язані між собою, матриця коваріацій залишків рівнянь цієї групи матиме блочно-діагональну структуру. Тоді відповідну групу поділяють на підгрупи й розглядають їх окремо.

4. Оцінюють параметри рівнянь для обох груп (чи підгруп) двокроковим МНК. (Якщо якась група містить лише одне рівняння, то обчислення закінчуються: у цьому разі ЗМНК перетворюється на 2МНК).
5. Будують матрицю коваріацій для кожної з виділених груп (підгруп) рівнянь за залишками оцінених рівнянь.
6. Виконують обчислення за узагальненим МНК.

6.6. Прогноз і загальні довірчі інтервали

Точковий прогноз залежних змінних виконують на підставі прогнозної (зведеної) системи рівнянь. Визначення довірчих інтервалів прогнозу залежить від способу, яким було отримано зведену форму моделі.

У загальному випадку довірчі інтервали для кожної ендогенної змінної задаються інтервалом

$$\left(\hat{Y}_i - t_{\text{табл}} S_{ii}, \hat{Y}_i + t_{\text{табл}} S_{ii} \right), \quad i = 1, 2, \dots, r,$$

де $t_{\text{табл}} = t(\alpha/2, k)$ — табличне значення розподілу Стюдента з рівнем значущості α і $k = n - l - r + 1$ степенями вільності; n — загальна кількість спостережень; l — кількість екзогенних змінних i -го рівняння системи; r — кількість рівнянь системи (кількість ендогенних змінних); S_{ii}^2 — дисперсія залишків i -го рівняння системи.

Довірчі інтервали для всіх ендогенних змінних визначають за формулами

$$\left(\hat{Y}_i - \sqrt{\frac{(1 + X_j^T (X^T X)^{-1} X_j)(n-l)r}{n-l-r+1}} S_{ii}^2, \right.$$

$$\hat{Y}_i + \sqrt{\frac{(1 + X_j^T (X^T X) X_j)(n-l)r}{n-l-r+1} S_{ii}^2}$$

де X_j — матриця спостережень l екзогенних змінних, які увійшли до i -го рівняння зведеної системи, розмірності $n \times l$; X — загальна матриця спостережень усіх ендогенних змінних системи розмірності $n \times r$.

Висновки

1. Для повного відображення реальних взаємозв'язків у складних економічних явищах потрібно застосовувати системи співвідношень, які можуть складатися з регресійних рівнянь і тотожностей.

2. Деякі показники в одних рівняннях є залежними, в інших — незалежними змінними, тобто водночас відіграють різні ролі, тому такі системи називають системами одночасних рівнянь.

3. Система рівнянь, що відображає структуру економічних зв'язків, називається структурною формою моделі.

4. У системах одночасних рівнянь існує кореляційна залежність між залишками рівнянь і змінними моделі, тобто порушується третя передумова застосування МНК. Оцінки параметрів, обчислені за класичним МНК для кожного рівняння структурної форми моделі, виявляються зміщеними і необґрунтованими.

5. Щоб усунути кореляційну залежність між залишками та змінними, систему рівнянь розв'язують відносно всіх спільно залежних змінних системи. Перетворена в такий спосіб система називається зведеною формою моделі.

6. Параметри зведеної системи рівнянь можна оцінювати класичним МНК. Вони будуть обґрунтованими. Але при цьому виникає питання повернення до параметрів структурної форми моделі, тобто постає проблема ідентифікованості системи.

7. Якщо за оцінками параметрів зведеної системи можна однозначно обчислити всі оцінки параметрів структурної системи, система рівнянь називається ідентифікованою. Якщо для якогось параметра структурної форми існує кілька оцінок, обчислених за оцінками параметрів зведеної, то система надідентифікована, якщо жодної — то система неідентифікована.

8. Умову ідентифікованості перевіряють для кожного рівняння структурної форми моделі. Якщо всі рівняння системи ідентифіковані, то система ідентифікована. Якщо є хоча б одне надідентифіко-

ване рівняння, то система надієтїфїкована, якщо є хоча б одне неїдентифїковане, то система неїдентифїкована.

9. Якщо система рівнянь ідентифїкована, її параметри оцїнюють непряим МНК. Якщо система рівнянь надієтїфїкована, її параметри оцїнюють двокроковим МНК, методом інструментальних змінних чи методом головних компонентів. Для одночасного оцїнювання параметрів усіх рівнянь системи застосовують трикроковий МНК.

Питання для самоперевїрки

1. Назвіть основні причини використання систем одночасних рівнянь.
2. У чому полягає основна відмінність між структурними рівняннями системи та рівняннями у зведеній формї?
3. Чому звичайний МНК практично не використовується для оцїнювання параметрів систем одночасних рівнянь?
4. У чому сутність непрямого методу найменших квадратів (НМК)?
5. Яка проблема виникає при обчислення оцїнок параметрів структурних рівнянь за оцїнками параметрів зведених рівнянь?
6. Назвіть причини неїдентифїкованостї та надієтїфїкованостї систем структурних рівнянь.
7. Вкажіть необхідні й достатні умови ідентифїкованостї систем.
8. Для оцїнювання параметрів яких систем можливе використання звичайного МНК?
9. У чому сутність двокрокового методу найменших квадратів?
10. Яку проблему дає змогу розв'язати метод інструментальних змінних?

Завдання

Завдання 1. За наведеними даними щодо ВВП (Y), споживання (C) і інвестицій (I) для економіки за 20 років:

Y	95,75	98,55	103,55	109,00	108,25
C	60,45	62,45	65,90	68,90	68,45
I	14,30	15,85	17,75	19,70	18,10

<i>Y</i>	107,40	112,70	117,75	123,45	126,55
<i>C</i>	70,00	73,55	76,55	79,70	81,60
<i>I</i>	14,60	17,35	20,00	22,15	22,30

<i>Y</i>	125,85	128,10	125,35	130,25	138,30
<i>C</i>	81,55	82,55	83,45	87,35	91,55
<i>I</i>	19,80	21,00	18,00	20,00	25,25

<i>Y</i>	142,65	146,80	151,30	157,40	161,25
<i>C</i>	95,50	99,00	101,75	105,40	107,45
<i>I</i>	24,85	24,50	25,00	25,80	26,15

а) оцінити за МНК параметри β_0 і β_1 функції споживання $c_t = \beta_0 + \beta_1 y_t + \varepsilon_t$;

б) оцінити ті самі параметри за схемою найпростішої кейнсіанської моделі формування прибутків на основі НМНК;

в) порівняти отримані результати. Зробити висновки про якість оцінок.

Вказівка. Варіанти завдань утворюють додаванням до кожного спостереження числа N , де N — останні дві цифри номера залікової книжки або порядковий номер у списку групи.

Завдання 2. Розглядається модель

$$r_t = \beta_0 + \beta_1 y_t + \beta_2 m_t + \varepsilon_t;$$

$$y_t = \alpha_0 + \alpha_1 r_t + v_t,$$

де r_t — процентна ставка в році t ; y_t — ВВП у році t ; m_t — грошова маса M_2 року t .

На підставі статистичних даних оцінити параметри ідентифікованих рівнянь. Чи збігаються знаки знайдених оцінок з передбачуваними теоретично?

r_t	6,55	4,55	4,45	7,00	7,50
y_t	95,75	98,5	103,55	109,00	108,25
m_t	58,30	60,00	60,55	64,50	65,00

r_t	8,75	9,70	10,00	11,50	7,75
y_t	107,40	112,70	117,75	123,45	126,55
m_t	63,45	67,60	70,50	74,00	76,50

r_t	6,00	6,10	5,90	9,80	8,00
y_t	125,85	128,10	125,35	130,25	138,30
m_t	75,00	77,25	74,00	78,45	85,50

r_t	7,50	7,00	6,50	7,40	5,50
y_t	142,65	146,80	151,30	157,40	161,25
m_t	87,00	88,00	90,50	94,40	96,50

Вказівка. Дані для різних варіантів утворити аналогічно до першого завдання.

МАУП

МАТЕМАТИЧНА БАЗА ЕКОНОМЕТРІЇ

7.1. Матриці

7.1.1. Основні види матриць

Матрицею називають прямокутну таблицю елементів, що містить m рядків і n стовпців:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix},$$

де a_{ij} – елементи матриці; i, j – індекси, $i = \overline{1, m}$, $j = \overline{1, n}$, що визначають номер рядка та номер стовпця, на яких розміщується зазначений елемент. Кількість рядків m і кількість стовпців n матриці визначають її розмірність (порядок) $m \times n$. Застосовують також скорочену форму запису матриці: $A = \|a_{ij}\|$, $i = \overline{1, m}$, $j = \overline{1, n}$, або $A_{m \times n} = \|a_{ij}\|$.

Якщо $m = 1$, одержимо матрицю-рядок (вектор-рядок) розмірності $1 \times n$

$$A = (a_{11} \ a_{12} \ \dots \ a_{1n}).$$

Якщо $n = 1$, одержимо матрицю-стовпець (вектор-стовпець) розмірності $m \times 1$

$$A = \begin{pmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{pmatrix}.$$

Матрицю називають нульовою, якщо в неї всі елементи дорівнюють нулю:

$$A' = A^T = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{m1} \\ a_{12} & a_{22} & \dots & a_{m2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{mn} \end{pmatrix}.$$

Транспонуючи матрицю-рядок, дістаємо матрицю-стовпець і навпаки:

якщо $A = \begin{pmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{pmatrix}$, то $A' = (a_{11} \ a_{12} \ \dots \ a_{1m})$

або

якщо $A = (a_{11} \ a_{12} \ \dots \ a_{1m})$, то $A' = \begin{pmatrix} a_{11} \\ a_{21} \\ \dots \\ a_{m1} \end{pmatrix}$.

Квадратну матрицю називають симетричною, якщо $A = A'$ або $a_{ij} = a_{ji}$.

7.1.2. Дії над матрицями

Дві матриці називають рівними, якщо вони однакової розмірності та відповідні елементи у них рівні:

$$A = B \Leftrightarrow a_{ij} = b_{ij} \quad i = \overline{1, m}, \quad j = \overline{1, n}.$$

Щоб додати дві матриці однакової розмірності, потрібно додати відповідні їхні елементи:

$$C = A + B \Leftrightarrow c_{ij} = a_{ij} + b_{ij}, \quad i = \overline{1, m}, \quad j = \overline{1, n}.$$

Щоб помножити матрицю на число, потрібно кожний її елемент помножити на це число: якщо $C = \alpha A$, то $c_{ij} = \alpha a_{ij}$, $i = \overline{1, m}$, $j = \overline{1, n}$.

Для того щоб матрицю A помножити на матрицю B , потрібно, щоб вони були узгоджені для множення: кількість стовпців матриці A має дорівнювати кількості рядків матриці B . У результаті множення дістаємо матрицю C , кількість рядків якої збігається з кількістю рядків матриці A , а кількість стовпців — з кількістю стовпців матриці B , тобто

$$\begin{matrix} A & B & = & C \\ n \times m & m \times k & & n \times k \end{matrix}.$$

При цьому кожний елемент c_{ij} матриці C дорівнює сумі добутків відповідних елементів i -го рядка матриці A та j -го стовпця матриці B :

$$c_{ij} = \sum_{p=1}^m a_{ip} b_{pj}, \quad i = \overline{1, n}; \quad j = \overline{1, k}.$$

Як правило, $AB \neq BA$. Хоча є й такі матриці, для яких виконується рівність $AB = BA$.

Для квадратних матриць $A^n = \underbrace{AA \dots A}_n$ разів.

Властивості операцій додавання матриць і множення матриці на число, множення, транспонування

Якщо A, B — матриці, а λ, β — числа, то:

- а) $A + B = B + A$, $(A + B) + C = A + (B + C) = A + B + C$;
- б) $\lambda A = A\lambda$, $\lambda(A + B) = \lambda A + \lambda B$, $(\lambda + \beta)A = \lambda A + \beta A$;
- в) $(\lambda\beta)A = \lambda(\beta A) = \beta(\lambda A)$;
- г) $A + O = A$, O — нуль-матриця.

Властивості операції множення погоджених матриць:

- а) $A(B + C) = AB + AC$;
- б) $(A + B)C = AC + BC$;
- в) $(AB)C = A(BC) = ABC$;
- г) $AE = A$, $EA = A$, E — одинична матриця.

Властивості операції транспонування:

- а) $(A^T)^T = A$;
- б) $(\lambda A)^T = \lambda A^T$;
- в) $(A + B)^T = A^T + B^T$;
- г) $(AB)^T = B^T A^T$.

7.1.3. Визначник матриці

Кожна квадратна матриця n -го порядку характеризується числом, яке називають визначником $|A|$ (або детермінантом $\det A$) і позначають так:

$$|A| = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

Визначник другого порядку задають рівністю

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21};$$

визначник третього порядку – рівністю

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - \\ - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}.$$

Знаки, з якими доданки входять до формули для обчислення визначника третього порядку, легко запам'ятати, скориставшись правилом трикутників (Сарруса):



Матрицю A називають виродженою, якщо її визначник дорівнює нулю $|A|=0$. У протилежному випадку, якщо $|A| \neq 0$, матриця є не-виродженою.

Основні властивості визначників

- При транспонуванні визначник не зміниться.
- Якщо елементи будь-якого рядка (стовпця) матриці помножити на число λ , то її визначник так само помножиться на це число λ .
- Якщо у визначника поміняти місцями два рядки (стовпці), то визначник лише змінить знак.
- Якщо всі елементи будь-якого рядка (стовпця) визначника дорівнюють нулю, то визначник дорівнює нулю.
- Якщо відповідні елементи двох рядків (стовпців) пропорційні, то визначник дорівнює нулю.
- Якщо всі елементи деякого рядка (стовпця) є сумою двох доданків, то визначник можна подати у вигляді суми двох визначників, елементи відповідних рядків (стовпців) яких дорівнюють відповідним доданкам.

Наприклад,

$$\begin{vmatrix} a'_{11} + a''_{11} & a_{12} & a_{13} \\ a'_{21} + a''_{21} & a_{22} & a_{23} \\ a'_{31} + a''_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a'_{11} & a_{12} & a_{13} \\ a'_{21} & a_{22} & a_{23} \\ a'_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a''_{11} & a_{12} & a_{13} \\ a''_{21} & a_{22} & a_{23} \\ a''_{31} & a_{32} & a_{33} \end{vmatrix}.$$

е. Визначник не зміниться, якщо до елементів деякого рядка (стовпця) додати елементи іншого рядка (стовпця), помноженого на довільний множник.

ж. Визначник добутку двох матриць дорівнює добутку визначників, тобто

$$|AB| = |A| \cdot |B|.$$

Мінором M_{ij} елемента a_{ij} називають визначник, який дістають з визначника матриці A вилученням i -го рядка та j -го стовпця. Якщо матриця A – матриця третього порядку:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

то

$$M_{11} = \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \quad M_{12} = \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix},$$

$$M_{13} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}, \dots, \quad M_{33} = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

Алгебраїчні доповнення елемента a_{ij} матриці A визначають за формулою

$$A_{ij} = (-1)^{i+j} M_{ij}.$$

Визначник дорівнює сумі добутків елементів будь-якого рядка або стовпця на їхні алгебраїчні доповнення, тобто

$$|A| = a_{i1}A_{i1} + a_{i2}A_{i2} + \dots + a_{in}A_{in}, \text{ або } |A| = a_{1r}A_{1r} + a_{2r}A_{2r} + \dots + a_{nr}A_{nr}.$$

Сума добутків елементів будь-якого рядка або стовпця на алгебраїчні доповнення відповідних елементів іншого рядка або стовпця дорівнює нулю, тобто

$$a_{i1}A_{j1} + a_{i2}A_{j2} + \dots + a_{in}A_{jn} = 0, \quad i \neq j, \text{ або } a_{1r}A_{1j} + a_{2r}A_{2j} + \dots + a_{nr}A_{nj} = 0, \quad i \neq j.$$

7.1.4. Ранг матриці

Мінором k -го порядку довільної матриці A називають визначник, що складається з елементів матриці, розміщених на перетині k рядків і k стовпців.

Рангом матриці A називають найбільший з порядків її відмінних від нуля мінорів і позначають $rg(A)$, $\text{rang}(A)$.

Елементарними перетвореннями матриці називають такі операції:

- перестановка місцями двох рядків (стовпців) матриці;
- множення рядка (стовпця) матриці на число, відмінне від нуля;
- додавання до елементів одного рядка (стовпця) відповідних елементів іншого рядка (стовпця).

При елементарних перетвореннях матриця перетворюється у східчасту (трикутну), ранг якої не змінюється.

Ранг східчастої матриці дорівнює кількості ненульових рядків.

Ранг матриці можна знайти за допомогою методу елементарних перетворень. Суть цього методу полягає в тому, що матрицю приводять до східчастого вигляду елементарними перетвореннями; кількість ненульових рядків отриманої східчастої матриці визначає ранг матриці.

Властивості рангу матриці:

а) $rgA = rgA^T$;

б) $rgAA^T = rgA$;

в) $rgAB \leq \min(rgA, rgB)$;

г) нехай A — будь-яка матриця; P , Q — будь-які невідроджені матриці, погоджені стосовно операції множення з матрицею A . Тоді $rgPAQ = rgA$.

Слідом матриці A називають суму всіх елементів матриці, розміщених на головній діагоналі, та позначають $\text{tr}A$, тобто

$$\text{tr}A = \sum_{i=1}^n a_{ii}.$$

Властивості сліду:

а) $\text{tr}A = \text{tr}A^T$;

б) $\text{tr}(\alpha A + \beta B) = \alpha \text{tr}A + \beta \text{tr}B$;

в) $\text{tr}(AB) = \text{tr}(BA)$;

г) $\text{tr}E_n = n$, E_n — одинична матриця порядку n .

Кожній системі рівнянь відповідають такі матриці:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}, \quad X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix},$$

де матрицю A називають матрицею системи, B – матрицею-стовпцем системи, а X – матрицею-стовпцем невідомих. Якщо до A приєднати матрицю-стовпець B , отриману матрицю називають $A|B$ розширеною.

Систему лінійних алгебраїчних рівнянь можна записати у матричній формі:

$$AX = B.$$

Система лінійних алгебраїчних рівнянь може не мати розв'язків (тоді її називають несумісною), може мати лише один розв'язок або багато (тоді її називають сумісною).

Питання про сумісність системи рівнянь розглядається в нижче наведеної теоремі.

Теорема Кронекера–Капеллі. Система лінійних алгебраїчних рівнянь сумісна тоді і тільки тоді, коли ранг матриці системи дорівнює рангу розширеної матриці: $r(A) = r(A|B)$.

Для сумісних систем справджуються твердження:

- 1) якщо ранг матриці сумісної системи збігається з кількістю невідомих ($r(A) = n$), система рівнянь має єдиний розв'язок;
- 2) якщо ранг матриці сумісної системи менший від кількості невідомих ($r(A) < n$), система рівнянь має безліч розв'язків. При цьому r змінних x_1, x_2, \dots, x_r називають основними (або базисними), якщо визначник матриці з коефіцієнтів при них (базисний мінор) відмінний від нуля. Решта $n - r$ змінних $x_{r+1}, x_{r+2}, \dots, x_n$ називають неосновними (або вільними).

Система лінійних рівнянь, в якій кількість відомих і кількість невідомих однакова ($m = n$), має єдиний розв'язок тоді і тільки тоді, коли її матриця невинроджена ($|A| \neq 0$).

7.1.7. Основні методи розв'язування системи рівнянь

Метод оберненої матриці. Якщо матриця системи невинроджена, то її розв'язок задається формулою

$$A^{-1}AX = EX = X = A^{-1}B.$$

Формули Крамера. Позначимо $\Delta = |A|$. Це основний визначник системи. Обчислимо допоміжні визначники Δ_j , які відповідають матрицям, отриманим заміною j -го стовпця даної матриці стовпцем B . Тоді розв'язок системи рівнянь можна представити рівностями

$$x_j = \frac{\Delta_j}{\Delta}, \quad j = \overline{1, n}.$$

Метод Гаусса. Це найпоширеніший метод розв'язування та дослідження системи лінійних рівнянь, який передбачає послідовне виключення невідомих. За допомогою елементарних перетворень розширену матрицю системи зводять до східчастого (трикутного) вигляду (прямий хід методу Гаусса):

$$\left(\begin{array}{ccccccc|c} a_{11} & a_{21} & \dots & a_{1r} & a_{1r+1} & \dots & a_{1n} & b_1 \\ 0 & a_{22}^{(1)} & \dots & a_{2r}^{(1)} & a_{2r+1}^{(1)} & \dots & a_{2n}^{(1)} & b_2^{(1)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{rr}^{(r-1)} & a_{rr+1}^{(r-1)} & \dots & a_{rn}^{(r-1)} & b_r^{(r-1)} \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & b_{r+1}^{(r-1)} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 & b_m^{(r-1)} \end{array} \right),$$

якому відповідає система рівнянь

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \dots + a_{1r}x_r + a_{1r+1}x_{r+1} + \dots + a_{1n}x_n = b_1, \\ a_{22}^{(1)}x_2 + \dots + a_{2r}^{(1)}x_r + a_{2r+1}^{(1)}x_{r+1} + \dots + a_{2n}^{(1)}x_n = b_2^{(1)}, \\ \dots \\ a_{rr}^{(r-1)}x_r + a_{rr+1}^{(r-1)}x_{r+1} + \dots + a_{rn}^{(r-1)}x_n = b_r^{(r-1)}, \\ 0 = b_{r+1}^{(r-1)}, \\ \dots \\ 0 = b_m^{(r-1)}. \end{array} \right.$$

Якщо хоча б одне із чисел $b_{r+1}^{(r-1)}, \dots, b_m^{(r-1)}$ відмінне від нуля, система рівнянь буде несумісною. У протилежному разі (якщо $b_{r+1}^{(r-1)} = \dots = b_m^{(r-1)} = 0$) вона буде сумісною.

Щоб помножити вектор на число, потрібно кожний її елемент помножити на це число:

$$c = \lambda a \Leftrightarrow c_i = \lambda a_i, \quad i = \overline{1, n}.$$

Два вектори називають колінеарними, якщо вони лежать на одній прямій або на паралельних прямих. У колінеарних векторів відповідні координати пропорційні:

$$a = \lambda b \quad \text{або} \quad \frac{a_1}{b_1} = \frac{a_2}{b_2} = \dots = \frac{a_n}{b_n} = \lambda.$$

7.2.2. Скалярний добуток. Базис

Скалярним добутком двох векторів a і b називають число, яке дорівнює добутку довжин цих векторів на косинус кута φ між ними (позначають $a^T b$ або (a, b)):

$$a^T b = |a||b|\cos\varphi.$$

Якщо вектори a і b задано своїми координатами, скалярний добуток цих векторів дорівнює сумі добутків їхніх відповідних координат:

$$(a, b) = a^T b = \sum_{i=1}^n a_i b_i,$$

Довжина (модуль, норма) вектора

$$|a| = \sqrt{a_1^2 + a_2^2 + \dots + a_n^2}.$$

З двох останніх формул випливає формула для визначення косинуса кута між двома векторами:

$$\cos\varphi = \frac{a^T b}{|a||b|} = \frac{a_1 b_1 + a_2 b_2 + \dots + a_n b_n}{\sqrt{a_1^2 + a_2^2 + \dots + a_n^2} \sqrt{b_1^2 + b_2^2 + \dots + b_n^2}},$$

де φ — кут між векторами a і b .

Властивості скалярного добутку:

- 1) $a^T b = b^T a$;
- 2) $(\lambda a)^T b = a^T (\lambda b) = \lambda a^T b$;
- 3) $(a + b)^T c = a^T c + b^T c$;
- 4) $a^T a = |a|^2$;
- 5) $a^T b = 0$ тоді і тільки тоді, коли $a \perp b$.

Якщо вектор $a = (a_x, a_y, a_z)$ утворює кути α, β, γ з осями координат Ox, Oy, Oz , то косинуси цих кутів називають напрямними і визначають за формулами

$$\cos \alpha = \frac{a_x}{\sqrt{a_x^2 + a_y^2 + a_z^2}}, \quad \cos \beta = \frac{a_y}{\sqrt{a_x^2 + a_y^2 + a_z^2}}, \quad \cos \gamma = \frac{a_z}{\sqrt{a_x^2 + a_y^2 + a_z^2}},$$

звідки

$$\cos^2 \alpha + \cos^2 \beta + \cos^2 \gamma = 1.$$

Систему векторів a_1, a_2, \dots, a_n називають лінійно незалежною, якщо рівність

$$k_1 a_1 + k_2 a_2 + \dots + k_n a_n = 0$$

виконується тоді і тільки тоді, коли $k_i = 0$ при всіх $i = 1, 2, \dots, n$. Якщо ця рівність виконується при хоча б одному значенні $k_j \neq 0$, то систему векторів називають *лінійно залежною*. У цьому разі вектор a_j лінійно виражається через інші вектори системи:

$$a_j = \frac{k_1}{k_j} a_1 + \dots + \frac{k_{j-1}}{k_j} a_{j-1} + \frac{k_{j+1}}{k_j} a_{j+1} + \dots + \frac{k_n}{k_j} a_n.$$

Два колінеарних вектори a і b лінійно залежні й один із них виражається через інший, наприклад: $b = \lambda a$, $\lambda \neq 0$. Два неколінеарних вектори лінійно незалежні.

Вектори називають компланарними, якщо вони лежать на одній площині або на паралельних площинах. Три компланарних вектори a, b і c лінійно залежні, а три некомпланарних — лінійно незалежні.

Базисом системи векторів a_1, a_2, \dots, a_n називають таку її підсистему, вектори якої лінійно незалежні, а будь-який вектор системи лінійно виражається через вектори підсистеми.

У просторі V_2 базисом є два неколінеарних вектори e_1 і e_2 . Довільний вектор \bar{a} , компланарний цим векторам, лінійно виражається через них: $\bar{a} = a_1 \bar{e}_1 + a_2 \bar{e}_2$, де числа a_1, a_2 — координати вектора \bar{a} в базисі \bar{e}_1, \bar{e}_2 .

У просторі V_3 базисом є три некомпланарних вектори e_1, e_2 і e_3 . Довільний вектор простору a лінійно виражається через них: $a = a_1 e_1 + a_2 e_2 + a_3 e_3$, де числа a_1, a_2 і a_3 — координати вектора a в базисі e_1, e_2 і e_3 .

Знаходження власного числа зведено до знаходження коренів многочлена, у цьому разі третього степеня. У загальному випадку характеристичне рівняння для власних чисел матриці є многочленом степеня n .

Проблема обчислення власних чисел для матриці зводиться до:

а) побудови характеристичного многочлена за елементами матриці;

б) знаходження коренів цього многочлена.

Матриця розмірності $n \times n$ має n власних чисел, серед яких можуть бути кратні й комплексні.

Сукупність усіх власних векторів матриці є системою лінійно незалежних векторів.

7.3. Функція однієї змінної

7.3.1. Означення, область визначення, область зміни

Нехай маємо множину X елементів x і множину Y елементів y . Якщо кожному значенню $x \in X$ відповідає за деяким правилом або законом f єдине значення $y \in Y$, то кажуть, що на множині X задано функцію $y = f(x)$.

Множину X , в якій визначено функцію, називають областю визначення. Якщо цю множину конкретно не вказано, під областю визначення розуміють множину, для кожного елемента якої правило f має зміст. Її ще називають областю допустимих значень аргументу.

Сукупність усіх значень y , яких може набувати функція $y = f(x)$ зі зміною аргументу x в області визначення X , називають областю зміни функції.

Функцію називають обмеженою зверху на відрізьку $[a, b]$, якщо для всіх значень $x \in [a, b]$ існує така константа M , що $f(x) \leq M$; функція буде обмеженою знизу, якщо існує $M \in R$, що для $x \in [a, b]$ виконується нерівність $f(x) \geq M$; функція буде обмеженою на відрізьку $[a, b]$, якщо для всіх значень $x \in [a, b]$ існує така константа $M > 0$, що $|f(x)| \leq M$.

Функцію $y = f(x)$ називають монотонно зростаючою на проміжку (a, b) , якщо для довільних $x_1 \in (a, b)$ і $x_2 \in (a, b)$, для таких, що $x_1 > x_2$, виконується нерівність $f(x_1) > f(x_2)$; функцію називають монотонно спадною, якщо для тих самих значень x_1, x_2 виконується нерівність $f(x_1) < f(x_2)$.

Графіком функції $y = f(x)$ є сукупність усіх точок (x, y) на площині.

Способи задання функції: 1) аналітичний, 2) табличний, 3) графічний.

Класифікація функцій. Основними елементарними функціями є:

- стала $y = C = \text{const}$;
- степеневі $y = x^a$;
- показникові $y = a^x$, $a > 0$;
- логарифмічні $y = \log_a x$, $a > 0$;
- тригонометричні $y = \sin x$, $y = \cos x$, $y = \text{tg } x = \frac{\sin x}{\cos x}$,
 $y = \text{ctg } x = \frac{\cos x}{\sin x}$;
- обернені тригонометричні $y = \arcsin x$, $y = \arccos x$, $y = \text{arctg } x$,
 $y = \text{arcctg } x$.

Функцію називають складною, якщо її аргументом є інша функція $y = f(\varphi(x))$; функцію $y = f(u)$ називають зовнішньою, а $u = \varphi(x)$ – внутрішньою. Наприклад: $y = \sin(e^x)$. Тут функція $y = \sin u$ – зовнішня, а $u = e^x$ – внутрішня.

Елементарними називають функції, які отримують із основних елементарних функцій за допомогою скінченного числа арифметичних дій (додавання, віднімання, множення і ділення) та суперпозиції кількох функцій.

Приклад неелементарної функції є така: $|x| = \begin{cases} x, & x \geq 0, \\ -x, & x < 0. \end{cases}$

Якщо кожному значенню $y \in Y$ відповідає за деяким правилом або законом g таке єдине значення $x \in X$, що $y = f(x)$, то функцію $x = g(y)$ називають оберненою до $y = f(x)$ і позначають $y = f^{-1}(x)$.

Обернена функція існує тільки для неперервних і строго монотонних функцій.

7.3.2. Похідна функції

Приростом функції $y = f(x)$ називатимемо різницю

$$\Delta y = f(x + \Delta x) - f(x),$$

де Δx – приріст аргументу x .

Розглянемо відношення

$$\frac{\Delta y}{\Delta x} = \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.$$

Якщо існує його скінченна границя, коли $\Delta x \rightarrow 0$ довільним чином, то її називають похідною функції $y = f(x)$ у точці x_0 і позначають

$$y' = f'(x) = \frac{dy}{dx} = \frac{df(x)}{dx} = \lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.$$

Якщо функція $y = f(x)$ має похідну в точці x_0 , її називають диференційованою в цій точці. Якщо функція має похідну в кожній точці деякого проміжку, її називають диференційованою у цьому проміжку.

Геометричний зміст похідної. Із рис. 7.1 видно, що $\operatorname{tg} \beta = \frac{\Delta y}{\Delta x}$, де кут β є кутом нахилу січної M . Якщо $\Delta x \rightarrow 0$, точка M прямує до точки M_0 , січна NM — до дотичної N_0M_0 , а відношення приростів функції й аргументу — до похідної функції в точці x_0 . Отже, похідна функції $y = f(x)$ у точці x_0 дорівнює кутовому коефіцієнту дотичної до графіка функції в тій самій точці.

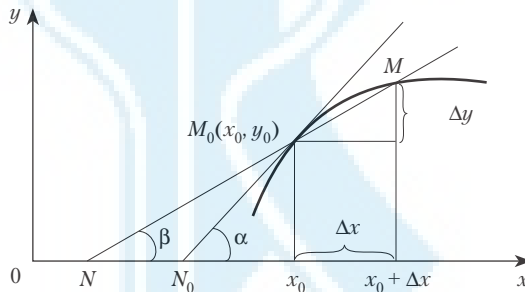


Рис. 7.1

Фізичний зміст похідної. Якщо функція $y = f(x)$ визначає шлях, пройдений точкою за час x , то відношення приросту функції до приросту аргументу $\Delta y / \Delta x$ є середньою швидкістю руху точки за час Δx , а при $\Delta x \rightarrow 0$ ця швидкість прямує до миттєвої швидкості в точці x_0 :

$$V(x_0) = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x} = f'(x_0).$$

Економічний зміст похідної. Якщо функції $V(x)$, $D(x)$, $P(x)$ визначають відповідно витрати, дохід і прибуток, а змінна x — обсяг продукції, то відношення приросту цих функцій до приросту аргументу характеризує приріст відповідних функцій на одиницю

продукції. Границя цих відношень при $\Delta x \rightarrow 0$ стає маргіальною, або граничною. Отже, маємо:

маргіальна вартість

$$V'(x_0) = \lim_{\Delta x \rightarrow 0} \frac{\Delta V}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{V(x_0 + \Delta x) - V(x_0)}{\Delta x};$$

маргіальний дохід

$$D'(x_0) = \lim_{\Delta x \rightarrow 0} \frac{\Delta D}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{D(x_0 + \Delta x) - D(x_0)}{\Delta x};$$

маргіальний прибуток

$$P'(x_0) = \lim_{\Delta x \rightarrow 0} \frac{\Delta P}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{P(x_0 + \Delta x) - P(x_0)}{\Delta x}.$$

Геометричний зміст похідної дає змогу записати рівняння прямої, що визначає дотичну до графіка функції $y = f(x)$, яка проходить через точку $(x_0; f(x_0))$:

$$y = f(x_0) + f'(x_0)(x - x_0),$$

оскільки $k = \operatorname{tg} \varphi = f'(x_0)$.

Правила диференціювання

Нехай $f = f(x)$ та $g = g(x)$ — деякі диференційовані функції. Тоді справджуються такі правила диференціювання:

$$(f \pm g)' = f' \pm g'; \quad (cf)' = cf';$$

$$(f \cdot g)' = f' \cdot g + f \cdot g'; \quad \left(\frac{f}{g}\right)' = \frac{f' \cdot g - f \cdot g'}{g^2}, \quad (g \neq 0).$$

Похідна складної функції. Якщо $y = f(u)$, $u = \varphi(x)$, тобто $y = f(\varphi(x))$ — складна функція, то її похідна дорівнює добутку похідних:

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}.$$

Похідна оберненої функції. Якщо функція $x = g(y)$ є оберненою функцією до $y = f(x)$, тобто $f(g(y)) = y$, то виконується така рівність:

$$x'_y(x_0) = \frac{1}{y'_x(x_0)}, \quad y_0 = f(x_0).$$

Похідна неявної функції. Якщо функція задана неявно $F(x, y) = 0$, то похідну функції $y'(x)$ шукають за формулою

$$\left[F(x, y(x)) \right]' = F'_x(x, y) + F'_y(x, y)y'(x) = 0 \Rightarrow y'(x) = -\frac{F'_x(x, y)}{F'_y(x, y)}.$$

7.3.3. Аналіз функції за допомогою похідних

Ознака монотонності. Якщо $f'(x) > 0$, $x \in (a, b)$, функція $f(x)$ зростає на (a, b) . Якщо $f'(x) < 0$, $x \in (a, b)$, функція $f(x)$ спадає на (a, b) .

Необхідна умова екстремуму. Якщо функція $y = f(x)$ досягає екстремуму в точці x_0 , її похідна в цій точці дорівнює нулю або не існує.

Достатні умови екстремуму. Назвемо точку x_0 критичною для $y = f(x)$, якщо в ній існує похідна і $f'(x_0) = 0$.

Перша достатня умова. Функція $y = f(x)$ досягає максимуму в критичній точці, якщо похідна функції змінює свій знак при переході через точку x_0 з “+” на “-”, і досягає мінімуму, якщо похідна змінює свій знак з “-” на “+”.

Друга достатня умова. Якщо функція двічі диференційована і в критичній точці $f''(x) < 0$, то в цій точці функція досягає мінімуму, а якщо $f''(x) > 0$ – максимуму.

7.4. Функція кількох змінних

7.4.1. Означення, область визначення

Сукупність усеможливих упорядкованих наборів (точок) $X = (x_1; x_2; \dots x_n)$, відстань між якими визначається за формулою

$$\rho(X, Y) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \dots + (y_n - x_n)^2},$$

називають *n*-вимірним евклідовим простором.

Функцією багатьох дійсних змінних називають закон, який кожній точці $X = (x_1; x_2; \dots x_n)$ множини $E \subset R_n$ ставить у відповідність єдине дійсне число y . При цьому записують: $y = f(X)$, $X \in E$, або $y = f(x_1; x_2; \dots x_n)$, $(x_1; x_2; \dots x_n) \in E$. Множину E називають областю визначення функції багатьох змінних.

Якщо явно не вказано область визначення, то під областю визначення розуміють множину всіх точок $X = (x_1; x_2; \dots x_n) \in R_n$, при яких закон має зміст.

Надалі розглядатимемо функції двох змінних. За традицією аргументи позначають літерами x та y , а функцію – літерою z : $z = f(x, y)$.

Лінією рівня називають лінію, яка визначається рівністю $f(x, y) = C$, $C = \text{const}$, тобто лінію, яку одержуємо при перетині графіка функції $z = f(x, y)$ площиною $Z = C$.

7.4.2. Частинні похідні, похідна за напрямком, диференціал

Частинною похідною функції $z = f(x, y)$ за змінною x називають границю $\lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x; y_0) - f(x_0; y_0)}{\Delta x}$, якщо ця границя існує. Символічно цю похідну позначають одним з таких способів:

$$f'_x(x_0, y_0) \text{ або } \frac{\partial f(x_0, y_0)}{\partial x} \text{ або } z'_x(x_0, y_0).$$

Аналогічно означають частинну похідну функції $z = f(x, y)$ за змінною y як границю $\lim_{\Delta y \rightarrow 0} \frac{f(x_0; y_0 + \Delta y) - f(x_0; y_0)}{\Delta y}$, якщо вона існує, і позначають

$$f'_y(x_0, y_0), \text{ або } \frac{\partial f(x_0, y_0)}{\partial y}, \text{ або } z'_y(x_0, y_0).$$

Похідна z'_l за напрямком, що визначається вектором $\bar{l} = (l_x, l_y)$, виражається через частинні похідні за формулою

$$z'_e = z'_x \cdot \cos \alpha + z'_y \cdot \cos \beta,$$

де $\cos \alpha = \frac{l_x}{\sqrt{l_x^2 + l_y^2}}$ і $\cos \beta = \frac{l_y}{\sqrt{l_x^2 + l_y^2}}$ — напрямні косинуси вектора \bar{l} .

Градiєнтом функції $z = f(x, y)$ називають вектор з координатами $(z'_x; z'_y)$ і позначають його $\nabla z = \overline{\text{grad}} z = (z'_x; z'_y)$. Градiєнт ∇z функції $z = f(x, y)$ у точці $(x_0; y_0)$ вказує напрям найшвидшої зміни функції у цій точці.

Диференціалом функції $z = f(x, y)$ називають суму добутків частинних похідних цієї функції на прирости відповідних незалежних змінних:

$$dz = z'_x \cdot \Delta x + z'_y \cdot \Delta y, \text{ або } dz = z'_x \cdot dx + z'_y \cdot dy$$

оскільки $dx = \Delta x$, $dy = \Delta y$.

7.4.3. Екстремум функції двох змінних. Найбільше та найменше значення функції двох змінних у замкненій області

Точку $M_0(x_0; y_0)$ називають точкою локального мінімуму функції $z = f(x, y)$, якщо існує окіл точки M_0 такий, що для всіх точок $(x; y)$ з цього околу виконується умова $f(x; y) > f(x_0; y_0)$.

Точку $M_0(x_0; y_0)$ називають точкою локального максимуму функції $z = f(x, y)$, якщо існує окіл точки M_0 такий, що для всіх точок $(x; y)$ з цього околу виконується умова $f(x; y) < f(x_0; y_0)$.

Точку, в якій всі частинні похідні перетворюються на нуль, називають стаціонарною (або критичною).

Необхідна умова екстремуму. Якщо $M_0(x_0; y_0)$ – точка екстремуму диференційовної функції $z = f(x, y)$, то в цій точці виконуються умови

$$\begin{cases} f'_x(x_0, y_0) = 0, \\ f'_y(x_0, y_0) = 0. \end{cases}$$

Достатня умова екстремуму. Нехай $M_0(x_0; y_0)$ – стаціонарна точка функції $z = f(x_0; y_0)$, причому в околі точки M_0 існують неперервні частинні похідні другого порядку цієї функції:

$$f''_{xx}(x_0; y_0) = A; \quad f''_{xy}(x_0; y_0) = f''_{yx}(x_0; y_0) = B; \quad f''_{yy}(x_0; y_0) = C.$$

Тоді якщо:

1) $\Delta = AC - B^2 > 0$, то функція $z = f(x, y)$ у точці $M_0(x_0; y_0)$ має екстремум:

а) при $A > 0$, M_0 – точка мінімуму,

б) при $A < 0$, M_0 – точка максимуму;

2) $\Delta < 0$, точка $M_0(x_0; y_0)$ не є точкою екстремуму функції $z = f(x, y)$;

3) $\Delta = 0$, потрібне додаткове дослідження.

7.5. Теорія ймовірностей

7.5.1. Означення ймовірності

Теорія ймовірностей вивчає випадкові (стохастичні) експерименти, які прийнятні теоретично можна повторювати безліч разів. Випадковим називають експеримент, умови якого не змінюються, а результати (наслідки) можуть бути різними і точно непередбачени-

ми. Випадковою називають подію, яка може відбутися чи не відбутися в умовах даного експерименту. Імовірність — це міра можливості події.

Класичне означення. Імовірність події A дорівнює відношенню кількості елементарних наслідків, які сприяють появі події A , до загальної кількості всіх єдиноможливих і рівноможливих елементарних наслідків.

Імовірність події A позначають $P(A)$, за означенням

$$P(A) = \frac{m}{n},$$

де m — кількість елементарних наслідків, що сприяють події A ; n — кількість усіх єдиноможливих і рівноможливих елементарних наслідків.

Це класичне означення має місце лише тоді, коли числа m , n — скінченні, усі елементарні наслідки рівноможливі.

Якщо множина елементарних наслідків нескінченна і займає деяку область G , а подія A — лише частину цієї області g , то ймовірність події A визначають згідно з геометричним означенням імовірності.

Геометричне означення. Імовірність події A дорівнює відношенню міри g до міри G :

$$P(A) = \frac{m(g)}{m(G)},$$

Як міру можна брати або довжину відрізків, або площу об'єктів, або об'єм тощо.

Відносною частотою події A називають відношення кількості випробувань, в якій подія A з'явилась, до кількості фактично виконаних випробувань.

Відносну частоту події A позначають $W(A)$ або $p_n(A)$

$$p_n(A) = W(A) = \frac{m}{n},$$

де m — кількість випробувань, в яких з'явилась подія A ; n — кількість усіх випробувань. Зазначимо, що ймовірність $P(A)$ обчислюють до випробування, а відносну частоту $p_n(A)$ — після випробування.

Статистичне означення. Статистична ймовірність події A — це відносна частота або число, близьке до неї.

7.5.2. Дискретні та неперервні випадкові величини

Випадковою називають величину, яка в результаті випробування набуває одного й тільки одного можливого значення, яке можна виміряти, але заздалегідь неможливо передбачити, тому що залежить від випадкових причин, які наперед неможливо врахувати.

Дискретною випадковою величиною називають таку величину, яка може набувати відокремлених ізольованих одне від одного числових значень з відповідними ймовірностями. Дискретні випадкові величини можуть бути як скінченні, так і злічені.

Законом розподілу дискретної випадкової величини називають співвідношення між можливими значеннями та їхніми ймовірностями. Ці співвідношення можна задати у вигляді таблиці, аналітично та графічно.

$X:$	x_1	X_2	...	X_n
$P:$	p_1	P_2	...	P_2

Також має виконуватись умова $p_1 + p_2 + \dots + p_n = 1$. Графічно дискретну випадкову величину задають полігоном частот або гістограмою.

Неперервною випадковою величиною називають таку величину, яка може набувати всіх значень із деякого скінченного або нескінченного інтервалу (a, b) . Для її повної характеристики вводять інтегральну та диференціальну функції розподілу.

Інтегральною функцією розподілу неперервної випадкової величини називають імовірність того, що випадкова величина X набуде значення, меншого від x :

$$F(x) = P(X < x).$$

Із означення випливає основна формула теорії ймовірності: якщо НВВ може набувати будь-якого значення з (a, b) , то

$$P(a < x < b) = F(b) - F(a).$$

Оскільки $P(x = a) = 0$, то остання формула може мати вигляд

$$P(a < x < b), P(a \leq x < b), P(a < x \leq b), P(a \leq x \leq b).$$

Властивості інтегральної функції розподілу $F(x)$:

- 1) $0 \leq F(x) \leq 1$;
- 2) $F(x)$ – неспадна функція,
- 3) $F(x) = 0, x < a; F(x) = 1, x > b$.

Властивість 3) виконується за умови, що випадкова величина X набуває можливого значення зі скінченного проміжку (a, b) . Якщо

вона може набувати будь-якого дійсного значення ($a = -\infty$, $b = +\infty$), то

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

Диференціальною функцією розподілу неперервної випадкової величини або щільністю ймовірностей називають похідну першого порядку від її інтегральної функції: $f(x) = F'(x)$.

Властивості щільності ймовірностей $f(x)$:

- 1) $f(x) \geq 0$;
- 2) $f(x) = 0$, $x < a$, $x > b$;
- 3) $\int_a^b f(x) dx = 1$.

Якщо $a = -\infty$, $b = \infty$, то друга властивість набуває вигляду:

$$\lim_{x \rightarrow \pm\infty} f(x) = 0, \quad \text{а третя} - \int_{-\infty}^{+\infty} f(x) dx = 1.$$

Графік щільності ймовірностей $f(x)$ називають кривою розподілу.

Якщо відомо щільність ймовірностей неперервної випадкової величини $f(x)$, то інтегральну функцію розподілу $F(x)$ визначають за формулою

$$F(x) = \int_{-\infty}^x f(y) dy.$$

При цьому ймовірність потрапляння випадкової величини в проміжок (a, b) визначається, очевидно, з рівності

$$P(a < x < b) = \int_a^b f(x) dx.$$

7.5.3. Числові характеристики випадкової величини

До числових характеристик випадкових величин належать:

- а) математичне сподівання;
- б) дисперсія;
- в) середньоквадратичне відхилення;
- г) мода й медіана;
- д) початкові й центральні моменти.

Математичне сподівання випадкової величини X є синонімом “середнє значення”.

Математичним сподіванням дискретної випадкової величини X називають число

$$M(X) = \sum_{i=1}^n x_i p_i.$$

Математичним сподіванням неперервної випадкової величини X називають число

$$M(X) = \int_a^b xf(x)dx \text{ для } X \in (a, b),$$

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx \text{ для } X \in (-\infty, +\infty).$$

Властивості математичного сподівання:

- 1) $M(C) = C$;
- 2) $M(X + Y) = M(X) + M(Y)$;
- 3) $M(XY) = M(X)M(Y)$, якщо X і Y – незалежні випадкові величини;
- 4) $M(CX) = CM(X)$;
- 5) $M(AX + B) = AM(X) + B$.

Дисперсія визначає розсіювання випадкової величини відносно математичного сподівання (середнього значення).

Дисперсією випадкової величини Z називають математичне сподівання квадрата відхилення випадкової величини від математичного сподівання:

$$D(X) = M\left(\left(X - M(X)\right)^2\right) = M\left(X^2\right) - \left(M(X)\right)^2.$$

Для дискретної випадкової величини

$$D(X) = \sum_{i=1}^n (x_i - M(X))^2 p_i = \sum_{i=1}^n x_i^2 p_i - \left(M(X)\right)^2,$$

для неперервної випадкової величини

$$D(X) = \int_{-\infty}^{+\infty} (x - M(X))^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - \left(M(X)\right)^2.$$

Властивості дисперсії:

- 1) $D(C) = 0$;
- 2) $D(X + Y) = D(X) + D(Y)$;
- 3) $D(CX) = C^2 D(X)$;
- 4) $D(AX + B) = A^2 D(X)$;
- 5) $D(X - Y) = D(X) + D(Y)$;
- 6) $D(X) \geq 0$.

Середнім квадратичним відхиленням випадкової величини називають величину

$$\sigma(X) = \sqrt{D(X)}.$$

Модюю (Mo) дискретної випадкової величини називають її можливе значення, якому відповідає найбільша ймовірність $Mo = x_k$, якщо $p_k = \max(p_1, p_2, \dots, p_n)$.

Модюю (Mo) неперервної випадкової величини називають її можливе значення, якому відповідає максимальне значення щільності ймовірності: $x = Mo$ $f(Mo) = \max f(x)$.

Якщо випадкова величина має одну моду, то такий розподіл ймовірностей називають одномодальним, якщо більше — мультимодальним.

Медіаною (Me) називають її значення, для якого виконується рівність

$$P(-\infty < Me) = P(Me < x < +\infty) \Rightarrow F(Me) = 0,5.$$

Початковим моментом i -го порядку випадкової величини називають число $v_i = M(X^i)$ ($i = 1, 2, \dots, n$).

Центральним моментом i -го порядку випадкової величини називають число $\mu_i = M((X - M(X))^i)$ ($i = 1, 2, \dots, n$).

Для $i = 1$

$$v_1 = M(X); \mu_1 = M(X - M(X)) = 0.$$

Для $i = 2$

$$v_2 = M(X^2); \mu_2 = M(X - M(X))^2 = D(X).$$

Центральний момент третього порядку характеризує асиметрію закону розподілу випадкової величини. Якщо $\mu_3 = 0$, випадкова величина X симетрично розподілена відносно $M(X)$. Безрозмірна випадкова величина $As = \frac{\mu_3}{\sigma^3}$ називається коефіцієнтом асиметрії.

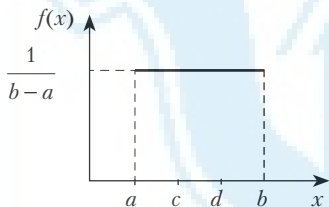
Центральний момент четвертого порядку використовують для визначення ексцесу, що характеризує плосковершинність або гостровершинність щільності випадкової величини $f(x)$. Ексцес обчислюють за формулою $Es = \frac{\mu_4}{\sigma^4} - 3$.

7.5.4. Закони розподілу неперервної випадкової величини

Рівномірний закон розподілу

$$f(x) = \begin{cases} \frac{1}{b-a}, & x \in (a, b), \\ 0, & x \notin (a, b). \end{cases} \quad F(x) = \int_a^x \frac{1}{b-a} dx = \begin{cases} 0, & x < a, \\ \frac{x-a}{b-a}, & a \leq x \leq b, \\ 1, & x > b. \end{cases}$$

$$M(x) = \frac{a+b}{2}, \quad D(x) = \frac{(b-a)^2}{12}, \quad \sigma(x) = \frac{b-a}{\sqrt{12}}.$$



Імовірність того, що випадкова величина, розподілена за рівномірним законом, потрапляє у проміжок (c, d) , визначають за формулою

$$P(c < x < d) = \frac{d-c}{b-a}.$$

Показниковий закон розподілу:

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

$$F(x) = \int_0^x \lambda e^{-\lambda x} dx = \begin{cases} 0, & x < 0, \\ 1 - e^{-\lambda x}, & x \geq 0. \end{cases}$$

$$M(x) = \frac{1}{\lambda}, \quad D(x) = \frac{1}{\lambda^2}, \quad \sigma(x) = \frac{1}{\lambda}.$$

Імовірність того, що випадкова величина, розподілена за показниковим законом, потрапляє у проміжок (c, d) , $c \geq 0$, визначають за формулою

$$P(c < x < d) = e^{-c\lambda} - e^{-d\lambda}.$$

Нормальний закон розподілу:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \quad F(x) = \int_0^x f(x) dx.$$

$$M(x) = a, \quad D(x) = \sigma^2, \quad \sigma(x) = \sigma.$$

Оскільки нормальний закон розподілу залежить лише від параметрів a і σ , часто цей закон розподілу позначають $N(a; \sigma)$.

Якщо $a = 0$, $\sigma = 1$, маємо нормальний нормований закон розподілу, який позначають $N(0; 1)$:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \text{ — локальна функція Лапласа;}$$

$$\Phi(x) = \int_0^x f(x) dx \text{ — інтегральна функція Лапласа.}$$

$$F(x) = \Phi(x) + 0,5.$$

Імовірність того, що випадкова величина, розподілена за нормальним законом, потрапляє у проміжок (c, d) , визначають за формулою

$$P(c < x < d) = \Phi\left(\frac{d-a}{\sigma}\right) - \Phi\left(\frac{c-a}{\sigma}\right).$$

Якщо проміжком є $(a - \delta, a + \delta)$, то

$$P(|x - a| < \delta) = P(a - \delta < x < a + \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right).$$

Правило трьох сигм для нормального закону. Якщо $\delta = 3\sigma$, то

$$P(|x - a| < 3\sigma) = 2\Phi\left(\frac{3\sigma}{\sigma}\right) = 2\Phi(3) \approx 1.$$

Імовірність того, що в результаті проведення експерименту випадкова величина X , яка має нормальний закон розподілу, потрапить у проміжок $(a - 3\sigma, a + 3\sigma)$, практично дорівнює одиниці, а тому її вважають вірогідною.

Логарифмічно нормальний закон розподілу. Це такий закон розподілу, за якого логарифм від нього є нормально розподілений. Якщо неперервна випадкова величина X нормально розподілена, то $Y = e^X$ логнормально розподілена.

$$f(y) = \begin{cases} 0, & y \leq 0, \\ \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(\ln y - a)^2}{2\sigma^2}}, & y > 0; \end{cases}$$

$$M(Y) = e^{a + \frac{\sigma^2}{2}}, \quad D(Y) = e^{2(2\sigma^2 + a)^2 - a^2} - e^{2a + \sigma^2}.$$

Розподіл χ^2 (хі-квадрат). Нехай X_i , $i = 1, 2, \dots, n$ — нормальні нормовані незалежні випадкові величини. Тоді сума квадратів цих величин

$$\chi^2 = \sum_{i=1}^n X_i^2$$

розподілена за законом χ^2 (хі-квадрат) з $k = n$ степенями вільності.

Диференціальна функція розподілу χ^2 має вигляд

$$f(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{2^{k/2} \cdot \Gamma\left(\frac{k}{2}\right)} e^{-\frac{x}{2}} \frac{k-1}{2x^2}, & x \geq 0, \sqrt{a^2 + b^2}, \end{cases}$$

де $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$ – гама-функція. Зокрема, $\Gamma(k+1) = k!$;

$$M(X) = k, \quad D(X) = 2k, \quad \sigma(X) = \sqrt{2k}.$$

Розподіл Стьюдента. Нехай X – нормальна нормована випадкова величина, а Y – незалежна від X величина, яка розподілена за законом χ^2 (хі-квадрат) з k степенями вільності. Тоді величину

$$T = \frac{X}{\sqrt{Y/k}}$$

називають T -розподілом, або розподілом Стьюдента з k степенями вільності. Його диференціальну та інтегральну функції задають формулами

$$f_T(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k} \Gamma\left(\frac{k}{2}\right)} \left(1 + \frac{x^2}{k}\right)^{-\frac{k+1}{2}}, \quad F_T(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k} \Gamma\left(\frac{k}{2}\right)} \int_{-\infty}^x \left(1 + \frac{z^2}{k}\right)^{-\frac{k+1}{2}} dz.$$

$$M(X) = 0, \quad D(X) = \frac{k}{k-2}, \quad \sigma(X) = \sqrt{\frac{k}{k-2}}.$$

Зі зростанням числа n до нескінченності розподіл Стьюдента наближається до нормального нормованого закону розподілу.

Розподіл Фішера–Снедекора. Неперервна випадкова величина X має розподіл Фішера–Снедекора, якщо

$$f(x) = \begin{cases} 0, & x < 0, \\ \frac{\Gamma\left(\frac{k_1 + k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)} \left(\frac{k_2}{k_1}\right)^{\frac{k_2}{k_1}} x^{\frac{k_2}{k_1}-1} \left(1 + \frac{k_2}{k_1} x\right)^{-\frac{k_1+k_2}{2}}, & x > 0; \end{cases}$$

$$F(x) = \begin{cases} 0, & x < 0, \\ \frac{\Gamma\left(\frac{k_1+k_2}{2}\right)}{\Gamma\left(\frac{k_1}{2}\right)\Gamma\left(\frac{k_2}{2}\right)} \left(\frac{k_2}{k_1}\right)^{\frac{k_2}{k_1}} \int_0^x z^{\frac{k_2}{k_1}-1} \left(1 + \frac{k_2}{k_1}z\right)^{-\frac{k_1+k_2}{2}} dz, & x > 0; \end{cases}$$

$$M(X) = \frac{k_1}{k_1-2}, \quad D(X) = \frac{2k_1^2(k_2+k_1-2)}{k_2(k_1-2)^2(k_1-4)} \quad \sigma(X) = \sqrt{2k}.$$

7.6. Математична статистика

7.6.1. Генеральна і вибіркова сукупності. Статистичний розподіл вибірки. Числові характеристики вибірки

Вибірковою сукупністю, або просто вибіркою, називають сукупність випадково вибраних об'єктів.

Генеральною сукупністю називають досліджувану сукупність, з якої проводять вибірку. Генеральна сукупність, як правило, має великий обсяг.

Обсягом сукупності називають кількість об'єктів цієї сукупності.

Якщо з генеральної сукупності проведена вибірка і величина x_1 спостерігалась n_1 разів, $x_2 - n_2$ разів, $x_k - n_k$ разів і $\sum n_i = n$ — обсяг вибірки, то значення x_i називають варіантами, відповідні їм числа спостережень n_i — частотами, а їх відношення до обсягу вибірки

$W_i = \frac{n_i}{n}$ — відносними частотами. Варіанти, що записані в порядку зростання, називають варіаційним рядом.

Статистичним розподілом вибірки називають перелік варіант і відповідних їм частот або відносних частот.

Генеральною середньою величиною \bar{x} називають середнє арифметичне значення ознаки генеральної сукупності. Якщо всі значення x_1, x_2, \dots, x_n ознаки генеральної сукупності обсягу N різні, то

$$\bar{x}_\Gamma = \frac{x_1 + x_2 + \dots + x_n}{N}.$$

Якщо значення ознаки генеральної сукупності x_1, x_2, \dots, x_n мають відповідні частоти N_1, N_2, \dots, N_k , де $N_1 + N_2 + \dots + N_k = N$, то

$$\bar{x}_\Gamma = \frac{N_1x_1 + N_2x_2 + \dots + N_kx_k}{N},$$

тобто генеральна середня є середньозважена величина значень ознаки за відповідними частотами.

Вибірковою середньою величиною \bar{x}_B називають середнє арифметичне значення ознаки вибіркової сукупності, яка дорівнює

$$\bar{x}_B = \frac{x_1 + x_2 + \dots + x_n}{n},$$

якщо всі значення x_1, x_2, \dots, x_n ознаки вибіркової сукупності обсягу n різні, та

$$\bar{x}_\Gamma = \frac{n_1x_1 + n_2x_2 + \dots + n_kx_k}{n},$$

якщо значення x_1, x_2, \dots, x_n ознаки вибіркової сукупності зустрічаються з відповідними частотами n_1, n_2, \dots, n_k , де $n_1 + n_2 + \dots + n_k = n$.

Генеральною дисперсією D_Γ називають середнє арифметичне квадратів відхилень значень ознаки генеральної сукупності від їх середнього значення:

$$D_\Gamma = \frac{\sum_{i=1}^N (x_i - \bar{x}_\Gamma)^2}{N} \quad \text{або} \quad D_\Gamma = \frac{\sum_{i=1}^k N_i (x_i - \bar{x}_\Gamma)^2}{N}.$$

Генеральним середнім квадратичним відхиленням називають квадратний корінь із генеральної дисперсії:

$$\sigma_\Gamma = \sqrt{D_\Gamma}.$$

Так само вводять вибіркoву дисперсію та вибіркoве середнє квадратичне відхилення:

$$D_B = \frac{\sum_{i=1}^N (x_i - \bar{x}_B)^2}{n}, \quad \text{або} \quad D_B = \frac{\sum_{i=1}^k n_i (x_i - \bar{x}_B)^2}{n};$$

$$\sigma_B = \sqrt{D_B}.$$

Існує ще інша формула для обчислення дисперсії:

$$D = \bar{x}^2 - (\bar{x})^2 = \frac{\sum n_i x_i^2}{n} - \left(\frac{\sum n_i x_i}{n} \right)^2.$$

Груповим середнім значенням називають середнє арифметичне значення ознаки, що належить групі.

Знаючи групові середні та обсяги груп, можна знайти загальне середнє значення, яке дорівнює середньому арифметичному, зваженому за обсягами груп:

$$\bar{x}_{\text{заг}} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + \dots + n_k\bar{x}_k}{n},$$

де n_i – обсяги i -ї групи; \bar{x}_i – групове середнє i -ї групи, $n = n_1 + n_2 + \dots + n_k$ – обсяг усіх груп.

Групова, внутрішньогрупова, міжгрупова та загальна дисперсії.

Груповою дисперсією називають дисперсію значень ознаки, що належить групі, відносно групового середнього значення:

$$D_{j\text{гр}} = \frac{\sum n_i(x_i - \bar{x}_j)^2}{N_j},$$

де j – номер групи; n_i – частота значень x_i ; \bar{x}_j – групове середнє значення групи j ; $N_j = \sum n_i$ – обсяг групи j .

Внутрішньогруповою дисперсією називають середнє арифметичне значення групових дисперсій, зважене за обсягами груп:

$$D_{\text{внгр}} = \frac{\sum N_j D_{j\text{гр}}}{n},$$

де j – номер групи; N_j – обсяг групи; $D_{j\text{гр}}$ – групова дисперсія групи j ; $n = \sum N_j$ – обсяг усієї сукупності.

Міжгруповою дисперсією називають дисперсію групових середніх значень відносно загального середнього значення:

$$D_{\text{міжгр}} = \frac{\sum N_j(\bar{x}_j - \bar{x})^2}{n},$$

де j – номер групи; N_j – обсяг групи; \bar{x}_j – середнє значення групи j ; \bar{x} – загальне середнє значення всієї сукупності; $n = \sum N_j$ – обсяг усієї сукупності.

Загальна дисперсія всієї сукупності дорівнює сумі внутрішньогрупової та міжгрупової дисперсій:

$$D_{\text{заг}} = D_{\text{внгр}} + D_{\text{міжгр}}.$$

7.6.2. Статистичні оцінки параметрів розподілу

Для оцінювання параметрів розподілу випадкової величини, що представлена вибіркою, застосовують числові характеристики вибірки, зокрема середнє вибіркоче і вибіркочу дисперсію.

Незмщеною називають оцінку, математичне сподівання якої дорівнює оцінюваному параметру. Зміщеною називають оцінку, математичне сподівання якої не дорівнює оцінюваному параметру.

Ефективною є статистична оцінка, яка при заданому обсязі вибірки має найменшу можливу дисперсію.

Обґрунтованою є статистична оцінка, яка при $n \rightarrow \infty$ прямує за ймовірністю до оцінюваного параметра.

Вибіркова дисперсія D_B є зміщеною і дає занижені значення для генеральної дисперсії D : $M(D_B) = \frac{n-1}{n}D$.

Виправлену вибіркочу дисперсію називають незмщеною дисперсією:

$$S^2 = \frac{n}{n-1} D_B = \frac{\sum n_i (x_i - \bar{x}_i)^2}{n-1}.$$

Точковими оцінками параметрів розподілу генеральної сукупності називають оцінки, які визначаються одним числом.

Інтервальною називають оцінку, яка визначається двома числами — кінцями інтервалу.

Надійністю (надійною ймовірністю) оцінки параметра θ за θ^* називають ймовірність γ , з якою виконується нерівність $|\theta - \theta^*| < \delta$:

$$\gamma = P(|\theta - \theta^*| < \delta).$$

Інтервал $(\theta^* - \delta, \theta^* + \delta)$ називають надійним, якщо він містить у собі невідомий параметр θ із заданою надійністю γ .

Якщо ознака X генеральної сукупності розподілена за нормальним законом з відомим середньоквадратичним відхиленням σ , надійний інтервал для невідомого математичного сподівання a за вибіркочим середнім значенням \bar{x} із заданою надійністю γ має вигляд

$$\left(\bar{x} - t \frac{\sigma}{\sqrt{n}}, \bar{x} + t \frac{\sigma}{\sqrt{n}} \right),$$

де параметр t визначають із рівності $\Phi(t) = \frac{\gamma}{2}$; $\Phi(t)$ — функція Лапласа; точність оцінки $\delta = t \frac{\sigma}{\sqrt{n}}$.

Якщо ознака X генеральної сукупності розподілена за нормальним законом з невідомим середньоквадратичним відхиленням σ , довірчий інтервал для невідомого математичного сподівання a за вибірковим середнім значенням \bar{x} із заданою надійністю γ має вигляд

$$\left(\bar{x} - t_\gamma \frac{s}{\sqrt{n}}, \bar{x} + t_\gamma \frac{s}{\sqrt{n}} \right),$$

де параметр t_γ визначають із рівності $F(t) = \frac{\gamma}{2}$; $F(t)$ – функція розподілу Стьюдента; s – виправлене середньоквадратичне відхилення; точність оцінки $\delta = t_\gamma \frac{s}{\sqrt{n}}$.

7.6.3. Статистична перевірка гіпотез

Основний результат, який потрібно перевірити на підставі заданої вибірки, називають основною (нульовою) гіпотезою і позначають H_0 .

Альтернативною (конкуруючою) називають гіпотезу H_1 , яка суперечить основній.

Похибка першого роду полягає в тому, що буде відхилена правильна гіпотеза H_0 .

Похибка другого роду полягає в тому, що буде прийнята неправильна гіпотеза H_0 .

Імовірність похибки першого роду позначають α і називають рівнем значущості.

Критерій узгодження для перевірки гіпотез

Статистичним критерієм узгодження перевірки гіпотези (критерієм) називають випадкову величину K , розподіл якої відомий і яку використовують для перевірки основної гіпотези.

Критичною областю називають сукупність значень критерію, при яких основну гіпотезу відхиляють.

Областю прийняття гіпотези називають множину значень критерію, при яких гіпотезу приймають.

Критичними точками критерію K називають точки $k_{кр}$, які відокремлюють критичну область від області прийняття гіпотези.

Правосторонньою називають критичну область, що визначається нерівністю $K > k_{кр} > 0$.

Лівосторонньою називають критичну область, що визначається нерівністю $K < k_{кр} < 0$.

Двосторонньою називають критичну область, що визначається нерівностями $K < k_{кр1} < 0$ і $K > k_{кр2} > 0$. Зокрема, якщо критичні точки симетричні відносно початку координат $k_{кр2} = -k_{кр1}$, то двостороння критична область визначається нерівністю $|K| > k_{кр} > 0$.

Щоб знайти односторонню критичну область, потрібно визначити критичну точку $k_{кр}$. Для цього задають достатньо малий рівень значущості α , а потім шукають $k_{кр}$ з рівностей: $P(K > k_{кр}) = \alpha$ для правосторонньої критичної області, $P(K < k_{кр}) = \alpha$ — для лівосторонньої, $P(K < k_{кр1}) + P(K > k_{кр2}) = \alpha$ — для двосторонньої.

Потужністю критерію називають імовірність належності критерію критичній області за умови, що правильною є альтернативна гіпотеза.

Якщо рівень значущості α уже обрано, критичну область доцільно будувати так, щоб потужність критерію була максимальною.

Порядок дій при перевірці статистичних гіпотез

1. Формулюють основну гіпотезу H_0 і визначають альтернативну до неї гіпотезу H_1 .

2. Обирають статистичну характеристику перевірки.

3. Визначають допустиму ймовірність похибки першого роду — рівень значущості α .

4. За певними таблицями знаходять критичні точки та критичну область для обраної статистичної характеристики.

5. Роблять статистичний висновок: якщо обрана характеристика потрапила до критичної області, основну гіпотезу відхиляють. Якщо ні, тобто обрана характеристика потрапила до області прийняття гіпотези, основну гіпотезу приймають.

7.7. Особливості застосування МНК

Якість оцінок параметрів регресійної моделі значною мірою залежить від особливостей статистичних даних і залишків моделі. До таких особливостей належать мультиколінеарність незалежних змінних, автокореляція і гетероскедастичність залишків.

7.7.1. Тестування наявності мультиколінеарності

Єдиного способу визначення мультиколінеарності не існує.

Для тестування мультиколінеарності здебільшого застосовують:

- *F-тест визначення мультиколінеарності* (запропонований Глобером і Фарраром тест має й іншу назву: *побудова допоміжної регресії*);
- характеристичні значення та умовний індекс.

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

F -тест перевіряє гіпотезу

$$H_0 : R_{x_i, x_1, x_2, \dots, x_m}^2 = 0$$

проти гіпотези

$$H_1 : R_{x_i, x_1, x_2, \dots, x_m}^2 > 0,$$

де $R_{x_i, x_1, x_2, \dots, x_m}^2$ — коефіцієнт детермінації в регресії, яка пов'язує фактор x_i з іншими факторами.

Згідно з F -тестом потрібно:

- для кожного коефіцієнта детермінації розрахувати F_i -відношення:

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

де n — кількість спостережень; m — кількість факторів;

- значення F_i порівняти з табличним значенням F -розподілу Фішера $F_{\text{табл}} = F(\alpha, m-1, n-m)$ з $(m-1)$ і $(n-m)$ степенями вільності при заданому рівні значущості α .

Якщо $F_i < F_{\text{табл}}$, приймають гіпотезу H_0 : фактор x_i — не є мультиколінеарним, тобто не залежить від інших факторів.

Якщо $F_i > F_{\text{табл}}$, гіпотезу H_0 відхиляють: x_i — мультиколінеарний фактор.

Другий тест застосовує *характеристичні значення (власні числа матриці спостережень) і умовний індекс R — відношення максимального власного числа матриці до її мінімального власного числа*.

Згідно з цим тестом розраховують не лише умовне число R , а й умовний індекс $CI = \sqrt{R}$. Якщо $100 \leq R \leq 1000$, мультиколінеарність *помірна*, при $R > 1000$ — *висока*. Аналогічно, якщо $10 \leq CI \leq 30$ — мультиколінеарність *помірна*, при $CI > 30$ — *висока*.

Жоден з розглянутих методів тестування не є універсальним. Обидва мають той самий недолік: жоден із них не проводить чіткої

межі між тим, що треба вважати “суттєвою” мультиколінеарністю, яку необхідно враховувати, і тим, коли мультиколінеарністю можна знехтувати.

7.7.2. Алгоритм Феррара–Глобера

Найповніше дослідження мультиколінеарності можна виконати за алгоритмом Феррара–Глобера, який застосовує три види статистичних критеріїв і дає змогу виявити мультиколінеарність:

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

Алгоритм Феррара–Глобера складається із семи кроків.

Крок 1. Нормалізуємо змінні x_1, x_2, \dots, x_m економетричної моделі, обчисливши

$$x_{ij}^* = \frac{(x_{ij} - \bar{x}_j)}{\sqrt{n\sigma_{x_j}^2}}, \text{ або } x_{ij}^* = \frac{(x_{ij} - \bar{x}_j)}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}}$$

де n — кількість спостережень ($i = 1, 2, \dots, n$); m — кількість незалежних змінних ($j = \overline{1, m}$); \bar{x}_j — середня арифметична j -ї незалежної змінної; $\sigma_{x_j}^2$ — дисперсія j -ї незалежної змінної (у цьому разі це може бути вибіркова дисперсія $\sigma_{x_j}^2 = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$).

Крок 2. Будуємо нову матрицю до X^* — матрицю, елементами якої є нормалізовані незалежні змінні x_{ij}^* , і обчислюємо кореляційну матрицю (матрицю моментів нормалізованої системи нормальних рівнянь):

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

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

Однак на основі цієї залежності не можна стверджувати, що отриманий зв'язок є явищем мультиколінеарності.

Зауваження. Якщо діагональні елементи матриці R не дорівнюють одиниці (зв'язок будь-якої незалежної змінної із собою має бути повний), то на діагоналі цієї матриці ми ставимо одиниці, а до решти елементів додаємо різницю між одиницею і значенням діагонального елемента.

Крок 3:

- обчислюємо $|R|$ — визначник кореляційної матриці R ;
- визначаємо експериментальне значення критерію χ^2 :

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

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

Крок 4. Визначаємо матрицю похибок

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

Крок 5:

- обчислюємо F -критерій:

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

де c_{kk} — діагональні елементи матриці C ;

- значення критеріїв F_k порівнюємо з табличним при $(n - m)$ і $(m - 1)$ степенях вільності й рівні значущості α (якщо $F_k > F_{\text{табл}}$, то відповідна k -та незалежна змінна мультиколінеарна з іншими);
- обчислюємо коефіцієнти детермінації для кожної змінної:

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

Крок 6. Знаходимо часткові коефіцієнти кореляції, які характеризують щільність зв'язку між двома змінними за умови, що інші змінні $x_{i1}, x_{i2}, \dots, x_{im}$ не впливають на цей зв'язок (існування парної мультиколінеарності):

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

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

Однак порівнявши конкретні числові значення часткових і парних коефіцієнтів, можна побачити, що перші значно менші за останні. Тому на основі знання парних коефіцієнтів кореляції висновок про мультиколінеарність робити неможливо. Для цього необхідно виконати сьомий крок.

Крок 7:

- розраховуємо t -критерії:

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

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

Висновки

1. Якщо $F_k > F_{\text{табл}}$, то x_k залежить від інших незалежних змінних.
2. Якщо $t_{kj} > t_{\text{табл}}$, то x_k і x_j щільно пов'язані між собою.
3. Аналіз F і t критеріїв дає змогу визначити, яку зі змінних вилучити з моделі.

7.7.3. Тестування наявності гетероскедастичності

Особливості залишків моделі здебільшого можна виявити за допомогою графічного аналізу.

Зокрема, наявність гетероскедастичності можна виявити за графіком залежності залишків u_i від теоретичних значень результативної ознаки \hat{y}_i . Якщо в розташуванні точок на такому графіку є певна закономірність, наприклад подібна до кореляційного поля парної лінійної регресії, можна висувати гіпотезу про наявність гетероскедастичності.

Для множинної регресії такий графічний аналіз є попереднім способом виявлення гетероскедастичності.

Однак візуальної перевірки недостатньо, потрібно підтвердити чи спростувати гіпотезу щодо наявності гетероскедастичності емпірич-

но, тобто за допомогою спеціальних тестів. Найпоширенішими з них є критерій μ , параметричний та непараметричний тести Гольдфельда–Квандта, тест Глейсера, тест рангової кореляції Спірмена та ін.

Розглянемо деякі з них.

Критерій μ

Цей тест, застосований до великих вибірок, дає змогу виявити гетероскедастичність за п'ять кроків.

Крок 1. Початкові дані залежної змінної $y = (y_1, y_2, \dots, y_n)'$ розбиваємо на k груп ($r = 1, \dots, k$) згідно зі зміною рівня величини y .

Крок 2. Для кожної групи спостережень обчислюємо суму квадратів відхилень:

$$S_r = \sum_{i=1}^{n_r} (y_i - \bar{y}_r)^2,$$

де \bar{y}_r — середнє арифметичне y_i для r -ї групи спостережень; n_r — кількість елементів цієї групи.

Крок 3. Обчислюємо суму квадратів відхилень за всією сукупністю спостережень:

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

Крок 4. Обчислюємо параметр β :

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

де n — загальна кількість спостережень; n_r — кількість спостережень r -ї групи; $\prod_{r=1}^k$ — знак добутку елементів, що мають індекс r .

Крок 5. Обчислюємо критерій μ :

$$\mu = -2 \ln \beta.$$

Розподіл величини μ наближається до відомого розподілу χ^2 (хі-квадрат) Пірсона.

Якщо $\mu < \chi_{\text{табл}}^2$ при заданому рівні значущості ($\alpha = 0,05$ чи $\alpha = 0,01$) і $k - 1$ степенях вільності, гіпотезу про наявність гетероскедастичності відхиляємо, тобто маємо гомоскедастичність.

Параметричний тест Гольдфельда–Квандта

Тест передбачає нормальний розподіл і незалежність випадкових величин u_i . Його можна застосовувати в разі невеликої кількості спос-

тережень. Це особливо важливо в ситуаціях, коли важко підібрати значну кількість статистичних даних щодо певних економічних процесів. Метод базується на припущенні, що дисперсія залишків зростає пропорційно до квадратів значень фактора.

Крок 1. Спостереження (вихідні дані) впорядковуємо за зростанням елементів вектора x_i , який може спричинити зміну дисперсії залишків.

Крок 2. Відкидаємо c спостережень, що перебувають усередині векторів вихідних даних, величину c обирають зі співвідношення

$$c \approx \frac{4n}{15}, \text{ причому } n - c > 2k,$$

де n – кількість спостережень (елементів вектора x_i); k – кількість параметрів моделі.

Крок 3. Будуємо дві моделі на основі звичайного МНК за двома сукупностями, що утворилися, і отримуємо відповідно обсяг n_1 і n_2 спостережень за умови, що $n_{1,2} \geq k$.

Крок 4. Обчислюємо суму квадратів залишків S_1 і S_2 для першої і другої моделей:

$$S_1 = \frac{1}{n_1 - m - 1} \sum_{i=1}^{n_1} u_{1i}^2, \quad S_2 = \frac{1}{n_2 - m - 1} \sum_{i=1}^{n_2} u_{2i}^2,$$

де u_1 і u_2 – залишки відповідно за першою та другою моделями.

Крок 5. Розраховуємо критерій $F^* = \frac{S_2}{S_1}$, який у разі виконання гіпотези про гомоскедастичність відповідатиме F -розподілу з $\gamma_1 = n_1 - k$ і $\gamma_2 = n_2 - k$ степенями вільності.

Значення критерію F^* порівнюємо з табличним значенням F -критерію при рівні значущості α і відповідних степенях вільності; якщо $F^* \leq F_{\text{табл}}$, то гетероскедастичність відсутня.

Зауваження. Чим більше F^* , тим більша гетероскедастичність залишків.

Непараметричний тест Гольдфельда–Квандта

Цей тест є візуальним способом вивчення гетероскедастичності. Хоча він не є цілком надійним, щоб робити обґрунтовані висновки, усе ж дуже простий і часто використовується для попереднього припущення про гетероскедастичність множини спостережень.

Графік залежності залишків моделі від упорядкованих спостережень x_{ij} фактора, що, найімовірніше, спричинює зміну дисперсії залишків (діаграма розсіювання), дає змогу висунути припущення щодо сталої дисперсії залишків. Якщо для всіх значень змінної x_{ij} залишки розподіляються приблизно однаково, їх дисперсія є однорідною і гетероскедастичність відсутня. Якщо вона змінюється, гетероскедастичність існує.

Тест Глейсера

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

Аналitiчна форма регресійних функцій може мати вигляд

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

Гіпотезу про відсутність гетероскедастичності залишків приймають на основі значущості коефіцієнтів a_0 і a_1 . Перевага цього методу визначається можливістю розрізнати чисту ($a_0 = 0, a_1 \neq 0$) і змішану ($a_0 \neq 0, a_1 \neq 0$) гетероскедастичність. Залежно від цього коригують оператор оцінювання параметрів моделі.

7.7.4. Трансформування початкової моделі

Розглянемо питання усунення гетероскедастичності через трансформування початкової моделі.

Припустимо, за статистичними даними побудовано початкову регресійну модель

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

і будь-яким тестом встановлено гетероскедастичність.

Щоб позбутися гетероскедастичності, початкову модель змінюють (трансформують) так, щоб похибки мали сталу дисперсію.

Трансформація моделі зводиться до зміни початкової форми моделі. Спосіб виконання перетворення залежить від специфічної форми гетероскедастичності, тобто від форми залежності між дисперсіями залишків і значеннями незалежних змінних:

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

Вважатимемо, що середнє значення залишків дорівнює нулю, а дисперсія змінюється для кожного значення фактора, причому про-

порційно до величини K_i — коефіцієнта пропорційності, який змінюється зі зміною фактора і зумовлює неоднорідність дисперсії, тобто

$$\sigma_{u_i}^2 = \sigma^2 K_i$$

де $\sigma_{u_i}^2$ — дисперсія залишків при конкретному значенні фактора в i -му спостереженні; σ^2 — стала дисперсія залишків за умови гомоскедастичності.

Зазвичай вважають, що величина σ^2 невідома (неспостережувана), а відносно коефіцієнта K_i висувають різні припущення, пов'язані зі структурою гетероскедастичності.

Розглянемо можливі випадки трансформації моделі на прикладі простої лінійної регресії. Нехай

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

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

Для того щоб від заданої моделі перейти до моделі з гомоскедастичними залишками, потрібно поділити всі змінні на $\sqrt{K_i}$. Тоді дисперсія залишків буде сталою величиною.

Така трансформація моделі означає, що відбувається перехід від початкової моделі до моделі, що описує залежність між перетвореними даними:

$$\frac{y_i}{\sqrt{K_i}} = \frac{a_0}{\sqrt{K_i}} + a_1 \frac{x_i}{\sqrt{K_i}} + \frac{u_i}{\sqrt{K_i}}.$$

Початкові дані для цього рівняння мають вигляд

$$y = \begin{pmatrix} \frac{y_1}{\sqrt{K_1}} \\ \frac{y_2}{\sqrt{K_2}} \\ \dots \\ \frac{y_n}{\sqrt{K_n}} \end{pmatrix}, \quad x = \begin{pmatrix} \frac{x_1}{\sqrt{K_1}} \\ \frac{x_2}{\sqrt{K_2}} \\ \dots \\ \frac{x_n}{\sqrt{K_n}} \end{pmatrix}.$$

Оцінювання параметрів нового рівняння з новими перетвореними змінними еквівалентне застосуванню зваженого методу наймен-

ших квадратів, оскільки змінні такої моделі мають вагові множники $\frac{1}{\sqrt{K_i}}$.

Якщо перетворені змінні розглядати у відхиленнях від середніх, то коефіцієнт регресії

$$a_1 = \frac{\sum_{i=1}^n \frac{1}{K_i} x_i y_i}{\sum_{i=1}^n \frac{1}{K_i} x_i^2},$$

а при застосуванні звичайного МНК цей коефіцієнт

$$a_1 = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

У разі застосування зваженого МНК спостереження з меншими значеннями перетворених змінних при обчисленні параметрів регресії мають відносно більшу вагу, ніж початкові змінні. Однак при цьому нові змінні мають інше економічне тлумачення, а їх регресія – інший економічний зміст.

Зауваження. Для моделі множинної регресії

$$y_i = a_0 + a_1 x_{1i} + a_2 x_{2i} + \dots + a_m x_{mi} + u$$

перетворена модель набуває вигляду

$$\frac{y_i}{K_i} = \frac{a_0}{K_i} + a_1 \frac{x_{1i}}{K_i} + a_2 \frac{x_{2i}}{K_i} + \dots + a_m \frac{x_{mi}}{K_i} + \frac{u_i}{K_i}.$$

Таке рівняння не містить вільного члена, але після перепозначення змінних набуває вигляду

$$\frac{y_i}{K_i} = A_0 + a_1 \frac{x_{1i}}{K_i} + a_2 \frac{x_{2i}}{K_i} + \dots + a_m \frac{x_{mi}}{K_i} + \frac{u_i}{K_i}$$

і його параметри можна оцінити за допомогою звичайного МНК.

Якщо припустити, що зміну дисперсії залишків спричинює змінування фактора X_1 , причому залишки пропорційні до значень його спостережень, то перетворена модель матиме вигляд

$$\frac{y_i}{x_1} = a_1 + a_2 \frac{x_2}{x_1} + \dots + a_m \frac{x_m}{x_1} + \epsilon_{1i}.$$

Якщо таку зміну зумовлює інший фактор X_p , модель буде така:

$$\frac{y_i}{x_p} = a_p + a_1 \frac{x_1}{x_p} + \dots + a_m \frac{x_m}{x_p} + \varepsilon_{pi}$$

Перехід до відносних величин суттєво знижує варіацію фактора, а отже, зменшує дисперсію залишків. Таке врахування гетероскедастичності є найпростішим порівняно з іншими методами. Встановлено також, що зважений МНК, застосований до початкової моделі, дає такі самі результати, що й МНК, застосований до трансформованої моделі.

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

Потрібно пам'ятати, що гетероскедастичність може виникати через невраховані фактори (погану специфікацію моделі). У цьому разі включення неврахованих факторів до моделі може усунути проблему змінної дисперсії залишків. Сліпе застосування трансформації (без аналізу причин гетероскедастичності) зробить гомоскедастичною випадкову складову, але оцінки параметрів залишаться неправильними (зміщеними і неефективними) через неврахування важливих факторів.

7.7.5. Оцінювання параметрів багатofакторної регресійної моделі на основі узагальненого МНК (УМНК)

Узагальнену багатofакторну регресійну модель подамо у матричному вигляді:

$$y = XA + u,$$

де y – вектор-стовпець залежної змінної розмірності $(n \times 1)$; X – матриця незалежних змінних розмірності $(n \times (m + 1))$; A – вектор-стовпець невідомих параметрів розмірності $((m + 1) \times 1)$; u – вектор-стовпець випадкових похибок розмірності $(n \times 1)$.

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

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

$$A = (X'S^{-1}X)^{-1}X'S^{-1}Y. \quad (7.1)$$

Вектор A у цьому разі містить незміщену лінійну оцінку параметрів моделі, яка має найменшу дисперсію і матрицю варіацій:

$$\sigma^2(A) = \sigma_u^2 (X' S^{-1} X)^{-1}.$$

Зауваження. Для отримання УМНК-оцінок необхідно знати коваріаційну матрицю S вектора похибок, яка зазвичай на практиці невідома. Тому природно спершу оцінити матрицю S , а потім застосувати її оцінку у формулі (7.1) замість S . Цей підхід є суттю узагальненого методу найменших квадратів.

Узагальнений МНК застосовується до перетворених даних; він дає змогу обчислювати оцінки, що мають найменшу дисперсію.

Визначення матриці S

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

$$S = \begin{pmatrix} \frac{1}{\lambda_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\lambda_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\lambda_n} \end{pmatrix}.$$

Тоді обернена до неї матриця матиме вигляд

$$S^{-1} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

Значимо, що матриця S залежить від специфічної форми гетероскедастичності й може бути розрахована згідно з припущенням про залежність залишків від однієї із незалежних змінних.

У матриці S значення λ_i , $i = 1, 2, \dots, n$, можна обчислити, спираючись на припущення:

1) $M(uu') = \sigma_u^2 x_{ij}$, тобто дисперсія залишків пропорційна до змін факторної змінної x_j ;

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

3) $M(uu') = \sigma_u^2 \{ |u| \}^2$, тобто дисперсія залишків пропорційна до зміни квадрата залишків.

Для першої гіпотези: $\lambda_i = \frac{1}{x_{ij}}$.

Для другої гіпотези: $\lambda_i = \frac{1}{x_{ij}^2}$.

Для третьої гіпотези: $\lambda_i = \{ |u| \}^2$ або $\lambda_i = (a_0 - a_1 x_{ij})^2$, або $\lambda_i = (a_0 - a_1 x_{ij}^{-1})^2$.

Отже, обчислення за узагальненим МНК передбачає виконання таких кроків:

1) оцінюють параметри початкової моделі звичайним МНК;

2) перевіряють залишки моделі на гетероскедастичність і визначають її форму;

3) формують матрицю S згідно з гіпотезою щодо форми гетероскедастичності й обчислюють обернену до неї матрицю S^{-1} ;

4) обчислюють параметри моделі за формулою (7.1).

Застосування узагальненого МНК, як і в разі зваженого МНК, призводить до того, що спостереження з меншими значеннями перетворених змінних набувають відносно більшої ваги, ніж початкові змінні. Крім того, нові змінні мають інший економічний зміст порівняно з початковими змінними.

Коефіцієнт детермінації не може слугувати задовільною мірою якості моделі в разі застосування УМНК (на відміну від класичної моделі). У загальному випадку значення коефіцієнта детермінації навіть може не належати інтервалу $[0, 1]$, а введення або вилучення незалежної змінної (фактора) не обов'язково призведе до його збільшення або зменшення.

Якщо припущення про нормальний закон розподілу випадкової змінної u порушується (або, як часто буває на практиці, ігнорується), оцінки параметрів залишаються найкращими, але визначити їх статистичну значущість (надійність) за допомогою класичних тестів значущості (t , F тощо) неможливо, тому що їх застосування в цьому разі призведе до неправильних статистичних висновків.

7.7.6. Тестування наявності автокореляції

Тестування наявності автокореляції зазвичай проводиться за d -тестом Дарбіна–Уотсона, хоча існують й інші не менш відомі тести: критерій фон Неймана, нециклічний коефіцієнт автокореляції, циклічний коефіцієнт автокореляції.

Критерій Дарбіна–Уотсона

Значення d -статистики обчислюють за формулою

$$DW = d = \frac{\sum_{t=2}^n (u_t - u_{t-1})^2}{\sum_{t=1}^n u_t^2}. \quad (7.2)$$

Доведено, що значення d -статистики Дарбіна–Уотсона перебуває в межах $0 \leq DW \leq 4$ і містить зони невизначеності (рис. 7.1).

За таблицею Дарбіна–Уотсона при заданому рівні значущості α , кількості факторів m і кількості спостережень n знаходимо два значення – DW_1 і DW_2 :

- якщо $0 < DW < DW_1$, наявна додатна автокореляція;
- якщо $DW_1 \leq DW \leq DW_2$ або $4 - DW_2 \leq DW \leq 4 - DW_1$, ми не можемо зробити висновки ані про наявність, ані про відсутність автокореляції (DW потрапляє до зони невизначеності);
- якщо $4 - DW_1 < DW < 4$, маємо від'ємну автокореляцію;
- якщо $DW_2 < DW < 4 - DW_2$, автокореляція відсутня.

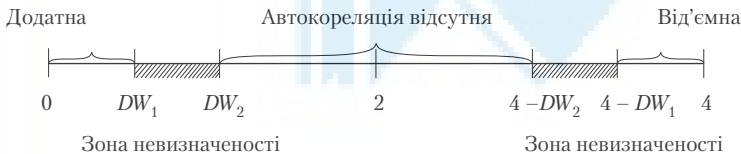


Рис. 7.1. Зони автокореляційного зв'язку за критерієм Дарбіна–Уотсона

Критерій фон Неймана

Формула для розрахунку:

$$Q = \frac{n}{n-1} \frac{\sum_{i=2}^n (u_i - u_{i-1})^2}{\sum_{i=1}^n u_i^2}.$$

Звідси $Q = DW \cdot \frac{n}{n-1}$. Отже, при $n \rightarrow \infty$ $Q = DW$.

Фактичне значення критерію фон Неймана порівнюють з табличним при вибраному рівні значущості α і заданій кількості спостережень: $Q_{\text{табл}} = Q_{(\alpha, n)}$.

Якщо $Q_{\text{факт}} < Q_{\text{табл}}$, то існує додатна автокореляція.

Крім статистик Дарбіна–Уотсона та фон Неймана, для перевірки автокореляції застосовують також нециклічний коефіцієнт автокореляції r^* , який відображає ступінь взаємозв'язку рядів u_1, u_2, \dots, u_{n-1} та u_2, u_3, \dots, u_n і обчислюється за формулою

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

Коефіцієнт r^* може набувати значень з інтервалу $(-1, 1)$. Від'ємні його значення свідчать про від'ємну автокореляцію залишків, а додатні — про додатну. Значення, що лежать в деякій критичній області поблизу нуля, підтверджують нульову гіпотезу про відсутність автокореляції в залишках. Розподіл імовірностей r^* встановити важко, тому на практиці замість r^* обчислюють *циклічний* коефіцієнт автокореляції r^0 . Якщо часовий ряд має циклічний характер, тобто припускають, що після значення u_τ загальний характер зміни членів ряду повторюється, автокореляцію вимірюють за допомогою коефіцієнта r^0 .

У цьому разі автокореляцію визначають між послідовностями, зсунутими на період τ :

$$u_1, u_2, \dots, u_\tau \text{ та } u_{\tau+1}, u_{\tau+2}, \dots, u_{2\tau}$$

Якщо період $\tau = 1$, маємо коефіцієнт циклічної автокореляції першого порядку, який відбиває інтенсивність взаємозв'язку між послідовностями $u_1, u_2, \dots, u_{n-1}, u_n$ та $u_2, u_3, \dots, u_n, u_1$.

Для досить довгих рядів вплив циклічних членів стає незначним, тому ймовірнісний розподіл коефіцієнта r^* наближається до ймовірнісного розподілу коефіцієнта циклічної автокореляції r^0 , який обчислюється за формулою

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

Якщо останній член ряду дорівнює першому, тобто $u_1 = u_n$, то нециклічний коефіцієнт автокореляції дорівнює циклічному. Очевидно, якщо залишки не містять тренда (тенденції), то припущення про рівність $u_1 = u_n$ недалеко від істини і циклічний коефіцієнт автокореляції близький до нециклічного. Крім того, якщо середня залишків дорівнює нулю, можна припустити, що

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

Тоді приблизна формула для обчислення циклічного коефіцієнта автокореляції така:

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

Значущість коефіцієнтів автокореляції перевіряють, як і значущість звичайного коефіцієнта лінійної кореляції. Якщо цей коефіцієнт суттєво відрізняється від нуля, то залишки автокорельовані.

7.7.7. Побудова моделі з автокорельованими залишками

Оцінити параметри моделі з автокорельованими залишками можна на основі чотирьох методів:

- 1) Ейткена (УМНК);
- 2) перетворення вихідної інформації;
- 3) Кочрена–Оркатта;
- 4) Дарбіна.

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

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

Ітераційні методи Кочрена–Оркатта і Дарбіна можна застосовувати для оцінювання параметрів економетричної моделі тоді, коли залишки описуються авторегресійною моделлю вищого порядку:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + v_t$$

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + v_t$$

Метод Ейткена (узагальнений МНК для моделі з автокорельованими залишками)

Оператор оцінювання УМНК має вигляд

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

або

$$\hat{A} = (X' S^{-1} X)^{-1} X' S^{-1} Y, \quad (7.3)$$

де Ω^{-1} — матриця, обернена до дисперсійно-коваріаційної матриці залишків Ω ; S^{-1} — матриця, обернена до матриці $S = \sigma_u^2 \cdot \Omega$.

Оскільки в Ω коваріація залишків $\rho^s \rightarrow 0$ при $s > 2$, то матриця Ω^{-1} матиме вигляд

$$\Omega^{-1} = \frac{1}{1-\rho^2} \cdot \begin{pmatrix} 1 & -\rho & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

На практиці для розрахунку ρ використовують співвідношення

$$\rho \approx r \approx \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2} \quad (7.4)$$

або

$$\rho \approx r' \approx \frac{\sum_{t=2}^n u_t u_{t-1}}{\sum_{t=1}^n u_t^2} \cdot \frac{n-m-1}{n-1}. \quad (7.5)$$

Розрахунки за узагальненим МНК виконують у такому порядку:

- 1) спершу оцінюють параметри моделі за звичайним МНК;

2) перевіряють наявність автокореляції залишків за критерієм Дарбіна–Уотсона (або фон Неймана); якщо залишки автокорельовані, то:

3) обчислюють значення ρ за формулами (7.4) або (7.5);

4) формують матрицю S і обчислюють S^{-1} ;

5) обчислюють параметри моделі за формулою (7.3).

Метод Кочрена–Оркатта

Метод Кочрена–Оркатта є ітераційним методом наближеного пошуку оцінок параметрів моделі з автокорельованими залишками, який базується на МНК і може бути застосований в разі авторегресійних моделей довільного порядку.

Крок 1. Довільно вибирають значення параметра ρ , наприклад $\rho = r_1$.

Підставивши його у формулу

$$\sum_{t=2}^n \varepsilon_t^2 = \sum_{t=2}^n [(y_t - \rho y_{t-1}) - a_0(1 - \rho) - a_1(x_t - \rho x_{t-1})]^2 \quad (7.6)$$

обчислюють $\hat{a}_0^{(1)}$ і $\hat{a}_1^{(1)}$.

Крок 2. Поклавши $\hat{a}_0 = \hat{a}_0^{(1)}$ і $\hat{a}_1 = \hat{a}_1^{(1)}$ та підставивши їх у рівняння (7.6), обчислюють $\rho = r_1$.

Крок 3. Підставивши у рівняння (7.6) значення $\rho = r_2$, знаходять $\hat{a}_0^{(2)}$ і $\hat{a}_1^{(2)}$.

Крок 4. Використовують $\hat{a}_0 = \hat{a}_0^{(2)}$ і $\hat{a}_1 = \hat{a}_1^{(2)}$ для мінімізації суми квадратів залишків (7.6) за невідомим параметром $\rho = r_3$. Процедуру повторюють доти, доки наступні значення параметрів \hat{a}_0 , \hat{a}_1 і ρ не відрізнятимуться менше ніж на задану величину від відповідних попередніх значень.

Зазначимо, що наведений метод завжди:

- забезпечує знаходження глобального оптимуму задачі (7.6);
- має досить високу збіжність, тобто дає змогу знаходити розв'язок за порівняно невелику кількість ітерацій.

ТАБЛИЦЯ СТАНДАРТИЗОВАНОГО НОРМАЛЬНОГО РОЗПОДІЛУ

 $A(z)$

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5703	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9700	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987									

Джерело: *Другерти К.* Введение в эконометрику: Пер. с англ. — М.: ИНФРА-М, 1997. — XIV. — С. 367.

$A(z)$ – це інтеграл щільності ймовірності стандартизованого нормального розподілу від $-\infty$ до z (площа під кривою зліва від z) і є ймовірність того, що величина нормально розподіленої випадкової змінної не перевищує середнього значення більш як на z стандартних відхилень.

Приклад:

$$p(z \leq 2,15) = 0,9842; p(z > 2,15) = 1 - p(z \leq 2,15) = 1 - 0,9842 = 0,0158.$$



МАУП

ТАБЛИЦЯ χ^2 -РОЗПОДІЛУ
(критичні значення χ^2 для рівня значущості α і k степенів вільності)

k	Рівень значущості α						
	0,1%	1%	2,5%	5%	9,5%	97,5%	99%
1	10,8	6,6	5,0	3,8	0,0039	0,001	0,0002
2	13,86	9,2	7,4	6,0	0,103	0,051	0,02
3	16,2	11,3	9,4	7,8	0,352	0,216	0,115
4	18,5	13,3	11,1	9,5	0,711	0,484	0,297
5	20,5	15,1	12,8	11,1	1,15	0,831	0,554
6	22,5	16,8	14,4	12,6	1,64	1,24	0,872
7	24,3	18,5	16,0	14,1	2,17	1,69	1,2
8	26,1	20,1	17,5	15,5	2,73	2,18	1,7
9	27,9	21,7	19,0	16,9	3,33	2,70	2,1
10	29,6	23,2	20,5	18,3	3,94	3,25	2,6
11	31,3	24,7	21,9	19,7	4,57	3,82	3,1
12	32,9	26,2	23,3	21,0	5,23	4,40	3,6
13	34,5	27,7	24,7	22,4	5,89	5,01	4,1
14	36,1	29,1	26,1	23,7	6,57	5,63	4,7
15	37,7	30,6	27,5	25,0	7,26	6,26	5,2
16	39,3	32,0	28,8	26,3	7,96	6,91	5,8
17	40,8	33,4	30,2	27,6	8,67	7,56	6,4
18	42,3	34,8	31,5	28,9	9,39	8,23	7,0
19	43,8	36,2	32,9	30,1	10,1	8,91	7,6
20	45,3	37,6	34,2	31,4	10,9	9,59	8,3
21	46,8	38,9	35,5	32,4	11,6	10,3	8,9
22	48,3	40,3	36,8	33,9	12,3	11,0	9,5
23	49,7	41,6	38,1	35,2	13,1	11,7	10,2
24	51,2	43,0	39,4	36,4	13,8	12,4	10,9
25	52,6	44,3	40,6	37,7	14,6	13,1	11,5
26	54,1	45,6	41,9	38,9	15,4	13,8	12,2
27	55,5	47,0	43,2	40,1	16,2	14,6	12,9
28	56,9	48,3	44,5	41,3	16,9	15,3	13,6
29	58,3	49,6	45,7	42,6	17,7	16,0	14,3
30	59,7	50,9	47,0	43,8	18,5	16,8	15,0

Джерела: *Доугерти К.* Введение в эконометрику: Пер. с англ. — М.: ИНФРА-М, 1997. — XIV. — С. 371.

Толбатов Ю. А. Эконометрика: Підруч. для студ. екон. спец. вищ. навч. закл. — К.: Четверта хвиля, 1997. — С. 301.

ТАБЛИЦЯ t -РОЗПОДІЛУ СТЬЮДЕНТА
(критичні значення $t(\alpha, k)$)

Тести	Рівень значущості α							
	50%	20%	10%	5%	2%	1%	0,2%	0,1%
Дво-сторонній	50%	20%	10%	5%	2%	1%	0,2%	0,1%
Одно-сторонній	25%	10%	5%	2,5%	1%	0,5%	0,1%	0,05%
k								
1	1,000	3,078	6,314	12,706	31,821	63,657	318,31	636,62
2	0,861	1,886	2,920	4,303	6,965	9,925	22,327	31,598
3	0,765	1,638	2,353	3,182	4,541	5,841	10,214	12,924
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893	6,869
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,697	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421	3,690
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,683	1,311	1,699	2,045	2,462	2,756	3,396	3,659
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,681	1,303	1,684	2,021	2,423	2,704	3,307	3,551
60	0,679	1,296	1,671	2,000	2,390	2,660	3,232	3,460
120	0,677	1,289	1,658	1,980	2,358	2,617	3,160	3,373
∞	0,674	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Джерела: *Дугерти К.* Введение в эконометрику: Пер. с англ. — М.: ИНФРА-М, 1997. — XIV. — С. 368.

Лук'яненко І. Г., Краснікова Л. І. Економетрика: Підручник. — К.: Т-во "Знання"; КОО, 1998. — С. 484.

ТАБЛИЦЯ $F(k_{1;k_2}, \alpha)$ РОЗПОДІЛУ ФІШЕРАдля рівня значущості $\alpha = 0,05$ (5%)

k_1	k_2												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	161	200	216	225	230	234	237	239	241	242	243	244	245
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,20	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,73
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,94	5,91	5,89
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,70	4,68	4,66
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00	3,98
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,60	3,57	3,55
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,31	3,28	3,26
9	5,12	4,26	3,68	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,05
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,94	2,91	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,82	2,79	2,76
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,72	2,69	2,66
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,63	2,60	2,58
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,57	2,53	2,51
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,51	2,48	2,45
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,46	2,42	2,40
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,41	2,38	2,35
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,37	2,34	2,31
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,34	2,31	2,28
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,31	2,28	2,25
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,28	2,25	2,22
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,26	2,23	2,20
23	4,48	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20	2,18
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,22	2,18	2,15
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,20	2,16	2,14
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,18	2,15	2,12
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,17	2,13	2,10
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,15	2,12	1,09
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,14	2,10	2,08
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,13	2,09	2,06
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,04	2,00	1,97
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,99	1,95	1,92
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,95	1,92	1,89
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,93	1,89	1,86
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,91	1,88	1,84
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,90	1,86	1,83
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,89	1,85	1,82
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,87	1,83	1,80
140	3,91	3,06	2,67	2,44	2,28	2,16	2,08	2,01	1,95	1,90	1,86	1,82	1,79
160	3,90	3,05	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,85	1,81	1,78
180	3,89	3,05	2,65	2,42	2,26	2,15	2,06	1,99	1,93	1,88	1,84	1,81	1,77
200	3,88	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,84	1,80	1,77
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,79	1,75	1,72

для рівня значущості $\alpha = 0,05$ (5%)

k_2	k_1												
	14	15	16	17	18	19	20	30	40	50	100	200	∞
1	245	246	246	247	247	248	248	250	251	252	253	254	254
2	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5	19,5
3	8,71	8,70	8,69	8,68	8,67	8,67	8,66	8,62	8,60	8,58	8,55	8,54	8,53
4	5,87	5,86	5,84	5,83	5,82	5,81	5,80	5,75	5,72	5,70	5,66	5,65	5,63
5	4,64	4,62	4,60	4,59	4,58	4,57	4,56	4,50	4,46	4,44	4,41	4,39	4,37
6	3,96	3,94	3,92	3,91	3,90	3,88	3,87	3,81	3,77	3,75	3,71	3,69	3,67
7	3,53	3,51	3,49	3,48	3,47	3,46	3,44	3,38	3,34	3,32	3,27	3,25	3,23
8	3,24	3,22	3,20	3,19	3,17	3,16	3,15	3,08	3,04	3,02	2,97	2,95	2,93
9	3,03	3,01	2,99	2,97	2,96	2,95	2,94	2,86	2,83	2,80	2,76	2,73	2,71
10	2,86	2,85	2,83	2,81	2,80	2,79	2,77	2,70	2,66	2,64	2,59	2,56	2,54
11	2,74	2,72	2,70	2,68	2,67	2,66	2,65	2,57	2,53	2,51	2,46	2,43	2,40
12	2,64	2,62	2,60	2,58	2,57	2,56	2,54	2,47	2,43	2,40	2,36	2,32	2,30
13	2,55	2,53	2,51	2,50	2,48	2,47	2,46	2,38	2,34	2,31	2,26	2,23	2,21
14	2,48	2,46	2,44	2,43	2,41	2,40	2,39	2,31	2,27	2,24	2,19	2,16	2,13
15	2,42	2,40	2,38	2,37	2,35	2,34	2,33	2,25	2,20	2,18	2,12	2,10	2,07
16	2,37	2,35	2,33	2,32	2,30	2,29	2,28	2,19	2,15	2,12	2,07	2,04	2,01
17	2,33	2,31	2,29	2,27	2,26	2,24	2,23	2,15	2,10	2,08	2,02	1,99	1,96
18	2,29	2,27	2,25	2,23	2,22	2,20	2,19	2,11	2,10	2,08	1,98	1,95	1,92
19	2,26	2,23	2,21	2,00	2,18	2,17	2,16	2,07	2,03	2,00	1,94	1,91	1,88
20	2,22	2,20	2,18	2,17	2,15	2,14	2,12	2,04	1,99	1,97	1,91	1,88	1,84
21	2,20	2,18	2,16	2,14	2,12	2,11	2,10	2,01	1,96	1,94	1,88	1,84	1,81
22	2,17	2,15	2,13	2,11	2,10	2,08	2,07	1,98	1,94	1,91	1,85	1,82	1,78
23	2,15	2,13	2,11	2,09	2,08	2,06	2,05	1,96	1,91	1,88	1,82	1,79	1,76
24	2,13	2,11	2,09	2,07	2,05	2,04	2,03	1,94	1,89	1,86	1,80	1,77	1,73
25	2,11	2,09	2,07	2,05	2,04	2,02	2,01	1,92	1,87	1,84	1,78	1,50	1,71
26	2,09	2,07	2,05	2,03	2,02	2,00	1,99	1,90	1,85	1,82	1,76	1,73	1,69
27	2,08	2,06	2,04	2,02	2,00	1,99	1,97	1,88	1,84	1,81	1,74	1,71	1,67
28	2,06	2,04	2,02	2,00	1,99	1,97	1,96	1,87	1,82	1,79	1,73	1,69	1,65
29	2,05	2,03	2,01	1,99	1,97	1,96	1,94	1,85	1,81	1,77	1,71	1,57	1,64
30	2,04	2,01	1,99	1,98	1,96	1,95	1,93	1,84	1,79	1,76	1,70	1,66	1,62
40	1,95	1,92	1,90	1,89	1,87	1,85	1,84	1,74	1,69	1,66	1,63	1,60	1,51
50	1,89	1,87	1,85	1,83	1,81	1,80	1,78	1,69	1,63	1,60	1,59	1,55	1,44
60	1,86	1,84	1,82	1,80	1,78	1,76	1,75	1,65	1,59	1,56	1,48	1,44	1,39
70	1,84	1,81	1,79	1,77	1,75	1,74	1,72	1,62	1,57	1,53	1,45	1,40	1,35
80	1,82	1,79	1,77	1,75	1,73	1,72	1,70	1,60	1,54	1,51	1,43	1,38	1,32
90	1,80	1,78	1,76	1,74	1,72	1,70	1,69	1,59	1,53	1,49	1,41	1,36	1,30
100	1,79	1,77	1,75	1,73	1,71	1,69	1,68	1,57	1,52	1,48	1,39	1,34	1,28
120	1,78	1,75	1,73	1,71	1,69	1,67	1,66	1,55	1,50	1,46	1,37	1,32	1,25
140	1,76	1,74	1,72	1,70	1,68	1,66	1,65	1,54	1,48	1,44	1,35	1,30	1,23
160	1,75	1,73	1,71	1,69	1,67	1,65	1,64	1,53	1,47	1,43	1,34	1,28	1,21
180	1,75	1,72	1,70	1,68	1,66	1,64	1,63	1,52	1,46	1,42	1,33	1,27	1,20
200	1,74	1,72	1,69	1,67	1,66	1,64	1,62	1,52	1,46	1,41	1,32	1,26	1,19
∞	1,69	1,67	1,64	1,62	1,60	1,59	1,57	1,46	1,39	1,35	1,24	1,17	1,00

для рівня значущості $\alpha = 0,01$ (1%)

k_2	k_1												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	4052	4999	5403	5625	5764	5859	5928	5981	6023	6056	6083	6106	6126
2	98,5	99,0	99,2	99,3	99,3	99,3	99,4	99,4	99,4	99,4	99,4	99,4	99,4
3	34,1	30,8	29,4	28,7	28,2	27,9	27,7	27,5	27,3	27,2	27,1	27,1	27,0
4	21,2	18,0	16,7	16,0	15,5	15,2	15,0	14,8	14,7	14,5	14,5	14,4	14,3
5	16,3	13,3	12,1	11,4	11,0	10,7	10,5	10,3	10,2	10,1	9,96	9,89	9,82
6	13,7	10,9	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72	7,66
7	12,2	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,54	6,47	6,41
8	11,3	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,73	5,67	5,61
9	10,6	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,18	5,11	5,05
10	10,0	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,77	4,71	4,65
11	9,64	7,20	6,21	5,67	5,31	5,07	4,88	4,74	4,63	4,54	4,46	4,39	4,34
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,22	4,16	4,10
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96	3,90
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80	3,75
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67	3,61
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,62	3,55	3,50
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,46	3,40
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,43	3,37	3,32
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,36	3,30	3,24
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,29	3,23	3,18
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,24	3,17	3,12
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,18	3,12	3,07
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,14	3,07	3,02
24	7,82	5,64	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,09	3,03	2,98
25	7,77	5,57	4,68	4,18	3,85	3,83	3,46	3,32	3,22	3,13	3,06	2,99	2,94
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	3,02	2,96	2,90
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,99	2,93	2,87
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,96	2,90	2,84
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,93	2,87	2,81
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,91	2,84	2,79
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,73	2,66	2,61
50	7,17	5,06	4,20	3,72	3,41	3,19	3,02	2,89	2,78	2,70	2,62	2,56	2,51
60	7,07	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,56	2,50	2,44
70	7,01	4,92	4,07	3,60	3,29	3,07	2,91	2,78	2,67	2,59	2,51	2,45	2,40
80	6,96	4,88	4,04	3,56	3,25	3,04	2,87	2,74	2,64	2,55	2,48	2,42	2,36
90	6,92	4,85	4,01	3,53	3,23	3,01	2,84	2,72	2,61	2,52	2,45	2,39	2,33
100	6,88	4,82	3,98	3,51	3,21	2,99	2,82	2,69	2,59	2,50	2,43	2,37	2,31
120	6,84	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,40	2,34	2,28
140	6,81	4,76	3,92	3,46	3,15	2,93	2,77	2,64	2,54	2,45	2,38	2,31	2,26
160	6,79	4,74	3,90	3,44	3,13	2,92	2,75	2,62	2,52	2,43	2,36	2,30	2,24
180	6,77	4,72	3,89	3,43	3,12	2,90	2,74	2,61	2,51	2,42	2,35	2,28	2,23
200	6,75	4,71	3,88	3,41	3,11	2,89	2,73	2,60	2,50	2,41	2,34	2,27	2,22
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,25	2,18	2,13

для рівня значущості $\alpha = 0,01$ (1%)

k_2	k_1												
	14	15	16	17	18	19	20	30	40	50	100	200	∞
1	6143	6157	6169	6182	6192	6201	6209	6261	6287	6303	6335	6350	6366
2	99,4	99,4	99,4	99,4	99,4	99,4	99,4	99,5	99,5	99,5	99,5	99,5	99,5
3	26,9	26,9	26,8	26,8	26,8	26,7	26,7	26,5	26,4	26,4	26,2	26,2	26,1
4	14,2	14,2	14,2	14,1	14,1	14,0	14,0	13,8	13,7	13,7	13,6	13,5	13,5
5	9,77	9,72	9,68	9,64	9,61	9,58	9,55	9,38	9,29	9,24	9,13	9,08	9,02
6	7,60	7,56	7,52	7,48	7,45	7,42	7,40	7,23	7,14	7,09	6,99	6,93	6,88
7	6,36	6,31	6,28	6,24	6,21	6,18	6,16	5,99	5,91	5,86	5,75	5,70	5,65
8	5,56	5,52	5,48	5,44	5,41	5,38	5,36	5,20	5,12	5,07	4,96	4,91	4,86
9	5,01	4,96	4,92	4,89	4,86	4,83	4,81	4,65	4,57	4,52	4,41	4,36	4,31
10	4,60	4,56	4,52	4,49	4,46	4,43	4,41	4,25	4,17	4,12	4,01	3,96	3,91
11	4,29	4,25	4,21	4,18	4,15	4,12	4,10	3,94	3,86	3,81	3,70	3,65	3,60
12	4,05	4,01	3,97	3,94	3,91	3,88	3,86	3,70	3,62	3,57	3,47	3,41	3,36
13	3,86	3,82	3,78	3,74	3,72	3,69	3,66	3,51	3,42	3,37	3,27	3,22	3,17
14	3,70	3,66	3,62	3,59	3,56	3,53	3,51	3,35	3,27	3,22	3,11	3,06	3,00
15	3,56	3,52	3,49	3,45	3,42	3,40	3,37	3,21	3,13	3,08	2,98	2,92	2,87
16	3,45	3,41	3,37	3,34	3,31	3,28	3,26	3,10	3,02	2,97	2,86	2,81	2,75
17	3,35	3,31	3,27	3,24	3,21	3,19	3,16	3,00	2,92	2,87	2,76	2,71	2,65
18	3,27	3,23	3,19	3,16	3,13	3,10	3,08	2,92	2,84	2,78	2,68	2,62	2,57
19	3,19	3,15	3,12	3,08	3,05	3,03	3,00	2,84	2,76	2,71	2,60	2,55	2,49
20	3,13	3,09	3,05	3,02	2,99	2,96	2,94	2,78	2,69	2,64	2,54	2,48	2,42
21	3,07	3,03	2,99	2,96	2,93	2,90	2,88	2,72	2,64	2,58	2,48	2,42	2,36
22	3,02	2,98	2,94	2,91	2,88	2,85	2,83	2,67	2,58	2,53	2,42	2,36	2,31
23	2,97	2,93	2,89	2,86	2,83	2,80	2,78	2,62	2,54	2,48	2,37	2,32	2,26
24	2,93	2,89	2,85	2,82	2,79	2,76	2,74	2,58	2,49	2,44	2,33	2,27	2,21
25	2,89	2,85	2,81	2,78	2,75	2,72	2,70	2,54	2,45	2,40	2,29	2,23	2,17
26	2,86	2,81	2,73	2,75	2,72	2,69	2,66	2,60	2,42	2,36	2,25	2,19	2,13
27	2,82	2,78	2,75	2,71	2,68	2,66	2,63	2,47	2,38	2,33	2,22	2,16	2,10
28	2,79	2,75	2,72	2,68	2,65	2,63	2,60	2,44	2,35	2,30	2,19	2,13	2,06
29	2,77	2,73	2,69	2,66	2,63	2,60	2,57	2,41	2,33	2,27	2,16	2,10	2,03
30	2,74	2,70	2,66	2,63	2,60	2,57	2,55	2,39	2,30	2,25	2,13	2,07	2,01
40	2,56	2,52	2,48	2,45	2,42	2,39	2,37	2,20	2,11	2,06	2,02	1,96	1,80
50	2,46	2,42	2,38	2,35	2,32	2,29	2,27	2,10	2,01	1,95	1,94	1,87	1,68
60	2,39	2,35	2,31	2,28	2,25	2,22	2,20	2,03	1,94	1,88	1,75	1,68	1,60
70	2,35	2,31	2,27	2,23	2,20	2,18	2,15	1,98	1,89	1,83	1,70	1,62	1,54
80	2,31	2,27	2,23	2,20	2,17	2,14	2,12	1,94	1,85	1,79	1,65	1,58	1,49
90	2,29	2,24	2,21	2,17	2,14	2,11	2,09	1,92	1,82	1,76	1,62	1,55	1,46
100	2,27	2,22	2,19	2,15	2,12	2,08	2,07	1,89	1,80	1,74	1,60	1,52	1,43
120	2,23	2,19	2,15	2,12	2,09	2,06	2,03	1,86	1,76	1,70	1,56	1,48	1,38
140	2,21	2,17	2,13	2,10	2,07	2,04	2,01	1,84	1,74	1,67	1,53	1,45	1,35
160	2,20	2,15	2,11	2,08	2,05	2,02	1,99	1,82	1,72	1,66	1,51	1,42	1,32
180	2,18	2,14	2,10	2,07	2,04	2,01	1,98	1,81	1,71	1,64	1,49	1,41	1,30
200	2,17	2,13	2,09	2,06	2,03	2,00	1,97	1,79	1,69	1,63	1,48	1,39	1,28
∞	2,08	2,04	2,00	1,97	1,93	1,90	1,88	1,70	1,59	1,52	1,36	1,25	1,00

Джерело: *Бронштейн И. Н., Семенов К. А.* Справочник по математике: для инженеров и учащихся ВТУЗов. — М.: Наука, 1981. — С. 82–87.

d-статистика Дарбіна–Уотсона: d_n і d_0 – рівні значущості
у $\alpha = 0,05$ (5 %); n – кількість спостережень

<i>n</i>	Кількість факторів									
	<i>m</i> = 1		<i>m</i> = 2		<i>m</i> = 3		<i>m</i> = 4		<i>m</i> = 5	
	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0
6	0,61	1,40								
7	0,70	1,36	0,47	1,90						
8	0,76	1,33	0,56	1,78	0,37	2,29				
9	0,82	1,32	0,63	1,70	0,46	2,13	0,30	2,59		
10	0,88	1,32	0,70	1,64	0,53	2,02	0,38	2,41	0,24	2,82
11	0,93	1,32	0,76	1,60	0,60	1,93	0,44	2,28	0,32	2,65
12	0,97	1,33	0,81	1,58	0,66	1,86	0,51	2,18	0,38	2,51
13	1,01	1,34	0,86	1,56	0,72	1,82	0,57	2,09	0,45	2,39
14	1,05	1,35	0,91	1,55	0,77	1,78	0,63	2,03	0,51	2,30
15	1,08	1,36	0,95	1,54	0,81	1,75	0,69	1,96	0,56	2,22
16	1,11	1,37	0,98	1,54	0,86	1,73	0,73	1,94	0,62	2,16
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,66	2,10
18	1,16	1,39	1,05	1,54	0,93	1,70	0,82	1,87	0,71	2,06
19	1,18	1,40	1,07	1,54	0,97	1,69	0,96	1,85	0,75	2,02
20	1,20	1,41	1,10	1,54	1,00	1,68	0,89	1,83	0,79	1,99
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90
25	1,29	1,45	1,21	1,55	1,12	1,65	1,04	1,77	0,95	1,89
26	1,30	1,46	1,23	1,55	1,14	1,65	1,06	1,76	0,98	1,87
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,75	1,00	1,86
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81
34	1,39	1,51	1,33	1,58	1,27	1,65	1,21	1,73	1,14	1,81
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80
36	1,41	1,53	1,35	1,59	1,30	1,65	1,24	1,72	1,18	1,80
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,20	1,79
39	1,44	1,54	1,38	1,60	1,33	1,66	1,37	1,72	1,22	1,79
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79
45	1,43	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,37	1,77
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77
70	1,58	1,64	1,55	1,67	1,53	1,70	1,49	1,74	1,46	1,77
75	1,60	1,65	1,57	1,68	1,54	1,71	1,52	1,79	1,49	1,77
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77
85	1,62	1,67	1,60	1,70	1,58	1,72	1,55	1,75	1,53	1,77
90	1,64	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78
95	1,65	1,69	1,62	1,71	1,60	1,73	1,58	1,76	1,56	1,78
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78
180	1,72	1,75	1,71	1,76	1,69	1,77	1,68	1,79	1,67	1,80
200	1,76	1,78	1,75	1,79	1,74	1,90	1,73	1,81	1,72	1,82

n	Кількість факторів									
	m = 6		m = 7		m = 8		m = 9		m = 10	
	d _н	d _о	d _н	d _о	d _н	d _о	d _н	d _о	d _н	d _о
11	0,12	2,89								
12	0,16	2,67	0,11	3,06						
13	0,21	2,49	0,14	2,64	0,09	3,18				
14	0,26	2,35	0,18	2,67	0,12	2,96	0,08	3,29		
15	0,30	2,24	0,23	2,53	0,16	2,82	0,11	3,10	0,07	3,37
16	0,35	2,15	0,27	2,42	0,20	2,68	0,14	2,94	0,09	3,20
17	0,39	2,08	0,31	2,31	0,24	2,57	0,18	2,81	0,13	3,05
18	0,44	2,02	0,37	2,24	0,28	2,47	0,22	2,70	0,16	2,93
19	0,48	1,96	0,40	2,17	0,32	2,36	0,26	2,60	0,20	2,81
20	0,52	1,92	0,44	2,11	0,36	2,31	0,29	2,51	0,23	2,71
21	1,55	1,88	0,47	2,06	0,40	2,24	0,33	2,43	0,27	2,63
22	1,57	1,85	0,51	2,02	0,44	2,19	0,37	2,37	0,30	2,55
23	1,62	1,82	0,55	1,98	0,47	2,14	0,40	2,31	0,34	2,48
24	1,65	1,79	0,58	1,94	0,51	2,10	0,44	2,26	0,38	2,42
25	1,68	1,78	0,61	1,92	0,54	2,06	0,47	2,21	0,41	2,37
26	0,71	1,76	0,64	1,89	0,57	2,03	0,51	2,17	0,44	2,31
27	0,74	1,74	0,67	1,87	0,60	2,00	0,54	2,13	0,47	2,27
28	0,76	1,73	0,70	1,85	0,66	1,97	0,57	2,10	0,50	2,23
29	0,79	1,72	0,72	1,83	0,69	1,95	0,60	2,07	0,53	2,19
30	0,81	1,71	0,76	1,82	0,69	1,93	0,62	2,04	0,56	2,16
31	0,83	1,70	0,77	1,80	0,71	1,91	0,65	2,02	0,59	2,13
32	0,86	1,69	0,79	1,78	0,73	1,89	0,67	2,00	0,62	2,10
33	0,88	1,68	0,82	1,78	0,76	1,87	0,70	1,98	0,64	2,00
34	0,90	1,68	0,84	1,77	0,78	1,86	0,72	1,96	0,67	2,06
35	0,91	1,67	0,86	1,76	0,80	1,85	0,74	1,94	0,69	2,04
36	0,93	1,67	0,88	1,75	0,82	1,84	0,77	1,93	0,71	2,01
37	0,95	1,66	0,90	1,74	0,84	1,83	0,79	1,91	0,73	2,00
38	0,97	1,66	0,91	1,74	0,86	1,82	0,81	1,90	0,75	1,99
39	0,98	1,66	0,93	1,73	0,88	1,81	0,83	1,89	0,77	1,98
40	1,00	1,65	0,95	1,72	0,90	1,80	0,84	1,88	0,79	1,96
45	1,07	1,64	1,02	1,71	0,97	1,77	0,93	1,83	0,86	1,90
50	1,12	1,64	1,06	1,69	1,04	1,75	1,00	1,81	0,96	1,96
55	1,17	1,64	1,13	1,69	1,10	1,73	1,06	1,79	1,02	1,84
60	1,21	1,64	1,18	1,68	1,14	1,73	1,11	1,77	1,07	1,82
65	1,25	1,64	1,22	1,68	1,19	1,72	1,15	1,76	1,12	1,80
70	1,28	1,65	1,25	1,68	1,22	1,72	1,19	1,75	1,16	1,79
75	1,31	1,65	1,28	1,68	1,26	1,72	1,23	1,75	1,20	1,79
80	1,34	1,65	1,31	1,68	1,29	1,71	1,26	1,75	1,23	1,78
85	1,36	1,66	1,34	1,69	1,31	1,71	1,29	1,74	1,26	1,77
90	1,38	1,66	1,36	1,69	1,34	1,71	1,31	1,74	1,29	1,77
95	1,40	1,67	1,38	1,69	1,36	1,72	1,34	1,74	1,31	1,77
100	1,42	1,67	1,40	1,69	1,38	1,72	1,36	1,74	1,34	1,77
150	1,54	1,71	1,53	1,72	1,52	1,74	1,50	1,75	1,49	1,77
200	1,61	1,74	1,60	1,75	1,59	1,76	1,58	1,77	1,57	1,78

n	Кількість факторів									
	m = 11		m = 12		m = 13		m = 14		m = 15	
	d_n	d_o	d_n	d_o	d_n	d_o	d_n	d_o	d_n	d_o
16	0,06	3,45								
17	0,08	3,29	0,05	3,51						
18	0,11	3,15	0,08	3,36	0,05	3,56				
19	0,15	3,02	0,10	3,23	0,07	3,42	0,04	3,60		
20	0,18	2,91	0,13	3,11	0,09	3,30	0,61	3,47	0,04	3,64
21	0,21	2,82	0,16	3,00	0,12	3,19	0,08	3,36	0,06	3,52
22	0,25	2,73	0,19	2,91	0,15	3,08	0,11	3,25	0,08	3,41
23	0,28	2,65	0,23	2,82	0,18	2,99	0,14	3,16	0,10	3,31
24	0,32	2,58	0,26	2,74	0,21	2,91	0,17	3,07	0,13	3,22
25	0,35	2,52	0,29	2,67	0,24	2,83	0,19	2,98	0,15	3,13
26	0,38	2,46	0,32	2,61	0,27	2,76	0,22	2,91	0,18	3,05
27	0,41	2,41	0,36	2,55	0,30	2,69	0,25	2,84	0,21	2,98
28	0,44	2,36	0,39	2,50	0,33	2,64	0,28	2,77	0,24	2,91
29	0,48	2,32	0,42	2,45	0,36	2,58	0,31	2,71	0,27	2,84
30	0,50	2,28	0,45	2,41	0,39	2,53	0,34	2,66	0,29	2,79
31	0,53	2,25	0,48	2,37	0,42	2,49	0,37	2,61	0,32	2,73
32	0,56	2,22	0,50	2,33	0,45	2,45	0,40	2,56	0,35	2,68
33	0,59	2,19	0,53	2,30	0,48	2,41	0,43	2,52	0,38	2,63
34	0,61	2,16	0,56	2,27	0,50	2,37	0,45	2,48	0,40	2,59
35	0,63	2,14	0,58	2,24	0,53	2,34	0,48	2,44	0,43	2,55
36	0,66	2,11	0,61	2,21	0,55	2,31	0,50	2,41	0,46	2,51
37	0,68	2,09	0,63	2,19	0,58	2,28	0,53	2,38	0,48	2,48
38	0,70	2,07	0,65	2,16	0,60	2,26	0,55	2,35	0,50	2,45
39	0,72	2,06	0,67	2,14	0,62	2,23	0,58	2,32	0,53	2,41
40	0,74	2,04	0,67	2,12	0,65	2,21	0,60	2,30	0,55	2,39
45	0,84	1,97	0,79	2,04	0,74	2,12	0,70	2,19	0,66	2,27
50	0,94	1,93	0,87	1,99	0,83	2,05	0,87	2,12	0,75	2,18
55	0,98	1,89	0,94	1,95	0,90	2,00	0,86	2,06	0,83	2,12
60	1,04	1,86	1,00	1,95	0,97	1,96	0,93	2,02	0,89	2,07
65	1,09	1,85	1,05	1,89	1,02	1,93	0,99	1,98	0,95	2,03
70	1,13	1,83	1,10	1,87	1,07	1,91	1,04	1,95	1,01	2,00
75	1,17	1,82	1,14	1,86	1,11	1,89	1,08	1,93	1,05	1,97
80	1,21	1,81	1,18	1,84	1,15	1,88	1,12	1,91	1,09	1,95
85	1,23	1,80	1,21	1,83	1,18	1,87	1,16	1,90	1,13	1,93
90	1,26	1,80	1,24	1,83	1,22	1,86	1,19	1,89	1,17	1,92
95	1,29	1,79	1,27	1,82	1,24	1,85	1,22	1,88	1,20	1,91
100	1,31	1,79	1,29	1,82	1,27	1,84	1,25	1,87	1,23	1,90
150	1,47	1,78	1,46	1,80	1,44	1,81	1,43	1,83	1,41	1,85
200	1,56	1,79	1,55	1,80	1,54	1,81	1,53	1,82	1,52	1,84

n	Кількість факторів									
	m = 16		m = 17		m = 18		m = 19		m = 20	
	d _н	d _о	d _н	d _о	d _н	d _о	d _н	d _о	d _н	d _о
21	0,04	3,67								
22	0,05	3,56	0,03	3,70						
23	0,07	3,46	0,05	3,60	0,03	3,73				
24	0,09	3,36	0,07	3,50	0,04	3,63	0,03	3,75		
25	0,12	3,27	0,09	3,41	0,06	3,54	0,04	3,66	0,03	3,77
26	0,14	3,19	0,11	3,33	0,08	3,45	0,06	3,57	0,04	3,68
27	0,17	3,11	0,13	3,25	0,10	3,37	0,07	3,49	0,05	3,60
28	0,19	3,04	0,16	3,17	0,12	3,30	0,09	3,41	0,07	3,52
29	0,22	2,97	0,18	3,10	0,15	3,22	0,11	3,34	0,09	3,45
30	0,25	2,91	0,21	3,03	0,17	3,15	0,14	3,27	0,11	3,38
31	0,28	2,85	0,23	2,97	0,20	3,09	0,16	3,20	0,13	3,31
32	0,30	2,80	0,26	2,91	0,22	3,03	0,18	3,14	0,15	3,25
33	0,33	2,75	0,29	2,86	0,25	2,97	0,21	3,08	0,17	3,14
34	0,36	2,70	0,31	2,81	0,27	2,92	0,23	3,02	0,20	3,13
35	0,38	2,66	0,34	2,76	0,30	2,87	0,26	2,97	0,22	3,07
36	0,41	2,61	0,36	2,72	0,32	2,82	0,28	2,92	0,24	3,02
37	0,43	2,58	0,39	2,68	0,35	2,77	0,31	2,87	0,27	2,97
38	0,46	2,54	0,41	2,64	0,37	2,73	0,33	2,83	0,29	2,92
39	0,48	2,51	0,44	2,60	0,40	2,69	0,35	2,79	0,32	2,88
40	0,51	2,48	0,46	2,57	0,42	2,66	0,38	2,75	0,34	2,84
45	0,61	2,35	0,57	2,42	0,53	2,50	0,49	2,58	0,45	2,66
50	0,71	2,25	0,67	2,32	0,63	2,39	0,59	2,46	0,55	2,53
55	0,79	2,18	0,75	2,24	0,71	2,30	0,67	2,36	0,64	2,42
60	0,86	2,12	0,82	2,17	0,79	2,23	0,75	2,28	0,72	2,34
65	0,92	2,08	0,89	2,12	0,85	2,17	0,82	2,22	0,79	2,27
70	0,97	2,04	0,94	2,06	0,91	2,13	0,88	2,17	0,85	2,22
75	1,02	2,01	0,99	2,05	0,96	2,09	0,93	2,13	0,91	2,17
80	1,07	1,96	1,04	2,02	1,01	2,06	0,99	2,10	0,96	2,14
85	1,11	1,97	1,08	2,00	1,05	2,03	1,03	2,07	1,01	2,10
90	1,14	1,95	1,12	1,48	1,09	2,01	1,07	2,04	1,04	2,08
95	1,17	1,93	1,15	1,96	1,13	1,99	1,10	2,02	1,08	2,05
100	1,20	1,92	1,18	1,95	1,16	1,98	1,14	2,01	1,11	2,03
150	1,40	1,86	1,39	1,86	1,37	1,90	1,36	1,91	1,35	1,93
200	1,51	1,85	1,50	1,86	1,48	1,87	1,47	1,88	1,46	1,90



d -статистика Дарбіна–Уотсона: d_n і d_0 – рівні значущості
у $\alpha = 0,01$ (5 %); n – кількість спостережень

n	Кількість факторів									
	$m = 1$		$m = 2$		$m = 3$		$m = 4$		$m = 5$	
	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0	d_n	d_0
6	0.39	1.14								
7	0.44	1.04	0.29	1.68						
8	0.50	1.00	0.35	1.49	0.23	2.10				
9	0.55	1.00	0.41	1.39	0.28	1.88	0.18	2.43		
10	0.60	1.00	0.47	1.33	0.34	1.73	0.23	2.19	0.15	2.69
11	0.65	1.01	0.52	1.29	0.40	1.64	0.29	2.03	0.19	2.45
12	0.70	1.02	0.57	1.27	0.45	1.58	0.34	1.91	0.24	2.28
13	0.74	1.04	0.62	1.26	0.50	1.53	0.39	1.83	0.29	2.15
14	0.78	1.05	0.66	1.25	0.55	1.49	0.46	1.76	0.34	2.05
15	0.81	1.07	0.70	1.25	0.59	1.46	0.49	1.70	0.39	1.97
16	0.84	1.09	0.74	1.25	0.63	1.45	0.53	1.66	0.44	1.90
17	0.87	1.11	0.77	1.26	0.67	1.43	0.57	1.63	0.48	1.85
18	0.90	1.12	0.81	1.26	0.71	1.42	0.61	1.60	0.52	1.80
19	0.93	1.13	0.84	1.27	0.74	1.42	0.65	1.58	0.56	1.77
20	0.95	1.15	0.87	1.27	0.77	1.41	0.69	1.57	0.60	1.74
21	0.98	1.16	0.89	1.28	0.80	1.41	0.72	1.55	0.63	1.71
22	1.00	1.17	0.92	1.28	0.83	1.41	0.75	1.54	0.67	1.69
23	1.02	1.19	0.94	1.29	0.86	1.41	0.78	1.53	0.70	1.67
24	1.04	1.20	0.96	1.30	0.88	1.41	0.81	1.53	0.73	1.66
25	1.06	1.21	0.98	1.31	0.91	1.41	0.83	1.52	0.76	1.65
26	1.07	1.22	1.00	1.31	0.93	1.41	0.86	1.52	0.78	1.64
27	1.09	1.23	1.02	1.32	0.95	1.41	0.88	1.52	0.81	1.63
28	1.10	1.24	1.04	1.33	0.97	1.42	0.90	1.51	0.83	1.62
29	1.12	1.25	1.05	1.33	0.99	1.42	0.92	1.51	0.86	1.61
30	1.13	1.26	1.07	1.34	1.01	1.42	0.94	1.51	0.88	1.61
31	1.15	1.27	1.09	1.35	1.02	1.43	0.96	1.51	0.90	1.60
32	1.16	1.28	1.10	1.35	1.04	1.43	0.98	1.51	0.92	1.60
33	1.17	1.29	1.11	1.36	1.06	1.43	1.00	1.51	0.94	1.59
34	1.18	1.30	1.13	1.36	1.07	1.44	1.01	1.51	0.95	1.59
5S	1.20	1.31	1.14	1.37	1.09	1.44	1.03	1.51	0.97	1.59
36	1.21	1.32	1.15	1.38	1.10	1.44	1.04	1.51	0.99	1.59
37	1.22	1.32	1.17	1.38	1.11	1.45	1.06	1.51	1.00	1.59
38	1.23	1.33	1.18	1.39	1.13	1.45	1.07	1.52	1.02	1.59
39	1.24	1.34	1.19	1.39	1.14	1.45	1.09	1.52	1.03	1.58
40	1.21	1.34	1.20	1.40	1.15	1.46	1.10	1.52	1.04	1.59
45	1.29	1.38	1.25	1.42	1.20	1.47	1.16	1.53	1.11	1.59
50	1.32	1.40	1.29	1.45	1.25	1.49	1.21	1.54	1.16	1.59
55	1.36	1.43	1.32	1.47	1.28	1.51	1.25	1.55	1.21	1.59
60	1.38	1.45	1.35	1.48	1.32	1.52	1.28	1.56	1.25	1.60
65	1.41	1.47	1.38	1.50	1.35	1.53	1.32	1.57	1.26	1.60
70	1.43	1.49	1.40	1.52	1.37	1.55	1.34	1.58	1.31	1.61
75	1.45	1.50	1.42	1.53	1.40	1.56	1.37	1.59	1.34	1.62
80	1.47	1.52	1.44	1.54	1.42	1.57	1.39	1.60	1.36	1.62
85	1.48	1.53	1.46	1.55	1.44	1.58	1.41	1.60	1.39	1.63
90	1.50	1.54	1.47	1.65	1.45	1.59	1.43	1.61	1.41	1.64
95	1.51	1.55	1.49	1.57	1.47	1.60	1.45	1.62	1.43	1.64
100	1.52	1.56	1.50	1.58	1.48	1.60	1.46	1.63	1.44	1.65
150	1.61	1.64	1.60	1.65	1.58	1.67	1.57	1.68	1.56	1.69
200	1.66	1.68	1.65	1.69	1.64	1.70	1.63	1.72	1.62	1.73

n	Кількість факторів									
	m = 6		m = 7		m = 8		m = 9		m = 10	
	d _н	d _о	d _н	d _о	d _н	d _о	d _н	d _о	d _н	d _о
11	0,12	2,89								
12	0,16	2,67	0,11	3,06						
13	0,21	2,49	0,14	2,64	0,09	3,18				
14	0,26	2,35	0,18	2,67	0,12	2,96	0,08	3,29		
15	0,30	2,24	0,23	2,53	0,16	2,82	0,11	3,10	0,07	3,37
16	0,35	2,15	0,27	2,42	0,20	2,68	0,14	2,94	0,09	3,20
17	0,39	2,08	0,31	2,31	0,24	2,57	0,18	2,81	0,13	3,05
18	0,44	2,02	0,37	2,24	0,28	2,47	0,22	2,70	0,16	2,93
19	0,48	1,96	0,40	2,17	0,32	2,36	0,26	2,60	0,20	2,81
20	0,52	1,92	0,44	2,11	0,36	2,31	0,29	2,51	0,23	2,71
21	1,55	1,88	0,47	2,06	0,40	2,24	0,33	2,43	0,27	2,63
22	1,57	1,85	0,51	2,02	0,44	2,19	0,37	2,37	0,30	2,55
23	1,62	1,82	0,55	1,98	0,47	2,14	0,40	2,31	0,34	2,48
24	1,65	1,79	0,58	1,94	0,51	2,10	0,44	2,26	0,38	2,42
25	1,68	1,78	0,61	1,92	0,54	2,06	0,47	2,21	0,41	2,37
26	0,71	1,76	0,64	1,89	0,57	2,03	0,51	2,17	0,44	2,31
27	0,74	1,74	0,67	1,87	0,60	2,00	0,54	2,13	0,47	2,27
28	0,76	1,73	0,70	1,85	0,66	1,97	0,57	2,10	0,50	2,23
29	0,79	1,72	0,72	1,83	0,69	1,95	0,60	2,07	0,53	2,19
30	0,81	1,71	0,76	1,82	0,69	1,93	0,62	2,04	0,56	2,16
31	0,83	1,70	0,77	1,80	0,71	1,91	0,65	2,02	0,59	2,13
32	0,86	1,69	0,79	1,78	0,73	1,89	0,67	2,00	0,62	2,10
33	0,88	1,68	0,82	1,78	0,76	1,87	0,70	1,98	0,64	2,00
34	0,90	1,68	0,84	1,77	0,78	1,86	0,72	1,96	0,67	2,06
35	0,91	1,67	0,86	1,76	0,80	1,85	0,74	1,94	0,69	2,04
36	0,93	1,67	0,88	1,75	0,82	1,84	0,77	1,93	0,71	2,01
37	0,95	1,66	0,90	1,74	0,84	1,83	0,79	1,91	0,73	2,00
38	0,97	1,66	0,91	1,74	0,86	1,82	0,81	1,90	0,75	1,99
39	0,98	1,66	0,93	1,73	0,88	1,81	0,83	1,89	0,77	1,98
40	1,00	1,65	0,95	1,72	0,90	1,80	0,84	1,88	0,79	1,96
45	1,07	1,64	1,02	1,71	0,97	1,77	0,93	1,83	0,86	1,90
50	1,12	1,64	1,06	1,69	1,04	1,75	1,00	1,81	0,96	1,96
55	1,17	1,64	1,13	1,69	1,10	1,73	1,06	1,79	1,02	1,84
60	1,21	1,64	1,18	1,68	1,14	1,73	1,11	1,77	1,07	1,82
65	1,25	1,64	1,22	1,68	1,19	1,72	1,15	1,76	1,12	1,80
70	1,28	1,65	1,25	1,68	1,22	1,72	1,19	1,75	1,16	1,79
75	1,31	1,65	1,28	1,68	1,26	1,72	1,23	1,75	1,20	1,79
90	1,34	1,65	1,31	1,68	1,29	1,71	1,26	1,75	1,23	1,78
85	1,36	1,66	1,34	1,69	1,31	1,71	1,29	1,74	1,26	1,77
90	1,38	1,66	1,36	1,69	1,34	1,71	1,31	1,74	1,29	1,77
95	1,40	1,67	1,38	1,69	1,36	1,72	1,34	1,74	1,31	1,77
100	1,42	1,67	1,40	1,69	1,38	1,72	1,36	1,74	1,34	1,77
150	1,54	1,71	1,53	1,72	1,52	1,74	1,50	1,75	1,49	1,77
200	1,61	1,74	1,60	1,75	1,59	1,76	1,58	1,77	1,57	1,78

<i>n</i>	Кількість факторів									
	<i>m</i> = 11		<i>m</i> = 12		<i>m</i> = 13		<i>m</i> = 14		<i>m</i> = 15	
	<i>d</i> _н	<i>d</i> _о	<i>d</i> _н	<i>d</i> _о	<i>d</i> _н	<i>d</i> _о	<i>d</i> _н	<i>d</i> _о	<i>d</i> _н	<i>d</i> _о
16	0,06	3,45								
17	0,08	3,29	0,05	3,51						
18	0,11	3,15	0,08	3,36	0,05	3,56				
19	0,15	3,02	0,10	3,23	0,07	3,42	0,04	3,60		
20	0,18	2,91	0,13	3,11	0,09	3,30	0,61	3,47	0,04	3,54
21	0,21	2,82	0,16	3,00	0,12	3,19	0,08	3,36	0,06	3,52
22	0,25	2,73	0,19	2,91	0,15	3,08	0,11	3,25	0,08	3,41
23	0,28	2,65	0,23	2,82	0,18	2,99	0,14	3,16	0,10	3,31
24	0,32	2,58	0,26	2,74	0,21	2,91	0,17	3,07	0,13	3,22
25	0,35	2,52	0,29	2,67	0,24	2,83	0,19	2,98	0,15	3,13
26	0,38	2,46	0,32	2,61	0,27	2,76	0,22	2,91	0,18	3,05
27	0,41	2,41	0,36	2,55	0,30	2,69	0,25	2,84	0,21	2,98
28	0,44	2,36	0,39	2,50	0,33	2,64	0,28	2,77	0,24	2,91
29	0,48	2,32	0,42	2,45	0,36	2,58	0,31	2,71	0,27	2,84
30	0,50	2,28	0,45	2,41	0,39	2,53	0,34	2,66	0,29	2,79
31	0,53	2,25	0,48	2,37	0,42	2,49	0,37	2,61	0,32	2,73
32	0,56	2,22	0,50	2,33	0,45	2,45	0,40	2,56	0,35	2,68
33	0,59	2,19	0,53	2,30	0,48	2,41	0,43	2,52	0,38	2,63
34	0,61	2,16	0,56	2,27	0,50	2,37	0,45	2,48	0,40	2,59
35	0,63	2,14	0,58	2,24	0,53	2,34	0,48	2,44	0,43	2,55
36	0,66	2,11	0,61	2,21	0,55	2,31	0,50	2,41	0,46	2,51
37	0,68	2,09	0,63	2,19	0,58	2,28	0,53	2,38	0,48	2,48
38	0,70	2,07	0,65	2,16	0,60	2,26	0,55	2,35	0,50	2,45
39	0,72	2,06	0,67	2,14	0,62	2,23	0,58	2,32	0,53	2,41
40	0,74	2,04	0,67	2,12	0,65	2,21	0,60	2,30	0,55	2,39
45	0,84	1,97	0,79	2,04	0,74	2,12	0,70	2,19	0,66	2,27
50	0,94	1,93	0,87	1,99	0,83	2,05	0,87	2,12	0,75	2,18
55	0,98	1,89	0,94	1,95	0,90	2,00	0,86	2,06	0,83	2,12
60	1,04	1,86	1,00	1,95	0,97	1,96	0,93	2,02	0,89	2,07
65	1,09	1,85	1,05	1,89	1,02	1,93	0,99	1,98	0,95	2,03
70	1,13	1,83	1,10	1,87	1,07	1,91	1,04	1,95	1,01	2,00
75	1,17	1,82	1,14	1,86	1,11	1,89	1,08	1,93	1,05	1,97
80	1,21	1,81	1,18	1,84	1,15	1,88	1,12	1,91	1,09	1,95
85	1,23	1,80	1,21	1,83	1,18	1,87	1,16	1,90	1,13	1,93
90	1,26	1,80	1,24	1,83	1,22	1,86	1,19	1,89	1,17	1,92
95	1,29	1,79	1,27	1,82	1,24	1,85	1,22	1,88	1,20	1,91
100	1,31	1,79	1,29	1,82	1,27	1,84	1,25	1,87	1,23	1,90
150	1,47	1,78	1,46	1,80	1,44	1,81	1,43	1,83	1,41	1,85
200	1,56	1,79	1,55	1,80	1,54	1,81	1,53	1,82	1,52	1,84

<i>n</i>	Кількість факторів									
	<i>m</i> = 16		<i>m</i> = 17		<i>m</i> = 18		<i>m</i> = 19		<i>m</i> = 20	
	<i>d</i> _н	<i>d</i> _о	<i>d</i> _н	<i>d</i> _о	<i>d</i> _н	<i>d</i> _о	<i>d</i> _н	<i>d</i> _о	<i>d</i> _н	<i>d</i> _о
21	0,04	3,67								
22	0,05	3,56	0,03	3,70						
23	0,07	3,46	0,05	3,60	0,03	3,73				
24	0,09	3,36	0,07	3,50	0,04	3,63	0,03	3,75		
25	0,12	3,27	0,09	3,41	0,06	3,54	0,04	3,66	0,03	3,77
26	0,14	3,19	0,11	3,33	0,08	3,45	0,06	3,57	0,04	3,68
27	0,17	3,11	0,13	3,25	0,10	3,37	0,07	3,49	0,05	3,60
28	0,19	3,04	0,16	3,17	0,12	3,30	0,93	3,41	0,07	3,52
29	0,22	2,97	0,18	3,10	0,15	3,22	0,11	3,34	0,09	3,45
30	0,25	2,91	0,21	3,03	0,17	3,15	0,14	3,27	0,11	3,38
31	0,28	2,85	0,23	2,97	0,20	3,09	0,16	3,20	0,13	3,31
32	0,30	2,80	0,26	2,91	0,22	3,03	0,18	3,14	0,15	3,25
33	0,33	2,75	0,29	2,86	0,25	2,97	0,21	3,08	0,17	3,14
34	0,36	2,70	0,31	2,81	0,27	2,92	0,23	3,02	0,20	3,13
35	0,38	2,66	0,34	2,76	0,30	2,87	0,26	2,97	0,22	3,07
36	0,41	2,61	0,36	2,72	0,32	2,82	0,28	2,92	0,24	3,02
37	0,43	2,58	0,39	2,68	0,35	2,77	0,31	2,87	0,27	2,97
38	0,46	2,54	0,41	2,64	0,37	2,73	0,33	2,83	0,29	2,92
39	0,48	2,51	0,44	2,60	0,40	2,69	0,35	2,79	0,32	2,88
40	0,51	2,48	0,46	2,57	0,42	2,66	0,38	2,75	0,34	2,84
45	0,61	2,35	0,57	2,42	0,53	2,50	0,49	2,58	0,45	2,66
50	0,71	2,25	0,67	2,32	0,63	2,39	0,59	2,46	0,55	2,53
55	0,79	2,18	0,75	2,24	0,71	2,30	0,67	2,36	0,64	2,42
60	0,86	2,12	0,82	2,17	0,79	2,23	0,75	2,28	0,72	2,34
65	0,92	2,08	0,89	2,12	0,85	2,17	0,82	2,22	0,79	2,27
70	0,97	2,04	0,94	2,06	0,91	2,13	0,88	2,17	0,85	2,22
75	1,02	2,01	0,99	2,05	0,96	2,09	0,93	2,13	0,91	2,17
80	1,07	1,96	1,04	2,02	1,01	2,06	0,99	2,10	0,96	2,14
85	1,11	1,97	1,08	2,00	1,05	2,03	1,03	2,07	1,01	2,10
90	1,14	1,95	1,12	1,48	1,09	2,01	1,07	2,04	1,04	2,08
95	1,17	1,93	1,15	1,96	1,13	1,99	1,10	2,02	1,08	2,05
100	1,20	1,92	1,18	1,95	1,16	1,98	1,14	2,01	1,11	2,03
150	1,40	1,86	1,39	1,86	1,37	1,90	1,36	1,91	1,35	1,93
200	1,51	1,85	1,50	1,86	1,48	1,87	1,47	1,88	1,46	1,90

Джерело: *Толбатов Ю. А.* Економетрика: Підруч. для студ. екон. спец. вищ. навч. закл. – К.: Четверта хвиля, 1997. – С. 293–300.

Приклад. Нехай для регресійного рівняння $\hat{y} = b_0 + b_1x_1 + b_2x_2$ при кількості спостережень $n = 23$ і рівні значущості $\alpha = 0,01$ ми визначили, що $d = 0,35$. Оскільки $k = 1 + m = 1 + 2 = 3$, за таблицями знаходимо $d_n = 0,86$ і $d_0 = 1,41$. Оскільки $0 < d = 0,35 < d_n = 0,86$, автокореляція додатна.

**ТАБЛИЦЯ КРИТИЧНИХ ЗНАЧЕНЬ
ДЛЯ ВІДНОШЕННЯ ФОН НЕЙМАНА**

$$p\left(\frac{\delta^2}{s^2} < k\right) = \int_0^k \omega\left(\frac{\delta^2}{s^2}\right) d\left(\frac{\delta^2}{s^2}\right)$$

<i>k</i>	<i>n</i>								
	4	5	6	7	8	9	10	11	12
0,25				0,00001	0,00001	0,00001	0,00001		0,00001
0,30				0,00007	0,00007	0,00005	0,00004	0,00002	0,00003
0,35			0,00006	0,00027	0,00021	0,00014	0,00009	0,00005	0,00007
0,40			0,00047	0,00065	0,00047	0,00031	0,00019	0,00012	0,00016
0,45			0,00126	0,00126	0,00088	0,00059	0,00038	0,00025	0,00031
0,50		0,00038	0,00246	0,00214	0,00150	0,00103	0,00069	0,00046	0,00055
0,55		0,00223	0,00409	0,00333	0,00237	0,00168	0,00116	0,00080	0,00094
0,60		0,00493	0,00615	0,00486	0,00355	0,00259	0,00185	0,00132	0,00152
0,65		0,00830	0,00865	0,00678	0,00511	0,00382	0,00282	0,00208	0,00235
0,70		0,01225	0,01161	0,00913	0,00710	0,00544	0,00414	0,00313	0,00351
0,75		0,01673	0,01505	0,01197	0,00958	0,00753	0,00587	0,00455	0,00508
0,80	0,00356	0,02171	0,01900	0,01534	0,01263	0,01015	0,00809	0,00642	0,00714
0,85	0,01302	0,02717	0,02348	0,01932	0,01631	0,01338	0,01089	0,00883	0,00980
0,90	0,02257	0,03310	0,02851	0,02403	0,02068	0,01729	0,01436	0,01188	0,01316
0,95	0,03223	0,03949	0,03412	0,02957	0,02579	0,02196	0,01858	0,01565	0,01733
1,00	0,04199	0,04634	0,04035	0,03598	0,03171	0,02745	0,02363	0,02025	0,02241
1,05	0,05186	0,05364	0,04728	0,04325	0,03849	0,03384	0,02959	0,02578	0,02852
1,10	0,06184	0,06140	0,05500	0,05137	0,04618	0,04120	0,03655	0,03232	0,03577
1,15	0,07194	0,06963	0,06361	0,06036	0,05482	0,04957	0,04458	0,03997	0,04425
1,20			0,07323	0,07020	0,06445	0,05901	0,05375	0,04882	0,05407
1,25						0,06956	0,06412	0,05894	0,06531
1,30								0,07040	

<i>k</i>	<i>n</i>						
	15	20	25	30	40	50	60
0,35	0,00001						
0,40	0,00002						
0,45	0,00004						
0,50	0,00009	0,00001					
0,55	0,00018	0,00002					
0,60	0,00033	0,00005	0,00001				
0,65	0,00059	0,00012	0,00002				
0,70	0,00100	0,00024	0,00005	0,00001			
0,75	0,00161	0,00044	0,00011	0,00003			
0,80	0,00250	0,00076	0,00023	0,00007	0,00001		
0,85	0,00375	0,00127	0,00044	0,00015	0,00002		
0,90	0,00547	0,00206	0,00079	0,00030	0,00004	0,00001	
0,95	0,00778	0,00323	0,00135	0,00057	0,00010	0,00002	
1,00	0,01079	0,00489	0,00222	0,00102	0,00022	0,00005	0,00001
1,05	0,01465	0,00720	0,00355	0,00176	0,00044	0,00012	0,00003
1,10	0,01950	0,01033	0,00550	0,00294	0,00085	0,00026	0,00008
1,15	0,02550	0,01448	0,00826	0,00474	0,00158	0,00054	0,00019
1,20	0,03280	0,01986	0,01208	0,00738	0,00280	0,00108	0,00043
1,25	0,04155	0,02670	0,01723	0,01117	0,00476	0,00206	0,00092
1,30	0,05189	0,03524	0,02402	0,01644	0,00780	0,00376	0,00185
1,35	0,06396	0,04571	0,03276	0,02357	0,01235	0,00656	0,00355
1,40	0,07787	0,05834	0,04379	0,03298	0,01892	0,01098	0,00649
1,45		0,07333	0,05743	0,04511	0,02810	0,01769	0,01133
1,50			0,07398	0,06038	0,04055	0,02750	0,01893
1,55				0,07920	0,05696	0,04131	0,03034
1,60					0,07797	0,06006	0,04675
1,65						0,08465	0,06942
1,70							0,09949

Значення *k*, для яких $p\left(\frac{\delta^2}{s^2} < k\right) = 0$

<i>n</i>	4	5	6	7	8	9	10	11	12	15
<i>k</i>	0,7811	0,4775	0,3215	0,2311	0,1740	0,1357	0,1088	0,0891	0,0743	0,0468

<i>n</i>	20	25	30	40	50	60
<i>k</i>	0,0259	0,0164	0,0113	0,0063	0,0040	0,0028

Джерело: *Наконечний С. І., Терещенко Т. О., Романюк Т. П.* Економетрія: Навч. посіб. — К.: КНЕУ, 1997. — С. 345–346.

Функції Excel, що застосовуються в економетричних розрахунках

Математична формула	Функція Excel	Аргументи	Результат
1	2	3	4
$\sum_{i=1}^n x_i$ (аналогічно $\sum_{i=1}^n y_i, \sum_{i=1}^n u_i$)	АВТОСУММ (матем.)	Діапазон комірок, де записано вектор $x = (x_1, x_2, \dots, x_n)$	Число
$\sum_{i=1}^n x_i^2$ (аналогічно $\sum_{i=1}^n y_i^2, \sum_{i=1}^n u_i^2$)	СУММКВ (матем.)	Діапазон комірок, де записано вектор $x = (x_1, x_2, \dots, x_n)$	Число
$\sum_{i=1}^n x_i y_i$	СУММПРОИЗВ (матем.)	Массив1: Діапазон комірок, де записано вектор $x = (x_1, x_2, \dots, x_n)$ Массив2: діапазон комірок, де записано вектор $y = (y_1, y_2, \dots, y_n)$	Число
\bar{x} (аналогічно \bar{y})	СРЗНАЧ (матем.)	Діапазон комірок, де записано вектор $x = (x_1, x_2, \dots, x_n)$	Число
$\overline{x^2}$ (\overline{xy})	СРЗНАЧ (матем.)	Діапазон комірок, де записано вектор $x = (x_1^2, x_2^2, \dots, x_n^2)$ $(xy = (x_1 y_1, x_2 y_2, \dots, x_n y_n))$	Число
$\overline{x^2} - (\bar{x})^2$	ДИСП (статистич.)	Діапазон комірок, де записано вектор $x = (x_1, x_2, \dots, x_n)$	Число
$\overline{xy} - \bar{x} \cdot \bar{y}$	КОВАРИАЦ (статистич.)	Массив1: Діапазон комірок, де записано вектор $x = (x_1, x_2, \dots, x_n)$ Массив2: Діапазон комірок, де записано вектор $y = (y_1, y_2, \dots, y_n)$	Число

1	2	3	4
X^T	ТРАНСП (ссылки и массивы)	Діапазон комірок, де записано матрицю X розмірності $(n \times k)$, $k = m + 1$	Діапазон комірок для запису транспонованої матриці X^T розмірності $(k \times n)$
$X^T X$	МУМНОЖ (матем.)	Массив1: Діапазон комірок, де записано матрицю X^T $(k \times n)$ Массив2: Діапазон комірок, де записано матрицю X $(n \times k)$	Діапазон комірок для запису добутку матриць $X^T X$ розмірності $(k \times k)$
$X^T Y$	МУМНОЖ (матем.)	Массив1: Діапазон комірок, де записано матрицю X^T $(k \times n)$, $k = m + 1$ Массив2: Діапазон комірок, де записано вектор $y = (y_1, y_2, \dots, y_n)$	Діапазон комірок для запису добутку матриць $X^T Y$ розмірності $(k \times 1)$ – вектор-стовпець
$(X^T X)^{-1}$	МОБР (матем.)	Массив1: Діапазон комірок, де записано квадратну матрицю $X^T X$ $(k \times k)$ $k = m + 1$	Діапазон комірок для запису оберненої матриці $(X^T X)^{-1}$ розмірності $(k \times k)$
$ (X^T X)^{-1} $	МОПРЕД (матем.)	Массив1: Діапазон комірок, де записано обернену квадратну матрицю $(X^T X)^{-1}$ $(k \times k)$, $k = m + 1$	Число
$A = (X^T X)^{-1} (X^T Y)$	ЛИНЕЙН (статистич.)	Известные значения Y: Діапазон комірок, де записано вектор $y = (y_1, y_2, \dots, y_n)$; Известные значения X: Діапазон комірок, де записано вектори спостережень незалежних змінних статистика: 0 – вивести лише значення параметрів; 1 – вивести додаткову статистичну інформацію	Діапазон комірок: Якщо статистика = 0 , то рядок значень параметрів $a_m, a_{m-1}, \dots, a_1, a_0$; Якщо статистика = 1 , то матриця $(5 \times (m + 1))$, у першому рядку якої записано параметри $a_m, a_{m-1}, \dots, a_1, a_0$;

1	2	3	4
		<p>Константа: 0 – модель не містить вільного члена a_0; 1 – модель містить a_0</p>	<p>у другому – їхні стандартні похибки $S_{a_m}, S_{a_{m-1}}, \dots, S_{a_0}$.</p> <p>В інших рядках заповнено лише по два перших елементи, в яких виведено відповідно: R^2 і $F_{\text{експ}}$, $n - m - 1$ і σ_u,</p> $S_{\hat{y}}^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ <p>і $\sum_{i=1}^n u_i^2$.</p> <p>В інших комірках немає даних (н/д)</p>
\hat{y}_i – модельні значення, обчислені за лінійною моделлю	ТЕНДЕНЦІЯ (статистич.)	<p>Известные значения Y: Діапазон комірок, де записано вектор $y = (y_1, y_2, \dots, y_n)$; інші запити ігнорувати</p> <p>Известные значения X: Діапазон комірок, де записано вектори спостережень незалежних змінних</p>	<p>Діапазон комірок для запису модельних значень $\hat{Y}^T = (\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n)^T$ (вектор-стовпець ($n \times 1$))</p>
$F_{\text{табл}} = F(\alpha, k_1, k_2)$	ФРАСПОБР (статистич.)	Вероятность: α; Степени свободы: k_1, k_2	Число
$t_{\text{табл}} = t(\alpha/2, k)$	СТЫОДРАСПОБР (статистич.)	Вероятность: $\alpha/2$; Степени свободы: k	Число
$\chi^2_{\text{табл}} = \chi^2(\alpha, k)$	ХИ2ОБР (статистич.)	Вероятность: $\alpha/2$; Степени свободы: k	

У формулах n – кількість спостережень, m – кількість факторів, $m + 1$ – кількість параметрів.

Порядок роботи з функціями EXCEL

1. Виділити місце під результат (комірку чи діапазон комірок).

2. Викликати функцію (натиснути кнопку f_x на панелі інструментів, вибрати категорію і потрібну функцію).

3. Визначити її аргументи (у вікні функції вказати мишкою діапазон комірок чи ввести з клавіатури потрібне значення аргументу).

Якщо результатом функції є число, то результат розміщено у відповідній комірці, якщо — діапазон комірок (вектор чи матриця), то необхідно виконати ще два кроки.

4. У рядку формул натиснути і відпустити ліву клавішу миші (встановити відповідність зв'язків: шрифт аргументу і рамка його значень одного кольору).

5. Одночасно натиснути три клавіші $\text{Ctrl}+\text{Shift}+\text{Enter}$.

МАУП

СПИСОК ВИКОРИСТАНОЇ ТА РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ

1. Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрики: Учебник для вузов. — М.: ЮНИТИ, 1998. — 1022 с.
2. Болч, Хуань К. Дж. Многомерные статистические методы для экономики. — М.: Статистика, 1979. — 318 с.
3. Варюхин А. М., Панкина О. Ю., Яковлева А. В. Эконометрика: Пособие для сдачи экзамена. — М.: Юрайт-Издат., 2005. — 191 с.
4. Венецкий И. Г., Венецкая В. И. Основные математико-статистические понятия и формулы в экономическом анализе: Справочник. — 2-е изд., перераб. и доп. — М.: Статистика, 1979. — 447 с.
5. Вини Р., Холден К. Введение в прикладной эконометрический анализ. — М.: Финансы и статистика, 1981. — 294 с.
6. Вучков И., Бояджијева Л., Солаков Е. Прикладной линейный регрессионный анализ. — М.: Финансы и статистика, 1987. — 239 с.
7. Грубер Й. Эконометрія: Вступ до множинної регресії та економетрії: У 2 т. — К.: Нічлава, 1998. — Т. I. — 384 с.; 1999. — Т. 2. — 308 с.
8. Демиденко Е. З. Линейная и нелинейная регрессии. — М.: Финансы и статистика, 1981. — 302 с.
9. Джонстон Дж. Эконометрические методы. — М.: Статистика, 1980. — 444 с.
10. Доугерти К. Введение в эконометрику: Пер. с англ. — М.: ИНФРА-М, 1997. — XIV. — 402 с.
11. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. — М.: Финансы и статистика, 1986. — Т. 1. — 365 с.; Т. 2. — 379 с.
12. Емельянов А. С. Эконометрия и прогнозирование. — М.: Экономика, 1985. — С. 82–89.
13. Эконометрика: Учебник / И. И. Елисеева, С. В. Курьшева, Г. В. Костышева и др.; Под ред. И. И. Елисеевой. — 2-е изд., перераб. и доп. — М.: Финансы и статистика, 2005. — 576 с.
14. Слейко В. Основи економетрії. — Львів: Марка Лтд, 1995. — 191 с.
15. Иванова В. М. Экономическая теория. Основы бизнеса: Ч. IV: Эконометрика / Ред. совет: А. Д. Смирнов, В. Ф. Максимова и др. — М.: СОМИНТЭК, 1991. — 158 с.
16. Корольов О. А. Эконометрія: Навч. посіб. — 2-ге вид., випр. та скор. — К.: КНТЕУ, 2005. — 416 с.

17. *Корольов О. А., Рязанцева В. В.* Эконометрия: Практикум: Навч. посіб. — К.: КНТЕУ, 2005. — 277 с.
18. *Ланге О.* Введение в эконометрию. — М.: Прогресс, 1964. — 7–60 с.
19. *Лещинський О. Л., Рязанцева В. В., Юнькова О. О.* Эконометрия: Навч. посіб. — К.: МАУП, 2003. — 208 с.
20. *Лук'яненко І. Г., Краснікова Л. І.* Эконометрика: Підручник. — К.: Т-во “Знання”; КОО, 1998. — 494 с.
21. *Магнус Я. Р., Катьшев П. К., Пересецкий А. А.* Эконометрика: Начальный курс. — М.: Дело, 1997. — 248 с.
22. *Маленко Э.* Статистические методы эконометрии. — М.: Статистика, 1975. — 423 с.
23. *Мартышов С.* Методологические проблемы построения и применения эконометрических моделей. — Вильнюс: Макслас, 1979. — 170 с.
24. *Назаренко О. М.* Основы эконометрики: Підручник. — 2-ге вид., переробл. — К.: Центр навч. л-ри, 2005. — 392 с.
25. *Наконечный С. І., Терещенко Т. О., Романюк Т. П.* Эконометрия: Навч. посіб. — К.: КНЕУ, 1997. — 352 с.
26. *Новак Э.* Введение в методы эконометрики: Сб. задач / Под ред. чл.-кор. РАН И. И. Елисейевой. — М.: Финансы и статистика, 2004. — 324 с.
27. *Пономаренко О. І., Перестюк М. О., Бурым В. М.* Основы математичної економіки. — К.: ИНФОРМТЕХНІКА, 1995. — 319 с.
28. *Сборник задач по эконометрике: Учеб. пособие для студ. экон. вузов / Сост.: Е. Ю. Дорохина, Л. Ф. Преснякова, Н. Ф. Тихомиров.* — М.: Экзамен, 2003. — 224 с.
29. *Статистическое моделирование и прогнозирование: Учеб. пособие / Г. М. Гамбаров, Н. М. Журавель, Ю. Г. Королев и др.; Под ред. А. Г. Громберга.* — М.: Финансы и статистика, 1990. — 383 с.
30. *Тинтнер Г.* Введение в эконометрию. — М.: Статистика, 1965. — 361 с.
31. *Толбатов Ю. А.* Эконометрика: Підруч. для студ. экон. спеціальн. вищ. навч. закл. — К.: Четверта хвиля, 1997. — 320 с.
32. *Устойчивые статистические методы оценки данных: Пер. с англ. / Под ред. Р. Л. Лонера, Г. Н. Уилкинсона.* — М.: Машиностроение, 1984. — 232 с.
33. *Фишер Ф.* Проблема идентификации в эконометрии. — М.: Статистика, 1978. — 223 с.
34. *Хейс Д.* Причинный анализ в статистических исследованиях. — М.: Финансы и статистика, 1981. — 225 с.
35. *Шаттелис Т.* Современные эконометрические методы. — М.: Статистика, 1975. — 161 с.
36. *Экономико-математические методы и прикладные модели: Учеб. пособие для вузов / Под ред. В. В. Федосеева.* — М.: ЮНИТИ, 1999. — 391 с.

Зміст

Вступ	3
Розділ 1. ПОРЯДОК ПОБУДОВИ І ДОСЛІДЖЕННЯ РЕГРЕСІЙНОЇ МОДЕЛІ	8
1.1. Основні етапи економетричного дослідження	8
1.2. Постановка задачі та статистична база даних	9
1.3. Специфікація моделі	11
1.4. Оцінювання параметрів за МНК	13
1.5. Основні припущення методу найменших квадратів	14
1.6. Показники якості регресійної моделі	17
1.7. Перевірка надійності регресійного рівняння	20
1.8. Прогнозування за багатofакторною регресійною моделлю	25
Розділ 2. ПАРНА ЛІНІЙНА РЕГРЕСІЙНА МОДЕЛЬ	27
2.1. Вибір форми залежності	27
2.2. Оцінювання параметрів	29
2.3. Дослідження парної регресійної моделі	31
2.4. Прогнозування за лінійною моделлю	35
2.5. Приклад побудови і дослідження моделі парної лінійної регресії	36
Розділ 3. МЕТОДИ ПОБУДОВИ МНОЖИННОЇ РЕГРЕСІЇ	47
3.1. Вибір факторів і форми рівняння	47
3.1.1. Виключення квазінезмінних змінних	48
3.1.2. Вектор і матриця коефіцієнтів кореляції	49
3.1.3. Метод аналізу матриці коефіцієнтів кореляції	50
3.1.4. Метод показників інформаційної місткості	53
3.1.5. Коефіцієнт множинної кореляції	56
3.1.6. Ефект каталізу в економетричній моделі	59
3.2. Процедури побудови моделей	61
3.2.1. Побудова моделі з коінциденцією	62
3.2.2. Процедура виключення <i>a posteriori</i>	64
3.2.3. Процедура покрокової селекції	65
3.3. Вибір аналітичної форми моделі	68
Розділ 4. ПАРАМЕТРИЗАЦІЯ ТА ДОСЛІДЖЕННЯ МНОЖИННОЇ РЕГРЕСІЇ	73
4.1. Постановка задачі. Вибір інструментальних засобів для її розв'язання	73
4.2. Специфікація моделі	75
4.3. Перевірка факторів моделі на мультиколінеарність	75
4.4. Оцінювання параметрів множинної регресійної моделі	81
4.5. Дослідження моделі та перевірка її на адекватність	83

4.6.	Особливі випадки поведінки залишків моделі: перевірка наявності гетероскедастичності та автокореляції	88
4.7.	Дослідження динамічних властивостей факторів.....	97
Розділ 5. ОДНОВИМІРНІ ЧАСОВІ РЯДИ ТА ЇХ МОДЕЛЮВАННЯ		
5.1.	Основні елементи часового ряду.....	105
5.2.	Виявлення структури часового ряду з використанням автокореляції рівнів.....	109
5.3.	Перевірка гіпотези про існування тенденції	115
5.3.1.	Перевірка наявності тенденції середнього рівня.....	116
5.3.2.	Метод Фостера–Стюарта	118
5.3.3.	Кумулятивний критерій	119
5.4.	Моделювання тенденції часового ряду	124
5.5.	Моделювання сезонних і циклічних коливань	128
5.5.1.	Адитивна модель часового ряду	129
5.5.2.	Мультиплікативна модель.....	133
5.6.	Використання бінарних змінних у сезонному аналізі	139
5.7.	Спектральний аналіз часових рядів.....	141
5.8.	Прогнозування на адаптивній моделі Хольта–Брауна.....	149
Розділ 6. СИСТЕМИ ОДНОЧАСНИХ РІВНЯНЬ		
6.1.	Поняття про системи одночасних рівнянь.....	158
6.2.	Структурна та зведена (прогнозна) форми системи рівнянь	159
6.3.	Проблеми оцінювання параметрів систем рівнянь.....	160
6.4.	Ідентифікація (ототожнення) системи рівнянь.....	161
6.5.	Методи оцінювання параметрів систем одночасних рівнянь	163
6.5.1.	Непрямий метод найменших квадратів оцінювання параметрів точно ідентифікованих систем	163
6.5.2.	Двокроковий МНК (2МНК)	168
6.5.3.	Метод інструментальних змінних	169
6.5.4.	Модифікований 2МНК	170
6.5.5.	Трикроковий МНК.....	171
6.6.	Прогноз і загальні довірчі інтервали.....	172
Розділ 7. МАТЕМАТИЧНА БАЗА ЕКОНОМЕТРІЇ.....		
7.1.	Матриці	177
7.1.1.	Основні види матриць.....	177
7.1.2.	Дії над матрицями	179
7.1.3.	Визначник матриці.....	180
7.1.4.	Ранг матриці	183
7.1.5.	Обернена матриця	184
7.1.6.	Системи лінійних алгебраїчних рівнянь	184
7.1.7.	Основні методи розв'язування системи рівнянь	185
7.2.	Вектори.....	187
7.2.1.	Означення вектора. Дії над векторами	187

7.2.2. Скалярний добуток. Базис.....	188
7.2.3. Власні числа. Власні вектори.....	190
7.3. Функція однієї змінної.....	191
7.3.1. Означення, область визначення, область зміни.....	191
7.3.2. Похідна функції.....	192
7.3.3. Аналіз функції за допомогою похідних.....	195
7.4. Функція кількох змінних.....	195
7.4.1. Означення, область визначення.....	195
7.4.2. Частинні похідні, похідна за напрямком, диференціал.....	196
7.4.3. Екстремум функції двох змінних. Найбільше та найменше значення функції двох змінних у замкненій області.....	197
7.5. Теорія ймовірностей.....	197
7.5.1. Означення ймовірності.....	197
7.5.2. Дискретні та неперервні випадкові величини.....	199
7.5.3. Числові характеристики випадкової величини.....	200
7.5.4. Закони розподілу неперервної випадкової величини.....	203
7.6. Математична статистика.....	206
7.6.1. Генеральна і вибірка сукупності. Статистичний розподіл вибірки. Числові характеристики вибірки ...	206
7.6.2. Статистичні оцінки параметрів розподілу.....	209
7.6.3. Статистична перевірка гіпотез.....	210
7.7. Особливості застосування МНК.....	211
7.7.1. Тестування наявності мультиколінеарності.....	211
7.7.2. Алгоритм Феррара–Глобера.....	213
7.7.3. Тестування наявності гетероскедастичності.....	215
7.7.4. Трансформування початкової моделі.....	218
7.7.5. Оцінювання параметрів багатофакторної регресійної моделі на основі узагальненого МНК (УМНК).	221
7.7.6. Тестування наявності автокореляції.....	224
7.7.7. Побудова моделі з автокорельованими залишками....	226
ДОДАТОК 1	229
ДОДАТОК 2	231
ДОДАТОК 3	232
ДОДАТОК 4	233
ДОДАТОК 5	237
ДОДАТОК 6	245
ДОДАТОК 7	247
СПИСОК ВИКОРИСТАНОЇ ТА РЕКОМЕНДОВАНОЇ ЛІТЕРАТУРИ	251

“The Practical Training Course on Econometrics” contains a necessary minimum of data on econometrics. It deals with the major tasks of econometric studies and the methods of their solution. The least-squares method (LSM) was chosen as the basic estimating procedure for the parameters of the regression equations. The manual describes the specifics of parameterization in case of violation of the key conditions of application of the classic LSM, as well as the alternative parameterization techniques intended for models (the principal-component method, LSM generalization and the like). The dynamics of economic processes is illustrated by examples of time-series models. To model economic processes with direct and back couplings, sets of simultaneous equations were used.

The manual includes parts covering higher mathematics, algebra, probability theory, and mathematical statistics essential for mastering the econometrics course.

The book is designed for the students majoring in economic professions.

Навчальне видання

Лещинський Олег Львович
Рязанцева Валентина Василівна
Юнькова Олена Олександрівна
Юртин Іван Іванович

ПРАКТИКУМ З ЕКОНОМЕТРІЇ

Навчальний посібник

Education publication

Leschynsky, Oleg L., Riazantseva, Valentyna V.
Yun'kova, Olena O., Yurtin, Ivan I.

THE PRACTICAL TRAINING COURSE ON ECONOMETRICS

Study guide

Відповідальний редактор *С. Г. Рогузько*

Редактор *Л. В. Логвиненко*

Коректор *Т. К. Валицька*

Комп'ютерне верстання *Т. Г. Замура*

Оформлення обкладинки *О. О. Стеценко*

Підп. до друку 11.04.08. Формат 60×84/16. Папір офсетний. Друк офсетний.

Ум. друк. арк. 14,88. Обл.-вид. арк. 15,75. Наклад 1200 пр.

Міжрегіональна Академія управління персоналом (МАУП)

03039 Київ-39, вул. Фрометівська, 2, МАУП

ДП «Видавничий дім «Персонал»

03039 Київ-39, просп. Червонозоряний, 119, літ. XX

*Свідоцтво про внесення до Державного реєстру
суб'єктів видавничої справи ДК № 3262 від 26.08.2008*