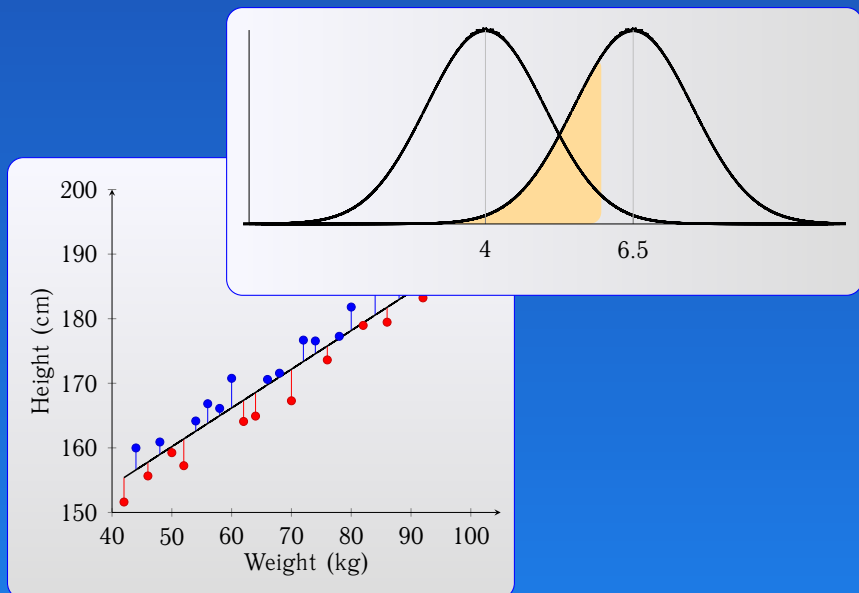


А. Т. Яровий, Є. М. Страхов

ЕКОНОМЕТРІЯ

Навчально-методичний посібник



А. Т. Яровий, Є. М. Страхов

ЕКОНОМЕТРІЯ

**Навчально-методичний посібник
для студентів математичних
та економічних спеціальностей**

**Освіта України
2017
Одеса**

УДК 330.43

ББК 65в6

Я76

Рекомендовано до друку рішенням Вченої ради ІМЕМ ОНУ імені І. І. Мечникова (протокол №3 від 23.12.2016).

Економетрія: навчально-методичний посібник

Автори-укладачі:

Яровий Анатолій Трохимович, кандидат фіз.-мат. наук, доцент;

Страхов Євген Михайлович, кандидат фіз.-мат. наук.

Рецензенти:

Кічмаренко О. Д., кандидат фіз.-мат. наук, доцент, завідувач кафедри оптимального керування і економічної кібернетики ОНУ імені І. І. Мечникова;

Мацкул В. М., кандидат фіз.-мат. наук, доцент, завідувач кафедри математичних методів аналізу економіки ОНЕУ;

Яшкіна О. І., кандидат екон. наук, доцент кафедри маркетингу ОНПУ.

ISBN 978-617-7366-24-8

У посібнику викладено деякі з основних положень побудови і дослідження економетричних моделей у випадку виконання передумов використання методу ІМНК, а також при наявності автокореляції і гетероскедастичності залишків, мультиколінеарності регресорів. Окрім методів побудови і дослідження моделей у посібнику розглядаються приклади їх застосування, що допоможе студентам краще опанувати основні положення дисципліни.

Посібник буде корисний студентам математичних та економічних напрямів, які вивчають дисципліну «Економетрія», а також всім, хто цікавиться застосуванням математичних методів для аналізу економічних даних.

© Яровий А. Т., Страхов Є. М.

Зміст

Вступ	8
I ПОБУДОВА БАГАТОФАКТОРНОЇ РЕГРЕСІЙНОЇ МОДЕЛІ ПРИ ВИКОНАННІ ВСІХ ПЕРЕДУМОВ 1МНК	10
1 Коефіцієнт кореляції	11
1.1 Парний коефіцієнт кореляції	11
1.2 Часткові коефіцієнти кореляції	13
2 Лінійна багаторфакторна модель	17
3 Стандартизовані коефіцієнти регресії. Коефіцієнти еластичності	22
3.1 Стандартизоване рівняння регресії	22
3.2 Коефіцієнти еластичності	23
3.3 Дисперсійно-коваріаційна матриця	24
3.4 Статистичні властивості 1МНК-оцінника $\hat{\beta}$	26
4 Значущість коефіцієнтів регресії та їх інтервали довіри. Прогноз регресанда. Прогнозні інтервали	27
4.1 Тестування значущості коефіцієнтів регресії	27
4.2 Інтервали довіри для коефіцієнтів регресії	29
4.3 Точкові та інтервальні прогнози регресанда	30
5 Показники адекватності регресійної моделі	34
5.1 Коефіцієнт детермінації	34
5.2 Коефіцієнт детермінації, скоригований за Тейлом	35
5.3 Коефіцієнт детермінації, скоригований за Амемією	37
5.4 Частковий коефіцієнт детермінації	37

6 Специфікація моделі. Побудова найкращої моделі	40
6.1 Специфікація моделі	40
6.2 Побудова «найкращої» регресії	42
II ОСОБЛИВІ ВИПАДКИ ПОБУДОВИ БАГАТОФАКТОРНОЇ РЕГРЕСІЙНОЇ МОДЕЛІ	48
7 Автокореляція залишків	49
7.1 Причини виникнення автокореляції залишків	49
7.2 Наслідки автокореляції залишків	50
7.3 Авторегресійний процес першого порядку	50
7.4 Тестування автокореляції залишків	50
7.4.1 Тест Дарбіна — Уотсона	50
7.4.2 Тест серій (Бреуша — Годфрі)	52
7.4.3 Тест на автокореляцію залишків для моделей з лаговою залежною змінною	52
7.5 Методи усунення автокореляції	53
7.5.1 Оцінка моделі, якщо параметр ρ невідомий	53
8 Гетероскедастичність залишків	57
8.1 Причини виникнення гетероскедастичності залишків	57
8.2 Наслідки гетероскедастичності залишків	58
8.3 Методи тестування гетероскедастичності	58
8.3.1 Тест Гольдфельда — Квандта	58
8.3.2 Тест рангової кореляції Спірмена	59
8.3.3 Тест Парка	60
8.3.4 Тест Уайта	61
8.3.5 Тест Глейзера	61
8.4 Методи усунення гетероскедастичності	61
9 Мультиколінеарність регресорів	66
9.1 Причини виникнення мультиколінеарності	66
9.2 Наслідки мультиколінеарності	66
9.3 Тестування мультиколінеарності	67
9.3.1 Рекомендації щодо виявлення мультиколінеарності	67
9.3.2 Метод інфляційних факторів визначення мультиколінеарності	68

9.3.3	Тест Фаррара — Глобера на наявність мультиколінеарності . . .	68
9.4	Методи усунення мультиколінеарності	70
9.4.1	Метод головних компонент	71
III	ДЕЯКІ ВИДИ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ	76
10	Лінійні регресійні моделі зі змінною структурою	77
10.1	Проблема неоднорідних даних	77
10.2	Фіктивні змінні у лінійній моделі регресії	78
10.3	Урахування ефекту взаємодії супровідних змінних	81
10.4	Перевірка регресійної однорідності двох груп спостережень (тест Г. Чоу)	82
11	Динамічні економетричні моделі	84
11.1	Інтерпретація параметрів моделей з розподіленим лагом	85
11.1.1	Приклад	86
11.2	Послідовна оцінка параметрів моделі з розподіленим лагом	87
11.3	Підхід Койка до моделей з розподіленим лагом	88
11.4	Модель часткових пристосувань (коригувань)	90
11.4.1	Приклад. Модель коригування величини дивідендів (модель Лінтнера)	91
11.5	Модель адаптивних очікувань	92
11.6	Оцінка параметрів моделі авторегресії	94
11.7	Лаги Алмона	95
12	Системи економетричних рівнянь	98
12.1	Непрямий метод найменших квадратів (НМНК) оцінки параметрів системи	100
12.2	Проблема ототожнення (ідентифікації) у системі рівнянь	100
12.3	Двокроковий метод найменших квадратів оцінки параметрів системи у структурній формі	103
12.3.1	Оцінка методом 2МНК статистичних характеристик системи рівнянь	105
12.4	Рекурсивні системи	106
	Рекомендована література	108

ДОДАТКИ	110
А Завдання для самостійної роботи	110
Завдання 1	110
Завдання 2	110
Завдання 3	110
Завдання 4	111
Завдання 5	111
Завдання 6	111
Завдання 7	111
Завдання 8	111
Завдання 9	112
Завдання 10	112
Завдання 11	112
Завдання 12	113
Б Варіанти завдань та початкові дані	114
До завдань 1–9	114
Варіанти 1–25	114
Варіанти 26–60	116
До завдання 10	118
До завдання 11	121
До завдання 12	121
В Таблиці математичної статистики	123
В.1 q -квантилі стандартного нормального розподілу	123
В.2 Розподіл Пірсона χ^2	124
В.3 F -розподіл Фішера	125
В.4 Розподіл Ст'юдента	127
В.5 z -перетворення Фішера	128
В.6 Статистика Дарбіна — Уотсона	130

Вступ

В умовах сучасності підготовка фахівців у галузі економіки направлена не лише на прищеплення їм знань щодо природи економічних явищ, а й на дослідження цих явищ з математичної точки зору з метою подальшого прогнозування їх розвитку. Економетрія — ефективний інструмент наукового аналізу та моделювання у професійній діяльності майбутнього економіста, менеджера, інженера.

Економетрія — наука, яка вивчає кількісні та якісні економічні взаємозв'язки і закономірності з використанням математичних і статистичних методів та моделей.

Економетрія як самостійний напрям в економічній теорії виникла на початку ХХ століття на фоні зростання складності соціально-економічних процесів, бурхливого розвитку науки та промисловості. Термін «економетрія» (або «економетрика»; англ. *econometrics*) був введений у 1926 році норвезьким економістом Рагнар Фрішем (Ragnar Frisch) і у буквальному перекладі означає «вимірювання економіки».

Основними завданнями дисципліни є побудова моделей, які виражають досліджувані економічні зв'язки, оцінка їх параметрів та перевірка гіпотез про закономірності зміни економічних показників.

У класичній праці «Економікс» Пола Самуельсона економетрія визначається як інструмент, який «дозволяє економісту просіяти гори даних, щоб виявити прості закономірності»¹.

У цьому посібнику викладено деякі з основних положень побудови і дослідження економетричних моделей у випадку виконання передумов використання методу найменших квадратів (частина I), а також при

¹Ориг. "to sift through mountains of data to extract simple relationships".

наявності автокореляції і гетероскедастичності залишків, мультиколінеарності регресорів (частина II). У частині III розглянуті деякі види економетричних моделей — регресійні моделі зі змінною структурою, динамічні моделі та системи економетричних рівнянь.

Викладення методів побудови і дослідження моделей проілюстровано конкретними прикладами їх застосування, що допоможе студентам краще опанувати основні положення дисципліни.

Посібник містить набір індивідуальних завдань для самостійного виконання студентами. У додатках наведені необхідні таблиці математичної статистики.

Частина I

ПОБУДОВА БАГАТОФАКТОРНОЇ РЕГРЕСІЙНОЇ МОДЕЛІ ПРИ ВИКОНАННІ ВСІХ ПЕРЕДУМОВ 1МНК

- ▣ Коефіцієнт кореляції
- ▣ Лінійна багатофакторна модель
- ▣ Стандартизовані коефіцієнти регресії.
Коефіцієнти еластичності
- ▣ Значущість коефіцієнтів регресії та їх інтервали довіри.
Прогноз регресанда
- ▣ Показники адекватності регресійної моделі
- ▣ Специфікація моделі. Побудова «найкращої» моделі

Коефіцієнт кореляції

При побудові економічних моделей виникає питання: які ж чинники включати до моделі? Зрозуміло, що ті чинники, які найбільш тісно пов'язані з результуючим показником. І тому необхідно мати правила, за допомогою яких можна визначати напрямок і тісноту зв'язку між результуючим показником і чинником. Для цього використовують парний коефіцієнт кореляції і частковий коефіцієнт кореляції.

1.1 Парний коефіцієнт кореляції

Коефіцієнт кореляції є одним з основних показників взаємозалежності випадкових величин. Його також називають парним коефіцієнтом кореляції або лінійним коефіцієнтом кореляції.

Отже, парний коефіцієнт кореляції характеризує тісноту і напрямок зв'язку між двома корелюючими ознаками у випадку наявності між ними лінійної залежності.

Ці дві ознаки повинні вести себе як двовимірна нормальна випадкова величина.

Вибіркова величина парного коефіцієнта кореляції визначається за формулою

$$\hat{r} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{\sqrt{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] \cdot \left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right]}}, \quad (1.1)$$

де n — довжина вибірки, а x і y — ознаки.

Розглянемо властивості парного коефіцієнта кореляції.

- 1) Значення парного коефіцієнта кореляції належать проміжку $[-1, 1]$.

Якщо $\hat{r} > 0$, то зв'язок між x і y **прямий** (обидві зростають або обидві змен-

шуються одночасно), якщо ж $\hat{r} < 0$, то зв'язок **обернений** (якщо одна ознака зростає, то друга — спадає, і навпаки).

- 2) Якщо випадкові величини x і y статистично незалежні, то $\hat{r} = 0$.
- 3) Із того, що $\hat{r} = 0$, випливає, що лінійний зв'язок відсутній.
- 4) Якщо $\hat{r} = 1$, то це означає, що між ознаками x і y існує функціональний лінійний зв'язок, і навпаки: якщо між x і y існує функціональна лінійна залежність, то $\hat{r} = 1$.

Вибірковий коефіцієнт кореляції, що визначений формулою (1.1), можна розраховувати для довільної двовимірної системи, яка має спільний нормальний розподіл випадкових величин.

Коефіцієнт кореляції разом з середніми і дисперсіями випадкових величин складає ті п'ять параметрів, що дають вичерпні відомості про стохастичну залежність величин, тому що однозначно визначають їх двовимірний закон розподілу.

У випадках, коли розподіл відхиляється від нормального, одна із величин не підпорядкована нормальному закону розподілу, величини не є випадковими, коефіцієнт кореляції можна використовувати лише як одну з можливих характеристик ступеня тісноти зв'язку. При цьому, не дивлячись на те, що у загальному випадку поки що не запропоновано характеристики лінійного зв'язку, яка мала б очевидні переваги у порівнянні з r , його інтерпретація досить часто є ненадійною. Якщо апіорі допускається можливість відхилення від лінійного вигляду залежності, то можна побудувати приклади, коли при $r = 0$ існує чисто функціональна залежність між ознаками. Тому про величини, для яких $r = 0$, кажуть, що вони **некорельовані**, і тільки після професійного аналізу можна сказати, чи є вони **незалежними**. І навпаки, з високого ступеня корельованості величин при великих відхиленнях їх розподілів від нормального ще не виходить їх досить тісна залежність.

Тест на перевірку значущості коефіцієнта кореляції

Після оцінки лінійного коефіцієнта кореляції виникає питання: яку величину вибіркового коефіцієнта кореляції можна вважати достатньою для статистично обґрунтованого висновку про наявність кореляційного зв'язку між змінними? Надійність статистичних характеристик, у тому числі і \hat{r} , зменшується зі зменшенням об'єму вибірки, а тому можливі випадки, коли відхилення від нуля отриманої величини вибіркового коефіцієнта кореляції \hat{r} є статистично незначущим, тобто цілком обумовленим випадковими коливаннями вибірки, за якою він розрахований. Відповіді на це питання допомагає знання закону імовірнісного розподілу \hat{r} . У випадку спільної нормальної розподіленості змінних і при достатньо великому об'ємі вибірки n розподіл \hat{r} можна вважати наближено нормальним із середнім, що дорівнює своєму теоретичному значенню r . Однак необхідно враховувати, що при малих значеннях n і r , близьких до ± 1 , це наближення є досить грубим. Крім того, при малих n слід брати до уваги, що величина \hat{r} є зміщеною оцінкою свого теоретичного значення r .

Відносно добрий ступінь наближення нормального розподілу при малих значен-

нях $|r|$ дозволяє отримати простий критерій перевірки гіпотези $r = 0$, тобто гіпотези про відсутність кореляційного зв'язку між змінними. Тест на перевірку значущості коефіцієнта кореляції має такий алгоритм:

Алгоритм 1.1 (перевірка значущості парного коефіцієнта кореляції).

- 1) Формулювання гіпотез: $H_0 : r = 0$, $H_A : r \neq 0$.
- 2) Задаємо рівень значущості α .
- 3) Розраховуємо $t_{CT} = |\hat{r}| \sqrt{\frac{n-2}{1-\hat{r}^2}}$.
- 4) За [таблицею Ст'юдента](#) визначаємо $t_{KP} = t\left(\frac{\alpha}{2}; n-2\right)$.
- 5) Якщо $t_{CT} > t_{KP}$, то гіпотеза відхиляється з ймовірністю $(1-\alpha) \cdot 100\%$, а це означає, що приймається гіпотеза H_A , тобто вибірковий коефіцієнт кореляції значимо відрізняється від нуля з ймовірністю $(1-\alpha) \cdot 100\%$.

1.2 Часткові коефіцієнти кореляції

Часткові коефіцієнти кореляції характеризують ступінь тісноти зв'язку між двома ознаками при умові, що всі інші фіксовані на певному рівні, тобто оцінюється зв'язок між ознаками у «чистому» вигляді.

Існує дві взаємопов'язані обставини, що перешкоджають широкому практичному використанню часткових характеристик статистичного зв'язку у загальному (негаусівському) випадку:

- а) часткові характеристики статистичного зв'язку залежать від рівнів x заважаючих змінних;
- б) для підрахунку вибіркових значень часткових характеристик статистичного зв'язку необхідно мати вибірку спеціальної структури, яка забезпечувала б наявність хоча би декількох спостережень при кожному з заданого ряду фіксованих значень x заважаючих змінних.

Але якщо випадкові змінні підпорядковуються багатовимірному нормальному закону, то згадані неподобства зникають, так як у цьому випадку часткові коефіцієнти кореляції не залежать від рівня заважаючих змінних x .

Має місце така оцінка часткового коефіцієнта кореляції:

$$\hat{r}_{ij|} = \frac{-R_{ij}}{(R_{ii} \cdot R_{jj})^{1/2}}, \quad (1.2)$$

де $\hat{r}_{ij|}$ означає оцінку часткового коефіцієнта кореляції між змінними x_i і x_j при фіксованих значеннях всіх інших змінних, R_{kl} — алгебраїчне доповнення елемента \hat{r}_{kl} у детермінанті кореляційної матриці R змінних.

Можна користуватися і такою формулою

$$\widehat{r}_{ij| \cdot} = \frac{-z_{ij}}{(z_{ii} \cdot z_{jj})^{1/2}}, \quad (1.3)$$

де z_{kl} — елементи матриці Z , що є оберненою до R .

У випадку залежності y від двох чинників x_1 і x_2 часткові коефіцієнти кореляції оцінюються наступним чином:

$$\widehat{r}_{yx_1|x_2} = \frac{\widehat{r}_{yx_1} - \widehat{r}_{yx_2} \cdot \widehat{r}_{x_1x_2}}{\sqrt{(1 - \widehat{r}_{yx_2}^2)(1 - \widehat{r}_{x_1x_2}^2)}}, \quad \widehat{r}_{yx_2|x_1} = \frac{\widehat{r}_{yx_2} - \widehat{r}_{yx_1} \cdot \widehat{r}_{x_1x_2}}{\sqrt{(1 - \widehat{r}_{yx_1}^2)(1 - \widehat{r}_{x_1x_2}^2)}}, \quad (1.4)$$

де \widehat{r}_{yx_1} , \widehat{r}_{yx_2} , $\widehat{r}_{x_1x_2}$ — оцінки парних коефіцієнтів кореляції.

Практика довела, що часткові коефіцієнти кореляції, що визначені формулами (1.2), (1.3) і (1.4), є, як правило, задовільними вимірниками очищеного лінійного зв'язку між x_i і x_j при фіксованих значеннях інших змінних і у випадку, коли розподіл показників y , x_1, \dots, x_p відрізняється від нормального. Ці показники зв'язку можна інтерпретувати як показники тісноти очищеного зв'язку, що усереднені по можливим значенням фіксованих на певному рівні «заважаючих» змінних.

Тест на значущість часткового коефіцієнта кореляції

При дослідженні статистичних властивостей вибіркового часткового коефіцієнта кореляції порядку k (тобто при виключенні опосередкованого впливу k «заважаючих» змінних) необхідно скористатися тим, що він розподілений так, як і парний вибірко-вий коефіцієнт кореляції між тими ж змінними, тільки скрізь необхідно замінити об'єм вибірки n на $n - k$.

Отже, t_{CT} розраховується за формулою

$$t_{CT} = |\widehat{r}_{yx| \cdot}| \sqrt{\frac{n - k - 2}{1 - \widehat{r}_{yx| \cdot}^2}},$$

$$\text{а } t_{KP} = t\left(\frac{\alpha}{2}; n - k - 2\right).$$

Якщо $t_{CT} > t_{KP}$, то з імовірністю $(1 - \alpha) \cdot 100\%$ частковий коефіцієнт кореляції значимо відрізняється від нуля.

Приклад 1.1. Розглянемо показники виробничо-господарської діяльності 25 підприємств машинобудування: y — продуктивність праці, x_2 — трудомісткість одиниці продукції, x_3 — середньорічний фонд заробітної платні підприємства, x_4 — фондо-озброєність праці, x_5 — невиробничі витрати. Вони мають такий вигляд:

Номер підприємства	y	x_2	x_3	x_4	x_5
1	9.26	0.23	47750	6.4	17.72
2	9.38	0.24	50391	7.8	18.39
3	12.11	0.19	43149	9.76	26.46
4	10.81	0.17	41089	7.9	22.37
5	9.35	0.23	14257	5.35	28.13
6	9.87	0.43	22661	9.9	17.55
7	8.17	0.31	52509	4.5	21.92
8	9.12	0.26	14903	4.88	19.52
9	5.88	0.49	25587	3.46	23.99
10	6.3	0.36	16821	3.6	21.76
11	6.22	0.37	19459	3.56	25.68
12	5.49	0.43	12973	5.65	18.13
13	6.5	0.35	50907	4.28	25.74
14	6.61	0.38	6920	8.85	21.21
15	4.32	0.42	5736	8.52	22.97
16	7.37	0.3	26705	7.19	16.38
17	7.02	0.32	20068	4.82	13.21
18	8.25	0.25	11487	5.46	14.48
19	8.15	0.31	32029	6.2	13.38
20	8.72	0.26	18946	4.25	13.69
21	6.64	0.37	28025	5.38	16.66
22	8.1	0.29	20968	5.88	15.06
23	5.52	0.34	11049	9.27	20.09
24	9.37	0.23	45893	4.36	15.98
25	13.17	0.17	99400	10.31	18.27

Необхідно розрахувати парні і часткові коефіцієнти кореляції між y та x_i , а також перевірити їх на значущість.

Скориставшись формулою (1.1), отримаємо:

$$r_{yx_2} = -0.801, \quad r_{yx_3} = 0.664, \quad r_{yx_4} = 0.400, \quad r_{yx_5} = -0.053.$$

Розрахуємо t -статистики. Маємо

$$t_{CT_{yx_2}} = 6.412, \quad t_{CT_{yx_3}} = 4.262, \quad t_{CT_{yx_4}} = 2.091, \quad t_{CT_{yx_5}} = 0.253.$$

Покладемо $\alpha = 0.2$, тоді $t_{кр} = t\left(\frac{\alpha}{2}; n - 2\right) = t(0.1; 23) = 1.714$.

Остаточно маємо, що всі парні коефіцієнти кореляції, крім останнього, значимо відрізняються від нуля з імовірністю 80%.

Тепер розрахуємо часткові коефіцієнти кореляції. Маємо

$$r_{yx_2|} = -0.718, \quad r_{yx_3|} = 0.467, \quad r_{yx_4|} = 0.373, \quad r_{yx_5|} = -0.0004.$$

Їх t -статистики такі:

$$t_{СТyx_2|} = 4.837, \quad t_{СТyx_3|} = 2.479, \quad t_{СТyx_4|} = 1.884, \quad t_{СТyx_5|} = 0.002.$$

Далі, $t_{КР} = t\left(\frac{\alpha}{2}; n - k - 2\right) = t(0.1; 20) = 1.725$.

Так як $t_{СТyx_i|} (i = 2, 3, 4) > t_{КР}$, то перші три часткові коефіцієнти кореляції значимо відрізняються від нуля з ймовірністю 80%, а останній — ні.

Зазначимо, що часткові коефіцієнти кореляції трохи менші за абсолютною величиною, ніж парні. А це означає, що існує невеликий зв'язок між x_i .

Дивлячись на величини парних і часткових коефіцієнтів кореляції, можна зробити висновок, що до економічної моделі, що показує зв'язок між y і x_i , недоцільно включати чинник x_5 . Найбільш впливовим чинником на продуктивність праці є трудомісткість одиниці продукції, потім — середньорічний фонд заробітної платні підприємства і потім фондоозброєність праці. Невиробничі витрати майже не впливають на продуктивність праці.

Лінійна багатofакторна модель

Узагальнену багатofакторну лінійну регресійну модель будемо записувати у вигляді

$$y = b_1x_1 + b_2x_2 + \dots + b_Nx_N + \varepsilon = \sum_{i=1}^N b_ix_i + \varepsilon, \quad (2.1)$$

де y — залежна змінна (регресанд), x_1, x_2, \dots, x_N — незалежні змінні (регресори), b_1, b_2, \dots, b_N — параметри моделі, які необхідно оцінити, ε — неспостережувана випадкова величина.

Зазначимо, що узагальнена регресійна модель — це модель, що дійсна для всієї генеральної сукупності. Так як випадкова величина ε — неспостережувана, то можна тільки робити припущення відповідно до закону її розподілу.

На відміну від узагальненої регресійної моделі, вибіркова модель будується для певної вибірки. Невідомі параметри вибіркової моделі є випадковими величинами, математичне очікування яких дорівнює параметрам узагальненої моделі, а випадкові величини можна оцінити, виходячи з вибірових даних.

Будемо вважати, що вибіркова лінійна багатofакторна модель має вигляд

$$y = \sum_{i=1}^N \beta_i x_i + u, \quad (2.2)$$

де y — задана залежна змінна, x_i ($i = \overline{1, N}$) — задані незалежні змінні, u — випадкова величина (помилка).

Будемо вважати модель лінійною за параметрами β_i .

Модель (2.2) можна записати у вигляді

$$y_t = \beta_1x_{1t} + \beta_2x_{2t} + \dots + \beta_Nx_{Nt} + u_t, \quad t = \overline{1, T}.$$

Зазначимо, що $x_{1t} \equiv 1$, $t = \overline{1, T}$, T — довжина вибірки (кількість спостережень).
 Модель (2.2) можна записати і у матричному вигляді

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U},$$

де

$$\begin{aligned} \mathbf{Y} &= (y_1, \dots, y_T)^\top — \text{вектор-стовпець значень } y_t; \\ \boldsymbol{\beta} &= (\beta_1, \dots, \beta_N)^\top — \text{вектор-стовпець параметрів}; \\ \mathbf{U} &= (u_1, \dots, u_T)^\top — \text{вектор-стовпець випадкової змінної } u; \\ \mathbf{X} &= \begin{pmatrix} 1 & x_{21} & \dots & x_{N1} \\ 1 & x_{22} & \dots & x_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{2T} & \dots & x_{NT} \end{pmatrix} — \text{матриця значень змінних } x_1, \dots, x_N. \end{aligned}$$

Через $\mathbf{D} = [\mathbf{Y}|\mathbf{X}]$ позначимо матрицю початкових даних, де Y і X визначені вище.

Основні припущення у багатофакторному регресійному аналізі

У регресійному аналізі елементи моделі (2.2) повинні задовольняти наступним умовам:

У1 Відсутність систематичних помилок спостережень y_t , тобто математичне очікування u дорівнює нулеві: $M(u) = 0$. Інколи випадковий член буде додатним, інколи від'ємним, але він не повинен мати систематичного зміщення у жодному з двох положень.

Фактично, якщо рівняння регресії включає постійний член, можна вважати, що ця умова виконується автоматично, так як роль константи полягає у визначенні довільної систематичної тенденції в Y , яку не враховують пояснюючі змінні, що включені до моделі.

У2 Дисперсійно-коваріаційна матриця помилок u має вигляд: $\boldsymbol{\Sigma}_u = \sigma_u^2 \mathbf{I}$, де σ_u^2 — дисперсія залишків, а \mathbf{I} — одинична матриця. Ця умова стверджує, що помилки мають постійну дисперсію (гомоскедастичні) і вільні від кореляції, тобто

$$\sigma_{u_i u_j} = \begin{cases} \sigma_u^2, & i = j, \\ 0, & i \neq j, \quad i, j = \overline{1, T}. \end{cases}$$

У3 Залишки u нормально розподілені: $\mathbf{U} \sim N(0, \sigma_u^2 \mathbf{I})$.

Якщо випадковий член u нормально розподілений, то таким же чином будуть розподілені і коефіцієнти регресії, а це дасть нам можливість робити тести і розраховувати інтервали довіри.

У4 Регресори вимірюються без помилок і утворюють лінійно незалежні вектори, тобто $\text{rang } \mathbf{X} = N$.

У5 Змінні x_i ($i = \overline{1, N}$) не корелюють з помилками u .

Оцінка параметрів багатофакторної і парної регресії

Якщо виконуються згадані припущення, то для оцінки невідомих параметрів β_i ($i = \overline{1, N}$) можна застосувати **метод найменших квадратів (МНК)**. Це означає, що β будемо знаходити з умови

$$\min_{\beta} \sum_{i=1}^T u_i^2,$$

тобто

$$\hat{\beta} = \arg \min_{\beta} [(\mathbf{Y} - \mathbf{X}\beta)^\top (\mathbf{Y} - \mathbf{X}\beta)] \quad (2.3)$$

($\hat{\beta}$ означає оцінку параметра β методом МНК). З (2.3) отримуємо

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}. \quad (2.4)$$

Підкладаючи β_i ($i = \overline{1, N}$) у регресійне рівняння (2.2), отримуємо оцінену за допомогою МНК емпіричну регресійну функцію

$$\hat{y} = \sum_{i=1}^N \hat{\beta}_i x_i, \quad \text{або} \quad \hat{\mathbf{Y}} = \mathbf{X}\hat{\beta}. \quad (2.5)$$

Емпіричний коефіцієнт $\hat{\beta}_i$ має таку **інтерпретацію**: збільшення величини i -го регресора (незалежної змінної x_i) на одиницю при інших рівних умовах викликає зміну величини y на $\hat{\beta}_i$ одиниць.

Якщо ж регресія парна: $y = \beta_1 x_1 + \beta_2 x_2 + u$, то для оцінки параметрів β_1 і β_2 можна користуватися формулами:

$$\hat{\beta}_2 = \frac{\sum_{t=1}^T x_{2t} y_t - \frac{1}{T} \sum_{t=1}^T x_{2t} \sum_{t=1}^T y_t}{\sum_{t=1}^T x_{2t}^2 - \frac{1}{T} \left(\sum_{t=1}^T x_{2t} \right)^2}; \quad \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \bar{x}_2, \quad \text{де} \quad \bar{y} = \frac{\sum_{t=1}^T y_t}{T}, \quad \bar{x}_2 = \frac{\sum_{t=1}^T x_{2t}}{T}.$$

Оцінка якості моделі

Дисперсію залишків оцінюємо таким чином:

$$\hat{\sigma}_u^2 = \frac{\sum_{t=1}^T \hat{u}_t^2}{T - N}, \quad \hat{u}_t = y_t - \hat{y}_t, \quad t = \overline{1, T}.$$

Для порівняння моделей використовується **середній квадрат модельної похибки**

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^T \hat{u}_t^2.$$

Якщо порівнюються декілька моделей, то найкращою вважається модель з найменшими значеннями $\hat{\sigma}_u^2$ і MSE. За $\hat{\sigma}_u^2$ порівнюються моделі з однаковою кількістю регресорів.

Для оцінки однієї моделі використовується **коефіцієнт апроксимації** або **середня відносна величина модельної помилки**

$$\text{MAPE} = \frac{100\%}{T} \sum_{t=1}^T \frac{|\hat{u}_t|}{y_t}.$$

Ця величина має такі порогові значення:

Оцінка MAPE, %	Характеристика якості регресійної моделі
< 10	висока точність
10 ÷ 20	добра точність
20 ÷ 50	задовільна точність
> 50	незадовільна точність

Для оцінки значущості моделі використовується **F-статистика**, яка розраховується за формулою

$$F_{\text{СТ}} = \frac{\hat{y}^T \hat{y} (T - N)}{\hat{u}^T \hat{u} (N - 1)}.$$

Якщо $F_{\text{СТ}} > F_{\text{КР}} = F(\alpha; N - 1; T - N)$ ($F_{\text{КР}}$ визначається за [таблицею Фішера](#)), то модель значимо відрізняється від нуля з ймовірністю $(1 - \alpha)100\%$.

Приклад 2.1. За даними [прикладу 1.1](#) досліджується залежність продуктивності праці від трудомісткості одиниці продукції, середньорічного фонду заробітної платні підприємства, фондоозброєності праці і невиробничих витрат. Необхідно:

1) Побудувати моделі, що виражають:

A: залежність продуктивності праці від трудомісткості одиниці продукції, середньорічного фонду заробітної платні підприємства, фондоозброєності

і невиробничих витрат;

Б: залежність продуктивності праці від трудомісткості одиниці продукції;

В: залежність продуктивності праці від середньорічного фонду заробітної платні підприємства;

Г: залежність продуктивності праці від фондоозброєності праці;

Д: залежність продуктивності праці від невиробничих витрат.

- 2) Для кожної з моделей оцінити $\hat{\sigma}_u^2$, MSE, MAPE, $M(\hat{u})$, де $M(\hat{u}) = \frac{1}{T} \sum_{t=1}^T u_t$.
- 3) Зробити порівняльну характеристику моделей і визначити найбільш впливовий чинник.

Отримані розрахунки занесемо до таблиці.

	Модель	$\hat{\sigma}_u^2$	MSE	MAPE	$M(\hat{u})$
А	$\hat{y} = 10.439 x_1 - 14.755 x_2 + 0.00003 x_3 + 0.198 x_4 - 0.00009 x_5$	1.28	1.024	9.4 %	$-9 \cdot 10^{-16}$
Б	$\hat{y} = 14.156 x_1 - 19.767 x_2$	1.682	1.547	10.6 %	$-4 \cdot 10^{-15}$
В	$\hat{y} = 6.056 x_1 + 0.00007 x_3$	2.622	2.412	17 %	$-9 \cdot 10^{-16}$
Г	$\hat{y} = 5.599 x_1 + 0.392 x_4$	3.942	3.627	21.2 %	$-6 \cdot 10^{-15}$
Д	$\hat{y} = 8.571 x_1 - 0.026 x_5$	4.679	4.305	22.2 %	$-2 \cdot 10^{-15}$

Висновки:

- 1) Перше припущення ($M(\hat{u}) = 0$) виконується у всіх п'яти моделях.
- 2) За показниками $\hat{\sigma}_u^2$ і MSE найкращою моделлю є модель А, далі моделі Б, В, Г і Д.
- 3) Так як серед парних регресій найкращою є модель Б, то трудомісткість одиниці продукції є найбільш впливовим чинником на продуктивність праці.
- 4) Коефіцієнт апроксимації так характеризує моделі: у моделі А відмінна апроксимація, Б і В — добра, а у Г і Д — задовільна.

Стандартизовані коефіцієнти регресії. Коефіцієнти еластичності

3.1 Стандартизоване рівняння регресії

Нагадаємо, що коефіцієнт $\hat{\beta}_k$ має таку інтерпретацію: якщо збільшити величину k -го регресора на одну одиницю при інших рівних умовах, то \hat{y} зміниться на $\hat{\beta}_k$ одиниць. Але звідси не випливає, що чим більший за модулем коефіцієнт при регресорі, тим впливовіший цей регресор. Це пов'язано з тим, що регресори мають різні виміри. Однак такий висновок можна робити, якщо перейти до стандартизованого рівняння. Для цього початкові дані необхідно стандартизувати за правилом

$$z_k^{\text{ст}} = \frac{z_k - \bar{z}}{\hat{\sigma}_z}, \quad \text{де} \quad \bar{z} = \frac{\sum_{i=1}^T z_i}{T}, \quad \hat{\sigma}_z = \sqrt{\frac{\sum_{i=1}^T (z_i - \bar{z})^2}{T-1}}.$$

Потім до стандартизованих (нормалізованих) початкових даних застосовується метод 1МНК, і ми отримуємо стандартизоване рівняння регресії.

Можна піти іншим шляхом. Якщо ми вже отримали регресійне рівняння, то оцінені значення стандартизованих регресійних коефіцієнтів можна обчислити за допомогою такої формули:

$$\hat{\beta}_k^{\text{ст}} = \hat{\beta}_k \frac{\hat{\sigma}_{x_k}}{\hat{\sigma}_y}, \quad (3.1)$$

де $\hat{\sigma}_{x_k}$ і $\hat{\sigma}_y$ — стандартні відхилення:

$$\hat{\sigma}_{x_k} = \sqrt{\frac{\sum_{i=1}^T (x_{ki} - \bar{x}_k)^2}{T-1}}, \quad \hat{\sigma}_y = \sqrt{\frac{\sum_{i=1}^T (y_i - \bar{y})^2}{T-1}}, \quad k = \overline{2, N}.$$

Стандартизований регресійний коефіцієнт β_k^{ct} вказує на те, як за великий при інших однакових умовах оцінений типовий ефект впливу k -го регресора порівняно з типовим ефектом зміни регресанда.

Чим більший за абсолютною величиною оцінений стандартизований коефіцієнт, тим більш впливовим є регресор.

Зазначимо, що $\hat{\sigma}_{x_1} = 0$, і тому $\beta_1^{ct} = 0$.

Приклад 3.1. Для моделі А із [прикладу 2.1](#)

$$\hat{y} = 10.439 x_1 - 14.755 x_2 + 0.00003 x_3 + 0.198 x_4 - 0.00009 x_5$$

розрахувати стандартизовані коефіцієнти і визначити найбільш впливовий регресор.

Розрахуємо середні значення регресорів і середньоквадратичні відхилення (стандартні відхилення):

$$\begin{aligned} \bar{x}_2 &= 0.308, & \bar{x}_3 &= 29587.28, & \bar{x}_4 &= 6.301, & \bar{x}_5 &= 19.550, & \bar{y} &= 8.068; \\ \hat{\sigma}_{x_1} &= 0, & \hat{\sigma}_{x_2} &= 0.086, & \hat{\sigma}_{x_3} &= 20719.923, & \hat{\sigma}_{x_4} &= 2.163, & \hat{\sigma}_{x_5} &= 4.340. \end{aligned}$$

Тоді за формулою (3.1) отримаємо

$$\hat{\beta}_1^{ct} = 0, \quad \hat{\beta}_2^{ct} = -0.598, \quad \hat{\beta}_3^{ct} = 0.306, \quad \hat{\beta}_4^{ct} = 0.202, \quad \hat{\beta}_5^{ct} = -0.0002.$$

Остаточо маємо

$$\hat{y}^{ct} = -0.598 x_2 + 0.306 x_3 + 0.202 x_4 - 0.0002 x_5.$$

Так як $|\hat{\beta}_2^{ct}| > |\hat{\beta}_3^{ct}| > |\hat{\beta}_4^{ct}| > |\hat{\beta}_5^{ct}|$, то найвпливовішим чинником на продуктивність праці є трудомісткість одиниці продукції, потім середньорічний фонд заробітної платні підприємства, далі фондоозброєність праці і невиробничі витрати. Такий же результат ми отримали раніше, коли порівнювали моделі між собою.

3.2 Коефіцієнти еластичності

Коли ми інтерпретували регресійні коефіцієнти, то приймали до уваги одиниці виміру регресорів і регресанда. Для визначення міри впливу регресора на регресанд без урахування одиниць їх виміру використовують коефіцієнт еластичності.

Коефіцієнт еластичності показує, на скільки відсотків зміниться регресанд, якщо при інших рівних умовах k -й регресор збільшити на один відсоток.

Оцінена еластичність регресанда y відносно регресора x_k розраховується за формулою

$$\hat{\varepsilon}_k = \hat{\beta}_k \frac{x_k^*}{y^*} \quad y^* \neq 0, \quad k = \overline{2, N}, \quad (3.2)$$

де x_k^* , y^* — значення k -го регресора і регресанда, що визначають точку регресійної

функції, для якої розраховується коефіцієнт еластичності. Частіше за все використовують значення арифметичних середніх \bar{x}_k, \bar{y} у базовому часовому або просторовому ряді.

Приклад 3.2. Для моделі А із [прикладу 2.1](#) розрахувати коефіцієнти еластичності $\hat{\varepsilon}_2, \hat{\varepsilon}_3, \hat{\varepsilon}_4, \hat{\varepsilon}_5$. Дати інтерпретацію отриманим результатам.

Користуючись формулою (3.2), розрахуємо коефіцієнти еластичності при

$$\bar{x}_2 = 0.308, \quad \bar{x}_3 = 29587.28, \quad \bar{x}_4 = 6.301, \quad \bar{x}_5 = 19.550, \quad \bar{y} = 8.068.$$

Маємо

$$\begin{aligned} \hat{\varepsilon}_2 &= \hat{\beta}_2 \frac{\bar{x}_2}{\bar{y}} = -14.755 \cdot \frac{0.308}{8.068} = -0.563; & \hat{\varepsilon}_3 &= \hat{\beta}_3 \frac{\bar{x}_3}{\bar{y}} = 0.00003 \cdot \frac{29587.28}{8.068} = 0.115; \\ \hat{\varepsilon}_4 &= \hat{\beta}_4 \frac{\bar{x}_4}{\bar{y}} = 0.198 \cdot \frac{6.301}{8.068} = 0.155; & \hat{\varepsilon}_5 &= \hat{\beta}_5 \frac{\bar{x}_5}{\bar{y}} = -0.00009 \cdot \frac{19.550}{8.068} = -0.0002. \end{aligned}$$

Коефіцієнт $\hat{\varepsilon}_2$ показує, що якщо трудомісткість одиниці продукції збільшити на один відсоток, то продуктивність праці зменшиться на 0.563%. Аналогічні висновки отримуємо і для інших коефіцієнтів еластичності. ☑

3.3 Дисперсійно-коваріаційна матриця

У класичній регресійній моделі вектор залишків \mathbf{U} і залежний від нього \mathbf{Y} є випадковими змінними. А так як вектор \mathbf{Y} у свою чергу входить до функції оцінювання коефіцієнтів регресії методом 1МНК, то ці коефіцієнти також є випадковими. Для характеристики випадкових змінних $\hat{\beta}_k$ ($k = \overline{1, N}$) разом з математичним очікуванням використовується дисперсія $\sigma_{\hat{\beta}_k}^2$ і коваріація $\sigma_{\hat{\beta}_i \hat{\beta}_k}$, ($k, i = \overline{1, N}, k \neq i$). Значення цих параметрів класичної регресійної моделі утворюють дисперсійно-коваріаційну матрицю $\hat{\Sigma}_{\hat{\beta}}$ розірності $N \times N$. Оцінена методом 1МНК дисперсійно-коваріаційна матриця має вигляд

$$\hat{\Sigma}_{\hat{\beta}} = \hat{\sigma}_u^2 (\mathbf{X}^T \mathbf{X})^{-1} = \begin{pmatrix} \hat{\sigma}_{\hat{\beta}_1}^2 & \hat{\sigma}_{\hat{\beta}_1 \hat{\beta}_2} & \cdots & \hat{\sigma}_{\hat{\beta}_1 \hat{\beta}_N} \\ \hat{\sigma}_{\hat{\beta}_2 \hat{\beta}_1} & \hat{\sigma}_{\hat{\beta}_2}^2 & \cdots & \hat{\sigma}_{\hat{\beta}_2 \hat{\beta}_N} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{\hat{\beta}_N \hat{\beta}_1} & \hat{\sigma}_{\hat{\beta}_N \hat{\beta}_2} & \cdots & \hat{\sigma}_{\hat{\beta}_N}^2 \end{pmatrix}. \quad (3.3)$$

Елементи цієї матриці використовуються при тестуванні гіпотез по окремим регресійним коефіцієнтам і для розрахунку інтервалу довіри.

На головній діагоналі матриці $\hat{\Sigma}_{\hat{\beta}}$ k -й елемент є 1МНК-оцінником дисперсії k -го оціненого регресійного коефіцієнта $\hat{\beta}_k$, а (k, i) -й елемент $\hat{\sigma}_{\hat{\beta}_k \hat{\beta}_i}$ є 1МНК-оцінником коваріації між $\hat{\beta}_k$ і $\hat{\beta}_i$.

Якщо оцінена коваріація $\hat{\sigma}_{\hat{\beta}_k \hat{\beta}_i} > 0$, то зв'язок між $\hat{\beta}_k$ і $\hat{\beta}_i$ прямий, тобто зі збільшенням (зменшенням) $\hat{\beta}_k$ збільшується (зменшується) $\hat{\beta}_i$, а якщо $\hat{\sigma}_{\hat{\beta}_k \hat{\beta}_i} < 0$, то зв'язок обернений — зі збільшенням (зменшенням) $\hat{\beta}_k$ зменшується (збільшується) $\hat{\beta}_i$ і навпаки.

Але коваріація має недолік: вона необмежена зверху і знизу: $-\infty < \sigma_{\beta_k \beta_i} < +\infty$. Тому до більш конкретного висновку про зв'язок між оціненими коефіцієнтами прийдемо, якщо розрахуємо коефіцієнти кореляції:

$$\hat{r}_{\hat{\beta}_k \hat{\beta}_i} = \frac{\hat{\sigma}_{\hat{\beta}_k \hat{\beta}_i}}{\hat{\sigma}_{\hat{\beta}_k} \cdot \hat{\sigma}_{\hat{\beta}_i}}, \quad (3.4)$$

де $\hat{\sigma}_{\hat{\beta}_k} = \sqrt{\hat{\sigma}_{\hat{\beta}_k}^2}$ — стандартне відхилення.

Приклад 3.3. Розрахувати оцінену дисперсійно-коваріаційну матрицю для моделі А із [прикладу 2.1](#).

Скориставшись формулою (3.3), отримаємо

$$\hat{\Sigma}_{\hat{\beta}} = \begin{pmatrix} 2.873 & -3.617 & -9.473 \cdot 10^{-6} & -0.081 & -0.047 \\ -3.617 & 10.235 & 2.127 \cdot 10^{-5} & 0.041 & -0.022 \\ -9.473 \cdot 10^{-6} & 2.127 \cdot 10^{-5} & 1.752 \cdot 10^{-10} & -1.996 \cdot 10^{-7} & -5.138 \cdot 10^{-8} \\ -0.081 & 0.041 & -1.996 \cdot 10^{-7} & 0.012 & -1.545 \cdot 10^{-4} \\ -0.047 & -0.022 & -5.138 \cdot 10^{-8} & -1.545 \cdot 10^{-4} & 0.003 \end{pmatrix},$$

де

$$\hat{\sigma}_{\hat{\beta}_1}^2 = 2.873, \quad \hat{\sigma}_{\hat{\beta}_2}^2 = 10.235, \quad \hat{\sigma}_{\hat{\beta}_3}^2 = 1.752 \cdot 10^{-10}, \quad \hat{\sigma}_{\hat{\beta}_4}^2 = 0.012, \quad \hat{\sigma}_{\hat{\beta}_5}^2 = 0.003, \\ \hat{\sigma}_{\hat{\beta}_1 \hat{\beta}_2} = \hat{\sigma}_{\hat{\beta}_2 \hat{\beta}_1} = -3.617, \quad \hat{\sigma}_{\hat{\beta}_1 \hat{\beta}_3} = \hat{\sigma}_{\hat{\beta}_3 \hat{\beta}_1} = -9.473 \cdot 10^{-6} \quad \text{і т. д.}$$

Оцінені коефіцієнти кореляції розрахуємо за формулою (3.4). Тоді кореляційна матриця оцінених коефіцієнтів має вигляд:

$$r = \begin{pmatrix} 1.0000 & -0.6669 & -0.4222 & -0.4306 & -0.5176 \\ -0.6669 & 1.0000 & 0.5023 & 0.1168 & -0.1267 \\ -0.4222 & 0.5023 & 1.0000 & -0.1366 & -0.0724 \\ -0.4306 & 0.1168 & -0.1366 & 1.0000 & -0.0261 \\ -0.5176 & -0.1267 & -0.0724 & -0.0261 & 1.0000 \end{pmatrix}.$$



3.4 Статистичні властивості 1МНК-оцінника $\hat{\beta}$

Оцінки регресійних коефіцієнтів, отримані за допомогою 1МНК, — випадкові величини. Тому їх статистичні властивості важливі для оцінки методу 1МНК.

Якщо виконуються всі умови У1–У5, то оцінник $\hat{\beta}$ має такі властивості:

- 1) Математичне очікування $\hat{\beta}$ дорівнює істинному значенню β , тобто $M(\hat{\beta}) = \beta$.
- 2) Оцінник $\hat{\beta}$ лінійний за Y : $\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T Y = AY$.
- 3) **Теорема Гауса — Маркова.** Оцінка $\hat{\beta}$ — ефективна, тобто має найменшу дисперсію серед усіх можливих лінійних незміщених оцінок:

$$\hat{\sigma}_{\hat{\beta}_k}^2 \leq \bar{\sigma}_{\bar{\beta}_k}^2, \quad i = \overline{1, N},$$

де $\bar{\beta}_k$ — можлива оцінка β_k за допомогою якогось іншого методу.

- 4) Оцінки $\hat{\beta}$ переконливі, тобто при зростанні кількості спостережень оцінки прямують за імовірністю до істинного значення:

$$\lim_{T \rightarrow +\infty} P \left\{ \left| \hat{\beta}_{iT} - \beta_i \right| < \varepsilon \right\} = 1.$$

Значущість коефіцієнтів регресії та їх інтервали довіри. Прогноз регресанда. Прогнозні інтервали

4.1 Тестування значущості коефіцієнтів регресії

Розглянемо t -тест, який використовується для перевірки гіпотез про істинні, але невідомі значення окремих або декількох коефіцієнтів регресії. При цьому статистично хочемо довести, що ці коефіцієнти не дорівнюють певному значенню, що сформульовано в якості нульової гіпотези, або що істинні, але невідомі значення двох або декількох коефіцієнтів рівняння не задовольняють певному співвідношенню величин. Інтервали довіри, які досить тісно пов'язані з t -тестом, також можуть використовуватися для перевірки гіпотез про числові значення окремих коефіцієнтів регресії або їх лінійної комбінації.

Умовою проведення t -тесту є наявність таких умов:

- 1) вектор $\mathbf{U} = (u_1, \dots, u_T)^\top$ є T -вимірним нормально розподіленим з нульовим вектором математичного очікування і коваріаційною матрицею $\sigma_u^2 \mathbf{I}$;
- 2) регресійна матриця \mathbf{X} детермінована і має повний ранг N .

У регресійному аналізі дуже важливим є питання: чи суттєво впливає k -й регресор у генеральній сукупності на регресанд, іншими словами, чи відрізняється істинне невідоме значення коефіцієнта регресії β_k від нуля? Для відповіді на це питання скористаємося тестом.

Алгоритм 4.1 (перевірка значущості коефіцієнтів регресії).

- 1) Формулюємо гіпотези: $H_A : \beta_k \neq 0$, $H_0 : \beta_k = 0$.
- 2) Обираємо рівень значущості α (0.1; 0.05; ...).

- 3) Розраховуємо $t_{CT} = \frac{\widehat{\beta}_k}{\widehat{\sigma}_{\widehat{\beta}_k}}$.
- 4) Знаходимо $t_{KP} = t\left(\frac{\alpha}{2}, T - N\right)$ за [таблицею Ст'юдента](#).
- 5) Якщо $|t_{CT}| > t_{KP}$, то гіпотеза H_0 відхиляється з ймовірністю $(1 - \alpha) 100\%$ і приймається гіпотеза H_A .

Розглянемо **більш загальний тест** з гіпотезами: $H_A : \widehat{\beta}_k \neq \beta_k^0$, $H_0 : \widehat{\beta}_k = \beta_k^0$, де β_k^0 є довільне, наперед встановлене і логічно обгрунтоване реальне число (випадок, коли $\beta_k^0 = 0$, розглянуто вище). t_{CT} розраховується таким чином:

$$t_{CT} = \frac{\widehat{\beta}_k - \beta_k^0}{\widehat{\sigma}_{\widehat{\beta}_k}}$$

Якщо $|t_{CT}| > t_{KP} = t\left(\frac{\alpha}{2}, T - N\right)$, то з ймовірністю $(1 - \alpha) 100\%$ гіпотеза H_0 відхиляється і приймається гіпотеза H_A .

***t*-тест гіпотез про лінійну комбінацію коефіцієнтів регресії**

У регресійному аналізі можуть виникнути такі питання:

- 1) Чи є рівновеликими два регресійні коефіцієнти?
- 2) Чи є сума декількох регресійних коефіцієнтів постійною величиною?

Для відповіді на ці питання можна скористатися тестом.

Алгоритм 4.2 (*t*-тест гіпотез про лінійну комбінацію коефіцієнтів регресії).

- 1) Формулюємо гіпотези:

$$H_A : c_1\beta_1 + c_2\beta_2 + \dots + c_N\beta_N \neq c^0, \quad \text{або} \quad \mathbf{c}^T \boldsymbol{\beta} \neq c^0;$$

$$H_0 : c_1\beta_1 + c_2\beta_2 + \dots + c_N\beta_N = c^0, \quad \text{або} \quad \mathbf{c}^T \boldsymbol{\beta} = c^0,$$

де c_i ($i = \overline{1, N}$) — константи, значення яких задаються; β_k — істинні невідомі значення регресійних коефіцієнтів; c^0 — скаляр, значення якого задається.

- 2) Задаємо рівень значущості α .
- 3) Розраховуємо $t_{CT} = \frac{\mathbf{c}^T \widehat{\boldsymbol{\beta}} - c^0}{\widehat{\sigma}}$, де $\widehat{\sigma} = \sqrt{\mathbf{c}^T \cdot \widehat{\boldsymbol{\Sigma}}_{\widehat{\boldsymbol{\beta}}} \cdot \mathbf{c}}$.
- 4) Визначаємо $t_{KP} = t\left(\frac{\alpha}{2}, T - N\right)$ за [таблицею Ст'юдента](#).
- 5) Якщо $|t_{CT}| > t_{KP}$, то з ймовірністю $(1 - \alpha) 100\%$ гіпотеза H_0 відхиляється.

Отримані оцінки коефіцієнтів регресії називають точковими, так як на числовій осі значення цього коефіцієнта представляє відповідна точка. Може виникнути питання: а наскільки ця точкова оцінка відрізняється від відповідного істинного значення? Для відповіді на це питання необхідно побудувати інтервал довіри для коефіцієнта.

4.2 Інтервали довіри для коефіцієнтів регресії

Інтервал довіри (інтервальна оцінка) для регресійного коефіцієнта β_k при рівні довіри $(1 - \alpha)$ є інтервалом з випадковими межами і включає істинне значення k -го регресійного коефіцієнта з ймовірністю $(1 - \alpha) 100\%$.

Він має такі межі:

$$\text{Нижня межа (НМ)} : \hat{\beta}_k - t\left(\frac{\alpha}{2}, T - N\right) \hat{\sigma}_{\hat{\beta}_k},$$

$$\text{Верхня межа (ВМ)} : \hat{\beta}_k + t\left(\frac{\alpha}{2}, T - N\right) \hat{\sigma}_{\hat{\beta}_k}.$$

Знаючи інтервали довіри для значень k -го коефіцієнта регресії, можна відповісти на питання про значущість цього коефіцієнта, а саме: якщо нуль належить інтервалу довіри, то коефіцієнт незначимо відрізняється від нуля з ймовірністю $(1 - \alpha) 100\%$.

Якщо деякі коефіцієнти моделі незначимо відрізняються від нуля, то цю модель не можна використовувати для прогнозу, а тільки для досліджень. Не варто користуватися моделлю, якщо всі коефіцієнти незначимо відрізняються від нуля.

Приклад 4.1. Визначити, чи значимо відрізняються від нуля коефіцієнти моделі А із [прикладу 2.1](#) і розрахувати інтервали довіри для коефіцієнтів цієї моделі. Візьмемо $\alpha = 0.2$.

Розглядається модель А:

$$\hat{y} = 10.439 x_1 - 14.755 x_2 + 0.00003 x_3 + 0.198 x_4 - 0.00009 x_5.$$

$$\text{Розрахуємо } t_{\text{кр}} = t\left(\frac{\alpha}{2}, T - N\right) = t\left(\frac{0.2}{2}, 25 - 5\right) = 1.725.$$

Далі розраховуємо $t_{\text{СТ}}$:

$$\begin{aligned} \beta_1 : t_{\text{СТ}} &= \frac{\hat{\beta}_1}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{10.439}{\sqrt{2.873}} = 6.159, \\ \beta_2 : t_{\text{СТ}} &= \frac{\hat{\beta}_2}{\hat{\sigma}_{\hat{\beta}_2}} = \frac{-14.755}{\sqrt{10.235}} = -4.612, \\ \beta_3 : t_{\text{СТ}} &= \frac{\hat{\beta}_3}{\hat{\sigma}_{\hat{\beta}_3}} = \frac{0.00003}{\sqrt{1.752 \cdot 10^{-10}}} = 2.364, \\ \beta_4 : t_{\text{СТ}} &= \frac{\hat{\beta}_4}{\hat{\sigma}_{\hat{\beta}_4}} = \frac{0.198}{\sqrt{0.012}} = 1.796, \\ \beta_5 : t_{\text{СТ}} &= \frac{\hat{\beta}_5}{\hat{\sigma}_{\hat{\beta}_5}} = \frac{-0.00009}{\sqrt{0.003}} = -0.002. \end{aligned}$$

Так як $|t_{\text{СТ}}|$ для $\beta_1, \beta_2, \beta_3, \beta_4$ більші за $t_{\text{кр}}$, то $\beta_1, \beta_2, \beta_3, \beta_4$ значимо відрізняються від нуля з ймовірністю 80%. У β_5 $|t_{\text{СТ}}| < t_{\text{кр}}$, тому коефіцієнт β_5 незначимо відрізняється

від нуля, тобто регресор x_5 не впливає на регресанд.

Визначимо інтервали довіри для коефіцієнтів моделі А:

$$\begin{aligned}
 \beta_1: \text{ НМ: } & \hat{\beta}_1 - t\left(\frac{\alpha}{2}, T - N\right) \hat{\sigma}_{\hat{\beta}_1} = 10.439 - 1.725\sqrt{2.873} = 7.515, \\
 & \text{ ВМ: } \hat{\beta}_1 + t\left(\frac{\alpha}{2}, T - N\right) \hat{\sigma}_{\hat{\beta}_1} = 10.439 + 1.725\sqrt{2.873} = 13.363, \\
 \beta_2: \text{ НМ: } & \hat{\beta}_2 - t\left(\frac{\alpha}{2}, T - N\right) \hat{\sigma}_{\hat{\beta}_2} = -14.755 - 1.725\sqrt{10.235} = -20.273, \\
 & \text{ ВМ: } \hat{\beta}_2 + t\left(\frac{\alpha}{2}, T - N\right) \hat{\sigma}_{\hat{\beta}_2} = -14.755 + 1.725\sqrt{10.235} = -9.236, \\
 \beta_3: \text{ НМ: } & \hat{\beta}_3 - t\left(\frac{\alpha}{2}, T - N\right) \hat{\sigma}_{\hat{\beta}_3} = 0.00003 - 1.725\sqrt{1.752 \cdot 10^{-10}} = 0.000008, \\
 & \text{ ВМ: } \hat{\beta}_3 + t\left(\frac{\alpha}{2}, T - N\right) \hat{\sigma}_{\hat{\beta}_3} = 0.00003 + 1.725\sqrt{1.752 \cdot 10^{-10}} = 0.00005, \\
 \beta_4: \text{ НМ: } & \hat{\beta}_4 - t\left(\frac{\alpha}{2}, T - N\right) \hat{\sigma}_{\hat{\beta}_4} = 0.198 - 1.725\sqrt{0.012} = 0.008, \\
 & \text{ ВМ: } \hat{\beta}_4 + t\left(\frac{\alpha}{2}, T - N\right) \hat{\sigma}_{\hat{\beta}_4} = 0.198 + 1.725\sqrt{0.012} = 0.389, \\
 \beta_5: \text{ НМ: } & \hat{\beta}_5 - t\left(\frac{\alpha}{2}, T - N\right) \hat{\sigma}_{\hat{\beta}_5} = -0.00009 - 1.725\sqrt{0.003} = -0.093, \\
 & \text{ ВМ: } \hat{\beta}_5 + t\left(\frac{\alpha}{2}, T - N\right) \hat{\sigma}_{\hat{\beta}_5} = -0.00009 + 1.725\sqrt{0.003} = 0.092.
 \end{aligned}$$

Інтерпретація для β_1 : істинне значення β_1 з ймовірністю 80% буде коливатися від 7.515 до 13.363.

Так як інтервалам довіри для $\beta_1, \beta_2, \beta_3, \beta_4$ нуль не належить, то ці коефіцієнти значимо відрізняються від нуля з ймовірністю 80%.

Якщо розглянути інтервал довіри для β_5 , то бачимо, що нуль належить йому, тому β_5 незначимо відрізняється від нуля з ймовірністю 80%, що ми раніше уже отримали. ☑

4.3 Точкові та інтервальні прогнози регресанда

Прогноз робиться для значень регресорів, що взяті з тієї ж генеральної сукупності, що й значення їх для оцінки β_k .

Якість прогнозу залежить від надійності оцінок коефіцієнтів моделі, повноти виконання передумов ІМНК і якості оцінок значень регресорів для прогнозної точки.

Прогноз при $\mathbf{x} = \tilde{\mathbf{x}}$ розраховується за формулою

$$\hat{y}^{\Pi} = \hat{\beta}^{\top} \tilde{\mathbf{x}}. \quad (4.1)$$

Точкове оцінене значення \hat{y}^{Π} , що розраховане за допомогою (4.1), може мати такі дві інтерпретації:

- оцінку математичного очікування регресанда;
- оцінку індивідуального значення регресанда.

Зазначимо, що в економічних дослідженнях важливе значення має не точкова оцінка прогнозу регресанда, а прогнозний інтервал.

Прогнозний інтервал величини математичного очікування регресанда має такі межі:

$$\text{НМ} : \hat{y}^{\text{П}} - t \left(\frac{\alpha}{2}, T - N \right) \hat{\sigma}_e,$$

$$\text{ВМ} : \hat{y}^{\text{П}} + t \left(\frac{\alpha}{2}, T - N \right) \hat{\sigma}_e,$$

де $\hat{y}^{\text{П}}$ розраховується за формулою (4.1); $\hat{\sigma}_e = \sqrt{\tilde{\mathbf{x}}^{\text{T}} \cdot \hat{\Sigma}_{\hat{\beta}} \cdot \tilde{\mathbf{x}}}$; $\tilde{\mathbf{x}}$ — прогнозна точка.

Інтервал прогнозу (прогнозний інтервал) для індивідуального значення регресанда розраховується так:

$$\text{НМ} : \hat{y}^{\text{П}} - t \left(\frac{\alpha}{2}, T - N \right) \hat{\sigma}_{e(i)},$$

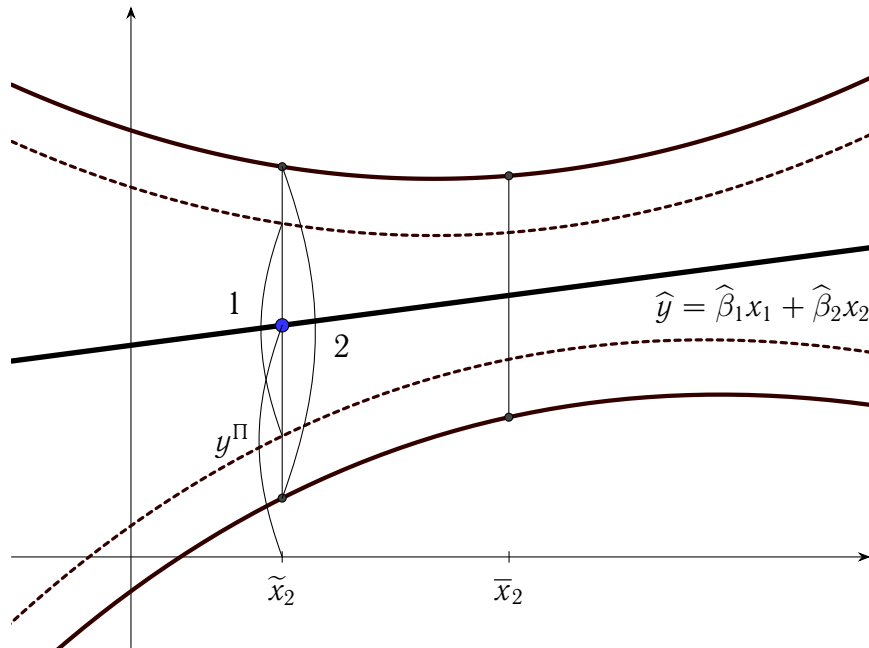
$$\text{ВМ} : \hat{y}^{\text{П}} + t \left(\frac{\alpha}{2}, T - N \right) \hat{\sigma}_{e(i)},$$

де $\hat{\sigma}_{e(i)} = \sqrt{\hat{\sigma}_u^2 + \tilde{\mathbf{x}}^{\text{T}} \cdot \hat{\Sigma}_{\hat{\beta}} \cdot \tilde{\mathbf{x}}}$.

Так як $\hat{\sigma}_u^2 \geq 0$, то прогнозний інтервал індивідуального значення регресанда завжди містить у собі прогнозний інтервал математичного очікування регресанда.

Зазначимо, що обидва інтервали будуть найменшими при $\tilde{\mathbf{x}} = \bar{\mathbf{x}}$ ($\bar{\mathbf{x}}$ — вектор середніх значень \mathbf{x}).

Розглянемо важливий зв'язок між точковим прогнозом і обома інтервальними прогнозами для гіпотетичної парної регресії у залежності від значення регресора x_2 у прогнозному періоді.



Позначення:

- межі прогнозного інтервалу індивідуального значення регресанда;
- - - - - межі прогнозного інтервалу математичного очікування регресанда;
- 1 величина прогнозного інтервалу математичного очікування регресанда;
- 2 величина прогнозного інтервалу індивідуального значення регресанда.

Як обирати прогнозні значення \tilde{x}_i ? Якщо відомо, що $a_i \leq x_i \leq b_i$ ($i = \overline{2, N}$), то прогнозні значення обирають так, щоб

$$a_i - \frac{b_i - a_i}{3} \leq \tilde{x}_i \leq b_i + \frac{b_i - a_i}{3}.$$

Приклад 4.2. Розрахувати точковий і інтервальний прогнози для моделі А із [прикладу 2.1](#) при $\tilde{\mathbf{x}} = \bar{\mathbf{x}}$ ($\bar{\mathbf{x}}$ — вектор середніх значень \mathbf{x}).

Нагадаємо, що $\bar{\mathbf{x}} = (1; 0.308; 29587.28; 6.301; 19.550)^\top$. Тоді

$$\begin{aligned} \hat{y}^\Pi &= 10.439 \cdot 1 - 14.755 \cdot 0.308 + 0.00003 \cdot 29587.28 + 0.198 \cdot 6.301 - \\ &\quad - 0.00009 \cdot 19.550 = 8.068. \end{aligned}$$

Прогнозне значення $\hat{y}^\Pi = 8.068$ має такі інтерпретації:

- 1) середня продуктивність праці на підприємствах, у яких трудомісткість продукції дорівнює 0.308, середньорічний фонд заробітної платні підприємства — 29587.28, фондоозброєність праці — 6.301, невиробничі витрати — 19.550, дорівнює 8.068;
- 2) продуктивність праці на підприємстві, у якого трудомісткість продукції дорівнює 0.308, середньорічний фонд заробітної платні підприємства — 29587.28, фондоозброєність праці — 6.301, невиробничі витрати — 19.550, дорівнює 8.068.

Оцінимо інтервал довіри для математичного очікування. Матрицю $\widehat{\Sigma}_{\beta}$ візьмемо із розрахунків [прикладу 3.3](#). Маємо

$$\widehat{\sigma}_e = \sqrt{\widetilde{\mathbf{x}}^T \cdot \widehat{\Sigma}_{\beta} \cdot \widetilde{\mathbf{x}}} = 0.226.$$

Тоді

$$\text{НМ : } \widehat{y}^{\text{П}} - t\left(\frac{\alpha}{2}, T - N\right) \widehat{\sigma}_e = 8.068 - 1.725 \cdot 0.226 = 7.678,$$

$$\text{ВМ : } \widehat{y}^{\text{П}} + t\left(\frac{\alpha}{2}, T - N\right) \widehat{\sigma}_e = 8.068 + 1.725 \cdot 0.226 = 8.458.$$

Отже, середня продуктивність праці на підприємствах, у яких трудомісткість продукції дорівнює 0.308, середньорічний фонд заробітної платні підприємства — 29587.28, фондоозброєність праці — 6.301, невиробничі витрати — 19.550, буде коливатися з ймовірністю 80 % у межах від 7.678 до 8.458.

Тепер оцінимо прогнозний інтервал для індивідуального значення регресанда. Розрахуємо

$$\widehat{\sigma}_{e(i)} = \sqrt{\widehat{\sigma}_u^2 + \widetilde{\mathbf{x}}^T \cdot \widehat{\Sigma}_{\beta} \cdot \widetilde{\mathbf{x}}} = \sqrt{1.280 + 0.226^2} = 1.154.$$

Тоді

$$\text{НМ : } \widehat{y}^{\text{П}} - t\left(\frac{\alpha}{2}, T - N\right) \widehat{\sigma}_{e(i)} = 8.068 - 1.725 \cdot 1.154 = 6.078,$$

$$\text{ВМ : } \widehat{y}^{\text{П}} + t\left(\frac{\alpha}{2}, T - N\right) \widehat{\sigma}_{e(i)} = 8.068 + 1.725 \cdot 1.154 = 10.058.$$

Отримали, що продуктивність праці на підприємстві, у якого трудомісткість продукції дорівнює 0.308, середньорічний фонд заробітної платні підприємства — 29587.28, фондоозброєність праці — 6.301, невиробничі витрати — 19.550, з ймовірністю 80 % буде коливатися від 6.078 до 10.058.

Як і очікували, прогнозний інтервал для математичного очікування вужчий за прогнозний інтервал для індивідуальних значень.

Показники адекватності регресійної моделі

5.1 Коефіцієнт детермінації

У класичному регресійному аналізі вважається, що функція регресії відома до оцінки значень параметрів, тобто лінійна регресійна модель правильно специфікована. Однак в емпіричних економічних і соціальних дослідженнях при застосуванні регресійного аналізу перш за все повинна бути обрана з множини варіантів рівнянь (які відрізняються регресорами) найбільш адекватна регресійна модель.

Для оцінки коефіцієнтів регресії використовується сума квадратів залишків. Її ми використовували для порівняння декількох моделей. Однак її не можна використовувати як показник адекватності однієї моделі, так як вона необмежена зверху. Відсутність верхньої межі у сумі квадратів залишків як недолік усувається за допомогою коефіцієнта детермінації.

Коефіцієнт детермінації R^2 можна ввести як квадрат емпіричного коефіцієнта кореляції між двома рядами спостережень: емпіричних даних y_t і розрахованих \hat{y}_t . Тобто

$$R^2 = r^2 = \frac{\left[\sum_{t=1}^T (y_t - \bar{y}) (\hat{y}_t - \bar{y}) \right]^2}{\sum_{t=1}^T (y_t - \bar{y})^2 \sum_{t=1}^T (\hat{y}_t - \bar{y})^2}. \quad (5.1)$$

Зазначимо, що $\bar{y} = \bar{\hat{y}}$. Друге рівноцінне визначення коефіцієнта детермінації робиться таким чином:

$$R^2 = \frac{\sum_{t=1}^T (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^T (y_t - \bar{y})^2}. \quad (5.2)$$

І нарешті третє визначення:

$$R^2 = 1 - \frac{\sum_{t=1}^T u_t^2}{\sum_{t=1}^T (y_t - \bar{y})^2}. \quad (5.3)$$

Коефіцієнт детермінації належить проміжку $[0; 1]$. Він характеризує долю варіації залежної змінної, обумовленої регресією.

Чим ближче R^2 до одиниці, тим краще регресійна модель апроксимує емпіричні дані. Якщо $\sum_{t=1}^T u_t^2 = 0$, то $R^2 = 1$. У цьому випадку $y_t = \hat{y}_t$, $t = \overline{1, T}$, а це означає, що всі емпіричні дані розташовані на регресійній гіперплощині. Випадок $R^2 = 0$ можливий при $\hat{y}_t = \bar{y}$ для $t = \overline{1, T}$.

З двох регресійних рівнянь, які відрізняються лише регресорами (кількість їх однакова), кращим вважається рівняння з найбільшим коефіцієнтом детермінації.

Коефіцієнт детермінації можна використати для визначення значущості регресійного рівняння, для цього розраховуємо

$$F_{\text{СТ}} = \frac{R^2(T - N)}{(1 - R^2)(N - 1)},$$

і якщо $F_{\text{СТ}} > F_{\text{КР}} = F(\alpha, N - 1, T - N)$ (значення $F_{\text{КР}}$ знаходимо за [таблицею Фішера](#)), то рівняння значимо відрізняється від нуля з ймовірністю $(1 - \alpha)$ 100%.

5.2 Коефіцієнт детермінації, скоригований за Тейлом

Звичайний коефіцієнт детермінації R^2 як критерій вибору функції регресії має недолік, який може призвести до того, що буде віддаватися перевага варіанту рівняння з великою кількістю регресорів. Цей недолік полягає у тому, що при включенні додаткового регресора до рівняння R^2 може тільки збільшитись. Якщо якість різних варіантів регресійного рівняння оцінювати тільки за допомогою R^2 , то рівняння з більшою кількістю регресорів, як правило, буде давати кращі результати, ніж з відносно малою їх кількістю. Однак з кожним додатковим регресором губиться одна ступінь вільностей, а цей недолік, на жаль, не враховується, коли R^2 виступає у якості критерію вибору. Кількість ступенів вільностей для регресійної моделі визначається як $T - N$, де T — кількість спостережень, а N — кількість регресорів. Якщо до регресійного рівняння включити додатковий регресор, то кількість регресорів N у рівнянні збільшується на 1, а кількість ступенів вільності зменшується на 1. При застосуванні t - і F -тестів, а також при побудові інтервалів довіри і прогнозних, бажано мати по можливості більшу кількість ступенів вільностей. Чим більша кількість ступенів вільностей, тим менші будуть інтервали. Тому у статистичному відношенні наявність додаткового регресора не завжди може бути бажана.

Що ж робити, коли ми хочемо порівнювати моделі з різною кількістю регресорів? У таких випадках необхідно коригувати коефіцієнт детермінації з урахуванням кількості регресорів x , які входять до різних моделей, тобто зменшити вплив залежності значення коефіцієнта детермінації від кількості регресорів. Для цього вводиться спеціальний скоригований за Тейлом коефіцієнт детермінації, який має вигляд:

$$\bar{R}_T^2 = 1 - \frac{\sum_{t=1}^T u_t^2}{T - N} \cdot \frac{T}{\sum_{t=1}^T (y_t - \bar{y})^2} \cdot (T - 1). \quad (5.4)$$

На відміну від простого коефіцієнта детермінації R^2 , скоригований за Тейлом коефіцієнт детермінації коригується з урахуванням ступенів вільності суми квадратів залишків та загальної суми квадратів.

Скоригований коефіцієнт детермінації за Тейлом пов'язаний з коефіцієнтом детермінації R^2 таким чином:

$$\bar{R}_T^2 = 1 - (1 - R^2) \frac{T - 1}{T - N}. \quad (5.5)$$

Зазначимо, що порівнювати значення двох або більше коефіцієнтів детермінації R^2 чи \bar{R}_T^2 можна лише за однакових залежних змінних, які можуть набирати різних функціональних форм.

Розглянемо деякі властивості скоригованого коефіцієнта детермінації за Тейлом \bar{R}_T^2 :

- 1) при $N > 1$ скоригований коефіцієнт детермінації за Тейлом не більший за звичайний коефіцієнт детермінації, тобто $\bar{R}_T^2 \leq R^2$;
- 2) якщо кількість регресорів зростає, то скоригований коефіцієнт детермінації за Тейлом зростає повільніше, ніж R^2 , тобто зменшується вплив кількості чинників на величину коефіцієнта детермінації;
- 3) скоригований за Тейлом коефіцієнт детермінації \bar{R}_T^2 може бути і від'ємним, тоді як R^2 завжди невід'ємний;
- 4) якщо до регресії додається новий чинник і його $|t_{CT}| > 1$, то скоригований за Тейлом коефіцієнт детермінації \bar{R}_T^2 збільшується, і тільки у цьому випадку.

Збільшення скоригованого за Тейлом коефіцієнта детермінації \bar{R}_T^2 при введенні нового регресора до моделі не обов'язково означає, що його коефіцієнт значимо відрізняється від нуля. Тому не потрібно вважати, що збільшення \bar{R}_T^2 означає покращення специфікації моделі. Це і є однією з причин, що заважає широкому використанню скоригованого за Тейлом коефіцієнта детермінації.

В останній час увага до коефіцієнта детермінації R^2 зменшилася, тому що у погано специфікованій моделі може бути великим коефіцієнт детермінації (наприклад, при наявності мультиколінеарності). Тепер дослідники розглядають коефіцієнт детермінації R^2 як один з діагностичних показників, що розраховуються при побудові моделі.

5.3 Коефіцієнт детермінації, скоригований за Аемією

Скоригований коефіцієнт детермінації за Аемією \bar{R}_A^2 розраховується таким чином:

$$\bar{R}_A^2 = 1 - (1 - R^2) \frac{T + N}{T - N}. \quad (5.6)$$

Скоригований коефіцієнт детермінації \bar{R}_A^2 відображає втрату ступеня свободи при включенні додаткового регресора більш чітко, ніж \bar{R}_T^2 . Це означає, що \bar{R}_A^2 змінюється на більшу величину, ніж \bar{R}_T^2 , при включенні додаткового регресора. Тому той, хто застосовує \bar{R}_A^2 замість \bar{R}_T^2 у якості критерію вибору, буде, при інших рівних умовах, віддавати перевагу рівнянню, що має меншу кількість регресорів.

5.4 Частковий коефіцієнт детермінації

Відомо, що $R^2(k+1) \geq R^2(k)$, тобто якщо включити до моделі регресор, то коефіцієнт детермінації не може зменшитися. Це означає, що

$$\begin{aligned} \Delta R^2 \text{ (від додаткового регресора)} &= \\ &= R^2 \text{ (з } k+1 \text{ регресором)} - R^2 \text{ (з } k \text{ регресорами)}. \end{aligned}$$

Щоб кількісно визначити ΔR^2 , необхідно оцінити дві регресії: одну з k , а другу з $(k+1)$ регресорами. А можна піти іншим шляхом: без значних обчислювальних витрат розрахувати коефіцієнт ΔR_k^2 , який називається частковим коефіцієнтом детермінації:

$$\Delta R_k^2 = \Delta R_{x_k}^2 = \frac{1 - R^2}{T - N} \left(\frac{\widehat{\beta}_k}{\widehat{\sigma}_{\widehat{\beta}_k}} \right)^2. \quad (5.7)$$

Інтерпретація ΔR_k^2 : він показує, на яку величину зменшиться коефіцієнт детермінації, якщо k -й регресор виключити з моделі.

Враховуючи суть коефіцієнта детермінації, можна зробити такий висновок: чим більший ΔR_k^2 , тим більш впливовим є у моделі k -й регресор.

Отже, ми отримали ще одну ознаку впливовості регресора у моделі.

Приклад 5.1. Для моделей А, Б, В, Г, Д із [прикладу 2.1](#) розрахувати коефіцієнти детермінації R^2 , \bar{R}_T^2 , \bar{R}_A^2 , а для моделі А — часткові коефіцієнти детермінації ΔR_2^2 , ΔR_3^2 , ΔR_4^2 , ΔR_5^2 . Визначити, які моделі значимо відрізняються від нуля, а які ні.

Розрахуємо всі параметри для моделі А.

Маємо:

$$R^2 = \frac{\sum_{t=1}^{25} (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^{25} (y_t - \bar{y})^2} = 0.7628.$$

Коефіцієнт детермінації досить великий. Розрахуємо \bar{R}_T^2 :

$$\bar{R}_T^2 = 1 - (1 - R^2) \frac{T - 1}{T - N} = 1 - (1 - 0.7628) \frac{25 - 1}{25 - 5} = 0.7154.$$

Тепер розрахуємо \bar{R}_A^2 :

$$\bar{R}_A^2 = 1 - (1 - R^2) \frac{T + N}{T - N} = 1 - (1 - 0.7628) \frac{25 + 5}{25 - 5} = 0.6442.$$

Як і писали вище, виконується співвідношення

$$R^2 > \bar{R}_T^2 > \bar{R}_A^2.$$

Визначимо, чи значимо відрізняється від нуля модель А. Для цього розрахуємо $F_{СТ}$ і $F_{КР}$:

$$F_{СТ} = \frac{R^2(T - N)}{(1 - R^2)(N - 1)} = \frac{0.7628(25 - 5)}{(1 - 0.7628)(5 - 1)} = 16.082,$$

$$F_{КР} = F(\alpha, N - 1, T - N) = F(0.05, 4, 20) = 2.87.$$

Так як $F_{СТ} > F_{КР}$, то з ймовірністю 95% модель А значимо відрізняється від нуля.

Такі ж параметри для інших моделей розраховуються аналогічно.

Тепер розрахуємо часткові коефіцієнти детермінації для моделі А.

Отримаємо:

$$\Delta R_2^2 = \frac{1 - R^2}{T - N} \left(\frac{\hat{\beta}_2}{\hat{\sigma}_{\hat{\beta}_2}} \right)^2 = \frac{1 - 0.7628}{25 - 5} \left(\frac{-14.755}{\sqrt{10.235}} \right)^2 = 0.252,$$

$$\Delta R_3^2 = \frac{1 - R^2}{T - N} \left(\frac{\hat{\beta}_3}{\hat{\sigma}_{\hat{\beta}_3}} \right)^2 = \frac{1 - 0.7628}{25 - 5} \left(\frac{0.00003}{\sqrt{1.752 \cdot 10^{-10}}} \right)^2 = 0.066,$$

$$\Delta R_4^2 = \frac{1 - R^2}{T - N} \left(\frac{\hat{\beta}_4}{\hat{\sigma}_{\hat{\beta}_4}} \right)^2 = \frac{1 - 0.7628}{25 - 5} \left(\frac{0.198}{\sqrt{0.012}} \right)^2 = 0.038,$$

$$\Delta R_5^2 = \frac{1 - R^2}{T - N} \left(\frac{\hat{\beta}_5}{\hat{\sigma}_{\hat{\beta}_5}} \right)^2 = \frac{1 - 0.7628}{25 - 5} \left(\frac{-0.00009}{\sqrt{0.003}} \right)^2 = 0.00000003.$$

Зважаючи на величини ΔR_i^2 , можемо зробити висновок, що найбільш впливовим регресором є x_2 , потім x_3 , x_4 і x_5 , що збігається з раніше отриманими результатами.

Отримані результати зведемо до однієї таблиці.

Модель	R^2	\bar{R}_T^2	\bar{R}_A^2	F_{CT}	F_{KP}
А $x_1 - x_5$	0.7628	0.7154	0.6442	16.082	2.87
Б x_1, x_2	0.6416	0.6260	0.4624	41.173	4.28
В x_1, x_3	0.4413	0.4170	0.362	18.166	4.28
Г x_1, x_4	0.1598	0.1233	-0.2603	4.374	4.28
Д x_1, x_5	0.0028	-0.0406	-0.4958	0.064	4.28

Усі моделі, крім моделі Д, значимо відрізняються від нуля.



Специфікація моделі. Побудова найкращої моделі

6.1 Специфікація моделі

Побудова економетричної моделі починається з питання про специфікацію моделі. Проблема специфікації моделі включає такі питання:

- 1) визначення кінцевої мети моделювання (прогноз, управління, імітація різних сценаріїв розвитку системи);
- 2) визначення списку регресорів і регресанда;
- 3) вибір виду рівняння регресії.

Відносно першого пункту. У залежності від кінцевої мети моделювання по-різному відносяться до кількості чинників у моделі, до виконання тих чи інших передумов використання 1МНК. Наприклад, якщо хочемо зробити за моделлю прогноз, то необхідно побільше включити до моделі чинників, а якщо проводити дослідження, то навпаки. Прогноз інколи можна робити і при наявності мультиколінеарності чинників, а от дослідження моделі робити недоцільно.

Отже, необхідно чітко сформулювати для себе мету побудови економетричної моделі і переходити до вибору змінних моделі.

Змінні, що включаються до моделі, повинні задовольняти таким вимогам:

- 1) їх можна кількісно вимірювати, а якщо до моделі включається якісний чинник, то необхідно надати йому кількісної визначеності;
- 2) незалежні змінні не повинні корелювати між собою.

Оптимальний вибір кількості чинників має велике значення при побудові моделі. Відбір чинників роблять за допомогою теоретико-економічного аналізу. Якщо це

неможливо зробити, то спочатку відбір роблять, виходячи із суті проблеми, а далі на підставі коефіцієнтів кореляції між чинниками і результатним показником. Тобто до моделі включають чинники, що мають найбільші за модулем часткові коефіцієнти кореляції з результатним показником. При цьому слідкуємо, щоб кореляція між чинниками не була суттєвою. Можна скористатися тестом Фаррара — Глобера для невиключення до моделі чинників, що викликають мультиколінеарність (див. [тему 9](#)).

При побудові моделі слід пам'ятати, що насичення моделі «зайвими» чинниками не тільки не збільшує суттєво коефіцієнт детермінації, але і призводить до зменшення точності оцінок. Елементи коваріаційної матриці для $\hat{\beta}$ будуть завищеними, що зменшує t -статистики і коефіцієнт детермінації, збільшуються інтервали довіри параметрів і прогнозний інтервал регресанда. Якщо з моделі видалити регресор і при цьому скоригований за Тейлом коефіцієнт детермінації збільшується, то цей регресор є зайвим.

Існують ще так звані «пропущені» змінні. Це ті суттєві змінні, що помилково не були включені до економетричної моделі. Якщо у регресійному рівнянні пропущена змінна, то оцінки коефіцієнтів регресії будуть зміщеними.

Ознакою, що вказує на пропущену змінну, є додатні знаки у добутку оцінки коефіцієнта при пропущеній змінній з коефіцієнтом парної кореляції пропущеної змінної з усіма змінними, включеними до моделі.

Наприклад, якщо у моделі

$$y = \beta_1 x_1 + \dots + \beta_{m-1} x_{m-1} + \beta_{m+1} x_{m+1} + \dots + \beta_N x_N + u$$

пропущена змінна x_m , то роблять висновок про її необхідність, якщо

$$\{\beta_m \cdot r_{x_i x_m}\} > 0, \quad i = \overline{2, N}, \quad i \neq m.$$

При дослідженні специфікації моделі необхідно дотримуватися таких правил:

- 1) за допомогою часткових коефіцієнтів кореляції виявити, які незалежні змінні суттєво корелюють з залежною;
- 2) перевірити значущість коефіцієнтів моделі;
- 3) виявити, чи збільшується скоригований за Тейлом коефіцієнт детермінації при включенні до моделі нової змінної;
- 4) чи суттєво впливає нова змінна на оцінки коефіцієнтів при інших змінних.

Існують статистичні критерії специфікації, розглянемо два найбільш відомі.

Критерій Рамсея Алгоритм його такий:

Алгоритм 6.1 (оцінка специфікації моделі за критерієм Рамсея).

1) оцінюємо коефіцієнти моделі за допомогою МНК

$$y = \sum_{i=1}^N \beta_i x_i + u_1; \quad (6.1)$$

2) розраховуємо $\hat{y}_t^2, \hat{y}_t^3, \hat{y}_t^4, t = \overline{1, T}$ і оцінюємо коефіцієнти моделі

$$y = \sum_{i=1}^N \alpha_i x_i + \alpha_{N+1} \hat{y}^2 + \alpha_{N+2} \hat{y}^3 + \alpha_{N+3} \hat{y}^4 + u_2 \quad (6.2)$$

методом МНК;

3) визначаємо $F_{СТ}$:

$$F_{СТ} = \frac{(RSS_1 - RSS_2)(T - 3)}{3 \cdot RSS_2},$$

де $RSS_i = \sum_{t=1}^T u_{it}^2, i = 1, 2, u_{i\bullet}$ — залишки у моделі (6.1) або (6.2);

4) якщо $F_{СТ} > F_{КР} = F(\alpha; 3; T - 3)$, то з ймовірністю $(1 - \alpha)100\%$ модель (6.1) є погано специфікованою.

Критерій Амемії Будуємо для конкретної моделі функцію

$$AF = \frac{\sum_{i=1}^T u_i^2 (T + N)}{T - N}.$$

Модель, для якої значення AF є меншим, вважається краще специфікованою.

6.2 Побудова «найкращої» регресії

Далі мова буде йти про побудову лінійної регресійної моделі. Існує дві протилежні думки для вибору регресії:

- 1) якщо за регресійним рівнянням необхідно робити прогноз, то до рівняння включається побільше незалежних змінних, що дає більш надійний прогноз;
- 2) оскільки збільшення кількості незалежних змінних призводить до небажаних наслідків (див. вище) і витрати на отримання інформації збільшуються, то намагаються включити до моделі як можна менше змінних.

Компромісним рішенням якраз і є побудова «найкращого» рівняння регресії. Існує декілька способів побудови такої регресії. Розглянемо один із них — метод усіх можливих регресій. Цей метод використовує три критерії оцінки регресії: скоригований

за Тейлом коефіцієнт детермінації \bar{R}_T^2 , залишковий середній квадрат $\hat{\sigma}_u^2$ і критерій Маллоуза — C_p -статистику.

Ідея методу дуже проста. Усі можливі регресії, що будуються на основі N регресорів, а їх буде 2^{N-1} , розбиваємо на групи. До кожної групи долучаються регресії з однаковою кількістю регресорів. За допомогою кожного з трьох критеріїв у групі знаходимо «найкращу» модель. Порівняльний аналіз результатів застосованих критеріїв дає можливість отримати «найкращу» регресію.

Зазначимо, що C_p -статистика Маллоуза розраховується за формулою

$$C_p = \frac{RSS_p}{\hat{\sigma}_u^2} - (T - 2N),$$

де p — кількість чинників у регресії, $\hat{\sigma}_u^2$ розраховано для найдовшої регресії.

Далі ідею методу проілюструємо на конкретному прикладі.

Приклад 6.1. Будуємо «найкращу модель» методом усіх регресій. Всього регресій буде 16. До групи А віднесемо регресію $y = f(x_1) = \bar{y}$. Зрозуміло, що ця модель не може скласти конкуренцію іншим і тому її надалі розглядати не будемо. До групи Б включаємо двофакторні регресії, В — трифакторні, Г — чотирифакторні і група Д складатиметься з однієї найдовшої регресії: $y = f(x_1, x_2, x_3, x_4, x_5)$. Спочатку порівнюємо регресії за допомогою \bar{R}_T^2 -критерію. Результати занесемо до таблиці.

Група	Вид регресії				\bar{R}_T^2			
Б	12	13	14	15	0.6260	0.4170	0.1233	-0.0406
В	123	124	125		0.6995	0.6686	0.6098	
		134	135	145	0.4616	0.3946	0.0871	
Г	1234	1235			0.7290	0.6852		
		1245	1345		0.6532	0.4407		
Д		12345				0.7154		

Відносно позначень: наприклад, 1345 — означає лінійну регресію з чинниками x_1, x_3, x_4, x_5 . З огляду отриманих результатів маємо: найкращою двофакторною моделлю є модель зі змінними x_1, x_2 ; трифакторною — x_1, x_2, x_3 ; чотирифакторною — x_1, x_2, x_3, x_4 .

Далі порівнюємо моделі за $\hat{\sigma}_u^2$ -критерієм: $\hat{\sigma}_u^2 = \frac{\mathbf{u}^T \mathbf{u}}{T - N}$. Найкращою моделлю в своїй групі є модель з найменшим $\hat{\sigma}_u^2$.

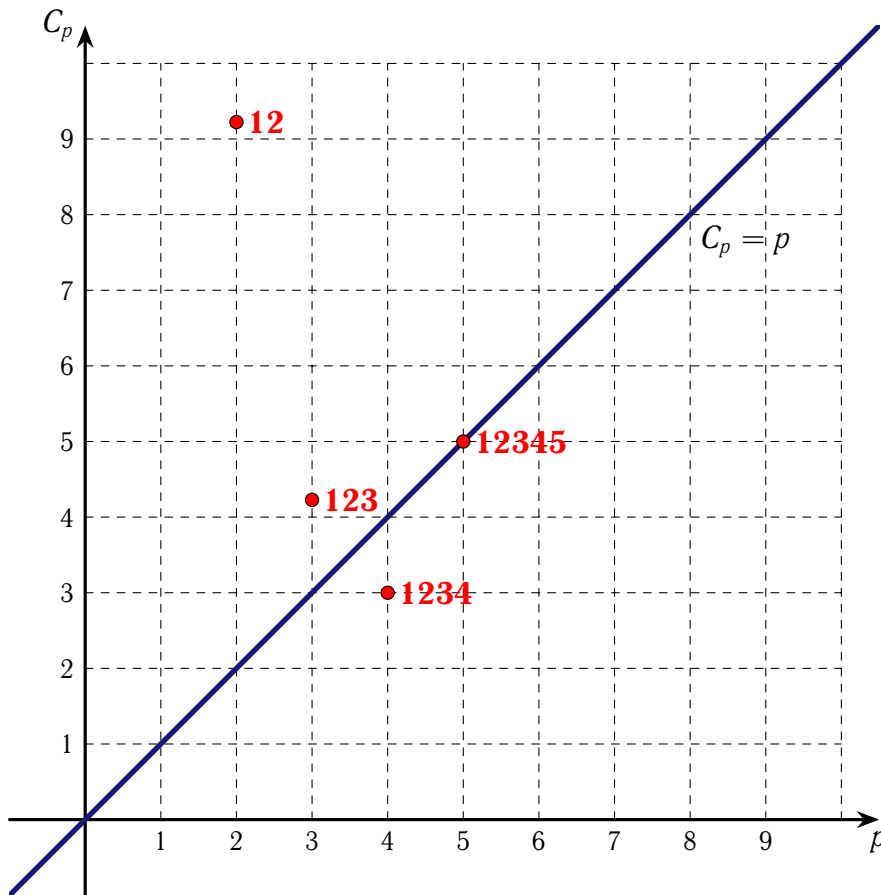
Група	Вид регресії	$\hat{\sigma}_u^2$
Б	12 13 14 15	1.6817 2.6216 3.9425 4.6792
В	123 124 125	1.3512 1.4901 1.7546
	134 135 145	2.4211 2.7223 4.1050
Г	1234 1235	1.2188 1.4154
	1245 1345	1.5593 2.5151
Д	12345	1.2798

Результати показують: найкращою двофакторною моделлю є модель з x_1, x_2 ; трифакторною — x_1, x_2, x_3 ; чотирифакторною — x_1, x_2, x_3, x_4 .

Залишилося зробити дослідження за допомогою C_p -статистики Маллоуза.

Група	Вид регресії	C_p -статистика
Б	12 13 14 15	9.2238 26.1151 49.8538 63.0944
В	123 124 125	4.2276 6.6163 11.1623
	134 135 145	22.6195 27.7973 51.5669
Г	1234 1235	3.0 6.2255
	1245 1345	8.5874 24.2707
Д	12345	5

Побудуємо графік залежності C_p -статистики від кількості чинників. Спочатку побудуємо пряму $C_p = p$. Вважається, що та модель краща у своїй групі, для якої C_p -статистика ближче знаходиться до прямої $C_p = p$.



Дивлячись на графік, можна зробити висновок, що найближче до прямої $C_p = p$ є точки, яким відповідають регресії з x_1x_2 ; $x_1x_2x_3$; $x_1x_2x_3x_4$.

Аналізуючи результати трьох критеріїв, можемо зробити висновок, що «найкращими» у своїй групі є моделі зі змінними x_1x_2 ; $x_1x_2x_3$; $x_1x_2x_3x_4$ та сама довга. Далі економетрист, враховуючи мету побудови регресії, обирає з них одну «найкращу».

Так як ми неодноразово показували, що регресор x_5 мало впливовий, то за «найкращу модель» оберемо модель з x_1, x_2, x_3, x_4 :

$$\hat{y} = 10.4377 x_1 - 14.7554 x_2 + 0.00003 x_3 + 0.1983 x_4. \quad (**)$$

Для з'ясування питання про специфікацію моделі скористаємося критерієм Рамсея.

Для моделі (**), розрахуємо \hat{y}_t і u_{1t} ($t = \overline{1, 25}$). Далі отримаємо, що $RSS_1 = \sum_{t=1}^{25} u_{1t}^2 = 25.5955$.

Будуємо модель

$$y^1 = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 \hat{y}^2 + \beta_6 \hat{y}^3 + \beta_7 \hat{y}^4 + u_2,$$

вона має вигляд

$$\hat{y}^1 = -80.7820 x_1 + 150.279 x_2 - 0.00033 x_3 - 2.0324 x_4 + \\ + 1.7887 \hat{y}^2 - 0.1234 \hat{y}^3 + 0.0032 \hat{y}^4.$$

Розраховуємо \hat{y}_t^1 , u_{2t} і $RSS_2 = \sum_{t=1}^{25} u_{2t}^2$. Отримуємо: $RSS_2 = 24.1516$. Тепер можна розрахувати $F_{СТ}$:

$$F_{СТ} = \frac{(RSS_1 - RSS_2)(T - 3)}{3RSS_2} = 0.4384.$$

$F_{КР} = F(\alpha; 3; T - 3) = F(0.05; 3; 22) = 3.05$. Так як $F_{СТ} < F_{КР}$, то обрана нами модель добре специфікована.

Скористаємося критерієм Амеїї для знаходження найкраще специфікованої моделі з найкращих моделей у групі.

У групі Б найкращою за трьома критеріями є модель $\hat{y} = 14.1564 x_1 - 19.7675 x_2$. Значення вирішальної функції дорівнює $AF = 45.4065$.

У групі В: $\hat{y} = 11.7905 x_1 - 15.4081 x_2 + 0.00003 x_3$, $AF = 37.8331$.

Далі, у групі Г:

$$\hat{y} = 10.4377 x_1 - 14.7554 x_2 + 0.00003 x_3 + 0.1983 x_4, \quad AF = 35.3462.$$

І нарешті у групі Д:

$$\hat{y} = 10.439 x_1 - 14.755 x_2 + 0.00003 x_3 + 0.198 x_4 - 0.00009 x_5, \quad AF = 38.3932.$$

Так як у моделі з групи Г AF найменше, то модель $\hat{y} = 10.4377 x_1 - 14.7554 x_2 + 0.00003 x_3 + 0.1983 x_4$ краще всіх специфікована.

Частина II

ОСОБЛИВИ ВИПАДКИ ПОБУДОВИ БАГАТОФАКТОРНОЇ РЕГРЕСІЙНОЇ МОДЕЛІ

- ▣ Автокореляція залишків
- ▣ Гетероскедастичність залишків
- ▣ Мультиколінеарність регресорів

Раніше ми будували і досліджували моделі при умові, що виконуються всі передумови 1МНК. Тепер розглядатимемо моделі, у яких умови 1МНК не виконуються. Ми розглянемо причини невиконання умов, наслідки, тестування наявності негативних явищ та їх усунення.

Розглянемо першу передумову: $M(u) = 0$ — математичне очікування залишків дорівнює нулеві.

Якщо ми маємо оцінену регресійну модель

$$\hat{y} = \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_N x_N \quad \text{і} \quad M(u) = a \neq 0,$$

то розглянемо нову регресійну модель

$$\hat{y} = (\hat{\beta}_1 - a) x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_N x_N,$$

у якій уже $M(u) = 0$.

Автокореляція залишків

Будемо вважати, що друга передумова 1МНК не виконується. А це означає, що дисперсійно-коваріаційна матриця залишків матиме вигляд $\widehat{\Sigma}_u = \widehat{\sigma}_u^2 \Omega$, де матриця Ω уже не є одиничною.

Розглянемо випадок, коли недиагональні елементи матриці $\widehat{\Sigma}_u$ не дорівнюють нулеві, тобто залишки u_t , $t = \overline{1, T}$ не є незалежними один від одного. У цьому випадку ми маємо справу з автокореляцією залишків.

7.1 Причини виникнення автокореляції залишків

Причинами появи автокореляції залишків можуть бути:

- 1) до регресійної моделі не включено чинники, що відіграють суттєву роль в економічній моделі;
- 2) специфікація моделі виявилася невдалою;
- 3) при дослідженні економічного явища числові дані отримано з великими похибками;
- 4) інерційність (наявність лагу) та циклічність економічного процесу;
- 5) перетворення початкової специфікації моделі до лінійної форми.

Зауважимо, що автокореляція залишків викликає більш суттєву проблему у випадку, коли інтервал між спостереженнями досить малий. Отже, якщо інтервал між спостереженнями збільшити, то вплив неврахованих змінних буде зменшуватися.

Якщо кореляція між послідовними значеннями залишків додатна, то автокореляція називається додатною, а якщо від'ємна, то — від'ємною.

У економічних явищах від'ємна автокореляція зустрічається досить рідко. Вона може з'явитися при перетворенні нелінійної форми моделі до лінійної.

7.2 Наслідки автокореляції залишків

Якщо параметри моделі оцінювали методом ІМНК і у залишків існує автокореляція, то матимемо такі погані наслідки:

- 1) $\hat{\beta}$ не буде кращою оцінкою, тобто теорема Гауса — Маркова не дійсна.
- 2) елементи коваріаційної матриці $\hat{\Sigma}_{\hat{\beta}}$ тенденційно «вниз» зміщені, а це негативно впливає на t - і F -тести, інтервали довіри і прогнозу (вони стають вузькими).

7.3 Авторегресійний процес першого порядку

Автокореляція залишків означає, що залишок u_t періоду t залежить від залишків більш ранніх періодів. Розглянемо простий вигляд залежності: u_t залежить від u_{t-1} .

Означення 7.1. Залишки u_t задовольняють **авторегресійному процесові першого порядку**, якщо виконуються такі умови:

$$u_t = \rho u_{t-1} + \varepsilon_t, \text{ де } |\rho| < 1, M(\varepsilon_t) = 0 \text{ для всіх } t, \quad (7.1)$$

$$\sigma_{\varepsilon_t \varepsilon_{t''}} = \begin{cases} 0, & t' \neq t'', \\ \sigma_{\varepsilon}^2, & t' = t''. \end{cases} \quad (7.2)$$

Умови (7.1) і (7.2) гарантують, що вплив далеких від u_t залишків буде малим і випадкова змінна ε_t вільна від автокореляції і гетероскедастичності.

Авторегресійні процеси більш високого порядку можна висувати як альтернативну гіпотезу, якщо у початкових даних є циклічні коливання. Це може виникати при обробці даних по півріччям при наявності у них сезонних коливань. Тоді у наявності буде авторегресійний процес **другого порядку**.

Якщо ж ми маємо справу з квартальними даними, то можемо мати справу з авторегресійним процесом **четвертого порядку**.

7.4 Тестування автокореляції залишків

7.4.1 Тест Дарбіна — Уотсона

Найбільш відомим і поширеним тестом перевірки моделі на наявність автокореляції першого порядку є d -тест Дарбіна — Уотсона.

Спочатку зробимо декілька зауважень відносно застосування цього тесту:

- 1) d -тест Дарбіна — Уотсона виявляє лише автокореляцію залишків першого порядку. Для інших порядків використовуються інші тести;

- 2) він не застосовується до моделей, у яких як незалежні змінні використовуються лагові значення залежної змінної. У цьому випадку використовують спеціальний h -тест Дарбіна.

Розглянемо етапи тестування автокореляції залишків тестом Дарбіна — Уотсона.

Алгоритм 7.1 (тест Дарбіна — Уотсона).

- 1) Формулюємо гіпотези:

$$H_A : \rho \neq 0, \quad \text{присутня автокореляція залишків,}$$

$$H_0 : \rho = 0, \quad \text{відсутня автокореляція залишків.}$$

- 2) Задаємо рівень значущості α .
 3) Розраховуємо значення d_{CT} за формулою

$$d_{CT} = \frac{\sum_{i=2}^T (u_i - u_{i-1})^2}{\sum_{i=1}^T u_i^2}. \quad (7.3)$$

Можна показати, що при великих вибірках між d_{CT} і параметром ρ існує зв'язок:

$$d_{CT} \approx 2(1 - \rho).$$

Якщо автокореляція залишків відсутня, то $\rho = 0$ і d_{CT} наближається до двох. При наявності додатної автокореляції d_{CT} буде менша двох і наближатиметься до нуля, а при від'ємній — більша двох і наближатиметься до чотирьох. Отже, $d_{CT} \in [0; 4]$.

- 4) Критичне значення d при довільному рівні значущості залежить від кількості пояснюючих змінних у рівнянні регресії і від кількості спостережень у вибірці. Але воно також залежить від конкретних значень пояснюючих змінних. Тому неможливо скласти таблицю з точними критичними значеннями для всіх можливих вибірок. Однак, можна обчислити верхню і нижню межі для критичного значення d .

Так як тест двосторонній, то d_{KP} знаходимо при $\alpha := \alpha/2$.

Далі представлено області прийняття рішень при d -тесті нульової гіпотези $H_0 : \rho = 0$.

0	d_0	d_u	2	$4 - d_u$	$4 - d_0$	4
-	?	+	?	-		

Умовні позначення:

–	області відхилення гіпотези H_0
?	області невизначеності
+	область прийняття гіпотези H_0

Якщо d_{CT} попадає в область невизначеності, то приймається гіпотеза про наявність автокореляції першого порядку.



Тест Дарбіна — Уотсона має певні недоліки: наявність зони невизначеності, визначає тільки автокореляцію першого порядку. Тому розглянемо і інші тести.

7.4.2 Тест серій (Бреуша — Годфрі)

Ідея тесту дуже проста і полягає у тому, що якщо існує автокореляція залишків першого порядку, то у рівнянні

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

коефіцієнт ρ буде значимо відрізнятися від нуля. Цей коефіцієнт оцінюємо ІМНК.

Перевага тесту Бреуша — Годфрі у порівнянні з тестом Дарбіна — Уотсона полягає в тому, що він перевіряється за допомогою статистичного критерію, а тест Дарбіна — Уотсона має зону невизначеності для значень статистики d . Другою перевагою тесту є можливість включати до регресорів залишки не тільки з лагом 1, а і з лагами 2, 3 і т. д. Це дозволяє виявити автокореляцію більш високих порядків.

Наприклад, розглянемо оцінену авторегресійну залежність залишків вигляду

$$u_t = 0.61 u_{t-1} - 0.15 u_{t-2} + 0.01 u_{t-3},$$

(0.10) (0.11) (0.10)

Внизу у дужках вписані стандартні відхилення коефіцієнтів. Зрозуміло, що значущим буде тільки перший коефіцієнт ($t_{CT} = 6.1$). Це означає, що існує авторегресійний процес першого порядку.

7.4.3 Тест на автокореляцію залишків для моделей з лаговою залежною змінною

При розгляді тесту Дарбіна — Уотсона ми записали, що його неможливо застосувати у присутності у рівнянні лагових залежних змінних. Тому Дарбін запропонував новий h -тест Дарбіна.

Нехай регресійне рівняння має вигляд

$$y_t = \alpha + \beta_1 x_t + \beta_2 y_{t-1} + u_t.$$

Далі обчислюється h_{CT} :

$$h_{CT} = \hat{\rho} \sqrt{\frac{T}{1 - T\hat{\sigma}_{\beta_2}^2}},$$

де $\hat{\rho}$ — оцінка ρ в означенні автокореляції першого порядку. Можна брати $\hat{\rho} \approx (1 - 0.5 d_{CT})$, де d_{CT} — оцінка статистики Дарбіна — Уотсона, T — довжина вибірки.

Розглядається гіпотеза: $H_0 : \rho = 0$. Вона відхиляється, якщо $h_{CT} > 1.96$ ($\alpha = 0.05$) або $h_{CT} > 2.58$ ($\alpha = 0.01$).

Зазначимо, що тест працює при великій вибірці.

7.5 Методи усунення автокореляції

Друга передумова 1МНК має вигляд: $\Sigma_u = \sigma_u^2 \mathbf{I}$. Якщо вона не виконується, то тоді $\hat{\Sigma}_u = \sigma_u^2 \mathbf{\Omega}$, і модель називається узагальненою регресійною. Якщо матриця $\mathbf{\Omega}$ нам відома, то параметри моделі можна оцінити за допомогою *узагальненого методу найменших квадратів* (УМНК). У випадку існування авторегресійного процесу першого порядку матриця $\mathbf{\Omega}$ має вигляд

$$\mathbf{\Omega} = \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \dots & \rho^{T-1} \\ \rho & 1 & \rho & \rho^2 & \dots & \rho^{T-2} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \rho^{T-4} & \dots & 1 \end{pmatrix}.$$

Узагальнена оцінка методом УМНК називається також оцінкою Ейткена і розраховується таким чином:

$$\begin{aligned} \tilde{\beta} &= (\mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{Y}, \\ \tilde{\Sigma}_{\tilde{\beta}} &= \tilde{\sigma}_u^2 (\mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{X})^{-1}, \\ \tilde{\sigma}_u^2 &= \frac{\tilde{\mathbf{u}}^\top \mathbf{\Omega}^{-1} \tilde{\mathbf{u}}}{T - N}, \\ \tilde{\mathbf{u}} &= \mathbf{y} - \mathbf{X}\tilde{\beta}. \end{aligned} \tag{A}$$

Знак $\tilde{}$ означає оцінку методом УМНК.

7.5.1 Оцінка моделі, якщо параметр ρ невідомий

Ситуації, коли параметр авторегресії ρ відомий, зустрічаються дуже рідко. Тому виникає необхідність у процедурах оцінення невідомого ρ . Як правило, вони мають ітеративний характер.

Метод Кохрейна — Оркатта

Розглядається регресійна модель

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_N x_N + u.$$

Алгоритм 7.2 (метод Кохрейна — Оркатта усунення автокореляції).

Спочатку оцінимо коефіцієнти моделі 1МНК. Далі

- 1) Розраховуємо залишки u_t , $t = \overline{1, T}$.
- 2) Застосовуємо тест Дарбіна — Уотсона. Якщо автокореляція залишків відсутня, то процес побудови моделі закінчено, а якщо присутня, то переходимо до п. 3.
- 3) Оцінюємо параметр $\hat{\rho}$:

$$\hat{\rho} = \frac{\sum_{t=2}^T u_t u_{t-1}}{\sum_{t=1}^T u_t^2}.$$

- 4) Перетворюємо початкові дані за допомогою матриці перетворень \mathbf{T}^A :

$$\mathbf{T}^A = \begin{pmatrix} \sqrt{1 - \hat{\rho}^2} & 0 & 0 & \dots & 0 & 0 \\ -\hat{\rho} & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\hat{\rho} & 1 \end{pmatrix}; \quad [\mathbf{T}^A \mathbf{Y} | \mathbf{T}^A \mathbf{X}] = [\mathbf{Y}^1 | \mathbf{X}^1].$$

- 5) За новими даними знаходимо 1МНК-оцінки $\hat{\beta}$ параметра β .
- 6) Переходимо до п. 1. ☑

Метод Дарбіна

Розглянемо регресію

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_N x_{Nt} + u_t. \quad (7.4)$$

Випишемо рівняння при $t := t - 1$:

$$y_{t-1} = \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1} + \dots + \beta_N x_{N,t-1} + u_{t-1}. \quad (7.5)$$

Домножимо (7.5) на (ρ) і віднімемо від (7.4):

$$y_t - \rho y_{t-1} = \beta_1 (x_{1t} - \rho x_{1,t-1}) + \beta_2 (x_{2t} - \rho x_{2,t-1}) + \dots + \beta_N (x_{Nt} - \rho x_{N,t-1}) + u_t - \rho u_{t-1}.$$

А так як присутній авторегресійний процес першого порядку, то:

$$u_t = \rho u_{t-1} + \varepsilon_t, \quad |\rho| < 1, \quad M(\varepsilon_t) = 0, \quad t = \overline{1, T},$$

$$\sigma_{\varepsilon_{t'} \varepsilon_{t''}} = \begin{cases} 0, & t' \neq t'', \\ \sigma_\varepsilon^2, & t' = t''. \end{cases}$$

Звідси маємо, що $u_t - \rho u_{t-1} = \varepsilon_t$, де залишки ε_t не мають автокореляції. Отже, отримали таку модель:

$$y_t = \rho y_{t-1} + \beta_1 x_{1t} (1 - \rho) + \beta_2 (x_{2t} - \rho x_{2,t-1}) + \dots + \beta_N (x_{Nt} - \rho x_{N,t-1}) + \varepsilon_t,$$

яка вільна від автокореляції залишків.

Далі застосовуємо 1МНК для оцінки коефіцієнтів цієї моделі. Отримаємо оцінку $\hat{\rho}$ (коефіцієнт при y_{t-1}). На наступному кроці оцінка $\hat{\rho}$ використовується для розрахунків перетворених змінних $(y_t - \hat{\rho} y_{t-1})$ і $(x_t - \hat{\rho} x_{t-1})$. Знову до нових змінних застосовуємо 1МНК. Тоді коефіцієнти при $(x_{it} - \hat{\rho} x_{i,t-1})$ будуть оцінками $\hat{\beta}_i$, а коефіцієнт при x_{1t} , поділений на $(1 - \hat{\rho})$, дає $\hat{\beta}_1$.

Зазначимо, що перетворення змінних можна робити за допомогою перетворення \mathbf{T}^A :

$$\mathbf{T}^A = \begin{pmatrix} -\hat{\rho} & 1 & 0 & \dots & 0 & 0 \\ 0 & -\hat{\rho} & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\hat{\rho} & 1 \end{pmatrix}, \quad \text{порядок матриці } \mathbf{T}^A : (T - 1) \times T.$$

Тоді перетворення робимо за правилом: $[\mathbf{T}^A \mathbf{Y} | \mathbf{T}^A \mathbf{X}]$. Тут ми губимо перше спостереження: $y_1, x_{11}, x_{21}, \dots, x_{N1}$.

Приклад 7.1. Для моделі А із [прикладу 2.1](#) необхідно:

- 1) дослідити залишки на автокореляцію першого порядку за допомогою тесту Дарбіна — Уотсона;
- 2) при наявності автокореляції залишків першого порядку оцінити коефіцієнти моделі за допомогою методу Кохрейна — Оркатта.

Нагадаємо, що модель А має вигляд:

$$\hat{y} = 10.439 x_1 - 14.755 x_2 + 0.00003 x_3 + 0.198 x_4 - 0.00009 x_5.$$

Формулюємо гіпотези: $H_A : \rho \neq 0, H_0 : \rho = 0$. Візьмемо $\alpha = 0.1$. За формулою

(7.3) розраховуємо d_{CT} :

$$d_{CT} = \frac{\sum_{i=2}^T (u_i - u_{i-1})^2}{\sum_{i=1}^T u_i^2} = 1.6368.$$

За таблицями Дарбіна — Уотсона при $\alpha = 0.1/2 = 0.05$, $T = 25$, $N = 5$ знаходимо: $d_0 = 1.038$, $d_u = 1.767$. Сформуємо таблицю

0	$d_0 = 1.038$	$d_u = 1.767$	2	2.233	2.962	4
-	?	+	?	-		

Ми бачимо, що $d_0 < d_{CT} < d_u$, тобто d_{CT} попадає в область невизначеності. Це означає, що нічого не можна сказати про наявність або відсутність автокореляції залишків першого порядку. Тому вважаємо, що автокореляція залишків першого порядку присутня з ймовірністю 90%.

Усунемо автокореляцію залишків методом Кохрейна — Оркатта. Розрахуємо параметр $\hat{\rho}$:

$$\hat{\rho} = \frac{\sum_{i=2}^T u_i u_{i-1}}{\sum_{i=1}^T u_i^2} = 0.1756.$$

Тепер побудуємо матрицю перетворень \mathbf{T}^A (вона має вимір 25×25):

$$\mathbf{T}^A = \begin{pmatrix} \sqrt{1 - \hat{\rho}^2} & 0 & 0 & \dots & 0 & 0 \\ -\hat{\rho} & 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -\hat{\rho} & 1 \end{pmatrix}.$$

Далі перетворимо початкові дані $\mathbf{D} = [\mathbf{Y}|\mathbf{X}]$ за допомогою матриці перетворень \mathbf{T}^A : $\mathbf{T}^A \mathbf{D} = [\mathbf{T}^A \mathbf{Y} | \mathbf{T}^A \mathbf{X}] = [\mathbf{Y}^1 | \mathbf{X}^1]$ і до нових даних застосуємо метод 1МНК. Отримаємо модель

$$A: \hat{y} = 10.712 x_1 - 14.428 x_2 + 0.00003 x_3 + 0.229 x_4 - 0.026 x_5.$$

Розрахуємо залишки і оцінимо за формулою (7.3) d_{CT} :

$$d_{CT} = 1.8775.$$

На цей раз d_{CT} попадає в область прийняття гіпотези H_0 , а це означає, що у отриманій моделі автокореляція залишків першого порядку вже відсутня.

Гетероскедастичність залишків

Друга передумова 1МНК передбачає однакову дисперсію у залишків u_t , $t = \overline{1, T}$. Таке твердження може здаватися дивним, адже у кожному спостереженні випадковий член має тільки одне значення, і тому виникає питання, а що тоді означає його дисперсія.

Мається на увазі його можлива поведінка до того, як буде зроблена вибірка. Тобто ймовірність того, що величина u_t прийме якесь значення, буде однаковою для всіх спостережень. Ця умова носить назву «гомоскедастичність», що означає «однаковий розкид».

Разом з тим для деяких вибірок більш доцільно вважати, що теоретичний розподіл випадкового члена є різним для різних спостережень у вибірці. Тобто апріорна ймовірність отримання великих відхилень буде досить великою. Це приклад гетероскедастичності, що означає «неоднаковий розкид».

Означення 8.1. Явище, коли дисперсії залишків неоднакові, називається *гетероскедастичністю залишків*, тобто $\sigma_{u_i}^2 \neq \sigma_u^2$ хоча б для деяких $i = \overline{1, T}$:

$$\Sigma_u = \sigma_u^2 \Omega = \sigma_u^2 \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_T^2 \end{pmatrix}.$$

8.1 Причини виникнення гетероскедастичності залишків

Гетероскедастичність залишків може виникнути, якщо:

- 1) об'єкти дослідження неоднорідні, тобто ми досліджуємо разом великі і малі

- об'єкти;
- 2) спостереження є членами часового ряду і якщо регресори і регресанд одночасно збільшуються або зменшуються;
 - 3) невключені змінні і похибки виміру, впливаючи на залишок, роблять його досить малим при малих значеннях регресорів і регресанда і досить великим при значних значеннях регресорів і регресанда;
 - 4) невдалою є специфікація моделі.

8.2 Наслідки гетероскедастичності залишків

Наслідки гетероскедастичності залишків можуть бути такі:

- 1) зменшується ефективність оцінок коефіцієнтів моделі, тобто можна знайти інші оцінки, що матимуть менші дисперсії і будуть незміщеними;
- 2) стандартні помилки будуть заниженими, а тому буде неправильне уявлення про точність оцінок рівняння регресії, інтервали довіри, значущість оцінок.

8.3 Методи тестування гетероскедастичності

Розглянемо декілька загальноновживаних тестів на виявлення гетероскедастичності залишків моделі

$$y = \sum_{i=1}^N \beta_i x_i + u.$$

Усі відомі тести можна розбити на дві групи. До першої групи відносяться тести Спірмена і Гольдфельда — Квандта, які дозволяють визначити наявність або відсутність гетероскедастичності, але вони не дають можливості дослідити кількісний характер залежності дисперсій помилок регресії від значень регресорів і тому не допомагають усунути гетероскедастичність залишків.

Друга група — це достатні умови існування гетероскедастичності. Це означає, що **якщо тест не виконується, то не можна вважати, що гетероскедастичність відсутня**. Ця група тестів допомагає усунути гетероскедастичність залишків.

8.3.1 Тест Гольдфельда — Квандта

Цей тест можна застосовувати як до малих, так і до великих вибірок, якщо справедливе припущення про пряму залежність дисперсії похибки від величини деякої незалежної змінної. Також необхідно, щоб помилки регресії були нормально розподіленими випадковими величинами.

Якщо ми не знаємо, від якої незалежної змінної залежать дисперсії похибки, то найбільш впливову змінну і обираємо. А ще краще застосувати тест до кожної із змінних (крім x_1).

Алгоритм 8.1 (тест Гольдфельда — Квандта).

Нехай дисперсія похибок залежить від змінної x_p . Тоді:

- 1) упорядковуємо початкові дані по мірі росту незалежної змінної x_p , відносно якої є підозра на гетероскедастичність;
- 2) видаляємо m середніх спостережень ($m = \frac{4}{15}T$);
- 3) на основі значень кожної із двох груп спостережень будуємо відповідно дві регресійні моделі методом МНК і розраховуємо залишки для кожної із них:

$$D = [Y|X] = \begin{bmatrix} \underline{Y1} & | & \underline{X1} \\ \hline & & \\ \hline \overline{Y2} & | & \overline{X2} \end{bmatrix} \begin{matrix} T_1 \times (N + 1) \\ m \\ T_2 \times (N + 1) \end{matrix},$$

$$T_1 + T_2 + m = T;$$

- 4) розраховуємо $F_{СТ}$:

$$F_{СТ} = \frac{\sum_{t=1}^{T_2} u_{2t}^2 / (T_2 - N)}{\sum_{t=1}^{T_1} u_{1t}^2 / (T_1 - N)},$$

де u_{kt} , $k = 1, 2$ — залишки k -ї групи спостережень;

- 5) Якщо $F_{СТ} > F_{КР} = F(\alpha, T_2 - N, T_1 - N)$, то гетероскедастичність залишків існує з ймовірністю $(1 - \alpha) 100\%$.

Зауважимо, що цей тест працює і без виключення m середніх спостережень, але при цьому його потужність зменшується.

8.3.2 Тест рангової кореляції Спірмена

Тест Спірмена застосовується як до малих, так і до великих вибірок при умові, що дисперсія похибок лінійно залежить від значень змінної x_p . Якщо ж залежність нелінійна, то доцільно користуватися тестом Кендалла.

Алгоритм 8.2 (тест Спірмена).

- 1) розраховуємо залишки моделі;
- 2) ранжуємо $|u_{pt}|$ та x_{pt} , $t = \overline{1, T}$ у зростаючому чи спадному порядку;

3) розраховуємо коефіцієнт рангової кореляції Спірмена за формулою

$$r^C = 1 - \frac{6 \sum_{i=1}^T d_i^2}{T(T^2 - 1)},$$

де d_i — різниця між рангами, що приписуються $|u_{pt}|$ і x_{pt} ;

4) перевіряємо значущість коефіцієнта Спірмена за критерієм Ст'юдента. Для цього розраховуємо t -статистику:

$$t_{CT} = \frac{|r^C| \sqrt{T-2}}{\sqrt{1-(r^C)^2}};$$

5) якщо $t_{CT} > t_{KP} = t\left(\frac{\alpha}{2}, T-2\right)$, то коефіцієнт Спірмена значимо відрізняється від нуля з ймовірністю $(1-\alpha) 100\%$.

Відомо, що $|r^C| \leq 1$. Якщо $|r^C| \geq 0.6$, то вважаємо, що гетероскедастичність залишків існує.

Розглянемо другу групу тестів, що дозволяють усунути гетероскедастичність, якщо вона присутня.

8.3.3 Тест Парка

Алгоритм 8.3 (тест Парка).

- 1) за допомогою 1МНК оцінюємо коефіцієнти регресії і розраховуємо залишки u_t ;
- 2) будуємо допоміжну модель

$$\ln |u| = b_1 z_1 + b_2 \ln |z_2| + \varepsilon, \quad (8.1)$$

де $z_{1t} \equiv 1$, $t = \overline{1, T}$, а z_2 — незалежна змінна, відносно якої ми вважаємо, що існує гетероскедастичність залишків;

- 3) за допомогою 1МНК оцінюємо коефіцієнти b_1 і b_2 ;
- 4) перевіряємо на значущість коефіцієнти b_1 і b_2 за допомогою t -критерію Ст'юдента (див. розділ 4.1). Якщо b_1 і b_2 значимо відрізняються від нуля, то існує змішана гетероскедастичність, а якщо b_1 незначимо відрізняється від нуля, а b_2 — значимо, то кажуть, що існує чиста гетероскедастичність залишків. Коли ж b_1 і b_2 незначимо відрізняються від нуля, то залежності (8.1) не існує між залишками і регресором x_p , і тому необхідно шукати іншу залежність, тобто користуватися іншим тестом.

8.3.4 Тест Уайта

Цей тест досить часто використовується. Будемо вважати, що дисперсія похибок регресії є одна і та ж функція від спостережених значень регресорів, тобто

$$\sigma_t^2 = f(x_t), \quad t = \overline{1, T}. \quad (8.2)$$

Досить часто функцію $f(x)$ обирають квадратичною, що відповідає тому, що середня квадратична похибка регресії залежить від спостережених значень регресорів приблизно лінійно.

Уайт пропонує оцінювати (8.2) за допомогою квадратів залишків:

$$u_t^2 = f(x_t) + \varepsilon_t, \quad t = \overline{1, T}, \quad (8.3)$$

де ε_t — випадковий член. Якщо регресія (8.3) значимо відрізняється від нуля, то гетероскедастичність залишків існує.

8.3.5 Тест Глейзера

Цей тест схожий на тест Уайта. Тільки у якості функції $f(x)$ обирається функція

$$\hat{f}(x) = \alpha + \gamma x_p^s$$

і розглядається регресія

$$|u_t| = \alpha + \gamma x_{pt}^s + \varepsilon_t, \quad (8.4)$$

де $s = 1, 2, \dots$. Регресія (8.4) оцінюється при різних значеннях s . Серед регресій, що значимо відрізняються від нуля, обираємо ту, що дає максимальну t_{CT} коефіцієнта γ .

8.4 Методи усунення гетероскедастичності

Вважаємо, що автокореляція залишків у моделі відсутня.

Якщо встановлено за будь-яким тестом наявність гетероскедастичності, то початкову модель перетворюємо так, щоб помилки мали постійну дисперсію. Потім невідомі параметри трансформованої моделі оцінюються за методом ІМНК. Вид перетворення первинної моделі залежить від форми гетероскедастичності залишків, тобто від форми залежності між дисперсією залишків та значеннями незалежних змінних.

Матриця перетворення \mathbf{T}^H при наявності гетероскедастичності залишків має

ВИГЛЯД:

$$\mathbf{T}^H = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{\sigma_2} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{\sigma_{T-1}} & 0 \\ 0 & 0 & 0 & \dots & 0 & \frac{1}{\sigma_T} \end{pmatrix},$$

де σ_i , $i = \overline{1, T}$ — стандартні відхилення залишків. Величини σ_i можуть бути невідомі. Визначати матрицю \mathbf{T}^H можна різними способами:

- 1) без використання статистичної оцінки;
- 2) визначення \mathbf{T}^H за принципом «гетероскедастичність між гомоскедастичними групами»;
- 3) σ_i або σ_i^2 можна отримати як оцінки $|u_i|$ і u_i^2 у регресіях (8.1), (8.3) і (8.4);
- 4) можна скористатися узагальненим методом найменших квадратів. Метод, що використовує матрицю \mathbf{T}^H , називають зваженим методом найменших квадратів.

1) **Визначення \mathbf{T}^H без статистичної оцінки.**

У цьому випадку достатньо зовнішньої інформації, щоб без статистичної оцінки визначити діагональні елементи матриці перетворень \mathbf{T}^H . Якщо відомо, що $\sigma_i = a x_{ki}$ або $\sigma_i^2 = a x_{ki}$, то елементи матриці \mathbf{T}^H такі: $\frac{1}{a x_{ki}}$ або $\frac{1}{\sqrt{a x_{ki}}}$, $i = \overline{1, T}$.

Константи a можна надати значення «1». Це не впливає на результат.

2) **Оцінка \mathbf{T}^H за принципом «гетероскедастичність між гомоскедастичними групами».**

Досить часто зовнішньої інформації недостатньо, щоб визначити діагональні елементи матриці \mathbf{T}^H без статистичної оцінки. Тому ці елементи повинні бути статистично оцінені. Насправді оцінка діагональних елементів матриці \mathbf{T}^H зводиться до розгляду гетероскедастичності між гомоскедастичними групами. Наприклад, якщо $\sigma_{u_t}^2 = \sigma_1^2$ для $x_{kt} < 20$ — перша група, $\sigma_{u_t}^2 = \sigma_2^2$ для $x_{kt} \geq 20$ — друга група, $1 \leq k \leq N$, це означає, що у групах існує гомоскедастичність, а між групами — гетероскедастичність. У цьому випадку необхідно оцінити σ_1^2 і σ_2^2 . Оцінки можна робити, як у тесті Гольдфелда — Квандта. Тоді діагональними елементами матриці \mathbf{T}^H будуть

$$\underbrace{\frac{1}{\widehat{\sigma}_1}, \frac{1}{\widehat{\sigma}_1}, \dots, \frac{1}{\widehat{\sigma}_1}}_{\text{для групи 1}}, \quad \underbrace{\frac{1}{\widehat{\sigma}_2}, \frac{1}{\widehat{\sigma}_2}, \dots, \frac{1}{\widehat{\sigma}_2}}_{\text{для групи 2}}.$$

3) На практиці значення σ_i майже ніколи не бувають відомі. У цьому випадку можна скористатися таким алгоритмом:

- оцінюємо модель методом 1МНК і розраховуємо залишки;

- будуюмо одну з моделей (8.1), (8.3), (8.4) або досить часто розглядають модель (8.2), де $f(x)$ є квадратичною функцією регресорів;
- розраховуємо $|\widehat{u}_i|$ або \widehat{u}_i^2 , $i = \overline{1, T}$, ці оцінки і будуть оцінками σ_i або σ_i^2 , $i = \overline{1, T}$.

4) Якщо існує гетероскедастичність залишків, то матриця Σ_u має вигляд

$$\Sigma_u = \sigma_u^2 \Omega = \sigma_u^2 \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 & 0 \\ 0 & \sigma_2^2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & \sigma_T^2 \end{pmatrix}.$$

Якщо σ_i^2 відомі, то можна скористатися узагальненим методом найменших квадратів, наприклад, $\widetilde{\mathbf{b}} = (\mathbf{X}^\top \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \Omega^{-1} \mathbf{Y}$.

Приклад 8.1. Перевірити на гетероскедастичність залишків модель А із [прикладу 7.1](#), яку звільнено від автокореляції залишків:

$$\widehat{y} = 10.712 x_1 - 14.428 x_2 + 0.00003 x_3 + 0.229 x_4 - 0.026 x_5.$$

Так як ми не знаємо, відносно якої змінної може існувати гетероскедастичність залишків, то застосовуємо тест Гольдфелда — Квандта, використовуючи послідовно усі змінні: x_2, x_3, x_4, x_5 . Якщо відносно декількох змінних існує підозра на гетероскедастичність, то обираємо ту, для якої F_{CT} найбільша.

У нашому прикладі відносно кожної із змінних x_2, x_3, x_4, x_5 упорядковуємо початкові дані по мірі зросту кожної із змінних. Видаляємо $m = \frac{4}{15} T = \frac{4}{15} \cdot 25 \approx 7$ середніх спостережень, і для кожної із отриманих двох груп будуюмо регресійні моделі, розраховуємо залишки і F_{CT} . Отримаємо такі F_{CT} :

$$F_{CT}^{x_2} = 3.8560, \quad F_{CT}^{x_3} = 1.5822, \quad F_{CT}^{x_4} = 5.8186, \quad F_{CT}^{x_5} = 1.3542.$$

Найбільшу F -статистику має змінна x_4 . Опишемо алгоритм побудови статистики для змінної x_4 . Так як $m = 7$, то після видалення семи спостережень залишиться тільки $25 - 7 = 18$ спостережень. Розбиваємо їх на дві групи по 9 спостережень у порядку збільшення значень змінної x_4 .

	y	x_1	x_2	x_3	x_4	x_5	
1	4.278589	0.824407	0.444346	22970.13	2.603104	20.56242	Перша група спостережень
2	8.400725	0.824407	0.170298	43952.87	2.73225	12.45233	
3	6.436894	0.824407	0.234495	48529.88	2.761626	18.83834	
4	5.113762	0.824407	0.306786	16505.34	2.927864	21.85909	
5	5.267511	0.824407	0.273959	12328.09	2.992447	17.54752	
6	7.288914	0.824407	0.205566	13321.92	3.161321	11.34056	
7	5.535992	0.824407	0.274495	48629.03	3.287898	22.55649	
8	5.725877	0.824407	0.267322	15378.78	3.557484	10.33378	
9	7.451836	0.824407	0.200149	7042.045	3.962812	24.20198	
17	8.683564	0.824407	0.136637	33512.32	6.186209	17.7238	Друга група спостережень
18	9.116125	0.984463	0.226426	47008.1	6.300562	17.44468	
19	7.754006	0.824407	0.199614	42006.42	6.676203	15.27849	
20	3.159328	0.824407	0.353275	4520.894	6.965999	19.24566	
21	5.468643	0.824407	0.318542	-2018.93	8.09846	16.69023	
22	4.097694	0.824407	0.289078	7367.159	8.237511	17.44556	
23	10.46293	0.824407	0.147858	34300.68	8.390372	23.23084	
24	8.228202	0.824407	0.389614	20157.57	8.960576	12.61056	
25	11.52469	0.824407	0.129614	91341.49	9.544413	15.46402	

За даними двох груп будуюмо дві моделі і розраховуємо залишки для цих моделей:

$$\begin{aligned}\hat{y}^1 &= 14.0164 x_1 - 15.5345 x_2 - 0.000002 x_3 - 0.3679 x_4 - 0.0056 x_5, \\ \hat{y}^2 &= 9.6385 x_1 - 16.8219 x_2 + 0.00003 x_3 + 0.6619 x_4 - 0.1393 x_5.\end{aligned}$$

$$\begin{aligned}\hat{\mathbf{u}}^1 &= (-0.7363, -0.6386, 0.2741, 0.4495, 0.8129, -0.1754, 0.3392, 0.2847, -0.6101)^\top, \\ \hat{\mathbf{u}}^2 &= (-0.4399, -0.3347, 0.3412, 0.9045, 0.0955, 2.2208, -1.6935, -2.0784, 1.0496)^\top.\end{aligned}$$

Розраховуємо $F_{СТ}$:

$$F_{СТ} = \frac{\sum_{i=1}^9 (u_i^2)^2 / (T_2 - N)}{\sum_{i=1}^9 (u_i^1)^2 / (T_1 - N)} = 5.8186.$$

Так як $F_{СТ} > F_{кр} = F(0.1; 9 - 5; 9 - 5) = F(0.1; 4; 4) = 4.11$, то гетероскедастичність залишків існує з ймовірністю 90%.

Так як доведено, що гетероскедастичність залишків існує, то тепер необхідно розрахувати σ_i ($i = \overline{1, 25}$) для матриці перетворень \mathbf{T}^H . Скористаємося третім способом розрахунку σ_i .

Вважаємо, що існує така залежність:

$$u_t^2 = a_1 x_{1t} + a_2 x_{2t} + a_3 x_{3t} + a_4 x_{4t} + a_5 x_{5t} + \varepsilon_t.$$

Коефіцієнти цієї регресії оцінимо методом 1МНК:

$$\widehat{u}_t^2 = -4.8967 x_{1t} + 8.2364 x_{2t} - 0.00001 x_{3t} + 0.4591 x_{4t} + 0.0495 x_{5t}. \quad (*)$$

Якщо доведемо, що ця модель значимо відрізняється від нуля, то це буде підтвердженням того, що гетероскедастичність залишків існує, і між u_t^2 і x_i ($i = \overline{1, 5}$) існує залежність (*). Знаходимо прогнозні значення \widehat{u}_t^2 і залишки $\widehat{\varepsilon}_t$. За ними розраховуємо $F_{CT} = \frac{\widehat{u}^{\top} \widehat{u}^2 (T - N)}{\widehat{\varepsilon}^{\top} \widehat{\varepsilon} (N - 1)} = 7.3182$. Розраховуємо $F_{KP} = F(\alpha; N - 1; T - N) = F(0.05; 5 - 1; 25 - 5) = 2.87$. Так як $F_{CT} > F_{KP}$, то з ймовірністю 95% модель (*) значимо відрізняється від нуля. Ми ще раз довели, що гетероскедастичність залишків існує.

Для усунення гетероскедастичності залишків оцінимо σ_i : $\sigma_i = \sqrt{|\widehat{u}_i^2|}$.

$$\mathbf{T}^H = \begin{pmatrix} \frac{1}{\sqrt{0.2799}} & 0 & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{\sqrt{0.9634}} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1}{\sqrt{1.2500}} & 0 \\ 0 & 0 & 0 & \dots & 0 & \frac{1}{\sqrt{1.1669}} \end{pmatrix}.$$

Початкові дані $[\mathbf{Y}|\mathbf{X}]$ перетворюємо за допомогою матриці перетворень \mathbf{T}^H :


$$[\mathbf{T}^H \mathbf{Y} | \mathbf{T}^H \mathbf{X}] = [\mathbf{Y}^1 | \mathbf{X}^1].$$

Далі використовуємо $[\mathbf{Y}^1 | \mathbf{X}^1]$ для побудови регресійної моделі методом 1МНК, у якій гетероскедастичність залишків уже відсутня:

$$\widehat{y} = 11.7747 x_1 - 18.7448 x_2 + 0.00001 x_3 + 0.1784 x_4 + 0.0216 x_5. \quad (**)$$

Для того, щоб показати відсутність гетероскедастичності у моделі (**), знову скористаємося тестом Гольдфелда — Квандта. Розрахуємо F_{CT} для кожної із змінних:

$$F_{CT}^{x_2} = 0.6558, \quad F_{CT}^{x_3} = 0.0746, \quad F_{CT}^{x_4} = 3.0079, \quad F_{CT}^{x_5} = 0.9325.$$

Як і раніше, $F_{KP} = F(0.1; 4; 4) = 4.11$. Так як всі $F_{CT} < F_{KP}$, то гетероскедастичність у моделі (**) відсутня. 

Мультиколінеарність регресорів

При побудові регресійної моделі намагаються включити до неї найбільш впливові чинники, які інколи дублюють один одного, тобто, четверта умова 1МНК не виконується. Це означає, що або стовпчиковий ранг матриці \mathbf{X} менше N , або ж регресори сильно корелюють між собою. Якщо хоча б один стовпчик матриці \mathbf{X} є лінійною комбінацією декількох, то має місце повна колінеарність. При цьому неможливо отримати оцінки коефіцієнтів моделі. Тоді необхідно уважно проаналізувати специфікацію моделі і зв'язки між чинниками регресії.

Повна колінеарність на практиці зустрічається досить рідко. Частіше зустрічається випадок, коли матриця \mathbf{X} має повний стовпчиковий ранг, але між регресорами існує суттєва кореляція. У цьому випадку мова йде про мультиколінеарність регресорів і формально можна розраховувати параметри моделі, але вони будуть мати «небажані» властивості.

9.1 Причини виникнення мультиколінеарності

Причинами мультиколінеарності можуть бути:

- 1) тенденція одночасної зміни економічних показників;
- 2) наявність трендів у динамічних рядах;
- 3) якщо декілька регресорів мають спільний часовий тренд, відносно якого вони здійснюють малі коливання;
- 4) використання в економетричних моделях лагових значень деяких чинників.

9.2 Наслідки мультиколінеарності

- 1) Оцінки дисперсій і коваріацій коефіцієнтів регресії досить великі.

Це означає, що будуть збільшуватися істинні і оцінники дисперсій і коваріацій коефіцієнтів регресії, а це призводить до таких наслідків: інтервали довіри коефіцієнтів регресії і прогнозні інтервали регресанда збільшуються, тестування стає неможливим, і це при великих значеннях R^2 .

- 2) Коефіцієнти регресії стають нестабільними.
Це означає, що невеликі зміни у специфікації моделі або у кількості спостережень призводять до великих змін значень коефіцієнтів регресії, до неправильних їхніх знаків і великих значень.
- 3) Труднощі в оцінці коефіцієнтів.
При повній колінеарності коефіцієнти регресії неможливо оцінити за методом ІМНК.
- 4) ІМНК-оцінники параметрів моделі втрачають властивість незміщеності.
- 5) Неможливо визначити внесок кожної незалежної змінної у дисперсію залежної змінної.
- 6) Оцінки параметрів при неколінеарних регресорах стають незначущими.

9.3 Тестування мультиколінеарності

9.3.1 Рекомендації щодо виявлення мультиколінеарності

Розглянемо декілька рекомендацій щодо виявлення мультиколінеарності регресорів.

- 1) Аналізують матрицю парних коефіцієнтів кореляції незалежних змінних. Якщо значення коефіцієнта кореляції за абсолютною величиною перевищує 0.8, то існує мультиколінеарність.
- 2) Якщо визначник матриці $\mathbf{X}^T \mathbf{X}$ близький до нуля, то це ознака мультиколінеарності.
- 3) Власні числа матриці $\mathbf{X}^T \mathbf{X}$ відіграють важливу роль при визначенні мультиколінеарності. Якщо мінімальне власне число близьке до нуля, то це вказує на мультиколінеарність.
- 4) Велике значення коефіцієнта детермінації R^2 і незначущість деяких коефіцієнтів регресії є ознакою мультиколінеарності.
- 5) Відомо, що детермінант кореляційної матриці регресорів задовольняє умові: $0 \leq \|r\| \leq 1$. Якщо $\|r\|$ близький до нуля, то це ознака мультиколінеарності регресорів.
- 6) Якщо модель двофакторна, то значення $r_{x_2x_3}$ достатнє для остаточного визначення мультиколінеарності.
- 7) Якщо коефіцієнт детермінації великий, а часткові коефіцієнти кореляції низькі, то мультиколінеарність можлива. Також можливо, що у моделі є надлишкові змінні. Але якщо коефіцієнт детермінації високий і часткові коефіцієнти кореляції високі, то мультиколінеарність не завжди можна виявити.

9.3.2 Метод інфляційних факторів визначення мультиколінеарності

- 1) Розглядається регресійне рівняння

$$y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_N x_N + u.$$

Для чинників x_2, \dots, x_N будемо багатфакторні моделі

$$x_i = \gamma_1 x_1 + \dots + \gamma_{i-1} x_{i-1} + \gamma_{i+1} x_{i+1} + \dots + \gamma_N x_N + \varepsilon, \quad i = \overline{2, N}. \quad (*)$$

- 2) Методом 1МНК оцінюємо коефіцієнти кожної з моделей (*) і розраховуємо коефіцієнти детермінації R_i^2 , далі обчислюємо інфляційні фактори

$$iF_i = \frac{1}{1 - R_i^2}.$$

Якщо $\min iF_i > 5$ ($i = \overline{2, N}$), то вважаємо, що існує мультиколінеарність регресорів.

Зрозуміло, що при дослідженні якогось явища доцільніше користуватися статистичними тестами. Тому розглянемо тест Фаррара — Глобера.

9.3.3 Тест Фаррара — Глобера на наявність мультиколінеарності

Найбільш повне дослідження мультиколінеарності регресорів можна здійснити за допомогою тесту Фаррара — Глобера. Тест складається з трьох статистичних критеріїв: χ^2 -критерій перевіряє весь масив регресорів на мультиколінеарність; F -критерій — кожного регресора з масивом інших; t -критерій — мультиколінеарність кожної пари регресорів.

Опишемо алгоритм тесту.

Алгоритм 9.1 (тест Фаррара — Глобера).

- 1) Нормалізація незалежних змінних.

Позначимо вектори незалежних змінних в економетричній моделі через $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$. Компоненти k -го вектора \mathbf{X}_k нормалізуються за формулою

$$x_{ki}^* = \frac{x_{ki} - \bar{x}_k}{\sigma_{x_k}},$$

де \bar{x}_k — середнє арифметичне значення компонент k -ї незалежної змінної, $i = \overline{1, T}$, σ_{x_k} — середнє квадратичне відхилення k -ї незалежної змінної, T — кількість спостережень, а N — кількість незалежних змінних.

- 2) Обчислення кореляційної матриці.

Кореляційна матриця обчислюється за формулою

$$\mathbf{r} = \frac{1}{T-1} \mathbf{X}^{*\top} \mathbf{X}^*,$$

де \mathbf{X}^* — матриця нормалізованих незалежних змінних.

3) Застосовуємо χ^2 -критерій.

Розраховуємо $\chi_{\text{СТ}}^2$ за формулою

$$\chi_{\text{СТ}}^2 = - \left[T - 1 - \frac{1}{6} (2N + 5) \right] \ln \|\mathbf{r}\|,$$

де $\|\mathbf{r}\|$ — детермінант кореляційної матриці \mathbf{r} . Якщо

$$\chi_{\text{СТ}}^2 > \chi_{\text{КР}}^2 = \chi^2(\alpha; 0.5N(N-1)),$$

то з ймовірністю $(1 - \alpha)$ 100% у масиві незалежних змінних існує мультиколінеарність. Значення $\chi_{\text{КР}}^2$ знаходимо за [таблицею \$\chi^2\$ -розподілу](#).

4) Визначаємо матрицю \mathbf{Z} .

Матриця \mathbf{Z} є оберненою до кореляційної матриці \mathbf{r} : $\mathbf{Z} = \mathbf{r}^{-1}$.

5) Застосування F_k -критеріїв.

Використовуючи елементи матриці \mathbf{Z} , розраховуємо F_k -статистики:

$$F_{k\text{СТ}} = (z_{kk} - 1) \frac{T - N}{N - 1}, \quad (k = \overline{1, N}),$$

де z_{kk} — діагональні елементи матриці \mathbf{Z} . Якщо $F_{k\text{СТ}} > F_{\text{КР}} = F(\alpha; N-1; T-N)$, то k -та незалежна змінна мультиколінеарна з масивом інших змінних.

6) Дослідження за t -критерієм.

Розраховуємо часткові коефіцієнти кореляції:

$$r_{ki|\bullet} = \frac{-z_{ki}}{\sqrt{z_{kk} \cdot z_{ii}}}$$

і розраховуємо t_{ki} -статистики:

$$t_{ki\text{СТ}} = |r_{ki|\bullet}| \sqrt{\frac{T - N}{1 - r_{ki|\bullet}^2}}.$$

Якщо $t_{ki\text{СТ}} > t_{\text{КР}} = t\left(\frac{\alpha}{2}; T - N\right)$, то коефіцієнт $r_{ki|\bullet}$ значимо відрізняється від нуля з ймовірністю $(1 - \alpha)$ 100% і по величині $|r_{ki|\bullet}|$ будемо судити про наявність чи відсутності зв'язку між \mathbf{X}_k і \mathbf{X}_i .

Зауваження 9.1. Елементи кореляційної матриці \mathbf{r} краще розраховувати за формулами [теми 1](#). У цьому випадку елементи кореляційної матриці будуть розраховані більш точно.

9.4 Методи усунення мультиколінеарності

Що ж робити при наявності мультиколінеарності? Однозначної відповіді на це питання не існує. Представники деяких шкіл вважають, що нічого не потрібно робити, а наявність мультиколінеарності це суть нашого буття.

Але все ж таки потрібно обережно підходити до проблеми усунення мультиколінеарності. Якщо ми бажаємо усунути «зайві» змінні, то у деяких ситуаціях модель втратить свою економічну суть, або ж отримаємо модель зі зміщеними оцінками. Якщо за побудованою моделлю бажаємо робити тільки прогноз, то можна використовувати модель і при наявності мультиколінеарності. Якщо коефіцієнт детермінації моделі великий, а коефіцієнти моделі значущі, то можна не звертати увагу на мультиколінеарність. Але якщо ми бажаємо аналізувати параметри моделі, то мультиколінеарність для нас уже проблема.

Розглянемо декілька рекомендацій по усуненню мультиколінеарності.

- 1) Змінити специфікацію моделі так, щоб знизити мультиколінеарність змінних до припустимої величини, тобто, щоб мультиколінеарність не була для нас проблемою.
- 2) Якщо між двома змінними \mathbf{X}_i і \mathbf{X}_j існує мультиколінеарність, то одну із них виключають з моделі. Яку змінну залишити у моделі, визначають економічною доцільністю. Якщо цього не можна зробити, то залишають ту, яка сильніше корелює з регресандом.
- 3) Інколи перетворення змінних знижує або і зовсім усуває мультиколінеарність, наприклад, від рівняння

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_N x_{Nt} + u_t, \quad t = \overline{1, T}$$

переходять до рівняння перших різниць

$$y_t - y_{t-1} = \beta_2 (x_{2t} - x_{2,t-1}) + \dots + \beta_N (x_{Nt} - x_{N,t-1}) + u_t, \quad t = \overline{1, T}.$$

- 4) Для зменшення мультиколінеарності можна перейти від незміщених оцінок, визначених методом МНК, до зміщених оцінок, що мають менше розсіювання відносно оціненого параметру.

Застосовують методи оцінки параметрів моделі, що враховують мультиколінеарність — «рідж-оцінки». При використанні «рідж-регресії» (або «гребневої регресії») замість незміщених оцінок розглядають зміщені оцінки, що задаються вектором $\beta_\tau = (\mathbf{X}^\top \mathbf{X} + \tau \mathbf{I})^{-1} \mathbf{X}^\top \mathbf{Y}$, де τ — деяке додатне число, \mathbf{I} — одинична матриця. Додача τ до діагональних елементів матриці $\mathbf{X}^\top \mathbf{X}$ робить оцінки параметрів моделі зміщеними, але при цьому збільшується детермінант матриці системи нормальних рівнянь, з якої визначається параметр β — замість $\|\mathbf{X}^\top \mathbf{X}\|$ він буде дорівнювати $\|\mathbf{X}^\top \mathbf{X} + \tau \mathbf{I}\|$.

- 5) Збільшення кількості спостережень може понизити або ж і усунути мультиколінеарність. Якщо ми використовуємо часові ряди, то можна зменшити довжину

кожного періоду. Але тут необхідно бути обережним, тому що може з'явитися автокореляція.

- 6) При кореляції \mathbf{X}_i і \mathbf{X}_j перевіряють наявність мультиколінеарності між \mathbf{X}_i і $\mathbf{X}^* = \mathbf{X}_i - \mathbf{X}_j$. При присутності її виключають з моделі одну із змінних, в іншому разі замість \mathbf{X}_j використовують змінну \mathbf{X}^* .
- 7) Інколи можна перейти до сумісних рівнянь регресії, тобто до рівнянь зі взаємодією чинників, тобто до добутоків чинників, якщо коефіцієнти при них будуть значимо відрізнятися від нуля.
- 8) Для усунення мультиколінеарності можна перейти до нових змінних, що ортогональні між собою і є лінійними комбінаціями початкових змінних. Це можна зробити за допомогою методу головних компонент.

9.4.1 Метод головних компонент

Метод головних компонент дає можливість перейти від змінних $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$, що корелюють між собою, до лінійно незалежних змінних. Для цього, використовуючи змінні $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N$, побудуємо усі можливі нормовано-центровані лінійні комбінації.

Означення 9.1. *Першою головною компонентою \mathbf{Z}_1 системи показників $\mathbf{X}_1, \dots, \mathbf{X}_N$ називається така нормовано-центрована лінійна комбінація цих показників, що має найбільшу дисперсію серед усіх нормовано-центрованих лінійних комбінацій.*

Означення 9.2. *k -ю головною компонентою \mathbf{Z}_k системи показників $\mathbf{X}_1, \dots, \mathbf{X}_N$ називається така нормовано-центрована лінійна комбінація цих показників, що не корелює з $k - 1$ попередніми головними компонентами і серед усіх нормовано-центрованих лінійних комбінацій, що не корелюють з $k - 1$ попередніми головними компонентами, має найбільшу дисперсію.*

Алгоритм 9.2 (алгоритм побудови головних компонент).

- 1) нормалізуємо змінні $\mathbf{X}_1, \dots, \mathbf{X}_N$ за правилом

$$x_{ij}^H = \frac{x_{ij} - \bar{x}_j}{\sigma_{x_j}}, \quad j = \overline{1, N},$$

отримаємо матрицю \mathbf{X}^H ;

- 2) будуємо кореляційну матрицю \mathbf{r} :

$$\mathbf{r} = \frac{1}{T - 1} (\mathbf{X}^H)^\top \mathbf{X}^H$$

або краще елементи (коефіцієнти кореляції між \mathbf{X}_i та \mathbf{X}_j) матриці розраховувати за формулами, описаними у [темі 1](#);

- 3) знаходимо власні числа і власні вектори матриці \mathbf{r} ;
- 4) ранжуємо власні числа у порядку зменшення їх величин: $\lambda_1, \lambda_2, \dots, \lambda_N$;

5) будуюмо головні компоненти

$$\mathbf{Z}_k = \mathbf{X}^\top \mathbf{l}_k = x_1 l_{k1} + x_2 l_{k2} + \dots + x_N l_{kN},$$

де \mathbf{l}_k — власний вектор матриці \mathbf{r} , що відповідає власному числу λ_k .

Зауважимо, що дисперсія k -ї головної компоненти \mathbf{Z}_k дорівнює λ_k . ☑

Головні компоненти характеризуються такими властивостями:

- 1) їх кількість дорівнює кількості вихідних ознак;
- 2) вони є ортогональними;
- 3) вони нормалізовані (середні значення рівні нулеві, дисперсії дорівнюють одиниці);
- 4) вони впорядковані таким чином, що перша головна компонента пояснює найбільшу частку дисперсії вихідних ознак, наступна — найбільшу частку дисперсії, що залишилася непоясненою першою компонентою, і т. д.;
- 5) якщо початкові дані нормалізовані, то коваріаційні і кореляційні матриці співпадають і $\sum_{k=1}^N \lambda_k = N$.

Відомо також, що $\sum_{i=1}^N \sigma_x^2 = \sum_{i=1}^N \sigma_z^2 = \sum_{k=1}^N \lambda_k = N$.

Враховуючи зменшення частки пояснюваної дисперсії вихідних ознак наступною головною компонентою, на практиці для аналізу беруться не всі компоненти, а лише ті, що пояснюють наперед задану сумарну частку цієї дисперсії. Ми користуємося критерієм інформативності:

$$I_p = \frac{\lambda_1 + \dots + \lambda_p}{N}.$$

На практиці прийнято, що якщо $I_p \geq 0.8$, то кількості головних компонент достатньо, щоб вони пояснювали дисперсію змінних не менше, ніж на 80%.

Розглянемо матрицю навантажень (факторних навантажень) $\mathbf{A} = (a_{ij})$, $i, j = \overline{1, N}$ головних компонент на початкові ознаки, яка є важливою характеристикою головних компонент. Якщо головні компоненти будуються для нормалізованих ознак, то елементи матриці навантажень a_{ij} визначають ступінь тісноти парного лінійного зв'язку між \mathbf{X}_i і \mathbf{Z}_j .

Матриця навантажень \mathbf{A} визначається співвідношенням $\mathbf{A} = \mathbf{L}^\top \mathbf{\Lambda}^{1/2}$, де матриця \mathbf{L} складається з рядків \mathbf{l}_j , $j = \overline{1, N}$, що є власними векторами матриці \mathbf{r} , а

$$\mathbf{\Lambda}^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\lambda_N} \end{pmatrix}.$$

Маємо

$$\mathbf{A} = \begin{matrix} & \mathbf{z}_1 & \mathbf{z}_2 & \dots & \mathbf{z}_N \\ \mathbf{x}_1 & \left(\begin{array}{cccc} a_{11} & a_{12} & \dots & a_{1N} \\ a_{21} & a_{22} & \dots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \dots & a_{NN} \end{array} \right) \\ \mathbf{x}_2 & & & & \\ \vdots & & & & \\ \mathbf{x}_N & & & & \end{matrix}.$$

Так як головні компоненти є лінійною комбінацією ознак, то досить важко дати економічне тлумачення головних компонент. Щоб все ж таки це зробити, виділяємо ті ознаки \mathbf{X}_i , які досить сильно корелюють ($r_{\mathbf{x}_i \mathbf{z}_j} \geq 0.6$) з головною компонентою \mathbf{Z}_j , і по назвам ознак намагаємося дати узагальнену назву головній компоненті \mathbf{Z}_j .

Приклад 9.1. Використовуючи метод Фаррара — Глобера, дослідити вектори \mathbf{x}_2 , \mathbf{x}_3 , \mathbf{x}_4 , \mathbf{x}_5 (див. [приклад 1.1](#)) на мультиколінеарність, і якщо вона існує, то усунути її за допомогою методу головних компонент.

Згідно з алгоритмом Фаррара — Глобера спочатку будемо кореляційну матрицю

$$\mathbf{r} = \begin{pmatrix} 1.0000 & -0.5228 & -0.2140 & 0.0993 \\ -0.5228 & 1.0000 & 0.2275 & 0.0131 \\ -0.2140 & 0.2275 & 1.0000 & 0.0141 \\ 0.0993 & 0.0131 & 0.0141 & 1.0000 \end{pmatrix}.$$

Для визначення мультиколінеарності у масиві змінних розрахуємо $\chi_{\text{СТ}}^2$: $\chi_{\text{СТ}}^2 = 8.7792$. Далі, $\chi_{\text{КР}}^2 = \chi^2(0.05; 6) = 12.59$. Так як $\chi_{\text{СТ}}^2 < \chi_{\text{КР}}^2$, то з ймовірністю 95% у масиві змінних мультиколінеарність відсутня.

Далі розрахуємо матрицю $\mathbf{Z} = \mathbf{r}^{-1}$:

$$\mathbf{Z} = \begin{pmatrix} 1.4171 & 0.7102 & 0.1438 & -0.1521 \\ 0.7102 & 1.4106 & -0.1678 & -0.0867 \\ 0.1438 & -0.1678 & 1.0693 & -0.0272 \\ -0.1521 & -0.0867 & -0.0272 & 1.0166 \end{pmatrix}.$$

Розглянемо питання про можливу мультиколінеарність між \mathbf{x}_i ($i = \overline{2, 5}$) і масивом інших змінних. Для цього розрахуємо $F_{i\text{СТ}}$:

$$F_{2\text{СТ}} = 2.6419, \quad F_{3\text{СТ}} = 2.6003, \quad F_{4\text{СТ}} = 0.4391, \quad F_{5\text{СТ}} = 0.1053;$$

$F_{\text{КР}} = 2.85$. Так як $F_{i\text{СТ}} < F_{\text{КР}}$ ($i = \overline{2, 5}$), то жодна зі змінних не корелює з відповідним масивом інших змінних.

Знайдемо часткові коефіцієнти кореляції:

$$\begin{aligned} r_{x_2x_3|\bullet} &= -0.5023, & r_{x_2x_4|\bullet} &= -0.1168, & r_{x_2x_5|\bullet} &= 0.1267, \\ r_{x_3x_4|\bullet} &= 0.1366, & r_{x_3x_5|\bullet} &= 0.0724, & r_{x_4x_5|\bullet} &= 0.0261. \end{aligned}$$

Далі розраховуємо t_{CT} для всіх $r_{x_i x_j|\bullet}$. Отримаємо:

$$\begin{aligned} t_{x_2x_3CT} &= 2.5320, & t_{x_2x_4CT} &= 0.5127, & t_{x_2x_5CT} &= 0.5567, \\ t_{x_3x_4CT} &= 0.6010, & t_{x_3x_5CT} &= 0.3163, & t_{x_4x_5CT} &= 0.1138; \end{aligned}$$

$t_{KP} = t(0.05; 21) = 1.721$, тобто з ймовірністю 95% коефіцієнт $r_{x_2x_3|\bullet}$ значимо відрізняється від нуля.

Отже, один з критеріїв вказує на те, що існує мультиколінеарність регресорів. Усувати її будемо методом головних компонент. Для цього розрахуємо власні числа і власні вектори матриці r . Проранжувавши власні числа, отримаємо:

$$\begin{aligned} \lambda_1 &= 1.6722, & \mathbf{l}_1 &= (0.6411, -0.6389, -0.4188, 0.0734)^\top, \\ \lambda_2 &= 1.0178, & \mathbf{l}_2 &= (0.1268, 0.0786, 0.2422, 0.9587)^\top, \\ \lambda_3 &= 0.8448, & \mathbf{l}_3 &= (-0.2803, 0.3190, -0.8751, 0.2320)^\top, \\ \lambda_4 &= 0.4653, & \mathbf{l}_4 &= (-0.7031, -0.6956, 0.0107, 0.1473)^\top. \end{aligned}$$

Тоді головні компоненти матимуть вигляд:

$$\begin{aligned} \mathbf{z}_1 &= 0.6411\mathbf{x}_2 - 0.6389\mathbf{x}_3 - 0.4188\mathbf{x}_4 + 0.0734\mathbf{x}_5, \\ \mathbf{z}_2 &= 0.1268\mathbf{x}_2 + 0.0786\mathbf{x}_3 + 0.2422\mathbf{x}_4 + 0.9587\mathbf{x}_5, \\ \mathbf{z}_3 &= -0.2803\mathbf{x}_2 + 0.3190\mathbf{x}_3 - 0.8751\mathbf{x}_4 + 0.2320\mathbf{x}_5, \\ \mathbf{z}_4 &= -0.7031\mathbf{x}_2 - 0.6956\mathbf{x}_3 + 0.0107\mathbf{x}_4 + 0.1473\mathbf{x}_5. \end{aligned}$$

Використовуючи критерій інформативності, в'ясимо питання про те, скільки головних компонент достатньо взяти, щоб вони найбільш інформативно пояснювали дисперсію змінних $\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3, \mathbf{z}_4$ (однаково, що і $\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5$). Практика показує, що якщо p перших головних компонент пояснюють на 80% і більше всю дисперсію головних компонент, то достатньо цих p головних компонент для подальшого дослідження об'єкта. Для цього розрахуємо критерій інформативності $I_p = \frac{\lambda_1 + \lambda_2 + \dots + \lambda_p}{N}$.

Отримали, що $I_3 = \frac{1.6722 + 1.0178 + 0.8448}{4} = 0.8837 > 0.8$. Отже, три перші головні компоненти на 88.37% пояснюють всю дисперсію головних компонент.

Далі розглянемо матрицю навантажень \mathbf{A} головних компонент на початкові озна-

ки: $\mathbf{A} = \mathbf{L}^\top \mathbf{\Lambda}^{1/2}$. Випишемо результати тільки для трьох перших головних компонент.

$$\mathbf{A} = \begin{matrix} & \mathbf{z}_1 & \mathbf{z}_2 & \mathbf{z}_3 \\ \begin{matrix} x_2 \\ x_3 \\ x_4 \\ x_5 \end{matrix} & \begin{pmatrix} \mathbf{0.8290} & 0.0740 & -0.3849 \\ -\mathbf{0.9092} & 0.1486 & 0.0098 \\ -0.3625 & 0.2341 & -\mathbf{0.8043} \\ 0.1640 & \mathbf{0.9672} & 0.2226 \end{pmatrix} \end{matrix}.$$

Отримали, що перша головна компонента досить сильно корелює зі змінними x_2 , x_3 . Друга головна компонента — з x_5 . Третя — з x_4 . Тепер досить легко дати економічне тлумачення першим трьом головним компонентам.

Частина III

ДЕЯКІ ВИДИ ЕКОНОМЕТРИЧНИХ МОДЕЛЕЙ

- ▣ Лінійні регресійні моделі зі змінною структурою
- ▣ Динамічні економетричні моделі
- ▣ Системи економетричних рівнянь

Лінійні регресійні моделі зі змінною структурою

Лінійні регресійні моделі зі змінною структурою розглядаються у випадках, коли під час збору початкових статистичних даних має місце непрямий вплив деяких **якісних чинників** (супровідних змінних), у результаті чого відбуваються стрибкоподібні зрушення у структурі зв'язків (тобто у значеннях коефіцієнтів регресії).

10.1 Проблема неоднорідних даних

Нехай функція щільності ймовірності результатного показника Y залежить не тільки від значень пояснюючих змінних X , але і від того, на яких конкретних рівнях (градаціях) зафіксовані так звані супровідні, як правило, якісні змінні Z , які визначають умови, при яких відбувається збір початкових статистичних даних. Тоді ця функція щільності тлумачиться як умовна щільність ймовірності випадкової величини Y при значеннях пояснюючих і супровідних змінних, зафіксованих відповідно на рівнях X і Z , — якщо пояснюючі і супровідні змінні випадкові за своєю суттю, — або як безумовна щільність ймовірності випадкової величини Y при значеннях не випадкових параметрів, рівних відповідно X і Z , якщо векторні змінні X і Z виконують роль не випадкових параметрів, від яких залежить закон розподілу ймовірності змінної Y .

Якщо виходити з того, що при довільних умовах (тобто при довільних значеннях супровідних змінних Z) функція регресії Y по X залишається лінійною, то природно вважати, що коефіцієнти цієї лінійної залежності, взагалі-то, залежать від Z .

Означення 10.1. Початкові статистичні дані називаються **однорідними**, якщо всі вони зафіксовані при одних і тих самих умовах (тобто при одних і тих самих значеннях супровідних змінних Z).

Означення 10.2. Якщо початкові статистичні дані зафіксовані при різних

значеннях супровідної змінної Z , то вони називаються **неоднорідними**.

Виникає проблема: яким найкращим чином використати дані для статистичного аналізу моделей. Здається, що відповідь дуже проста: розбити початкові дані на однорідні групи (тобто у кожній із груп значення супровідної змінної Z не змінюється), а потім оцінити значення коефіцієнтів по кожній із груп окремо. Тоді для кожного фіксованого значення Z буде визначена своя однорідна вибірка об'єму $n(Z) < n$ і кожній з них відповідає своя регресія. При цьому отримані по двом різним однорідним підвибіркам оцінки регресій можуть статистично значимо відрізнятися одна від одної.

На жаль, розбиття вибірки на декілька однорідних не завжди дає бажані задовільні результати:

- по-перше: супровідні змінні Z не можна спостерігати або їх значення своєчасно не були зафіксовані при зборі початкових статистичних даних. В цьому випадку використовують методи кластер-аналізу;
- по-друге: супровідні змінні Z можна спостерігати (тобто їх значення відомі), але розбиття на однорідні підвибірки призводить до дуже малих підвибірок, тобто до таких об'ємів підвибірок, яких недостатньо для статистично надійної оцінки функцій регресії.

В цьому випадку для оцінки функцій регресії використовують методи, пов'язані з введенням **фіктивних змінних**.

10.2 Фіктивні змінні у лінійній моделі регресії

У лінійну модель регресії фіктивні змінні вводяться для того, щоб відобразити вплив на показник Y супровідних якісних змінних, якщо присутні неоднорідні початкові дані.

Використання такого способу у цій ситуації обумовлене двома причинами:

- 1) статистична надійність отриманих при цьому оцінок коефіцієнтів регресії буде вищою за ту, яку б ми отримали, якщо оцінювали б ці коефіцієнти окремо по кожній однорідній вибірці;
- 2) під час побудови регресійної моделі з фіктивними змінними ми отримуємо можливість одночасно перевіряти гіпотези про наявність або відсутність статистично значущого впливу супровідних змінних на структуру моделі.

Урахування впливу супровідних змінних на структуру моделі відбувається, як правило, за допомогою адитивно-лінійного введення у праву частину регресійного рівняння деякої кількості дихотомічних (бінарних, булевих) змінних, тобто таких змінних, які можуть приймати одне з двох можливих значень (нуль або одиниця). При цьому, якщо супровідна якісна змінна z_j має k_j градацій (тобто може приймати k_j можливих значень), то для відображення її впливу на структуру регресійного

зв'язку необхідно ввести $k_j - 1$ дихотомічних змінних $z_{j,1}, \dots, z_{j,k_j-1}$. Конкретна форма, в якій ці змінні будуть представлені в рівнянні, буде залежати від наших припущень про характер впливу супровідної змінної на коефіцієнти моделі регресії.

Приклад 10.1. [1] Дослідити залежність питомого (тобто у розрахунку на одного члена родини в один і той самий момент часу) споживання прохолодних напоїв (y) від величини прибутку на підставі результатів вибірових досліджень бюджетів домашніх господарств.

Тоді векторна супровідна якісна змінна \mathbf{Z} складається з двох компонент $\mathbf{Z} = (z_1, z_2)^\top$, де z_1 — номер соціально-економічної групи, до якої належить домашнє господарство, а z_2 — номер кварталу (сезону). Будемо вважати, що родини поділяються по рівню прибутку на 3 групи: 1 — низькоприбуткові, 2 — середньоприбуткові і 3 — високоприбуткові (тобто число градацій k_1 якісної змінної z_1 дорівнює трьом). З означення z_2 маємо, що кількість її градацій (k_2) дорівнює чотирьом (1 — зима, 2 — весна, 3 — літо, 4 — осінь).

Щоб врахувати вплив змінної z_1 на структуру моделі, введемо $k_1 - 1 = 2$ фіктивні змінні:

$$z_{1,1}^i = \begin{cases} 1, & \text{якщо } i\text{-е спостереження відноситься до господарства 2-ї групи,} \\ 0 & \text{в іншому випадку;} \end{cases}$$

$$z_{1,2}^i = \begin{cases} 1, & \text{якщо } i\text{-е спостереження відноситься до господарства 3-ї групи,} \\ 0 & \text{в іншому випадку;} \end{cases}$$

Для врахування впливу сезонності на структуру моделі вводиться $k_2 - 1 = 3$ фіктивні змінні:

$$z_{2,1}^i = \begin{cases} 1, & \text{якщо } i\text{-е спостереження робилося навесні,} \\ 0 & \text{в іншому випадку;} \end{cases}$$

$$z_{2,2}^i = \begin{cases} 1, & \text{якщо } i\text{-е спостереження робилося влітку,} \\ 0 & \text{в іншому випадку;} \end{cases}$$

$$z_{2,3}^i = \begin{cases} 1, & \text{якщо } i\text{-е спостереження робилося восени,} \\ 0 & \text{в іншому випадку.} \end{cases}$$

Конкретна форма застосування цих фіктивних змінних у рівнянні залежить від припущень про характер впливу z_1 і z_2 на структуру моделі. Розглянемо два варіанти таких припущень.

Варіант 1. Чинники z_1 і z_2 впливають на кількість спожитого продукту, але не впливають на схильність до споживання (x).

Тоді рівняння регресії матиме вигляд:

$$y_i = \beta_0 + \beta_1 x^i + \beta_{01} z_{11}^i + \beta_{02} z_{12}^i + \beta_{03} z_{21}^i + \beta_{04} z_{22}^i + \beta_{05} z_{23}^i + \varepsilon_i.$$

Далі, у залежності від припущень про природу регресійних залишків ε_i використовуємо звичайний метод найменших квадратів для отримання оцінок коефіцієнтів регресії. Тепер можемо вписати вид моделі для довільних сполучень значень супровідних змінних. А саме:

$$\begin{aligned} y_i &= \beta_0 + \beta_1 x^i + \varepsilon_i \text{ — для родин 1 групи у зимовий період;} \\ y_i &= (\beta_0 + \beta_{03}) + \beta_1 x^i + \varepsilon_i \text{ — для родин 1 групи у весняний період;} \\ y_i &= (\beta_0 + \beta_{04}) + \beta_1 x^i + \varepsilon_i \text{ — для родин 1 групи у літній період;} \\ y_i &= (\beta_0 + \beta_{05}) + \beta_1 x^i + \varepsilon_i \text{ — для родин 1 групи в осінній період;} \\ y_i &= (\beta_0 + \beta_{01}) + \beta_1 x^i + \varepsilon_i \text{ — для родин 2 групи у зимовий період;} \\ y_i &= (\beta_0 + \beta_{01} + \beta_{03}) + \beta_1 x^i + \varepsilon_i \text{ — для родин 2 групи у весняний період;} \\ &\dots \\ y_i &= (\beta_0 + \beta_{02} + \beta_{05}) + \beta_1 x^i + \varepsilon_i \text{ — для родин 3 групи в осінній період.} \end{aligned}$$

Зрозуміло, що у рамках припущень варіанту 1 структурні змінні моделі при переході з однієї групи до другої і з одного сезону в інший відбуваються лише у зміні значень вільного члена регресійного рівняння і не торкаються значень параметра, що визначає нахил до споживання.

Зробимо деякі зауваження.

Зауваження 10.1 (про статистичну надійність методу оцінки, який використовує фіктивні змінні). *Надійність статистичних висновків при побудові моделі суттєво залежить від співвідношення об'єму початкових статистичних даних (n) і кількості параметрів моделі (N). Чим більше відношення $n : N$, тим точніші відповідні оцінки.*

Зауваження 10.2. *Якщо виявиться, що жодна з фіктивних змінних, що введені до моделі з метою врахування впливу супровідної змінної z_j , не впливає на Y , то зміна значень цієї супровідної змінної не спричиняє за собою неоднорідності відповідних початкових статистичних даних.*

Зауваження 10.3 (про необхідність ухилятися від «пасток», пов'язаних із введенням фіктивних змінних). *Якщо супровідна змінна z_j має k_j градацій і введемо k_j фіктивних змінних, то стовпчики матриці початкових даних X будуть лінійно залежними. Таким чином, отримаємо строгу мультиколінеарність.*

Варіант 2. Нехай чинник належності домашнього господарства до тієї чи іншої соціально-економічної групи z_1 впливає на характеристику нахилу до споживання β_1 (чинник сезонності z_2 , як і раніше, впливає лише на кількість спожитого). Тоді впроваджені вище фіктивні змінні слід представити у рівнянні у такому вигляді:

$$y_i = \beta_0 + \beta_1 x^i + \beta_{11} (z_{11}^i x^i) + \beta_{12} (z_{12}^i x^i) + \beta_{01} z_{21}^i + \beta_{02} z_{22}^i + \beta_{03} z_{23}^i + \varepsilon_i.$$

При цьому зміниться спосіб і результат комплектування загальної матриці спостережень X і відповідно результати методу найменших квадратів. У цьому випадку

за оцінки «нахилу до споживання» отримаємо:

$$\begin{aligned} \text{для групи 1: } & \beta_1; \\ \text{для групи 2: } & \beta_1 + \beta_{11}; \\ \text{для групи 3: } & \beta_1 + \beta_{12}. \end{aligned}$$

Зроблені вище зауваження відносяться також і до результатів аналізу варіанту 2.



10.3 Урахування ефекту взаємодії супровідних змінних

Раніше ми вважали, що фіктивні змінні відображали взаємозалежний вплив кожної із супровідних змінних на структуру моделі. Проте у ряді випадків може статися суттєвим вплив, викликаний взаємодією різних супровідних змінних. Цю взаємодію можна врахувати у рівнянні регресії введенням до нього додаткових фіктивних змінних. А саме: ефект взаємодії супровідної змінної z_j (представленої в рівнянні фіктивними змінними $z_{j,1}, \dots, z_{j,k_j-1}$) і супровідної змінної z_l (представленої в рівнянні фіктивними змінними $z_{l,1}, \dots, z_{l,k_l-1}$) може бути врахований введенням у рівняння додатково $P = (k_j - 1)(k_l - 1)$ фіктивних змінних z_1, z_2, \dots, z_P , що утворюються можливими попарними добутками виду $\tilde{z} = z_{j,q} z_{l,s}$, де $q = 1, 2, \dots, k_j - 1$ і $s = 1, 2, \dots, k_l - 1$. Звичайно, при статистичному аналізі отриманої таким чином лінійної множинної регресії оцінки коефіцієнтів регресії при деяких змінних — «взаємодіях» \tilde{z}_t можуть виявитися статистично незначущими, що буде означати відсутність впливу відповідних взаємодій на структуру моделі.

Приклад 10.2. [1] Нехай y — заробітна платня робітника; $\mathbf{X} = (x_1, \dots, x_p)^\top$ — набір кількісних ознак, від яких може залежати величина y ; z_1 — супровідна змінна, що визначає рівень освіти робітника ($k_1 = 3$: початкова, середня, вища) і z_2 — супровідна змінна, що визначає стать робітника ($k_2 = 2$: чоловіча, жіноча). Згідно з вищенаписаним, вводимо фіктивні змінні:

$$\begin{aligned} z_{1,1}^i &= \begin{cases} 1, & \text{якщо } i\text{-е спостереження відноситься до робітника з середньою освітою,} \\ 0 & \text{в іншому випадку;} \end{cases} \\ z_{1,2}^i &= \begin{cases} 1, & \text{якщо } i\text{-е спостереження відноситься до робітника з вищою освітою,} \\ 0 & \text{в іншому випадку;} \end{cases} \\ z_{2,1}^i &= \begin{cases} 1, & \text{якщо } i\text{-е спостереження відноситься до жінки,} \\ 0 & \text{в іншому випадку;} \end{cases} \end{aligned}$$

$$\tilde{z}_1^i = z_{1,1}^i \cdot z_{2,1}^i; \quad \tilde{z}_2^i = z_{1,2}^i \cdot z_{2,1}^i.$$

У таблиці відображені правила побудови елементів матриці спостережень \mathbf{X} у тій її частині, яка відноситься до «значень» фіктивних змінних.

Рівень освіти	Стать	Фіктивні змінні				
		$z_{1,1}^i$	$z_{1,2}^i$	$z_{2,1}^i$	\tilde{z}_1^i	\tilde{z}_2^i
початкова	чоловіча	0	0	0	0	0
початкова	жіноча	0	0	1	0	0
середня	чоловіча	1	0	0	0	0
середня	жіноча	1	0	1	1	0
вища	чоловіча	0	1	0	0	0
вища	жіноча	0	1	1	0	1

Лінійна модель регресії у цьому випадку має вигляд:

$$y_i = \beta_0 + \beta_1 x_1^i + \dots + \beta_p x_p^i + \beta_{01} z_{1,1}^i + \beta_{02} z_{1,2}^i + \beta_{03} z_{2,1}^i + \beta_{04} \tilde{z}_1^i + \beta_{05} \tilde{z}_2^i + \varepsilon_i.$$

Далі застосовуємо метод ІМНК для оцінок коефіцієнтів отриманого регресійного рівняння. Знаючи оцінки $\hat{\beta}_k$ ($k = 0, \dots, p$), $\hat{\beta}_{0j}$ ($j = 1, 2, \dots, 5$) і їх середньоквадратичні помилки, можна зробити важливі у прикладному плані висновки відносно характеру залежності. Наприклад, якщо величина $\hat{\beta}_{03}$ статистично значимо відрізняється від нуля і є від'ємною, то робимо висновок, що існуюча система оплати праці характеризується деякими дискримінаційними властивостями. Якщо ж $\hat{\beta}_{04}$ і $\hat{\beta}_{05}$ статистично не значимо відрізняються від нуля, то це означає, що взаємодія двох супровідних факторів (рівень освіти — стать) ніяк не впливає на структуру моделі.

10.4 Перевірка регресійної однорідності двох груп спостережень (тест Г. Чоу)

Нехай супровідна змінна z приймала у процесі збору регресійних спостережень

$$\mathbf{A} = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$$

всього два значення: z_1 і z_2 . Тоді загальна вибірка \mathbf{A} може бути поділена на дві підвибірки

$$\begin{aligned} \mathbf{A}_1 &= \{(x_1, y_1), \dots, (x_{n_1}, y_{n_1}) \mid z = z_1\}, \\ \mathbf{A}_2 &= \{(x_1, y_1), \dots, (x_{n_2}, y_{n_2}) \mid z = z_2\} \end{aligned}$$

таким чином, що всі спостереження у кожній із підвбірок були зроблені при одному і тому самому значенні супровідної змінної z : перша підвбірка при $z = z_1$, а друга —

при $z = z_2$.

Нам необхідно отримати відповідь на питання: чи дійсно, що підвибірки \mathbf{A}_1 і \mathbf{A}_2 неоднорідні у регресійному сенсі, чи перехід від градації z_1 до градації z_2 в умовах збору початкових статистичних даних ніяк не впливає на структуру лінійної моделі регресії y по x і, отже, підвибірки \mathbf{A}_1 і \mathbf{A}_2 можна об'єднати і будувати функцію регресії по об'єднаній вибірці \mathbf{A} ? Тобто необхідно статистично перевірити гіпотезу

$$H_0: \beta^1 = \beta^2, \quad \sigma_{\varepsilon_1}^2 = \sigma_{\varepsilon_2}^2 = \sigma^2,$$

де $\beta^j = (\beta_1^j, \dots, \beta_p^j)$ і ε_j відповідно коефіцієнти і випадкові залишки регресій, побудованих на основі підвбірок \mathbf{A}_1 і \mathbf{A}_2 .

Якщо $n_1 > p$ і $n_2 > p$, то можна запропонувати декілька можливостей перевірки гіпотези H_0 (наприклад, для прийняття гіпотези H_0 достатньо перевірити той факт, що точкові оцінки коефіцієнтів, отриманих по першій підвбірці, попадають всередину інтервальних оцінок коефіцієнтів регресії, що побудована по другій підвбірці).

Якщо ж об'єм однієї з підвбірок (наприклад, другої) не дозволяє зробити надійну оцінку невідомих коефіцієнтів, то необхідно користуватися тестом Г. Чоу, який має такий алгоритм.

Алгоритм 10.1 (тест Г. Чоу).

- 1) За вибірками \mathbf{A}_1 і \mathbf{A} будуюмо дві моделі і розраховуємо похибки ε_1 і ε .
- 2) Розраховуємо

$$F_{CT} = \frac{(\varepsilon^\top \varepsilon - \varepsilon_1^\top \varepsilon_1) (n_1 - p)}{\varepsilon_1^\top \varepsilon_1 n_2}.$$

- 3) Якщо $F_{CT} > F(\alpha, n_2, n_1 - p)$, то гіпотеза H_0 про однорідність вибірок відхиляється з ймовірністю $(1 - \alpha) 100\%$.

Зауваження 10.4. Якщо об'єм другої вибірки n_2 досить великий, то можна розраховувати таку статистику:

$$F_{CT} = \frac{(\varepsilon^\top \varepsilon - \varepsilon_1^\top \varepsilon_1 - \varepsilon_2^\top \varepsilon_2) (n_1 + n_2 - 2p)}{(\varepsilon_1^\top \varepsilon_1 + \varepsilon_2^\top \varepsilon_2) p}.$$

Якщо $F_{CT} > F(\alpha, p, n_1 + n_2 - 2p)$, то з ймовірністю $(1 - \alpha) 100\%$ гіпотеза H_0 про однорідність відхиляється.

Динамічні економетричні моделі

Економетрична модель є динамічною, якщо у даний момент часу t вона враховує значення змінних, які відносяться як до поточного, так і до попередніх моментів часу.

Можна виділити два основних типи динамічних економетричних моделей. До моделей першого типу відносяться моделі авторегресії і моделі з розподіленим лагом. Моделі другого типу враховують динамічну інформацію у неявному вигляді. До цих моделей включені змінні, що характеризують очікуваний або бажаний рівень результату y або одного з факторів у момент часу t . Цей рівень вважається невідомим і визначається економічними одиницями з урахуванням інформації, яку вони мають у момент $(t - 1)$.

У залежності від способу визначення очікуваних значень показників розрізняють моделі неповного коригування і адаптивних очікувань. Оцінка параметрів цих моделей зводиться до оцінки параметрів моделей авторегресії.

При дослідженні економічних процесів нерідко доводиться моделювати ситуації, коли значення результативної ознаки у поточний момент часу t формується під впливом ряду чинників, що діяли у попередні моменти часу $t - 1, t - 2, \dots, t - l$. Величину l , що характеризує запізнення у впливі чинника на результат, називають в економетрії *лагом*, а часові ряди самих змінних, що зсунуті на один або декілька моментів часу — *лаговими змінними*.

Економетричне моделювання охарактеризованих вище процесів відбувається із застосуванням моделей, що включають не тільки поточні, але і лагові значення незалежних змінних. Ці моделі називаються моделями з розподіленим лагом. Вони мають вигляд

$$y_t = a + b_0 x_t + b_1 x_{t-1} + \dots + b_k x_{t-k} + \varepsilon_t. \quad (11.1)$$

До моделі може включатися нескінченна кількість лагових змінних.

Разом з лаговими значеннями незалежних змінних на величину залежної змінної

поточного періоду можуть мати вплив її значення у попередні моменти або періоди часу. Ці процеси описують за допомогою моделей регресії, що мають лагові значення залежної змінної і мають вигляд

$$y_t = a + b_0 x_t + c_1 y_{t-1} + \dots + c_k y_{t-k} + \varepsilon_t \quad (11.2)$$

і називаються моделями авторегресії.

Побудова моделей з розподіленими лагами і моделей авторегресії має свою специфіку. По-перше, оцінку параметрів моделей авторегресії і, у більшості випадків, моделей з розподіленим лагом не можна зробити за допомогою МНК, зважаючи на порушення його передумов, і тому необхідно застосовувати інші методи. По-друге, дослідникам доводиться розв'язувати проблеми вибору оптимальної величини лагу. І по-третє, між моделями з розподіленим лагом і моделями авторегресії існує певна взаємозалежність, і у деяких випадках необхідно здійснювати перехід від одного типу моделі до іншого.

11.1 Інтерпретація параметрів моделей з розподіленим лагом

Розглянемо модель з розподіленим лагом у загальному вигляді, вважаючи, що максимальна величина лагу скінченна

$$y_t = a + b_0 x_t + b_1 x_{t-1} + \dots + b_p x_{t-p} + \varepsilon_t. \quad (11.3)$$

Ця модель стверджує про те, що якщо у деякий момент часу t відбувається зміна незалежної змінної x , то ця зміна буде впливати на значення змінної y на протязі p наступних моментів часу.

Коефіцієнт регресії b_0 при змінній x_t характеризує середню абсолютну зміну y_t при зміні x_t на 1 од. свого виміру у деякий фіксований момент часу t без врахування впливу лагових значень чинника x . Цей коефіцієнт називають *короткостроковим мультиплікатором*.

У момент часу $(t + 1)$ сукупний вплив змінної x_t на результат y_t складає $(b_0 + b_1)$ ум. од., у момент $(t + 2)$ цей вплив можна охарактеризувати сумою $(b_0 + b_1 + b_2)$ і т. д. Отримані таким чином суми називають *проміжними мультиплікаторами*.

З урахуванням скінченної величини лагу можна сказати, що зміна змінної x у момент t на 1 од. призводить до загальної зміни y через p моментів на $(b_0 + b_1 + \dots + b_p)$ одиниць.

Введемо позначення:

$$b = b_0 + b_1 + \dots + b_p. \quad (11.4)$$

Величину b називають *довгостроковим мультиплікатором*. Він показує абсолютну зміну у довгостроковому періоді $(t + p)$ результату y під впливом зміни на 1 од.

чинника x .

Нехай

$$\beta_j = \frac{b_j}{b}, \quad j = \overline{0, p}. \quad (11.5)$$

Отримані коефіцієнти називають стандартизованими. Якщо всі коефіцієнти b_j мають однакові знаки, то для довільного j : $0 < \beta_j < 1$ і $\sum_{j=0}^p \beta_j = 1$. У цьому випадку стандартизовані коефіцієнти β_j є вагами для відповідних коефіцієнтів b_j . Кожен з них вимірює частку загальної зміни результативної ознаки у момент часу $(t + j)$.

Знаючи величини β_j , можна визначити ще дві важливі характеристики моделі множинної регресії: величину середнього лагу і медіанного лагу.

Середній лаг визначається за формулою $l = \sum_{j=0}^p j \beta_j$ і визначає середній період, протягом якого буде відбуватися зміна результату під впливом чинника у момент часу t . Невелика величина середнього лагу свідчить про відносно швидке реагування результату y на зміну чинника, тоді як велике його значення свідчить про те, що вплив чинника на результат буде відчуватися на протязі довгого періоду часу.

Медіанний лаг — це величина лагу, для якого $\sum_{j=0}^k \beta_j = 0.5$. Це той період часу, на протязі якого з моменту часу t буде реалізована половина загального впливу чинника на результат.

11.1.1 Приклад

Приклад 11.1. За результатами вивчення залежності об'ємів продажу компанії у середньому за місяць від витрат на рекламу була отримана така модель з розподіленим лагом:

$$\hat{y}_t = 7 + 4x_t + 2x_{t-1} + 1.5x_{t-2} + 0.5x_{t-3}.$$

У цій моделі короткостроковий мультиплікатор дорівнює 4. Це означає, що збільшення витрат на рекламу на 1 млн грн. веде у середньому до зросту об'єму продажу компанії на 4 млн грн. у тому ж періоді. Під впливом витрат на рекламу продаж компанії у період $(t + 1)$ зростає на $4 + 2 = 6$ млн грн., $(t + 2)$ — на $6 + 1.5 = 7.5$ млн грн. Довгостроковий мультиплікатор для моделі дорівнює: $4 + 2 + 1.5 + 0.5 = 8$. Тобто у довгостроковій перспективі (через три місяці) збільшення витрат на рекламу на 1 млн грн. у поточний час призведе до загального росту об'єму продаж на 8 млн грн.

Стандартизовані коефіцієнти регресії дорівнюють:

$$\beta_0 = \frac{4}{8} = 0.5, \quad \beta_1 = \frac{2}{8} = 0.25, \quad \beta_2 = \frac{1.5}{8} = 0.1875, \quad \beta_3 = \frac{0.5}{8} = 0.0625.$$

Отже, 50% загального збільшення об'єму продаж, викликаних ростом витрат на рекламу, відбувається у поточному часі; 25% — у момент $(t + 1)$; 18.75% — у момент

$(t + 2)$ і 6.25 % цього збільшення відбувається на момент часу $(t + 3)$.

Визначимо середній лаг:

$$l = 0 \cdot 0.5 + 1 \cdot 0.25 + 2 \cdot 0.1875 + 3 \cdot 0.0625 = 0.8125 \text{ місяця.}$$

Величина $l = 0.8125$ показує, що більша частина ефекту зросту витрат на рекламу проявляється за перший місяць. Медіанний лаг складає один місяць.

Викладені прийоми аналізу параметрів моделі з розподіленим лагом дійсні тільки при умові, що всі коефіцієнти при поточному і лагових значеннях чинника мають однаковий знак. Ця вимога виправдана з економічної точки зору: вплив одного і того самого чинника на результат повинен бути однонаправленим незалежно від того, з яким часовим лагом вимірюється сила або тіснота зв'язку між цими ознаками. Однак, на практиці отримати статистично значущу модель, параметри якої мали б однакові знаки, особливо при великому значенні лагу p , дуже складно.

Застосування звичайного ІМНК до таких моделей у більшості випадків складно за таких причин:

- 1) поточні і лагові значення незалежної змінної, як правило, тісно пов'язані одне з одним. Тому оцінка параметрів моделі відбувається при досить високій мультиколінеарності регресорів;
- 2) при великому значенні лагу зменшується число ступенів вільності;
- 3) у моделях з розподіленим лагом досить часто виникає проблема автокореляції залишків.

Вищевказані обставини призводять до зниження точності при оцінці параметрів, їх неефективності і т. д. Тому на практиці параметри моделей з розподіленим лагом визначають при умові певних обмежень на коефіцієнти регресії і в умовах обраної структури лагу.

11.2 Послідовна оцінка параметрів моделі з розподіленим лагом

Оскільки припускається, що x_t нестохастичні (або принаймні не корелюють з похибкою), то x_{t-1} , x_{t-2} , ... теж нестохастичні. Тоді до моделі

$$y_t = a + b_0 x_t + b_1 x_{t-1} + \dots + \varepsilon_t$$

можна застосувати метод ІМНК. Для цього діємо послідовно: тобто спочатку необхідно побудувати регресію y_t за x_t та оцінити невідомі параметри, потім — регресію y_t за x_t і x_{t-1} , потім — за x_t , x_{t-1} і x_{t-2} і т. д. Ця послідовна процедура припиняється, коли параметри при лагових змінних x починають бути статистично незначущими або коефіцієнт хоча б однієї змінної змінює свій знак.

Приклад 11.2. Досліджується залежність споживання напою y від надходження нових замовлень x . Отримано такі результати:

$$\hat{y}_t = 9.0 + 0.237 x_t;$$

$$\hat{y}_t = 9.0 + 0.215 x_t + 0.071 x_{t-1};$$

$$\hat{y}_t = 9.0 + 0.213 x_t + 0.081 x_{t-1} - 0.04 x_{t-2};$$

$$\hat{y}_t = 9.0 + 0.212 x_t + 0.070 x_{t-1} + 0.03 x_{t-2} - 0.03 x_{t-3}.$$

Обираємо друге рівняння, тому що у останніх двох знак при x_{t-2} змінюється.

Метод послідовних оцінок має такі недоліки:

- 1) невідома максимальна величина лагу;
- 2) при оцінці послідовних лагів залишається все менше ступенів вільності, що робить економічні висновки непевними;
- 3) в економічних даних послідовні значення змінних мають високу кореляцію і тому з'являється проблема мультиколінеарності.

11.3 Підхід Койка до моделей з розподіленим лагом

Койк запропонував свій метод оцінки параметрів моделей з розподіленим лагом. Нехай розглядається модель

$$y_t = a + b_0 x_t + b_1 x_{t-1} + \dots + u_t. \quad (11.6)$$

Будемо вважати, що b_i мають однаковий додатний знак. Койк припустив, що вони змінюються у геометричній прогресії: $b_k = b_0 \lambda^k$, $k = 0, 1, \dots$, де λ — темп зменшення лагу, $0 < \lambda < 1$, $(1 - \lambda)$ — швидкість пристосування. Отже, з кожним наступним кроком у минуле вплив лагу на змінну y зменшується, що є досить слушним. Чим ближче λ до 1, тим повільніший темп зменшення b_k , а чим ближче воно до 0, тим швидше спадає b_k . Підхід Койка має такі переваги: параметр λ може бути і від'ємним і тоді b_i будуть різних знаків; так як $\lambda < 1$, то віддалені за часом значення x стали менш впливовими; довгостроковий мультиплікатор є скінченною величиною:

$\sum_{k=0}^{\infty} b_k = \frac{b_0}{1 - \lambda}$. Враховуючи припущення Койка, модель (11.6) матиме вигляд:

$$y_t = a + \beta_0 x_t + \beta_0 \lambda x_{t-1} + \beta_0 \lambda^2 x_{t-2} + \dots + \varepsilon_t. \quad (11.7)$$

У рівнянні (11.7) маємо всього три параметри a , β_0 , λ . Для їх оцінки не можна застосовувати МНК, тому що виникла б проблема мультиколінеарності і з отриманих оцінок не вдалося б однозначно отримати оцінки параметрів.

Однак можна уникнути цих серйозних проблем, якщо використати нелінійний метод найменших квадратів (НМНК).

Алгоритм методу починається із задання меж можливих значень λ , наприклад, $\lambda \in (0, 1)$ і розглядаємо всі можливі значення λ всередині цих меж з достатньо малим кроком. Чим менший крок, тим кращий ми отримуємо результат. Для кожного значення λ розраховуємо

$$z_t = x_t + \lambda x_{t-1} + \lambda^2 x_{t-2} + \dots + \lambda^p x_{t-p} \quad (11.8)$$

з таким значенням p , при якому подальші лагові значення x не впливають на зміну z_t . Потім оцінюється рівняння регресії

$$y_t = a + \beta_0 z_t + \varepsilon_t. \quad (11.9)$$

Ці розрахунки робимо для всіх значень λ і обираємо таке значення λ , що забезпечує найбільший коефіцієнт детермінації R^2 при оцінці рівняння (11.9). У якості оцінок a і β_0 обирають їх оцінки у «найкращому» рівнянні (11.9). Рівняння (11.8) і (11.9) в сукупності еквівалентні рівнянню (11.7).

Койк запропонував інший підхід. Рівняння (11.7) виконується для довільних t , а отже воно справедливе для періоду $(t - 1)$:

$$y_{t-1} = a + \beta_0 x_{t-1} + \beta_0 \lambda x_{t-2} + \dots + \varepsilon_{t-1}. \quad (11.10)$$

Домножимо обидві частини цього рівняння на λ і віднімемо від (11.7); отримаємо:

$$y_t - \lambda y_{t-1} = a(1 - \lambda) + \beta_0 x_t + \varepsilon_t - \lambda \varepsilon_{t-1}, \quad \text{або} \\ y_t = a(1 - \lambda) + \beta_0 x_t + \lambda y_{t-1} + \varepsilon_t - \lambda \varepsilon_{t-1}. \quad (11.11)$$

Модель (11.11) називається *моделлю Койка*.

Отримана модель є авторегресійною. Визначивши її параметри, ми знайдемо оцінки a , β_0 і λ початкової моделі. Зазначимо, що застосування звичайного 1МНК для оцінки параметрів моделі (11.11) дає зміщені оцінки, так як у цій моделі у правій частині присутнє y_{t-1} , а воно корелює з ε_{t-1} і передумови 1МНК не виконуються.

Форма моделі (11.11) дозволяє аналізувати коротко- і довгострокові динамічні властивості моделі. У короткостроковому аспекті (у поточному періоді) значення y_{t-1} потрібно розглядати як фіксоване, і вплив x_t на y_t відображається коефіцієнтом β_0 . У довгостроковому періоді, якщо x_t прямує до деякого свого рівноважного значення \bar{x} , то y_t і y_{t-1} також будуть прямувати до рівноважного значення \bar{y} , тобто

$$\bar{y} = a(1 - \lambda) + \beta_0 \bar{x} + \lambda \bar{y},$$

звідки $\bar{y} = a + \frac{\beta_0}{1 - \lambda} \bar{x}$. Отже, довгостроковий вплив x на y виражається коефіцієнтом $\frac{\beta_0}{1 - \lambda}$. Якщо $0 < \lambda < 1$, то цей коефіцієнт більший за β_0 , тобто довгостроковий вплив сильніший короткострокового.

Незважаючи на нескінченну кількість лагових змінних у моделі (11.7), геометри-

чна структура лагу дозволяє визначити середній і медіанний лаги у моделі Койка. Отримаємо, що середній лаг у моделі Койка дорівнює $l = \frac{\lambda}{1 - \lambda}$, а медіанний лаг — $\frac{\ln 0.5}{\ln \lambda}$.

11.4 Модель часткових пристосувань (коригувань)

Розглянемо одну з модифікацій моделі Койка. Зазначимо, що модель Койка отримана чисто алгебраїчним шляхом і не має під собою ніякого теоретичного підґрунтя.

У моделі часткових пристосувань передбачається, що рівняння поведінки визначає не фактичне значення залежної змінної y_t , а її бажаний рівень y_t^* :

$$y_t^* = a + \beta x_t + \varepsilon_t. \quad (11.12)$$

Вважаємо також, що фактичний приріст залежної змінної $y_t - y_{t-1}$ пропорційний різниці між її бажаним рівнем і значенням у попередній період, тобто $y_t^* - y_{t-1}$:

$$y_t - y_{t-1} = \lambda (y_t^* - y_{t-1}), \quad (0 \leq \lambda \leq 1), \quad (11.13)$$

λ відоме під назвою коефіцієнт пристосування, $y_t^* - y_{t-1}$ — бажана зміна, $y_t - y_{t-1}$ — фактична зміна. Вираз (11.13) можна переписати так:

$$y_t = \lambda y_t^* + (1 - \lambda) y_{t-1}, \quad (11.14)$$

звідки видно, що y_t є зважене середнє бажаного рівня і фактичного значення цієї змінної у попередній період. Чим більше значення λ , тим швидше відбувається процес пристосувань. Якщо $\lambda = 1$, то $y_t = y_t^*$ і повне коригування відбувається за один період. Якщо ж $\lambda = 0$, то коригування взагалі не відбувається.

Підкладемо вираз (11.12) у (11.14) і отримаємо:

$$y_t = a \lambda + \beta \lambda x_t + (1 - \lambda) y_{t-1} + \lambda \varepsilon_t. \quad (11.15)$$

Параметри a , β , λ моделей поведінки (11.12) і (11.13) можна оцінити з рівняння (11.15). Коефіцієнт при y_{t-1} дає оцінку $(1 - \lambda)$, а отже, і λ , коефіцієнт при x_t , поділений на оцінку λ , дає оцінку β . Постійний член, поділений на оцінку λ , дає оцінку a .

Ця модель включає стохастичну змінну y_{t-1} . На відміну від моделі Койка, ця змінна не корелює з залишком ε_t . При таких умовах ІМНК дозволяє отримувати асимптотично незміщені і ефективні оцінки; однак оцінки не матимуть таких властивостей при малих вибірках.

Хоча модель часткового коригування на перший погляд і не відноситься до моделі Койка, покажемо, що це не так. Рівняння (11.15) виконується для довільних

t , покладемо $t := t - 1$.

$$y_{t-1} = a\lambda + \beta\lambda x_{t-1} + (1 - \lambda)y_{t-2} + \lambda\varepsilon_{t-1}. \quad (11.16)$$

Підкладемо вираз (11.16) у (11.15), отримаємо

$$y_t = a\lambda(1 + (1 - \lambda)) + \beta\lambda x_t + (1 - \lambda)\beta\lambda x_{t-1} + (1 - \lambda)^2 y_{t-2} \quad (11.17)$$

(випадковий член для простоти не пишемо). Тепер у (11.15) покладемо $t := t - 2$ і підкладемо у (11.17) і т. д.

Отримавши вираз для y_t через поточні і лагові значення x з геометрично зменшувальними вагами у вигляді моделі Койка, замінимо $(1 - \lambda)$ на δ , а $\beta\lambda$ на b і отримаємо:

$$y_t = a + b(x_t + \delta x_{t-1} + \delta^2 x_{t-2} + \dots), \quad (11.18)$$

що по формі співпадає з рівнянням Койка.

11.4.1 Приклад. Модель коригування величини дивідендів (модель Лінтнера)

Лінтнер [3] використав модель часткового коригування у дослідженнях розподілу дивідендів. Звичайно виробничі компанії розподіляють прибуток, що залишився після виплати податків, частково на виплату дивідендів акціонерам, а залишки направляють на фінансування інвестицій. Коли прибуток росте, то і дивіденди теж зростають, але не у такій пропорції. Причиною цього в основному є обережність керівництва компаній. Збільшення прибутку може бути тимчасовим, і якщо дивіденди будуть збільшуватися досить швидко, то можливо пізніше їх доведеться зменшувати. Керівництво компанії вважає, що ніщо не наносить такий сильний удар по репутації фірми, як скорочення дивідендів, і тому воно проявляє обережність. Другим доказом проти негайного збільшення дивідендів у тій же пропорції, що і збільшення прибутку, є міркування про те, що збільшення прибутку може свідчити про покращення інвестиційних можливостей, що вимагають фінансування.

Лінтнер вважав, що у фірм існує цільова довгострокова доля виплат γ і що бажаний об'єм дивідендів D_t^* співвідноситься з поточним прибутком Π_t як

$$D_t^* = \gamma \Pi_t. \quad (11.19)$$

Однак реальний об'єм дивідендів підпорядковується процесові часткового коригування:

$$\Delta D_t = \lambda(D_t^* - D_{t-1}) + \varepsilon_t. \quad (11.20)$$

Тоді

$$\begin{aligned} D_t - D_{t-1} &= \lambda (D_t^* - D_{t-1}) + \varepsilon_t, \\ D_t - D_{t-1} &= \lambda (\gamma \Pi_t - D_{t-1}) + \varepsilon_t, \\ D_t &= \lambda \gamma \Pi_t + (1 - \lambda) D_{t-1} + \varepsilon_t. \end{aligned} \quad (11.21)$$

Використовуючи дані про діяльність корпоративного сектору США за період 1918–1941 рр., Лінтнер побудував таке рівняння регресії

$$\widehat{D}_t = 352.3 + 0.15 \Pi_t + 0.70 D_{t-1}.$$

Так як $1 - \lambda = 0.70$, то $\lambda = 0.3$ — коефіцієнт швидкості коригування, а $\gamma\lambda = 0.15$ і тому отримуємо, що доля виплат дивідендів $\gamma = 0.15/0.3 = 0.5$.

11.5 Модель адаптивних очікувань

Моделювання очікувань часто стає найбільш відповідальною і складною задачею у прикладній економіці. Це особливо справедливо для макроекономіки, де інвестиції, збереження і попит на активи виявляється дуже чутливим до очікувань відносно майбутнього. На жаль, у теперішній час відсутні задовільні методи виміру очікувань для розв'язання макроекономічних задач. Як наслідок, макроекономічні моделі не дозволяють отримувати достатньо точні прогнози, що утруднює керування економікою.

У якості напівміри розв'язання проблем у деяких моделях використовується метод, відомий під назвою «процес адаптивних очікувань».

Розглянемо модель вигляду

$$y_t = a + b x_{t+1}^* + \varepsilon_t, \quad (11.22)$$

де y_t — фактичне значення результативної ознаки, x_{t+1}^* — очікуване значення чинника.

Механізм формування очікувань у цій моделі такий:

$$x_{t+1}^* - x_t^* = \alpha (x_t - x_t^*) \quad (11.23)$$

або

$$x_{t+1}^* = \alpha x_t + (1 - \alpha) x_t^*, \quad 0 < \alpha < 1. \quad (11.24)$$

Таким чином, очікуване значення змінної x_{t+1}^* є середня арифметична зважена її фактичного і очікуваного значень у попередній період. Іншими словами, у кожний період $t + 1$ очікування коригуються на деяку долю α різниці між фактичним значенням ознаки і її очікуваним значенням у попередній період. Параметр α у цій моделі називають коефіцієнтом очікувань. Чим ближче коефіцієнт очікування до 1, тим у більшій мірі реалізується очікування економічних агентів. І, навпаки, наближення величини α до нуля свідчить про стійкість існуючих тенденцій. При $\alpha = 0$

з (11.23) або (11.24) отримаємо, що $x_{t+1}^* = x_t^*$, тобто умови, що домінують сьогодні, зберуться і на всі майбутні періоди часу. Очікувані майбутні значення показників співпадають з їх значеннями поточних періодів.

Підкладемо у (11.22) замість x_{t+1}^* співвідношення (11.24):

$$y_t = a + b(\alpha x_t + (1 - \alpha) x_t^*) + \varepsilon_t = a + \alpha b x_t + (1 - \alpha) b x_t^* + \varepsilon_t. \quad (11.25)$$

Модель (11.22) справедлива і для періоду $t - 1$. Отримаємо

$$y_{t-1} = a + b x_t^* + \varepsilon_{t-1}. \quad (11.26)$$

Помножимо (11.26) на $(1 - \alpha)$ і віднімемо від (11.25):

$$y_t - (1 - \alpha) y_{t-1} = a - (1 - \alpha) a + \alpha b x_t + \varepsilon_t - (1 - \alpha) \varepsilon_{t-1}$$

або

$$y_t = \alpha a + \alpha b x_t + (1 - \alpha) y_{t-1} + u_t, \quad u_t = \varepsilon_t - (1 - \alpha) \varepsilon_{t-1}. \quad (11.27)$$

Зазначимо різницю між (11.22) і (11.27). У першій моделі b оцінює середню зміну y у відповідь на одиничну зміну x^* . У (11.27) b оцінює середню зміну y у відповідь на одиничну зміну фактичного, тобто спостережуваного значення x . На практиці спочатку оцінюють коефіцієнти моделі (11.27), а потім уже параметри моделі (11.22).

Модель адаптивних очікувань і модель Койка схожі між собою, хоча у них різні інтерпретації коефіцієнтів. Виникає питання: наскільки модель адаптивних очікувань є реалістичною? Вона забезпечує доволі прості способи моделювання очікувань в економічній теорії.

Гіпотеза про те, що люди вчаться на попередньому досвіді, є набагато розумнішою, ніж припущення про те, що вони позбавлені пам'яті.

Співвідношення (11.24) справедливе для довільних t , отже, буде справедливим і при $t := t - 1$. Отримаємо:

$$x_t^* = \alpha x_{t-1} + (1 - \alpha) x_{t-1}^*. \quad (11.28)$$

У (11.28) присутня величина x_{t-1}^* . Замінімо її з (11.24) при $t := t - 2$. Повторюючи цю процедуру нескінченну кількість разів, отримаємо

$$x_{t+1}^* = \alpha \left[x_t + (1 - \alpha) x_{t-1} + (1 - \alpha)^2 x_{t-2} + \dots \right].$$

Підкладемо отриманий вираз у (11.22) і замінімо $(1 - \alpha)$ на δ , отримаємо

$$y_t = a + b \alpha \left[x_t + \delta x_{t-1} + \delta^2 x_{t-2} + \dots \right] + u_t,$$

а це і є модель Койка.

11.6 Оцінка параметрів моделі авторегресії

Описані вище перетворення Койка, модель адаптивних очікувань і модель часткового коригування зводяться до моделі авторегресії

$$y_t = a + b x_t + c y_{t-1} + \varepsilon_t. \quad (11.29)$$

Однак при побудові моделей авторегресії виникають дві серйозні проблеми.

Перша проблема пов'язана з вибором методу оцінки параметрів рівняння авторегресії. Наявність лагових стохастичних значень результуючої ознаки у правій частині рівняння приводить до порушення передумови 1МНК про поділ змінних на результативну (стохастичну) і нестохастичні чинники.

Друга проблема полягає у тому, що присутній зв'язок між чинником правої частини і залишком. Тому застосування методу 1МНК для оцінки параметрів моделі призводить до отримання зміщеної оцінки параметра при y_{t-1} .

Одним із можливих методів розрахунку параметрів рівняння авторегресії є метод інструментальних змінних. Суть цього методу полягає у тому, щоб замінити змінну правої частини моделі, для якої порушується передумова 1МНК, новою змінною, включення якої до моделі не порушує передумов 1МНК. Інструментальна змінна повинна мати дві властивості: вона тісно корелює з y_{t-1} і не корелює із залишками.

Існує декілька способів отримання такої інструментальної змінної. Так як в моделі (11.29) змінна y_t залежить не тільки від y_{t-1} , але і від x_t , то можна вважати, що існує залежність між y_{t-1} і x_{t-1} , тобто

$$y_{t-1} = d_0 + d_1 x_{t-1} + u_t. \quad (11.30)$$

Таким чином, змінну y_{t-1} можна виразити так:

$$y_{t-1} = \hat{y}_{t-1} + u_t, \quad (11.31)$$

де

$$\hat{y}_{t-1} = \hat{d}_0 + \hat{d}_1 x_{t-1}. \quad (11.32)$$

Знайдена за допомогою (11.32) оцінка \hat{y}_{t-1} може бути інструментальною змінною для чинника y_{t-1} . Ця змінна, по-перше, тісно корелює з y_{t-1} (див. (11.31)), по-друге, вона пов'язана лінійно з x_{t-1} , яке не корелює із залишками і тому вона теж не корелює з ними, тобто з u_t .

Таким чином, оцінки параметрів рівняння (11.29) можна знайти із співвідношення

$$y_t = a + b x_t + c \hat{y}_{t-1} + \nu_t, \quad (11.33)$$

попередньо оцінивши \hat{y}_{t-1} .

Можна використовувати модифікацію цього методу. Підкладемо у модель (11.29)

замість y_{t-1} його вираз з рівняння (11.30):

$$y_t = a + b x_t + c (d_0 + d_1 x_{t-1} + u_t) + \varepsilon_t$$

або

$$y_t = a + c d_0 + b x_t + c d_1 x_{t-1} + (c u_t + \varepsilon_t). \quad (11.34)$$

Рівняння (11.34) є моделлю з розподіленим лагом, для якої не порушуються передумови 1МНК. Оцінивши параметри моделей (11.30) і (11.34), можна розрахувати параметри початкової моделі (11.29). Модель (11.34) демонструє важливу властивість викладеного вище методу інструментальних змінних для оцінки параметрів моделей авторегресії: метод дозволяє перейти від моделі авторегресії до моделі з розподіленими лагами.

11.7 Лаги Алмона

Хоча модель Койка широко використовується на практиці, вона базується на припущенні, що коефіцієнти спадають у геометричній прогресії у міру зростання довжини лагу. Це припущення може бути занадто строгим у деяких ситуаціях, і схема моделей Койка не спрацьовує. У складніших випадках параметри β_i моделі можна виразити як функцію від i , тривалості лагу, і підібрати відповідні криві, які відображатимуть цю функціональну залежність. Саме цей підхід і запропонував Алмон.

Розглянемо загальну модель з розподіленим лагом з максимальною величиною лагу p , яка описується співвідношенням

$$y_t = a + b_0 x_t + b_1 x_{t-1} + \dots + b_p x_{t-p} + \varepsilon_t. \quad (11.35)$$

Нехай було встановлено, що у досліджуваній моделі має місце поліноміальна структура лагу, тобто залежність коефіцієнтів моделі b_j від величини лагу описується поліномом k -го ступеня:

$$b_j = c_0 + c_1 j + c_2 j^2 + \dots + c_k j^k. \quad (11.36)$$

Тоді кожний з коефіцієнтів b_j моделі (11.35) можна виразити таким чином:

$$\begin{aligned} b_0 &= c_0, \\ b_1 &= c_0 + c_1 + \dots + c_k, \\ b_2 &= c_0 + 2 c_1 + 4 c_2 \dots + 2^k c_k, \\ &\dots \\ b_p &= c_0 + p c_1 + p^2 c_2 \dots + p^k c_k. \end{aligned} \quad (11.37)$$

Підкладемо у (11.35) співвідношення для b_j і отримаємо

$$y_t = a + c_0 x_t + (c_0 + c_1 + \dots + c_k) x_{t-1} + \dots + (c_0 + p c_1 + p^2 c_2 + \dots + p^k c_k) x_{t-p} + \varepsilon_t. \quad (11.38)$$

Перегрупуємо доданки у (11.38):

$$y_t = a + c_0 (x_t + x_{t-1} + \dots + x_{t-p}) + c_1 (x_{t-1} + 2 x_{t-2} + \dots + p x_{t-p}) + \dots + c_k (x_{t-1} + 2^k x_{t-2} + \dots + p^k x_{t-p}) + \varepsilon_t. \quad (11.39)$$

Позначимо доданки при c_i як нові змінні:

$$\begin{aligned} z_0 &= x_t + x_{t-1} + \dots + x_{t-p}, \\ z_1 &= x_{t-1} + 2 x_{t-2} + \dots + p x_{t-p}, \\ &\dots \\ z_k &= x_{t-1} + 2^k x_{t-2} + \dots + p^k x_{t-p}. \end{aligned} \quad (11.40)$$

Перепишемо модель (11.38) з урахуванням (11.40):

$$y_t = a + c_0 z_0 + c_1 z_1 + \dots + c_k z_k + \varepsilon_t. \quad (11.41)$$

Алгоритм застосування методу Алмона для розрахунку параметрів моделі з розподіленим лагом виглядає так:

Алгоритм 11.1 (метод Алмона).

- 1) визначається максимальна величина лагу p ;
- 2) визначається ступінь поліному k , що описує структуру лагу;
- 3) використовуючи (11.40), розраховуємо змінні z_0, \dots, z_k ;
- 4) методом 1МНК визначаємо параметри моделі (11.41);
- 5) за допомогою (11.37) розраховуємо параметри моделі (11.35).

Застосування методу Алмона спряжене з рядом проблем.

По-перше, величина лагу p повинна бути відомою. При її визначенні краще використовувати максимально можливий лаг, ніж обмежитись лагами невеликої довжини. Якщо обрати менший лаг, ніж його реальне значення, то у моделі не буде враховано впливовий чинник. Вплив цього чинника у моделі буде виражено в залишках. Тоді у моделі не буде виконуватися передумова 1МНК про випадковість залишків, а отримані оцінки її параметрів будуть неефективними і зміщеними. Вибір більшої величини лагу порівняно з її реальним значенням буде означати включення до моделі статистично незначущого чинника і зниження ефективності оцінок, однак ці оцінки все ж таки будуть незміщеними.

Існує декілька практичних підходів для визначення реальної величини лагу, наприклад, побудова декількох регресій і вибір найкращої з них. Однак, краще за все скористатися кореляційною функцією. Крім того, оптимальну величину лагу

можна наближено визначити на основі апіорної інформації, економічної теорії або проведення емпіричних досліджень.

По-друге, необхідно встановити ступінь поліному k . У загальному випадку ступінь поліному k має бути принаймні на одиницю більшим за кількість точок екстремуму кривої, що показує залежність b_i від i . Тобто заздалегідь потрібно знати кількість точок екстремуму, таким чином, вибір k є великою мірою суб'єктивним. Але у деяких випадках теорія може допомогти знайти потрібний вигляд кривої. На практиці припускають, що за допомогою поліному низького ступеня (наприклад, 2 або 3) можна отримати добрі результати. Якщо ми обрали певне значення k і хочемо з'ясувати, чи не буде кращим поліном вищого ступеня, потрібно діяти таким чином. Припустимо, нам потрібно зробити вибір між поліномом другого та третього ступеня. Рівняння матимуть такий вигляд:

$$\text{для другого ступеня: } y_t = a + c_0 z_0 + c_1 z_1 + c_2 z_2 + \varepsilon_1,$$

$$\text{для третього ступеня: } y_t = b + b_0 z_0 + b_1 z_1 + b_2 z_2 + b_3 z_3 + \varepsilon_2.$$

Якщо після знаходження параметрів другої регресії ми отримаємо, що b_2 статистично значуще, а b_3 — ні, то можемо вважати, що достатньою буде апроксимація поліномом другого ступеня.

Слід бути обережним щодо проблеми мультиколінеарності, яка може виникнути через те, що значення z_i були отримані через значення x_j . У випадку мультиколінеарності \hat{b}_3 може стати статистично незначущою не тому, що дійсно значення b_3 дорівнює нулеві, а просто тому, що вибірка не дозволяє оцінити окремий вплив z_3 на y . Отже, у нашому прикладі перед тим, як дійти висновку, що не можна обирати поліном третього ступеня, необхідно впевнитись, що мультиколінеарність відсутня.

По-третє, змінні z , що визначаються як лінійні комбінації змінної x , будуть корелювати між собою у випадках, коли спостерігається великий зв'язок між самими початковими змінними. Тому оцінку параметрів моделі (11.41) роблять в умовах мультиколінеарності чинників. Однак мультиколінеарність чинників z_0, \dots, z_k у моделі (11.41) впливає на оцінки параметрів b_0, b_1, \dots, b_p дещо у меншій мірі, ніж якби ці оцінки були отримані шляхом застосування звичайного ІМНК безпосередньо до моделі (11.35) в умовах мультиколінеарності чинників $x_t, x_{t-1}, \dots, x_{t-p}$. Це пов'язано з тим, що у моделі (11.41) мультиколінеарність веде до пониження ефективності оцінок c_0, \dots, c_k , тому кожний з параметрів b_0, \dots, b_p , які визначаються як лінійні комбінації оцінок c_0, \dots, c_k , буде представляти собою більш точну оцінку.

Розглянемо переваги методу Алмона.

По-перше, він забезпечує гнучкий спосіб залучення до моделі цілого ряду лагових структур, у той час як модель Койка досить суворо вимагає від коефіцієнтів, щоб вони спадали у геометричній прогресії.

По-друге, на відміну від методу Койка, у моделі Алмона не потрібно турбуватися по те, що серед пояснювальних змінних є залежні, а отже, ми позбавляємося проблем, які можуть виникнути у зв'язку з цим. Нарешті, якщо обрано поліном досить низького ступеня, кількість оцінюваних коефіцієнтів (c_i) буде набагато меншою, ніж початкова їх кількість (b_j).

Системи економетричних рівнянь

При моделюванні складних економічних об'єктів досить часто розглядається не одне, а декілька зв'язаних між собою рівнянь, тобто модель описується системою рівнянь. І тому виникає питання про оцінку параметрів системи рівнянь. У загальному випадку система взаємозалежних рівнянь має вигляд:

$$\begin{aligned}y_1 &= b_{12} y_2 + \dots + b_{1M} y_M + \gamma_{11} x_1 + \dots + \gamma_{1N} x_N + \varepsilon_1, \\y_2 &= b_{21} y_1 + \dots + b_{2M} y_M + \gamma_{21} x_1 + \dots + \gamma_{2N} x_N + \varepsilon_2, \\&\dots \\y_M &= b_{M1} y_1 + \dots + b_{M,M-1} y_{M-1} + \gamma_{M1} x_1 + \dots + \gamma_{MN} x_N + \varepsilon_M,\end{aligned}\tag{12.1}$$

де y_i ($i = \overline{1, M}$) — це ендogenous змінні, що залежать від раніше визначених змінних x_i ($i = \overline{1, N}$). До раніше визначених змінних x_i належать незалежні екogenous змінні і лагові ендogenous змінні.

Система рівнянь у вигляді (12.1) називається *системою взаємозалежних рівнянь у структурній формі*. Параметри b_{ij} називаються структурними параметрами чи коефіцієнтами.

З математичної точки зору, головна відмінність між ендogenousними і раніше визначеними змінними полягає у тому, що раніше визначені змінні не корелюють із залишками, а ендogenousні — корелюють.

Структурна форма моделі дозволяє бачити вплив змін довільної раніше визначеної змінної на значення ендogenousної змінної.

Крім регресійних рівнянь, а вони називаються рівняннями поведінки, модель може включати тотожності. У тотожностях відсутні залишки ε . Тотожності дозволяють виключити деякі ендogenousні змінні і розглядати систему регресійних рівнянь меншого виміру.

Так як у системі (12.1) ендogenous змінні, що стоять справа у рівнянні, корелюють із залишками, то передумови 1МНК не виконуються, і тому оцінки коефіцієнтів методом 1МНК будуть зміщеними і необґрунтованими. Тому для оцінки структурних коефіцієнтів структурна форма системи рівнянь перетворюється до скороченої форми.

Скороченою формою структурної моделі є модель, у якій ендogenous змінні виражені як функції лише раніше визначених змінних. Скорочена форма записується двома способами. Перший — згорнуте вираження ендogenous змінних як функцій раніше визначених змінних. Наприклад, маємо систему у структурній формі

$$\begin{aligned} y_1 &= b_{12} y_2 + \gamma_{11} x_1 + \gamma_{12} x_2 + \varepsilon_1, \\ y_2 &= b_{21} y_1 + \gamma_{21} x_1 + \gamma_{22} x_2 + \varepsilon_2, \end{aligned} \quad (12.2)$$

тоді система у скороченій формі матиме вигляд

$$\begin{aligned} y_1 &= \pi_{11} x_1 + \pi_{12} x_2 + u_1, \\ y_2 &= \pi_{21} x_1 + \pi_{22} x_2 + u_2. \end{aligned} \quad (12.3)$$

Другий спосіб запису скороченої форми — розгорнуте вираження ендogenous змінних через раніше визначені змінні, структурні параметри та випадкові величини. У цьому випадку скорочена форма матиме вигляд:

$$\begin{aligned} y_1 &= \frac{b_{12}\gamma_{21} + \gamma_{11}}{1 - b_{21}b_{12}} x_1 + \frac{b_{12}\gamma_{22} + \gamma_{12}}{1 - b_{21}b_{12}} x_2 + \frac{b_{12}\varepsilon_2 + \varepsilon_1}{1 - b_{21}b_{12}}, \\ y_2 &= \frac{b_{21}\gamma_{11} + \gamma_{21}}{1 - b_{21}b_{12}} x_1 + \frac{b_{21}\gamma_{12} + \gamma_{22}}{1 - b_{21}b_{12}} x_2 + \frac{b_{21}\varepsilon_1 + \varepsilon_2}{1 - b_{21}b_{12}}. \end{aligned}$$

Для збігу двох типів запису скороченої форми необхідно, щоб виконувалося таке співвідношення між π_{ij} та структурними параметрами:

$$\begin{aligned} \pi_{11} &= \frac{b_{12}\gamma_{21} + \gamma_{11}}{1 - b_{21}b_{12}}, & \pi_{12} &= \frac{b_{12}\gamma_{22} + \gamma_{12}}{1 - b_{21}b_{12}}, \\ \pi_{21} &= \frac{b_{21}\gamma_{11} + \gamma_{21}}{1 - b_{21}b_{12}}, & \pi_{22} &= \frac{b_{21}\gamma_{12} + \gamma_{22}}{1 - b_{21}b_{12}}. \end{aligned} \quad (12.4)$$

Параметри скороченої форми вимірюють загальний (прямий та непрямий) вплив попередньо визначених змінних на ендogenous змінні, у той час як структурні параметри вимірюють тільки прямий вплив.

Таким чином, параметри скороченої форми можна застосовувати для прогнозування та аналізу економічної діяльності, тому що вони дають одночасно оцінку прямого та непрямого впливу екзogenous змінних на ендogenous змінні.

Далі розглянемо метод оцінки параметрів структурної форми.

12.1 Непрямий метод найменших квадратів (НМНК) оцінки параметрів системи

Ідея методу дуже проста:

- 1) будуємо скорочену форму системи (12.1);
- 2) окремо до кожного рівняння скороченої форми застосовуємо метод МНК, а це можливо, тому що у правій частині скороченого рівняння знаходяться раніше визначені змінні, що не корелюють із залишками;
- 3) за оцінками коефіцієнтів скороченої форми отримуємо оцінки коефіцієнтів структурної форми.

Тобто якщо розглядати систему (12.2), то, оцінивши методом МНК коефіцієнти π_{ij} скороченої форми (12.3), знайдемо із системи (12.4) оцінки коефіцієнтів b_{ij} , γ_{kl} структурної форми.

12.2 Проблема ототожнення (ідентифікації) у системі рівнянь

Якщо ми будемо оцінювати коефіцієнти структурної форми методом НМНК, то можливі такі ситуації при розв'язанні системи (12.4):

- 1) коефіцієнти b_{ij} і γ_{kl} знаходяться однозначно (кількість коефіцієнтів структурної і скороченої форм співпадають), тоді система рівнянь називається *точно ототожненою* або *точно ідентифікованою*;
- 2) якщо кількість коефіцієнтів скороченої форми більша за кількість структурних коефіцієнтів, то у цьому випадку можна отримати два або і більше значень одного структурного коефіцієнта. У цьому випадку система є *переототожненою* або *зверхідентифікованою*;
- 3) якщо ж кількість коефіцієнтів скороченої форми менша за кількість структурних коефіцієнтів, то структурні коефіцієнти не можуть бути оцінені через коефіцієнти скороченої форми. Тоді ми кажемо, що система є *неототожненою* або *неідентифікованою*.

Якщо система точно ототожнена або переототожнена, то кажемо, що система *ототожнена*.

Вважаємо, що система ототожнена, якщо кожне із рівнянь ототожнене. Коли хоч одне із рівнянь неототожнене, то і система неототожнена. Якщо всі рівняння точно ототожені, а одне — переототожнене, то і система переототожнена.

Розглянемо правила ототожнення рівнянь системи. Введемо такі позначення:

M — кількість ендогенних змінних у системі;

m — кількість ендогенних змінних у окремому рівнянні;

N — кількість раніше визначених змінних у системі;

n — кількість раніше визначених змінних у окремому рівнянні.

Необхідна умова ототожнення (умова порядку)

Для ототожнення рівняння необхідно, щоб кількість відсутніх раніше визначених змінних у цьому рівнянні була не меншою за кількість включених до нього ендогенних змінних мінус одиниця, тобто

$$N - n \geq m - 1. \quad (12.5)$$

Приклад 12.1. Розглянемо систему (12.2). Перевіримо виконання необхідної умови ототожнення.

Нагадаємо, що система (12.2) має вигляд:

$$y_1 = b_{12} y_2 + \gamma_{11} x_1 + \gamma_{12} x_2 + \varepsilon_1,$$

$$y_2 = b_{21} y_1 + \gamma_{21} x_1 + \gamma_{22} x_2 + \varepsilon_2.$$

Розглянемо перше рівняння системи. Для нього: $M = 2$, $m = 2$, $N = 2$, $n = 2$. Перевіримо умову (12.5):

$$2 - 2 \stackrel{?}{\geq} 2 - 1.$$

Так як нерівність (12.5) не виконується, то перше рівняння системи (12.2) неототожене. Зрозуміло, що і друге рівняння неототожене, тому що необхідна умова ототожнення не виконується.

Необхідна і достатня умова ототожнення (рангова умова ототожнення)

У системі, що містить M рівнянь з M ендогенними змінними, рівняння буде ототожненим тоді і тільки тоді, коли ранг матриці \mathbf{A} , утвореної з коефіцієнтів, котрі відповідають пропущеним змінним рівняння, що розглядається, у всіх інших рівняннях моделі, крім даного, дорівнює $M - 1$.

Приклад 12.2. Розглянемо систему рівнянь і вивіримо, чи ототожене перше рівняння.

$$y_1 - \beta_{12} y_2 - \beta_{13} y_3 - \gamma_{11} x_1 = \varepsilon_1, \quad (12.6a)$$

$$y_2 - \beta_{23} y_3 - \gamma_{21} x_1 - \gamma_{22} x_2 = \varepsilon_2, \quad (12.6b)$$

$$y_3 - \beta_{31} y_1 - \gamma_{31} x_1 - \gamma_{32} x_2 = \varepsilon_3, \quad (12.6c)$$

$$y_4 - \beta_{41} y_1 - \beta_{42} y_2 - \gamma_{43} x_3 = \varepsilon_4. \quad (12.6d)$$

Для полегшення пояснення запишемо попередню систему у вигляді таблиці 1.

Таблиця 1

№ рівняння					+	+	+
	y_1	y_2	y_3	y_4	x_1	x_2	x_3
(12.6a)	1	$-\beta_{12}$	$-\beta_{13}$	0	$-\gamma_{11}$	0	0
(12.6b)	0	1	$-\beta_{23}$	0	$-\gamma_{21}$	$-\gamma_{22}$	0
(12.6c)	$-\beta_{31}$	0	1	0	$-\gamma_{31}$	$-\gamma_{32}$	0
(12.6d)	$-\beta_{41}$	$-\beta_{42}$	0	1	0	0	$-\gamma_{43}$

Перевіримо виконання необхідної умови ототожнення для рівняння (12.6a). Маємо: $M = 4$, $m = 3$, $N = 3$, $n = 1$. Тоді нерівність $N - n \geq m - 1$ має вигляд: $3 - 1 \stackrel{?}{\geq} 3 - 1$, тобто $2 = 2$. Отже, перше рівняння може бути ототожненим.

Розглянемо рангову умову ототожнення для рівняння (12.6a). У таблиці 1 знаком (+) позначені змінні, що пропущені у рівнянні (12.6a). Тоді матриця \mathbf{A} має вигляд:

$$\mathbf{A} = \begin{pmatrix} 0 & -\gamma_{22} & 0 \\ 0 & -\gamma_{32} & 0 \\ 1 & 0 & -\gamma_{43} \end{pmatrix}$$

Зрозуміло, що ранг матриці \mathbf{A} менший трьох, тому що перший і третій стовпчики лінійно залежні (детермінант матриці \mathbf{A} дорівнює нулеві). Маємо, що $\text{rang } \mathbf{A} < 3$, а $M - 1 = 4 - 1 = 3$. Отже, необхідна і достатня умова ототожнення першого рівняння не виконується. Це означає, що рівняння (12.6a) — неототожене.

Аналогічно можна показати, що рівняння (12.6b) і (12.6c) неототожені, а рівняння (12.6d) — ототожене.

Використовуючи умови порядку і рангу, можна сформулювати загальне правило ототожнення структурного рівняння системи, що має M рівнянь.

- 1) якщо $N - n > m - 1$ і ранг матриці \mathbf{A} дорівнює $M - 1$, то відповідне рівняння переототожене;
- 2) якщо $N - n = m - 1$ і ранг матриці \mathbf{A} дорівнює $M - 1$, то рівняння точно ототожене;
- 3) якщо $N - n \geq m - 1$ і ранг матриці \mathbf{A} менший за $M - 1$, то рівняння неототожене;
- 4) якщо $N - n < m - 1$, то структурне рівняння неототожене. У цьому випадку ранг матриці \mathbf{A} менший за $M - 1$.

12.3 Двокроковий метод найменших квадратів оцінки параметрів системи у структурній формі

Якщо система точно ототожнена, то коефіцієнти системи у структурній формі рівнянь оцінюються непрямим методом найменших квадратів. При умові, що система переототожнена, для оцінки коефіцієнтів системи у структурній формі використовують двокроковий метод найменших квадратів.

Розглянемо систему (12.1):

$$\begin{aligned} y_1 &= b_{12} y_2 + \dots + b_{1M} y_M + \gamma_{11} x_1 + \dots + \gamma_{1N} x_N + \varepsilon_1, \\ &\dots \\ y_M &= b_{M1} y_1 + \dots + b_{M,M-1} y_{M-1} + \gamma_{M1} x_1 + \dots + \gamma_{MN} x_N + \varepsilon_M. \end{aligned} \quad (12.7)$$

Так як ендогенні змінні правої частини системи корелюють із залишками, то застосовувати метод 1МНК не можна. Тому необхідно «очистити» ендогенні змінні правої частини, що без запізнень, від впливу залишків. Для цього можна скористатися інструментальними змінними, які досить сильно корелюють з ендогенними змінними правої частини системи, але не корелюють із залишками, або скористатися двокроковим методом найменших квадратів (2МНК).

Розглянемо алгоритм двокрокового методу найменших квадратів.

Алгоритм 12.1 (двокроковий метод найменших квадратів).

- 1) Систему рівнянь переписуємо у скороченій формі і оцінюємо коефіцієнти кожного рівняння методом 1МНК, розраховуємо теоретичні значення ендогенних змінних.

Тобто розглядаємо систему рівнянь:

$$\begin{aligned} y_1 &= \alpha_{11} x_1 + \dots + \alpha_{1N} x_N + u_1, \\ &\dots \\ y_M &= \alpha_{M1} x_1 + \dots + \alpha_{MN} x_N + u_M, \end{aligned}$$

коефіцієнти оцінюємо методом 1МНК і розраховуємо $\hat{y}_1, \dots, \hat{y}_M$.

- 2) У кожному рівнянні системи (12.7) у структурній формі замінюємо значення ендогенних (без запізнень) змінних правої частини на розраховані у першому пункті теоретичні значення і до кожного рівняння застосовуємо метод 1МНК.

Отже, отримаємо таку систему:

$$y_1 = b_{12}\hat{y}_2 + \dots + b_{1M}\hat{y}_M + \gamma_{11}x_1 + \dots + \gamma_{1N}x_N + \varepsilon_1,$$

...

$$y_M = b_{M1}\hat{y}_1 + \dots + b_{M,M-1}\hat{y}_{M-1} + \gamma_{M1}x_1 + \dots + \gamma_{MN}x_N + \varepsilon_M,$$

коефіцієнти якої оцінюємо методом 1МНК.

Зауваження 12.1. При побудові системи у скороченій формі не обов'язково у рівняннях використовувати усі раніше визначені змінні. Можна обирати тільки найбільш впливові змінні.

Зауваження 12.2. Метод 2МНК розроблений для переототожнених систем, але його можна використовувати і для точно ототожнених систем, і тоді оцінки НМНК і 2МНК співпадають.

Приклад 12.3. Розглянемо систему

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 x_1 + u_1,$$

$$y_2 = \beta_3 + \beta_4 y_1 + \beta_5 x_2 + \beta_6 x_3 + u_2$$

при таких початкових даних:

	x_1	x_2	x_3	y_1	y_2
1	0.23	0.78	0.4	9.26	13.26
2	0.24	0.75	0.26	9.38	10.16
3	0.19	0.68	0.4	12.11	13.72
4	0.17	0.7	0.5	10.81	12.85
5	0.23	0.62	0.4	9.35	10.63
6	0.43	0.76	0.19	9.87	9.12
7	0.31	0.73	0.25	8.17	25.83
8	0.26	0.71	0.44	9.12	23.39
9	0.49	0.69	0.17	5.88	14.68
10	0.36	0.73	0.39	6.3	10.05

Необхідно оцінити коефіцієнти системи.

Використовуючи умови порядку і рангу ототожнення систем, легко показати, що перше рівняння переототожнене, а друге — точно ототожнене. Тому для оцінки коефіцієнтів скористаємося методом 2МНК. Будуємо таку систему рівнянь:

$$y_1 = \alpha_0 + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \varepsilon_1,$$

$$y_2 = \alpha_4 + \alpha_5 x_1 + \alpha_6 x_2 + \alpha_7 x_3 + \varepsilon_2$$

і оцінюємо коефіцієнти кожного рівняння методом 1МНК. Отримали такі два регресійні рівняння:

$$\begin{aligned}y_1 &= 16.644 - 18.196 x_1 - 0.305 x_2 - 6.194 x_3, \\y_2 &= 15.671 - 4.124 x_1 + 0.637 x_2 - 1.639 x_3.\end{aligned}$$

Розраховуємо \hat{y}_1, \hat{y}_2 :

$$\begin{aligned}\hat{y}_1 &= (9.743, 10.438, 10.502, 10.240, 9.792, 7.411, 9.232, 8.971, 6.465, 7.455)^\top, \\ \hat{y}_2 &= (14.564, 14.733, 14.665, 14.596, 14.462, 14.070, 14.448, 14.330, 13.811, 14.012)^\top.\end{aligned}$$

Далі розглядаємо систему

$$\begin{aligned}y_1 &= \beta_0 + \beta_1 \hat{y}_2 + \beta_2 x_1 + u_1, \\y_2 &= \beta_3 + \beta_4 \hat{y}_1 + \beta_5 x_2 + \beta_6 x_3 + u_2,\end{aligned}$$

і до кожного рівняння застосовуємо метод 1МНК.

Отримаємо таку систему:

$$\begin{aligned}\hat{y}_1 &= -37.986 + 3.352 \hat{y}_2 - 3.979 x_1, \\ \hat{y}_2 &= 11.899 + 0.227 \hat{y}_1 + 0.706 x_2 - 0.236 x_3.\end{aligned}$$



12.3.1 Оцінка методом 2МНК статистичних характеристик системи рівнянь

Нехай у i -е регресійне рівняння крім n_i раніше визначених змінних входять також m_i незапізнених ендогенних змінних, тоді для t -го спостереження запишемо

$$y_{it} = b_{1i} y_{1t} + \dots + b_{m_i i} y_{m_i t} + \gamma_{1i} x_{1t} + \dots + \gamma_{n_i i} x_{n_i t} + u_{it}. \quad (12.8)$$

Тепер рівняння (12.8) запишемо у матричній формі:

$$\mathbf{y}_i = \mathbf{Y}_i \cdot \mathbf{b}_i + \mathbf{X}_i \cdot \boldsymbol{\gamma}_i + \mathbf{u}_i \quad (12.9)$$

$T \times 1$ $T \times m_i$ $m_i \times 1$ $T \times n_i$ $n_i \times 1$ $T \times 1$

Зазначимо, що у i -му рівнянні знаходиться $n_i + m_i$ змінних (регресорів). У рівнянні (12.9) об'єднаємо матриці даних у загальну матрицю, а вектори структурних коефіцієнтів — у загальний вектор:

$$\mathbf{Z}_i = [\mathbf{Y}_i | \mathbf{X}_i], \quad \boldsymbol{\alpha}_i^\top = [\mathbf{b}_i^\top | \boldsymbol{\gamma}_i^\top],$$

тоді (12.9) матиме вигляд:

$$\mathbf{y}_i = \mathbf{Z}_i \boldsymbol{\alpha}_i + \mathbf{u}_i. \quad (12.10)$$

Тепер перейдемо до оцінки статистичних характеристик системи рівнянь.

- 1) *Прогноз регресанда (теоретичне значення) по окремому рівнянню.*

Такий прогноз по i -му рівнянню, що оцінене методом 2МНК, можна зробити за формулою:

$$\hat{\mathbf{y}}_i = \mathbf{Y}_i \hat{\mathbf{b}}_i + \mathbf{X}_i \hat{\boldsymbol{\gamma}}_i = \mathbf{Z}_i \hat{\boldsymbol{\alpha}}_i.$$

Варто зазначити, що використовуються «неочищені» матриці \mathbf{Y}_i і \mathbf{Z}_i .

- 2) *Розрахунок вектора залишків.*

Вектор залишків $\hat{\mathbf{u}}_i$ розраховується таким чином:

$$\hat{\mathbf{u}}_i = \mathbf{y}_i - \hat{\mathbf{y}}_i = \mathbf{y}_i - \mathbf{Y}_i \hat{\mathbf{b}}_i - \mathbf{X}_i \hat{\boldsymbol{\gamma}}_i = \mathbf{y}_i - \mathbf{Z}_i \hat{\boldsymbol{\alpha}}_i. \quad (12.11)$$

Тут теж використовуються «неочищені» матриці \mathbf{Y}_i і \mathbf{Z}_i .

Вектор залишків, розрахований за формулою (12.11), використовується у якості вхідної інформації у тестах на автокореляцію і гетероскедастичність, а також при оцінці дисперсії залишків і розрахунку коефіцієнта детермінації.

- 3) *Оцінка дисперсії залишків.*

Дисперсія залишків $\sigma_{u_i}^2$ у i -му структурному рівнянні розраховується таким чином:

$$\hat{\sigma}_{u_i}^2 = \frac{\hat{\mathbf{u}}_i^\top \hat{\mathbf{u}}_i}{T - n_i - m_i}, \quad (12.12)$$

де $\hat{\mathbf{u}}_i$ взяте з виразу (12.11).

- 4) *Оцінка асимптотичної коваріаційної матриці коефіцієнтів.*

Оцінник асимптотичної коваріаційної матриці 2МНК-оцінника для $\boldsymbol{\alpha}_i$, заданого виразом

$$\hat{\boldsymbol{\alpha}}_i = \begin{pmatrix} \hat{\boldsymbol{\gamma}}_i \\ \hat{\boldsymbol{\beta}}_i \end{pmatrix} = \left(\hat{\mathbf{Z}}_i^\top \hat{\mathbf{Z}}_i \right)^{-1} \hat{\mathbf{Z}}_i^\top \mathbf{y}_i,$$

має вигляд:

$$\hat{\boldsymbol{\Sigma}}_{\hat{\boldsymbol{\alpha}}_i} = \hat{\sigma}_{u_i}^2 \left(\hat{\mathbf{Z}}_i^\top \hat{\mathbf{Z}}_i \right)^{-1},$$

де $\hat{\sigma}_{u_i}^2$ розраховується за формулою (12.12).

- 5) *Точковий прогноз.*

Точковий прогноз отримується просто підстановкою значень екзогенних змінних до скороченої форми рівнянь.

12.4 Рекурсивні системи

Особливим випадком системи рівнянь є рекурсивні системи.

Модель називається рекурсивною, якщо її структурні рівняння можна впорядкувати таким чином, що перше містить у правій частині лише раніше визначені змінні;

друге містить раніше визначені змінні та першу ендогенну змінну і так далі.

Наприклад:

$$\begin{aligned}y_1 &= f_1(x_1, x_2, \dots, x_k, u_1), \\y_2 &= f_2(x_1, x_2, \dots, x_k, y_1, u_2), \\y_3 &= f_3(x_1, x_2, \dots, x_k, y_1, y_2, u_3), \\&\dots\end{aligned}$$

Особливість рекурсивної моделі — це те, що є можливість оцінити її рівняння послідовно за допомогою ІМНК. Так як у першому рівнянні справа тільки екзогенні змінні, то можна застосувати ІМНК і отримати оцінене значення \hat{y}_1 для першої ендогенної змінної. Маючи оцінку \hat{y}_1 , можна оцінювати коефіцієнти другого рівняння, тому що x не корелюють з u_2 і \hat{y}_1 теж не корелює з u_2 , і т. д.

Рекурсивні системи називають ще трикутними системами, тому що коефіцієнти при ендогенних змінних утворюють трикутну матрицю, головна діагональ якої містить одиниці.

Рекомендована література

- [1] Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрики. — М.: ЮНИТИ, 1998. — 1022 с.
- [2] Грубер Й. Эконометрия. Введение в эконометрию. Т. 1. — К.: 1996. — 397 с.
- [3] Доугерти К. Введение в эконометрику. — М.: МГУ, 1999. — 402 с.
- [4] Под ред. Н. Н. Елисеевой. Эконометрика. — М.: Финансы и статистика, 2002. — 242 с.
- [5] Кремер Н. Ш., Путко Б. А. Эконометрика. — М.: ЮНИТИ, 2002. — 311 с.
- [6] Лук'яненко І., Краснікова Л. Економетрика. — К.: Знання, 1998. — 494 с.
- [7] Магнус Я. Р., Катышев П. К., Перестецкий А. А. Эконометрика. Начальный курс. — М.: Дело, 1998. — 246 с.

Завдання для самостійної роботи

Завдання 1

1. За **початковими даними** розрахувати парні коефіцієнти кореляції між залежною (результативною) ознакою Y і незалежними (факторними) ознаками X_1 – X_5 . Перевірити значущість отриманих коефіцієнтів. Надати економічного тлумачення результатам.
2. Розрахувати часткові коефіцієнти кореляції $r_{y|x_i}$ та перевірити їх значущість. Надати економічного тлумачення результатам.
3. Зробити висновки за результатами пп. 1–2.

Завдання 2

1. За **початковими даними** побудувати методом ІМНК
 - багатофакторну модель А, яка включає всі факторні ознаки X_1 – X_5 ;
 - однофакторні моделі Б–Е, які виражають залежність результативної ознаки від кожної факторної ознаки окремо.
2. Для кожної з моделей оцінити $\hat{\sigma}_u^2$, MSE, MAPE, $M(\hat{u})$. Результати представити у вигляді таблиці.
3. Зробити порівняльну характеристику моделей і визначити найбільш впливовий чинник. Порівняти отримані результати із результатами [завдання 1](#).

Завдання 3

1. Для моделі А (див. [завдання 2](#)) розрахувати стандартизовані коефіцієнти регресії і визначити найбільш впливовий регресор.

2. Розрахувати коефіцієнти еластичності для моделі А у середній точці. Надати економічного тлумачення отриманим результатам.
3. Розрахувати оцінену дисперсійно-коваріаційну матрицю $\hat{\beta}$ -коефіцієнтів для моделі А.

Завдання 4

1. Перевірити значущість коефіцієнтів моделі А і розрахувати їх інтервали довіри.
2. Розрахувати точковий прогноз регресанда та інтервальні прогнози для математичного очікування і для індивідуального значення регресанда для моделі А. У якості прогнозної точки взяти вектор середніх значень регресорів. Надати економічного тлумачення цим прогнозам.

Завдання 5

1. Для моделей А–Е, побудованих у [завданні 2](#), розрахувати коефіцієнти детермінації R^2 , \bar{R}_T^2 , \bar{R}_A^2 . Перевірити значущість цих моделей. Результати представити у вигляді таблиці.
2. Для моделі А розрахувати часткові коефіцієнти детермінації. Зробити висновки про впливовість чинників на основі отриманих результатів і порівняти ці висновки із попередніми.

Завдання 6

1. Побудувати «найкращу» модель методом усіх регресій (див. [приклад 6.1](#)).
2. Перевірити, чи добре специфікована обрана модель за критерієм Рамсея.
3. Оцінити специфікацію «найкращих» моделей кожної групи за критерієм Аменії.

Завдання 7

1. Для «найкращої» моделі (за результатами [завдання 6](#)) дослідити залишки на автокореляцію першого порядку за допомогою тесту Дарбіна — Уотсона.
2. У випадку наявності автокореляції залишків першого порядку усунути її за допомогою методу Кохрейна — Оркатта.

Завдання 8

1. Перевірити модель, отриману в результаті виконання [завдання 7](#) (вільну від автокореляції), на гетероскедастичність залишків за допомогою тесту Гольд-

фельда — Квандта (послідовно для кожної незалежної змінної).

2. Оцінивши σ_i , побудувати матрицю перетворень \mathbf{T}^H . Перетворити початкові дані і розрахувати методом ІМНК коефіцієнти нової моделі. Показати, що у побудованій моделі гетероскедастичність відсутня.

Завдання 9

1. Дослідити вектори X_1 – X_5 **початкових даних** на мультиколінеарність, використовуючи метод Фаррара — Глобера.
2. У випадку наявності мультиколінеарності усунути її за допомогою методу головних компонент.

Завдання 10

1. За **початковими даними** побудувати модель залежності вартості квартири у м. Одеса (2000 р.) від деяких чинників. Використати супровідні змінні: одну, що відображає місцезнаходження будинку і дозволяє розділити всю сукупність квартир на дві групи (квартири у Суворовському та Київському районах і квартири у Малиновському та Приморському районах), і другу, що відображає квартири новобудов і вторинного ринку.
2. Із отриманого регресійного рівняння виокремити регресійні рівняння для квартир у новобудовах і квартир вторинного ринку.

Завдання 11

1. Розглядається модель $y_t^* = \beta_0 + \beta_1 x_t + u_t$, де y_t^* — бажані або довгострокові витрати на товар, x — обсяг продажу. Використовуючи **дані таблиці** і модель часткових пристосувань, оцінити параметри довгострокової і короткострокової функцій витрат.
2. Використовуючи **дані таблиці**, оцінити параметри моделі $y_t = \beta_0 + \beta_1 x_t^* + u_t$, де x^* — бажаний або довгостроковий обсяг продажу, а y — витрати на товар.
3. Порівняти між собою отримані моделі.

Завдання 12

1. Застосувавши умови порядку і рангу, визначити, чи ототожене кожне з рівнянь системи

$$\beta_{11}y_1 + \beta_{12}y_2 + \beta_{13}y_3 + \gamma_{11}x_1 + \gamma_{12}x_2 + \gamma_{14}x_4 + \gamma_{15}x_5 + a_1 + u_1 = 0,$$

$$\beta_{21}y_1 + \beta_{22}y_2 + \beta_{23}y_3 + \gamma_{21}x_1 + \gamma_{22}x_2 + \gamma_{24}x_4 + \gamma_{25}x_5 + a_2 + u_2 = 0,$$

$$\beta_{31}y_1 + \beta_{32}y_2 + \beta_{33}y_3 + \gamma_{31}x_1 + \gamma_{32}x_2 + \gamma_{34}x_4 + \gamma_{35}x_5 + a_3 + u_3 = 0.$$

2. Якщо система ототожнена, то оцінити її параметри методом 2МНК.

Варіанти завдань та початкові дані

До завдань 1–9

Варіанти 1–25

Розглядаються наступні показники виробничо-господарської діяльності 53 підприємств:

- Y_1 Продуктивність праці
- Y_2 Індекс зниження собівартості продукції
- Y_3 Рентабельність

- X_1 Трудомісткість одиниці продукції
- X_2 Питома вага робітників у складі промислово-виробничого персоналу
- X_3 Питома вага покупних виробів
- X_4 Коефіцієнт змінності устаткування
- X_5 Премії та винагороди на одного робітника
- X_6 Питома вага втрат від браку
- X_7 Фондовіддача
- X_8 Середньорічна чисельність промислово-виробничого персоналу
- X_9 Середньорічна вартість основних виробничих фондів
- X_{10} Середньорічний фонд заробітної платні промислово-виробничого персоналу
- X_{11} Фондоозброєність праці
- X_{12} Оборотність нормованих оборотних коштів
- X_{13} Оборотність ненормованих оборотних коштів
- X_{14} Невиробничі витрати


 Таблицю початкових даних можна завантажити [тут](#).

Табл. Б.1. Номери факторних і результативних ознак за варіантами

Номер варіанту	Номер регресан- да Y	Номери регресорів X					Номер варіанту	Номер регресан- да Y	Номери регресорів X				
1	1	3	5	8	9	14	14	1	2	4	7	11	14
2	1	3	5	8	10	14	15	1	2	3	7	11	14
3	1	5	8	9	10	14	16	3	5	7	12	13	14
4	1	3	5	10	11	14	17	3	2	3	7	12	14
5	1	5	8	10	11	14	18	3	2	3	4	8	9
6	1	3	5	9	10	14	19	3	5	6	7	8	14
7	1	4	8	9	10	14	20	3	5	6	7	9	14
8	1	4	6	9	10	14	21	2	1	2	3	5	6
9	1	5	8	9	10	14	22	2	1	2	3	4	6
10	1	5	6	10	11	14	23	2	1	2	3	5	14
11	1	2	3	4	6	14	24	2	1	2	5	6	14
12	1	2	4	6	8	14	25	2	1	2	4	6	14
13	1	2	3	9	10	14							

Табл. Б.2. Показники діяльності підприємств

№ з/п	Y_1	Y_2	Y_3	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
1	9.26	204.2	13.26	0.23	0.78	0.4	1.37	1.23	0.23	1.45	26006	167.69	47750	6.4	166.32	10.08	17.72
2	9.38	209.6	10.16	0.24	0.75	0.26	1.49	1.04	0.39	1.3	23935	186.1	50391	7.8	92.88	14.76	18.39
3	12.11	222.6	13.72	0.19	0.68	0.4	1.44	1.8	0.43	1.37	22589	220.45	43149	9.76	158.04	6.48	26.46
4	10.81	236.7	12.85	0.17	0.7	0.5	1.42	0.43	0.18	1.65	21220	169.3	41089	7.9	93.96	21.96	22.37
5	9.35	62	10.63	0.23	0.62	0.4	1.35	0.88	0.15	1.91	7394	39.53	14257	5.35	173.88	11.88	28.13
6	9.87	53.1	9.12	0.43	0.76	0.19	1.39	0.57	0.34	1.68	11586	40.41	22661	9.9	162.3	12.6	17.55
7	8.17	172.1	25.83	0.31	0.73	0.25	1.16	1.72	0.38	1.94	26609	102.96	52509	4.5	88.56	11.52	21.92
8	9.12	56.5	23.39	0.26	0.71	0.44	1.27	1.7	0.09	1.89	7801	37.02	14903	4.88	101.16	8.28	19.52
9	5.88	52.6	14.68	0.49	0.69	0.17	1.16	0.84	0.14	1.94	11587	45.74	25587	3.46	166.32	11.52	23.99
10	6.3	46.6	10.05	0.36	0.73	0.39	1.25	0.6	0.21	2.06	9475	40.07	16821	3.6	140.76	32.4	21.76
11	6.22	53.2	13.99	0.37	0.68	0.33	1.13	0.82	0.42	1.96	10811	45.44	19459	3.56	128.52	11.52	25.68
12	5.49	30.1	9.68	0.43	0.74	0.25	1.1	0.84	0.05	1.02	6371	41.08	12973	5.65	177.84	17.28	18.13
13	6.5	146.4	10.03	0.35	0.66	0.32	1.15	0.67	0.29	1.85	26761	136.14	50907	4.28	114.48	16.2	25.74
14	6.61	18.1	9.13	0.38	0.72	0.02	1.23	1.04	0.48	0.88	4210	42.39	6920	8.85	93.24	13.32	21.21
15	4.32	13.6	5.37	0.42	0.68	0.06	1.39	0.66	0.41	0.62	3557	37.39	5736	8.52	126.72	17.28	22.97
16	7.37	89.8	9.86	0.3	0.77	0.15	1.38	0.86	0.62	1.09	14148	101.78	26705	7.19	91.8	9.72	16.38
17	7.02	62.5	12.62	0.32	0.78	0.08	1.35	0.79	0.56	1.6	9872	47.55	20068	4.82	69.12	16.2	13.21
18	8.25	46.3	5.02	0.25	0.78	0.2	1.42	0.34	1.76	1.53	5975	32.61	11487	5.46	66.24	24.84	14.48
19	8.15	103.5	21.18	0.31	0.81	0.2	1.37	1.6	1.31	1.4	16662	103.25	32029	6.2	67.68	14.76	13.38
20	8.72	73.3	25.17	0.26	0.79	0.3	1.41	1.46	0.45	2.22	9166	38.95	18946	4.25	50.4	7.56	13.69
21	6.64	76.6	19.4	0.37	0.77	0.24	1.35	1.27	0.5	1.32	15118	81.32	28025	5.38	70.56	8.64	16.66
22	8.1	73	21	0.29	0.78	0.1	1.48	1.58	0.77	1.48	11429	67.26	20968	5.88	72	8.64	15.06
23	5.52	32.3	6.57	0.34	0.72	0.11	1.24	0.68	1.2	0.68	6462	59.92	11049	9.27	97.2	9	20.09
24	9.37	199.6	14.19	0.23	0.79	0.47	1.4	0.86	0.21	2.3	24628	107.34	45893	4.36	80.28	14.76	15.98
25	13.17	598.1	15.81	0.17	0.77	0.53	1.45	1.98	0.25	1.37	49727	512.6	99400	10.31	51.48	10.08	18.27
26	6.67	71.2	5.23	0.29	0.8	0.34	1.4	0.33	0.15	1.51	11470	53.81	20719	4.69	105.12	14.76	14.42
27	5.68	90.8	7.99	0.41	0.71	0.2	1.28	0.45	0.66	1.43	19448	80.83	36813	4.16	128.52	10.44	22.76
28	5.22	82.1	17.5	0.41	0.79	0.24	1.33	0.74	0.74	1.82	18963	59.42	33956	3.13	94.68	14.76	15.41
29	10.02	76.2	17.16	0.22	0.76	0.54	1.22	0.03	0.32	2.62	9185	36.96	17016	4.02	85.32	20.52	19.35
30	8.16	119.5	14.54	0.29	0.78	0.4	1.28	0.99	0.89	1.75	17478	91.43	34873	5.23	76.32	14.4	16.83
31	3.78	21.9	6.24	0.51	0.62	0.2	1.47	0.24	0.23	1.54	6265	17.16	11237	2.74	153	24.84	30.53
32	6.48	48.4	12.08	0.36	0.75	0.64	1.27	0.57	0.32	2.25	8810	27.29	17306	3.1	107.64	11.16	17.98
33	10.44	173.5	9.49	0.23	0.71	0.42	1.51	1.22	0.54	1.07	17659	184.33	39250	10.44	90.72	6.48	22.09

Табл. Б.2. Показники діяльності підприємств

№ з/п	Y_1	Y_2	Y_3	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}
34	7.65	74.1	9.28	0.26	0.74	0.27	1.46	0.68	0.75	1.44	10342	58.42	19074	5.65	82.44	9.72	18.29
35	8.77	68.6	11.42	0.27	0.65	0.37	1.27	1	0.16	1.4	8901	59.4	18452	6.67	79.92	3.24	26.05
36	7	60.8	10.31	0.29	0.66	0.38	1.43	0.81	0.24	1.31	8402	49.63	17500	5.91	120.96	6.48	26.2
37	11.06	355.6	8.65	0.01	0.84	0.35	1.5	1.27	0.59	1.12	32625	391.27	7888	11.99	84.6	5.4	17.26
38	9.02	264.8	10.94	0.02	0.74	0.42	1.35	1.14	0.56	1.16	31160	258.62	58947	8.3	85.32	6.12	18.83
39	13.28	526.6	9.87	0.18	0.75	0.32	1.41	1.89	0.63	0.88	46461	75.66	94697	1.63	101.52	8.64	19.7
40	9.27	118.6	6.14	0.25	0.75	0.33	1.47	0.67	1.1	1.07	13833	123.68	29626	8.94	107.64	11.88	16.87
41	6.7	37.1	12.93	0.31	0.79	0.29	1.35	0.96	0.39	1.24	6391	37.21	11688	5.82	85.32	7.92	14.63
42	6.69	57.7	9.78	0.38	0.72	0.3	1.4	0.67	0.73	1.49	11115	53.37	21955	4.8	131.76	10.08	22.17
43	9.42	51.6	13.22	0.24	0.7	0.56	1.2	0.98	0.28	2.03	6555	32.87	12243	5.01	116.64	18.72	22.62
44	7.24	64.7	17.29	0.31	0.66	0.42	1.15	1.16	0.1	1.84	11085	45.63	20193	4.12	138.24	13.68	26.44
45	5.39	48.3	7.11	0.42	0.69	0.26	1.09	0.54	0.68	1.22	9484	48.41	20122	5.1	156.96	16.56	22.26
46	5.61	15	22.49	0.51	0.71	0.16	1.26	1.23	0.87	1.72	3967	13.58	7612	3.49	137.52	14.76	19.13
47	5.59	87.5	12.14	0.31	0.73	0.45	1.36	0.78	0.49	1.75	15283	63.99	27404	4.19	135.72	7.92	18.28
48	6.57	108.4	15.25	0.37	0.65	0.31	1.15	1.16	0.16	1.46	20874	104.55	39648	5.01	155.52	18.36	28.23
49	6.54	267.3	31.34	0.16	0.82	0.08	1.87	4.44	0.85	1.6	19418	222.11	43799	11.44	48.6	8.28	12.39
50	4.23	34.2	11.56	0.18	0.8	0.68	1.17	1.06	0.13	1.47	3351	25.76	6235	7.67	42.84	14.04	11.64
51	5.22	26.8	30.14	0.43	0.83	0.03	1.61	2.13	0.49	1.38	6338	29.52	11524	4.66	142.2	16.92	8.62
52	18	43.6	19.71	0.4	0.7	0.02	1.34	1.21	0.09	1.41	9756	41.99	17309	4.3	145.8	11.16	20.1
53	11.03	72	23.56	0.31	0.74	0.22	1.22	2.2	0.79	1.39	11795	78.11	22225	6.62	120.52	14.76	19.41

Варіанти 26–60

Розглядаються наступні соціально-економічні показники 52 країн світу:

- Y_1 Тривалість життя чоловіків (у роках)
 Y_2 Тривалість життя жінок (у роках)
 Y_3 ВВП на душу населення (у дол. США за купівельною спроможністю валют)
 Y_4 Приріст населення (% на рік)
- X_1 Чисельність населення (тис. чол.)
 X_2 Народжуваність (на 1000 чол.)
 X_3 Смертність (на 1000 чол.)
 X_4 Смертність серед малюків (на 1000 чол.)
 X_5 Середнє число дітей у родині
 X_6 ВВП на душу населення (у дол. США за купівельною спроможністю валют)
 X_7 Густина населення (кількість чол. на кв. км)
 X_8 Відсоток міського населення
 X_9 Відсоток грамотних
 X_{10} Приріст населення (% на рік)

Таблицю початкових даних можна завантажити [тут](#).

Табл. Б.3. Номери факторних і результативних ознак за варіантами

Номер варіанту	Номер регресанда Y	Номери регресорів X					Номер варіанту	Номер регресанда Y	Номери регресорів X				
26	1	1	2	3	4	6	44	3	1	2	3	4	5

Табл. Б.3. Номери факторних і результативних ознак за варіантами

Номер варіанту	Номер регресанда Y	Номери регресорів X					Номер варіанту	Номер регресанда Y	Номери регресорів X				
27	1	1	2	3	6	7	45	3	1	2	3	5	7
28	1	1	2	3	6	8	46	3	1	2	3	5	8
29	1	1	2	3	5	8	47	3	1	2	3	4	9
30	1	1	2	3	6	9	48	3	1	2	3	4	10
31	1	1	2	5	6	9	49	3	1	2	3	5	9
32	1	1	2	5	6	7	50	3	1	2	3	7	9
33	1	1	2	4	6	9	51	3	1	2	3	5	10
34	1	1	2	5	8	9	52	3	1	2	3	4	9
35	2	1	2	3	4	6	53	4	1	2	3	4	5
36	2	1	2	3	6	7	54	4	1	2	3	5	7
37	2	1	2	3	6	8	55	4	1	2	3	4	9
38	2	1	2	3	5	8	56	4	1	2	3	4	6
39	2	1	2	3	6	9	57	4	1	2	3	5	9
40	2	1	2	5	6	9	58	4	1	2	3	7	9
41	2	1	2	5	6	7	59	4	1	2	3	5	6
42	2	1	2	4	6	9	60	4	1	2	3	5	8
43	2	1	2	5	8	9							

Табл. Б.4. Соціально-економічні показники країн світу

№ з/п	Країна	Y ₁	Y ₂	Y ₃	Y ₄	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀
1	Австралія	74	80	16 848	1.38	17 800	15	8	7.3	1.9	16 848	2.3	85	100	1.38
2	Австрія	73	79	18 396	0.2	8 000	12	11	6.7	1.5	18 396	94	58	99	0.2
3	Аргентина	68	75	3 408	1.3	33 900	20	9	25.6	2.8	3 408	12	86	95	1.3
4	Бангладеш	53	53	202	2.4	125 000	35	11	106	4.7	202	800	16	35	2.4
5	Бельгія	73	79	17 912	0.2	10 100	12	11	7.2	1.7	17 912	329	96	99	0.2
6	Білорусь	66	76	6 500	0.32	10 300	13	11	19	1.88	6 500	50	65	99	0.32
7	Бразилія	57	67	2 354	1.28	156 600	21	9	66	2.7	2 354	18	75	81	1.28
8	Буркіна-Фасо	47	50	357	2.81	10 000	47	18	118	6.94	357	36	15	18	2.81
9	Велика Британія	74	80	15 974	0.2	58 400	13	11	7.2	1.83	15 974	237	89	99	0.2
10	В'єтнам	63	68	230	1.78	73 100	27	8	46	3.33	230	218	20	88	1.78
11	Гаїті	43	47	383	1.63	6 500	40	19	109	5.94	383	231	29	53	1.63
12	Гондурас	65	70	1 030	2.73	5 600	35	6	45	4.9	1 030	46	44	73	2.73
13	Гонконг	75	80	14 641	-0.09	5 800	13	6	5.8	1.4	14 641	5494	94	77	-0.09
14	Ефіопія	51	54	122	3.1	55 200	45	14	110	6.81	122	47	12	24	3.1
15	Єгипет	60	63	748	1.95	60 000	29	9	76.4	3.77	748	57	44	48	1.95
16	Замбія	44	45	573	2.8	9 100	46	18	85	6.68	573	11	42	73	2.8
17	Індія	58	59	275	1.9	911 600	29	10	79	4.48	275	283	26	52	1.9
18	Ірландія	73	78	12 170	0.3	3 600	14	9	7.4	1.99	12 170	51	57	98	0.3
19	Іспанія	74	81	13 047	0.25	39 200	11	9	6.9	1.4	13 047	77	78	95	0.25
20	Італія	74	81	17 500	0.21	58 100	11	10	7.6	1.3	17 500	188	69	97	0.21
21	Канада	74	81	19 904	0.7	29 100	14	8	6.8	1.8	19 904	2.8	77	97	0.7
22	Китай	67	69	377	1.1	1 205 200	21	7	52	1.84	377	124	26	78	1.1
23	Колумбія	69	75	1 538	2	35 600	24	6	28	2.47	1 538	31	70	87	2
24	Коста-Рика	76	79	2 031	2.3	3 300	26	4	11	3.1	2 031	64	47	93	2.3
25	Куба	74	78	1 382	0.95	11 100	17	7	10.2	1.9	1 382	99	74	94	0.95
26	Малайзія	66	72	2 995	2.3	19 500	29	5	25.6	3.51	2 995	58	43	78	2.3
27	Марокко	66	70	1 062	2.12	28 600	29	6	50	3.83	1 062	63	46	50	2.12

Табл. Б.4. Соціально-економічні показники країн світу

№ з/п	Країна	Y_1	Y_2	Y_3	Y_4	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
28	Мексика	69	77	3 604	1.9	91 800	28	5	35	3.2	3 604	46	73	87	1.9
29	Нідерланди	75	81	17 245	0.58	15 400	13	9	6.3	1.58	17 245	366	89	99	0.58
30	Німеччина	73	79	17 539	0.36	81 200	11	11	6.5	1.47	17 539	227	85	99	0.36
31	Нова Зеландія	73	80	14 381	0.57	3 524	16	8	8.9	2.03	14 381	13	84	99	0.57
32	Норвегія	74	81	17 755	0.4	4 300	13	10	6.3	2	17 755	11	75	99	0.4
33	ОАЕ	70	74	14 193	4.8	2 800	28	3	22	4.5	14 193	32	81	68	4.8
34	ПАР	62	68	3 128	2.6	43 900	34	8	47.1	4.37	3 128	35	49	76	2.6
35	Південна Корея	68	74	6 627	1	45 000	16	6	21.7	1.65	6 627	447	72	96	1
36	Північна Корея	67	73	1 000	1.83	23 100	24	6	27.7	2.4	1 000	189	60	99	1.83
37	Польща	69	77	4 429	0.3	38 600	14	10	13.8	1.94	4 429	123	62	99	0.3
38	Португалія	71	78	9 000	0.36	10 500	12	10	9.2	1.5	9 000	108	34	85	0.36
39	Росія	64	74	6 680	0.2	149 200	13	11	27	1.83	6 680	8.8	74	99	0.2
40	Саудівська Аравія	66	70	6 651	3.2	18 000	38	6	52	6.67	6 651	7.7	77	62	3.2
41	Сінгапур	73	79	14 990	1.2	2 900	16	6	5.7	1.88	14 990	4456	100	88	1.2
42	США	73	79	23 474	0.99	260 800	15	9	8.11	2.06	23 474	26	75	97	0.99
43	Таїланд	65	72	1 800	1.4	59 400	19	6	37	2.1	1 800	115	22	93	1.4
44	Туреччина	69	73	3 721	2.02	62 200	26	6	49	3.21	3 721	79	61	81	2.02
45	Україна	65	75	2 340	0.05	51 800	12	13	20.7	1.82	2 340	87	67	97	0.05
46	Філіппіни	63	68	867	1.92	69 800	27	7	51	3.35	867	221	43	90	1.92
47	Фінляндія	72	80	15 877	0.3	5 100	13	10	5.3	1.8	15 877	39	60	100	0.3
48	Франція	74	82	18 944	0.47	58 000	13	9	6.7	1.8	18 944	105	73	99	0.47
49	Чилі	71	78	2 591	1.7	14 000	23	6	14.6	2.5	2 591	18	85	93	1.7
50	Швейцарія	75	82	22 384	0.7	7 000	12	9	6.2	1.6	22 384	170	62	99	0.7
51	Швеція	75	81	16 900	0.52	8 800	14	11	5.7	2.1	16 900	19	84	99	0.52
52	Японія	76	82	19 860	0.3	125 500	11	7	4.4	1.55	19 860	330	77	99	0.3

До завдання 10

Розглядаються чинники, що впливають на вартість квартири (Y):

- X_1 Кількість кімнат
- X_2 Район міста:
К — Київський, М — Малиновський, П — Приморський, С — Суворовський
- X_3 Загальна площа квартири
- X_4 Житлова площа квартири
- X_5 Площа кухні
- X_6 Тип будинку: 0 — цегляний, 1 — інший
- X_7 Наявність балкону: 0 — ні, 1 — так
- X_8 Н — квартира у новобудові, В — квартира вторинного ринку

Початкові дані наведені в таблиці Б.5. Для i -го варіанту обираються рядки з номерами від i до $(i + 39)$.

Табл. Б.5. Характеристики квартир у м. Одеса

№ з/п	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
1	15.9	1	С	39.0	20.0	8.2	0	1	Н
2	27.0	3	К	68.4	40.5	10.7	0	1	Н
3	21.1	2	С	54.7	28.0	10.7	0	1	Н
4	24.5	4	С	90.0	64.0	15.0	0	0	В

Табл. Б.5. Характеристики квартир у м. Одеса

№ з/п	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
5	13.5	1	K	34.8	16.0	10.7	0	0	H
6	22.5	2	K	48.0	29.0	8.0	1	1	B
7	15.5	3	C	68.1	44.4	7.2	0	0	B
8	75.9	4	П	132.0	89.6	11.0	1	1	H
9	15.1	1	K	39.0	20.0	8.5	0	1	H
10	26.0	2	K	55.5	35.0	8.0	0	1	B
11	38.0	3	П	107.0	58.0	24.0	0	0	B
12	27.3	4	K	102.0	66.0	11.8	0	1	B
13	28.7	3	K	74.7	46.3	10.7	0	1	H
14	20.0	1	M	37.0	21.5	6.5	0	0	B
15	28.0	2	M	53.0	31.1	10.0	1	1	H
16	27.2	3	C	71.7	45.9	10.7	0	0	H
17	41.0	4	M	87.0	56.5	12.5	0	1	B
18	28.3	3	K	74.5	47.5	10.4	0	0	H
19	22.0	1	П	40.0	17.7	11.0	1	1	H
20	34.4	2	M	62.6	21.4	11.0	1	1	H
21	31.0	4	C	114.8	74.0	25.6	0	0	B
22	35.6	4	K	114.3	74.7	12.0	1	0	B
23	45.0	3	П	86.0	48.7	14.0	1	1	H
24	34.7	2	П	62.6	21.4	11.0	1	1	H
25	24.7	1	П	45.3	20.6	10.4	1	1	H
26	46.0	4	M	90.0	62.0	8.0	1	0	B
27	34.0	3	П	96.4	58.0	12.6	1	0	B
28	18.5	2	K	48.0	28.0	8.0	0	0	B
29	15.9	1	C	37.0	17.8	8.3	0	1	H
30	52.3	4	П	137.7	87.2	14.6	0	1	H
31	30.0	3	C	100.0	58.0	20.0	0	0	B
32	13.2	2	C	44.1	30.0	6.0	0	0	B
33	13.0	1	C	37.0	21.5	6.5	0	0	B
34	35.0	4	K	116.0	81.0	16.5	1	1	B
35	29.0	3	K	67.5	43.5	8.6	0	1	H
36	15.4	1	K	37.0	17.8	8.3	0	1	H
37	34.1	2	M	68.1	35.4	13.0	1	1	H
38	37.7	2	П	75.3	41.4	12.1	1	1	H
39	34.0	3	M	96.4	58.0	12.6	1	1	B
40	42.7	4	M	107.0	75.5	9.5	0	1	B
41	15.6	1	K	40.0	20.0	8.3	0	0	H
42	24.4	1	П	48.7	22.3	12.4	1	1	H
43	32.6	3	M	68.0	42.5	8.3	0	1	H
44	25.8	2	C	80.0	51.0	13.0	0	1	B
45	27.0	4	K	93.0	66.0	10.0	0	0	B
46	16.5	1	M	60.0	27.0	22.4	0	0	B
47	40.8	4	M	91.6	55.2	9.4	0	1	H
48	30.8	2	M	59.2	31.9	11.2	1	1	H
49	19.5	3	C	79.0	50.3	9.1	1	0	B
50	75.0	4	П	176.0	129.0	15.0	1	1	B
51	16.0	2	K	50.1	31.0	6.0	0	0	B
52	21.3	1	П	39.9	18.0	8.1	1	0	H
53	53.9	3	П	98.0	56.0	22.0	1	1	H

Табл. Б.5. Характеристики квартир у м. Одеса

№ з/п	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈
54	38.0	4	K	96.0	69.4	9.0	0	0	B
55	23.0	3	C	69.7	42.0	10.8	0	1	B
56	17.0	2	K	60.0	38.0	10.0	0	0	B
57	17.0	1	M	60.0	30.0	15.0	0	0	B
58	21.5	1	П	39.0	20.0	9.2	1	0	H
59	26.4	2	M	48.6	31.0	8.0	1	1	H
60	24.0	3	K	71.0	52.0	7.5	1	0	B
61	23.5	4	C	92.0	72.5	9.5	0	0	B
62	15.0	1	П	53.0	26.2	13.0	0	1	B
63	34.0	3	M	68.0	41.0	8.0	1	1	H
64	46.6	2	П	93.2	49.5	14.0	1	1	H
65	75.0	4	П	130.0	88.0	12.0	1	1	H
66	14.2	1	C	35.0	19.0	9.0	0	0	B
67	10.5	1	K	30.3	17.5	5.6	1	0	B
68	18.0	2	K	50.0	30.0	8.7	1	0	B
69	51.0	4	П	98.0	65.8	13.0	1	1	H
70	28.6	3	K	69.0	42.4	8.3	0	1	H
71	24.2	2	C	42.0	21.0	10.2	1	0	H
72	23.0	1	M	43.0	25.5	8.5	0	1	B
73	32.5	3	M	98.0	51.0	15.0	0	1	B
74	65.0	4	П	176.0	110.0	33.0	1	0	B
75	30.8	2	M	56.4	29.7	9.4	1	1	H
76	19.0	1	П	38.0	19.0	7.4	1	1	H
77	35.7	2	П	62.0	35.0	11.0	1	0	H
78	22.0	3	C	62.0	37.0	10.2	1	1	B
79	23.0	4	C	74.0	49.0	6.5	0	1	B
80	35.5	4	K	106.0	73.7	9.0	0	1	H
81	43.0	3	П	100.0	45.0	35.0	1	1	B
82	12.0	1	C	30.0	17.8	5.5	1	0	B
83	21.0	2	C	54.6	32.0	10.0	1	1	B
84	40.0	4	M	89.0	50.4	9.2	0	1	H
85	35.2	3	M	79.1	42.4	15.5	0	1	H
86	34.2	2	П	68.5	30.7	8.3	1	1	H
87	15.6	1	M	35.0	18.0	5.3	1	0	B
88	12.5	1	C	32.0	17.0	6.0	1	0	B
89	34.0	4	C	88.0	61.7	9.0	0	1	B
90	58.5	3	П	117.0	55.2	25.0	1	1	H
91	31.9	2	M	60.2	36.3	10.9	1	1	H
92	21.2	1	П	40.8	19.2	10.1	1	1	H
93	11.3	1	K	31.0	18.0	5.5	1	0	B
94	14.5	2	C	43.0	27.0	5.5	0	1	B
95	23.0	4	C	74.0	45.8	9.0	0	0	B
96	34.0	3	M	65.4	38.9	9.3	1	1	H
97	36.7	2	П	68.6	35.5	17.0	1	1	H
98	35.6	2	M	71.1	36.2	13.3	1	1	H
99	19.8	1	M	40.5	16.0	11.0	0	1	H
100	43.0	4	M	110.0	79.5	10.0	0	1	B

До завдання 11

Початкові дані наведені в таблиці Б.6. Для i -го варіанту обираються рядки, що відповідають значенням t від i до $(i + 29)$.

Табл. Б.6. Початкові дані для завдання 11

t	y	x	t	y	x	t	y	x	t	y	x	t	y	x
1	15.6	31.2	19	30.2	35.8	37	80.9	103.8	55	103.7	126.0	73	135.6	162.4
2	16.0	31.6	20	30.7	40.1	38	90.2	104.1	56	104.5	129.4	74	138.7	166.5
3	16.3	31.9	21	30.9	40.5	39	90.3	105.6	57	105.8	131.5	75	140.9	171.4
4	16.8	32.0	22	40.2	40.9	40	90.6	107.3	58	106.1	132.6	76	141.6	172.5
5	16.9	32.2	23	40.5	50.0	41	90.9	107.9	59	108.2	134.4	77	143.4	177.6
6	17.3	32.4	24	40.9	50.2	42	100.2	108.2	60	113.4	135.8	78	151.5	178.8
7	17.5	32.6	25	50.1	50.8	43	100.4	109.3	61	113.9	136.1	79	156.7	179.3
8	18.0	32.7	26	50.3	51.2	44	100.6	109.9	62	114.5	137.2	80	156.9	181.5
9	18.2	32.9	27	50.8	51.7	45	101.1	110.1	63	115.6	139.4	81	158.3	186.7
10	18.9	33.0	28	60.0	52.0	46	101.4	110.9	64	117.3	141.7	82	161.2	188.3
11	19.0	33.3	29	60.4	52.5	47	101.7	111.5	65	120.5	143.4	83	162.3	190.5
12	20.5	33.7	30	60.7	52.9	48	101.8	112.3	66	122.4	144.5	84	162.9	192.6
13	24.3	34.0	31	70.3	53.2	49	102.0	114.5	67	126.3	146.7	85	165.4	193.8
14	25.6	34.3	32	70.5	53.8	50	102.3	117.6	68	127.1	148.3	86	166.7	195.6
15	26.4	34.8	33	70.9	54.0	51	102.5	118.3	69	127.9	150.4	87	169.8	197.3
16	27.2	35.0	34	80.1	54.4	52	102.8	120.2	70	130.5	155.6	88	170.1	198.4
17	27.6	35.1	35	80.5	54.8	53	103.0	121.1	71	130.6	158.4	89	172.4	199.6
18	27.9	35.6	36	80.7	55.0	54	103.4	125.0	72	132.4	160.6	90	175.6	201.3

До завдання 12

Підклавши у систему значення параметрів β_{ij} , γ_{ij} з відповідного варіанту, отримаємо модель для дослідження. У системі y_1, y_2, y_3 — ендогенні змінні, x_i ($i = \overline{1, 5}$) — раніше визначені змінні.

Значення y_i ($i = \overline{1, 3}$), x_i ($i = \overline{1, 5}$) обираються з таблиці Б.2.

Табл. Б.7. Варіанти для завдання 12

$\beta_{11} = 0$	Вар. 1	Вар. 2	Вар. 3	Вар. 4	Вар. 5	Вар. 6	Вар. 7	Вар. 8	Вар. 9	Вар. 10
$\beta_{22} = 0$	$\gamma_{11} = 0$	$\gamma_{11} = 0$	$\gamma_{12} = 0$	$\gamma_{13} = 0$	$\gamma_{14} = 0$	$\gamma_{15} = 0$	$\gamma_{15} = 0$	$\gamma_{13} = 0$	$\gamma_{14} = 0$	$\gamma_{15} = 0$
$\beta_{33} = 0$	$\gamma_{22} = 0$	$\gamma_{12} = 0$	$\gamma_{14} = 0$	$\gamma_{15} = 0$	$\gamma_{23} = 0$	$\gamma_{25} = 0$	$\gamma_{24} = 0$	$\gamma_{22} = 0$	$\gamma_{23} = 0$	$\gamma_{21} = 0$
	$\gamma_{33} = 0$	$\gamma_{31} = 0$	$\gamma_{23} = 0$	$\gamma_{25} = 0$	$\gamma_{24} = 0$	$\gamma_{21} = 0$	$\gamma_{31} = 0$	$\gamma_{31} = 0$	$\gamma_{25} = 0$	$\gamma_{22} = 0$
	$\gamma_{34} = 0$	$\gamma_{32} = 0$	$\gamma_{31} = 0$	$\gamma_{34} = 0$	$\gamma_{35} = 0$	$\gamma_{31} = 0$	$\gamma_{34} = 0$	$\gamma_{35} = 0$	$\gamma_{35} = 0$	$\gamma_{35} = 0$
$\beta_{11} = 0$	Вар. 11	Вар. 12	Вар. 13	Вар. 14	Вар. 15	Вар. 16	Вар. 17	Вар. 18	Вар. 19	Вар. 20
$\beta_{23} = 0$	$\gamma_{12} = 0$	$\gamma_{12} = 0$	$\gamma_{13} = 0$	$\gamma_{14} = 0$	$\gamma_{13} = 0$	$\gamma_{15} = 0$	$\gamma_{13} = 0$	$\gamma_{14} = 0$	$\gamma_{12} = 0$	$\gamma_{14} = 0$
$\beta_{32} = 0$	$\gamma_{13} = 0$	$\gamma_{15} = 0$	$\gamma_{21} = 0$	$\gamma_{22} = 0$	$\gamma_{23} = 0$	$\gamma_{24} = 0$	$\gamma_{15} = 0$	$\gamma_{24} = 0$	$\gamma_{24} = 0$	$\gamma_{24} = 0$
	$\gamma_{24} = 0$	$\gamma_{21} = 0$	$\gamma_{25} = 0$	$\gamma_{24} = 0$	$\gamma_{24} = 0$	$\gamma_{25} = 0$	$\gamma_{21} = 0$	$\gamma_{31} = 0$	$\gamma_{33} = 0$	$\gamma_{31} = 0$
	$\gamma_{35} = 0$	$\gamma_{34} = 0$	$\gamma_{31} = 0$	$\gamma_{32} = 0$	$\gamma_{34} = 0$	$\gamma_{35} = 0$	$\gamma_{32} = 0$	$\gamma_{35} = 0$	$\gamma_{34} = 0$	$\gamma_{34} = 0$

$\beta_{12} = 0$	Вар. 21	Вар. 22	Вар. 23	Вар. 24	Вар. 25	Вар. 26	Вар. 27	Вар. 28	Вар. 29	Вар. 30
$\beta_{21} = 0$	$\gamma_{11} = 0$	$\gamma_{12} = 0$	$\gamma_{13} = 0$	$\gamma_{15} = 0$	$\gamma_{13} = 0$	$\gamma_{14} = 0$	$\gamma_{15} = 0$	$\gamma_{11} = 0$	$\gamma_{14} = 0$	$\gamma_{15} = 0$
$\beta_{33} = 0$	$\gamma_{12} = 0$	$\gamma_{13} = 0$	$\gamma_{14} = 0$	$\gamma_{14} = 0$	$\gamma_{21} = 0$	$\gamma_{22} = 0$	$\gamma_{23} = 0$	$\gamma_{21} = 0$	$\gamma_{24} = 0$	$\gamma_{21} = 0$
	$\gamma_{22} = 0$	$\gamma_{23} = 0$	$\gamma_{21} = 0$	$\gamma_{23} = 0$	$\gamma_{25} = 0$	$\gamma_{23} = 0$	$\gamma_{25} = 0$	$\gamma_{25} = 0$	$\gamma_{32} = 0$	$\gamma_{31} = 0$
	$\gamma_{31} = 0$	$\gamma_{35} = 0$	$\gamma_{34} = 0$	$\gamma_{32} = 0$	$\gamma_{31} = 0$	$\gamma_{34} = 0$	$\gamma_{35} = 0$	$\gamma_{34} = 0$	$\gamma_{34} = 0$	$\gamma_{35} = 0$
$\beta_{12} = 0$	Вар. 31	Вар. 32	Вар. 33	Вар. 34	Вар. 35	Вар. 36	Вар. 37	Вар. 38	Вар. 39	Вар. 40
$\beta_{23} = 0$	$\gamma_{11} = 0$	$\gamma_{11} = 0$	$\gamma_{12} = 0$	$\gamma_{12} = 0$	$\gamma_{13} = 0$	$\gamma_{13} = 0$	$\gamma_{14} = 0$	$\gamma_{15} = 0$	$\gamma_{14} = 0$	$\gamma_{12} = 0$
$\beta_{31} = 0$	$\gamma_{12} = 0$	$\gamma_{12} = 0$	$\gamma_{13} = 0$	$\gamma_{14} = 0$	$\gamma_{14} = 0$	$\gamma_{15} = 0$	$\gamma_{21} = 0$	$\gamma_{21} = 0$	$\gamma_{22} = 0$	$\gamma_{23} = 0$
	$\gamma_{21} = 0$	$\gamma_{22} = 0$	$\gamma_{23} = 0$	$\gamma_{24} = 0$	$\gamma_{23} = 0$	$\gamma_{21} = 0$	$\gamma_{22} = 0$	$\gamma_{25} = 0$	$\gamma_{23} = 0$	$\gamma_{24} = 0$
	$\gamma_{31} = 0$	$\gamma_{31} = 0$	$\gamma_{31} = 0$	$\gamma_{31} = 0$	$\gamma_{32} = 0$	$\gamma_{33} = 0$	$\gamma_{33} = 0$	$\gamma_{34} = 0$	$\gamma_{35} = 0$	$\gamma_{31} = 0$
$\beta_{13} = 0$	Вар. 41	Вар. 42	Вар. 43	Вар. 44	Вар. 45	Вар. 46	Вар. 47	Вар. 48	Вар. 49	Вар. 50
$\beta_{21} = 0$	$\gamma_{11} = 0$	$\gamma_{12} = 0$	$\gamma_{13} = 0$	$\gamma_{15} = 0$	$\gamma_{14} = 0$	$\gamma_{15} = 0$	$\gamma_{11} = 0$	$\gamma_{12} = 0$	$\gamma_{13} = 0$	$\gamma_{14} = 0$
$\beta_{32} = 0$	$\gamma_{21} = 0$	$\gamma_{23} = 0$	$\gamma_{23} = 0$	$\gamma_{24} = 0$	$\gamma_{23} = 0$	$\gamma_{23} = 0$	$\gamma_{21} = 0$	$\gamma_{22} = 0$	$\gamma_{23} = 0$	$\gamma_{24} = 0$
	$\gamma_{31} = 0$	$\gamma_{24} = 0$	$\gamma_{25} = 0$	$\gamma_{25} = 0$	$\gamma_{24} = 0$	$\gamma_{25} = 0$	$\gamma_{31} = 0$	$\gamma_{32} = 0$	$\gamma_{33} = 0$	$\gamma_{34} = 0$
	$\gamma_{32} = 0$	$\gamma_{31} = 0$	$\gamma_{32} = 0$	$\gamma_{33} = 0$	$\gamma_{34} = 0$	$\gamma_{34} = 0$	$\gamma_{32} = 0$	$\gamma_{34} = 0$	$\gamma_{34} = 0$	$\gamma_{35} = 0$
$\beta_{13} = 0$	Вар. 51	Вар. 52	Вар. 53	Вар. 54	Вар. 55	Вар. 56	Вар. 57	Вар. 58	Вар. 59	Вар. 60
$\beta_{23} = 0$	$\gamma_{11} = 0$	$\gamma_{12} = 0$	$\gamma_{13} = 0$	$\gamma_{14} = 0$	$\gamma_{15} = 0$	$\gamma_{11} = 0$	$\gamma_{12} = 0$	$\gamma_{13} = 0$	$\gamma_{14} = 0$	$\gamma_{15} = 0$
$\beta_{31} = 0$	$\gamma_{21} = 0$	$\gamma_{22} = 0$	$\gamma_{21} = 0$	$\gamma_{22} = 0$	$\gamma_{21} = 0$	$\gamma_{22} = 0$	$\gamma_{23} = 0$	$\gamma_{24} = 0$	$\gamma_{25} = 0$	$\gamma_{21} = 0$
	$\gamma_{22} = 0$	$\gamma_{23} = 0$	$\gamma_{25} = 0$	$\gamma_{24} = 0$	$\gamma_{31} = 0$	$\gamma_{32} = 0$	$\gamma_{33} = 0$	$\gamma_{31} = 0$	$\gamma_{31} = 0$	$\gamma_{22} = 0$
	$\gamma_{31} = 0$	$\gamma_{35} = 0$	$\gamma_{33} = 0$	$\gamma_{34} = 0$	$\gamma_{32} = 0$	$\gamma_{33} = 0$	$\gamma_{34} = 0$	$\gamma_{35} = 0$	$\gamma_{35} = 0$	$\gamma_{34} = 0$

ДОДАТОК В

Таблиці математичної статистики

В.1 q -квантилі стандартного нормального розподілу

Табл. В.1. Значення q -квантилей u_q стандартного нормального розподілу

q	u_q	q	u_q	q	u_q	q	u_q
0,50	0,000000	0,70	0,524401	0,90	1,281552	0,983	2,120072
0,51	0,025069	0,71	0,553385	0,91	1,340755	0,984	2,144411
0,52	0,050154	0,72	0,582842	0,92	1,405072	0,985	2,170090
0,53	0,075270	0,73	0,612813	0,93	1,475791	0,986	2,197286
0,54	0,100434	0,74	0,643345	0,94	1,554774	0,987	2,226212
0,55	0,125661	0,75	0,674490	0,95	1,644854	0,988	2,257129
0,56	0,150969	0,76	0,706303	0,96	1,750686	0,989	2,290368
0,57	0,176374	0,77	0,738847	0,97	1,880794	0,990	2,326348
0,58	0,201893	0,78	0,772193	0,971	1,895698	0,991	2,365618
0,59	0,227545	0,79	0,806421	0,972	1,911036	0,992	2,408916
0,60	0,253347	0,80	0,841621	0,973	1,926837	0,993	2,457263
0,61	0,279319	0,81	0,877896	0,974	1,943134	0,994	2,512144
0,62	0,305481	0,82	0,915365	0,975	1,959964	0,995	2,575829
0,63	0,331853	0,83	0,954165	0,976	1,977368	0,996	2,652070
0,64	0,358459	0,84	0,994458	0,977	1,995393	0,997	2,747781
0,65	0,385320	0,85	1,036433	0,978	2,014091	0,998	2,878162
0,66	0,412463	0,86	1,080319	0,979	2,033520	0,999	3,090232
0,67	0,439913	0,87	1,126391	0,980	2,053749		
0,68	0,467699	0,88	1,174987	0,981	2,074855		
0,69	0,495850	0,89	1,226528	0,982	2,096927		

Примітки.

1. При знаходженні q -квантилей для значень $q < 0.5$ слід скористатися співвідношенням $u_q = -u_{1-q}$. Наприклад, $u_{0,4} = -u_{1-0,4} = -u_{0,6} = -0,253347$.
2. Значення q -квантилей u_q можна розрахувати у Microsoft Excel за допомогою функції НОРМ.ОБР(q ; 0; 1) категорії «Статистичні». Наприклад, щоб розрахувати $u_{0,95}$, слід задати формулу '=НОРМ.ОБР(0.95; 0; 1)'.

В.2 Розподіл Пірсона χ^2

Табл. В.2. Критичні значення χ^2 -розподілу Пірсона $\chi^2(\alpha; \nu)$ для рівнів значущості α та ν ступенів вільності

$\alpha \backslash \nu$	0,995	0,99	0,975	0,95	0,9	0,1	0,05	0,025	0,01	0,005
1	0,00004	0,00016	0,00098	0,00393	0,01579	2,70554	3,84146	5,02389	6,63490	7,87944
2	0,01003	0,02010	0,05064	0,10259	0,21072	4,60517	5,99146	7,37776	9,21034	10,59663
3	0,07172	0,11483	0,21580	0,35185	0,58437	6,25139	7,81473	9,34840	11,34487	12,83816
4	0,20699	0,29711	0,48442	0,71072	1,06362	7,77944	9,48773	11,14329	13,27670	14,86026
5	0,41174	0,55430	0,83121	1,14548	1,61031	9,23636	11,07050	12,83250	15,08627	16,74960
6	0,67573	0,87209	1,23734	1,63538	2,20413	10,64464	12,59159	14,44938	16,81189	18,54758
7	0,98926	1,23904	1,68987	2,16735	2,83311	12,01704	14,06714	16,01276	18,47531	20,27774
8	1,34441	1,64650	2,17973	2,73264	3,48954	13,36157	15,50731	17,53455	20,09024	21,95495
9	1,73493	2,08790	2,70039	3,32511	4,16816	14,68366	16,91898	19,02277	21,66599	23,58935
10	2,15586	2,55821	3,24697	3,94030	4,86518	15,98718	18,30704	20,48318	23,20925	25,18818
11	2,60322	3,05348	3,81575	4,57481	5,57778	17,27501	19,67514	21,92005	24,72497	26,75685
12	3,07382	3,57057	4,40379	5,22603	6,30380	18,54935	21,02607	23,33666	26,21697	28,29952
13	3,56503	4,10692	5,00875	5,89186	7,04150	19,81193	22,36203	24,73560	27,68825	29,81947
14	4,07467	4,66043	5,62873	6,57063	7,78953	21,06414	23,68479	26,11895	29,14124	31,31935
15	4,60092	5,22935	6,26214	7,26094	8,54676	22,30713	24,99579	27,48839	30,57791	32,80132
16	5,14221	5,81221	6,90766	7,96165	9,31224	23,54183	26,29623	28,84535	31,99993	34,26719
17	5,69722	6,40776	7,56419	8,67176	10,08519	24,76904	27,58711	30,19101	33,40866	35,71847
18	6,26480	7,01491	8,23075	9,39046	10,86494	25,98942	28,86930	31,52638	34,80531	37,15645
19	6,84397	7,63273	8,90652	10,11701	11,65091	27,20357	30,14353	32,85233	36,19087	38,58226
20	7,43384	8,26040	9,59078	10,85081	12,44261	28,41198	31,41043	34,16961	37,56623	39,99685
21	8,03365	8,89720	10,28290	11,59131	13,23960	29,61509	32,67057	35,47888	38,93217	41,40106
22	8,64272	9,54249	10,98232	12,33801	14,04149	30,81328	33,92444	36,78071	40,28936	42,79565
23	9,26042	10,19572	11,68855	13,09051	14,84796	32,00690	35,17246	38,07563	41,63840	44,18128
24	9,88623	10,85636	12,40115	13,84843	15,65868	33,19624	36,41503	39,36408	42,97982	45,55851
25	10,51965	11,52398	13,11972	14,61141	16,47341	34,38159	37,65248	40,64647	44,31410	46,92789
26	11,16024	12,19815	13,84390	15,37916	17,29188	35,56317	38,88514	41,92317	45,64168	48,28988
27	11,80759	12,87850	14,57338	16,15140	18,11390	36,74122	40,11327	43,19451	46,96294	49,64492
28	12,46134	13,56471	15,30786	16,92788	18,93924	37,91592	41,33714	44,46079	48,27824	50,99338
29	13,12115	14,25645	16,04707	17,70837	19,76774	39,08747	42,55697	45,72229	49,58788	52,33562
30	13,78672	14,95346	16,79077	18,49266	20,59923	40,25602	43,77297	46,97924	50,89218	53,67196
40	20,70654	22,16426	24,43304	26,50930	29,05052	51,80506	55,75848	59,34171	63,69074	66,76596
50	27,99075	29,70668	32,35736	34,76425	37,68865	63,16712	67,50481	71,42020	76,15389	79,48998
60	35,53449	37,48485	40,48175	43,18796	46,45889	74,39701	79,08194	83,29767	88,37942	91,95170
70	43,27518	45,44172	48,75756	51,73928	55,32894	85,52704	90,53123	95,02318	100,4252	104,2149
80	51,17193	53,54008	57,15317	60,39148	64,27784	96,57820	101,8795	106,6286	112,3288	116,3211
90	59,19630	61,75408	65,64662	69,12603	73,29109	107,5650	113,1453	118,1359	124,1163	128,2989
100	67,32756	70,06489	74,22193	77,92947	82,35814	118,4980	124,3421	129,5612	135,8067	140,1695

Примітка. Критичні значення $\chi^2(\alpha; \nu)$ можна розрахувати у Microsoft Excel за допомогою функції `ХИ2.ОБР(1 - α ; ν)` категорії «Статистичні». Наприклад, для розрахунку $\chi^2(0,1; 30)$ слід задати формулу `'=ХИ2.ОБР(0,9; 30)'`.

В.3 F -розподіл Фішера

Табл. В.3. Критичні значення F -розподілу Фішера $F(\alpha; \nu_1; \nu_2)$ для рівнів значущості α із числом ступенів вільності чисельника ν_1 та знаменника ν_2

$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	11	12	15	20	25	30	40	60	120	
$\alpha = 0.1$																				
1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86	60,19	60,47	60,71	61,22	61,74	62,05	62,26	62,53	62,79	63,06	
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,40	9,41	9,42	9,44	9,45	9,46	9,47	9,47	9,48	
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,22	5,22	5,20	5,18	5,17	5,17	5,16	5,15	5,14	
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,91	3,90	3,87	3,84	3,83	3,82	3,80	3,79	3,78	
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,28	3,27	3,24	3,21	3,19	3,17	3,16	3,14	3,12	
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,92	2,90	2,87	2,84	2,81	2,80	2,78	2,76	2,74	
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,68	2,67	2,63	2,59	2,57	2,56	2,54	2,51	2,49	
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,52	2,50	2,46	2,42	2,40	2,38	2,36	2,34	2,32	
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,40	2,38	2,34	2,30	2,27	2,25	2,23	2,21	2,18	
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,30	2,28	2,24	2,20	2,17	2,16	2,13	2,11	2,08	
11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27	2,25	2,23	2,21	2,17	2,12	2,10	2,08	2,05	2,03	2,00	
12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21	2,19	2,17	2,15	2,10	2,06	2,03	2,01	1,99	1,96	1,93	
13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16	2,14	2,12	2,10	2,05	2,01	1,98	1,96	1,93	1,90	1,88	
14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12	2,10	2,07	2,05	2,01	1,96	1,93	1,91	1,89	1,86	1,83	
15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09	2,06	2,04	2,02	1,97	1,92	1,89	1,87	1,85	1,82	1,79	
16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06	2,03	2,01	1,99	1,94	1,89	1,86	1,84	1,81	1,78	1,75	
17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03	2,00	1,98	1,96	1,91	1,86	1,83	1,81	1,78	1,75	1,72	
18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00	1,98	1,95	1,93	1,89	1,84	1,80	1,78	1,75	1,72	1,69	
19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98	1,96	1,93	1,91	1,86	1,81	1,78	1,76	1,73	1,70	1,67	
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94	1,91	1,89	1,84	1,79	1,76	1,74	1,71	1,68	1,64	
21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95	1,92	1,90	1,87	1,83	1,78	1,74	1,72	1,69	1,66	1,62	
22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93	1,90	1,88	1,86	1,81	1,76	1,73	1,70	1,67	1,64	1,60	
23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92	1,89	1,87	1,84	1,80	1,74	1,71	1,69	1,66	1,62	1,59	
24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91	1,88	1,85	1,83	1,78	1,73	1,70	1,67	1,64	1,61	1,57	
25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89	1,87	1,84	1,82	1,77	1,72	1,68	1,66	1,63	1,59	1,56	
26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88	1,86	1,83	1,81	1,76	1,71	1,67	1,65	1,61	1,58	1,54	
27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87	1,85	1,82	1,80	1,75	1,70	1,66	1,64	1,60	1,57	1,53	
28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87	1,84	1,81	1,79	1,74	1,69	1,65	1,63	1,59	1,56	1,52	
29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86	1,83	1,80	1,78	1,73	1,68	1,64	1,62	1,58	1,55	1,51	
30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85	1,82	1,79	1,77	1,72	1,67	1,63	1,61	1,57	1,54	1,50	
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	1,76	1,74	1,71	1,66	1,61	1,57	1,54	1,51	1,47	1,42	
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	1,71	1,68	1,66	1,60	1,54	1,50	1,48	1,44	1,40	1,35	
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	1,65	1,63	1,60	1,55	1,48	1,44	1,41	1,37	1,32	1,26	
$\alpha = 0.05$																				
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,40	19,41	19,43	19,45	19,46	19,46	19,47	19,48	19,49	
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,76	8,74	8,70	8,66	8,63	8,62	8,59	8,57	8,55	
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	5,86	5,80	5,77	5,75	5,72	5,69	5,66	
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	4,62	4,56	4,52	4,50	4,46	4,43	4,40	
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	3,94	3,87	3,83	3,81	3,77	3,74	3,70	
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	3,51	3,44	3,40	3,38	3,34	3,30	3,27	
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	3,22	3,15	3,11	3,08	3,04	3,01	2,97	
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,10	3,07	3,01	2,94	2,89	2,86	2,83	2,79	2,75	
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	2,85	2,77	2,73	2,70	2,66	2,62	2,58	
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	2,72	2,65	2,60	2,57	2,53	2,49	2,45	
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	2,62	2,54	2,50	2,47	2,43	2,38	2,34	
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	2,53	2,46	2,41	2,38	2,34	2,30	2,25	
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	2,46	2,39	2,34	2,31	2,27	2,22	2,18	
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	2,40	2,33	2,28	2,25	2,20	2,16	2,11	
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	2,35	2,28	2,23	2,19	2,15	2,11	2,06	
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	2,31	2,23	2,18	2,15	2,10	2,06	2,01	
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	2,27	2,19	2,14	2,11	2,06	2,02	1,97	

Табл. В.3. Критичні значення F -розподілу Фішера $F(\alpha; \nu_1; \nu_2)$ для рівнів значущості α із числом ступенів вільності чисельника ν_1 та знаменника ν_2

$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9	10	11	12	15	20	25	30	40	60	120
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	2,23	2,16	2,11	2,07	2,03	1,98	1,93
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	2,20	2,12	2,07	2,04	1,99	1,95	1,90
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	2,18	2,10	2,05	2,01	1,96	1,92	1,87
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	2,15	2,07	2,02	1,98	1,94	1,89	1,84
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,24	2,20	2,13	2,05	2,00	1,96	1,91	1,86	1,81
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	2,11	2,03	1,97	1,94	1,89	1,84	1,79
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	2,09	2,01	1,96	1,92	1,87	1,82	1,77
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	2,07	1,99	1,94	1,90	1,85	1,80	1,75
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	2,06	1,97	1,92	1,88	1,84	1,79	1,73
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	2,04	1,96	1,91	1,87	1,82	1,77	1,71
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	2,03	1,94	1,89	1,85	1,81	1,75	1,70
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	2,01	1,93	1,88	1,84	1,79	1,74	1,68
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	1,92	1,84	1,78	1,74	1,69	1,64	1,58
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	1,84	1,75	1,69	1,65	1,59	1,53	1,47
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	1,75	1,66	1,60	1,55	1,50	1,43	1,35

Примітки.

1. При обчисленні критичних значень для $\alpha \geq 0.9$ слід скористатися тотожністю

$$F(\alpha; \nu_1; \nu_2) = (F(1 - \alpha; \nu_1; \nu_2))^{-1}.$$

2. Критичні значення $F(\alpha; \nu_1; \nu_2)$ можна розрахувати у Microsoft Excel за допомогою функції $\text{F.ОБР}(1 - \alpha; \nu_1; \nu_2)$ категорії «Статистичні». Наприклад, для розрахунку $F(0.1; 5; 10)$ слід задати формулу $'=\text{F.ОБР}(0.9; 5; 10)'$.

В.4 Розподіл Ст'юдента

Табл. В.4. Критичні значення $t(\alpha; \nu)$ розподілу Ст'юдента для рівнів значущості α з ν ступенями вільності

$\nu \backslash \alpha$	0,2	0,1	0,05	0,025	$\nu \backslash \alpha$	0,2	0,1	0,05	0,025
1	3,0777	6,3138	12,7062	25,4517	31	1,3095	1,6955	2,0395	2,3556
2	1,8856	2,9200	4,3027	6,2053	32	1,3086	1,6939	2,0369	2,3518
3	1,6377	2,3534	3,1824	4,1765	33	1,3077	1,6924	2,0345	2,3483
4	1,5332	2,1318	2,7764	3,4954	34	1,3070	1,6909	2,0322	2,3451
5	1,4759	2,0150	2,5706	3,1634	35	1,3062	1,6896	2,0301	2,3420
6	1,4398	1,9432	2,4469	2,9687	36	1,3055	1,6883	2,0281	2,3391
7	1,4149	1,8946	2,3646	2,8412	37	1,3049	1,6871	2,0262	2,3363
8	1,3968	1,8595	2,3060	2,7515	38	1,3042	1,6860	2,0244	2,3337
9	1,3830	1,8331	2,2622	2,6850	39	1,3036	1,6849	2,0227	2,3313
10	1,3722	1,8125	2,2281	2,6338	40	1,3031	1,6839	2,0211	2,3289
11	1,3634	1,7959	2,2010	2,5931	41	1,3025	1,6829	2,0195	2,3267
12	1,3562	1,7823	2,1788	2,5600	42	1,3020	1,6820	2,0181	2,3246
13	1,3502	1,7709	2,1604	2,5326	43	1,3016	1,6811	2,0167	2,3226
14	1,3450	1,7613	2,1448	2,5096	44	1,3011	1,6802	2,0154	2,3207
15	1,3406	1,7531	2,1314	2,4899	45	1,3006	1,6794	2,0141	2,3189
16	1,3368	1,7459	2,1199	2,4729	46	1,3002	1,6787	2,0129	2,3172
17	1,3334	1,7396	2,1098	2,4581	47	1,2998	1,6779	2,0117	2,3155
18	1,3304	1,7341	2,1009	2,4450	48	1,2994	1,6772	2,0106	2,3139
19	1,3277	1,7291	2,0930	2,4334	49	1,2991	1,6766	2,0096	2,3124
20	1,3253	1,7247	2,0860	2,4231	50	1,2987	1,6759	2,0086	2,3109
21	1,3232	1,7207	2,0796	2,4138	60	1,2958	1,6706	2,0003	2,2990
22	1,3212	1,7171	2,0739	2,4055	70	1,2938	1,6669	1,9944	2,2906
23	1,3195	1,7139	2,0687	2,3979	80	1,2922	1,6641	1,9901	2,2844
24	1,3178	1,7109	2,0639	2,3909	90	1,2910	1,6620	1,9867	2,2795
25	1,3163	1,7081	2,0595	2,3846	100	1,2901	1,6602	1,9840	2,2757
26	1,3150	1,7056	2,0555	2,3788	110	1,2893	1,6588	1,9818	2,2725
27	1,3137	1,7033	2,0518	2,3734	120	1,2886	1,6577	1,9799	2,2699
28	1,3125	1,7011	2,0484	2,3685	130	1,2881	1,6567	1,9784	2,2677
29	1,3114	1,6991	2,0452	2,3638	140	1,2876	1,6558	1,9771	2,2658
30	1,3104	1,6973	2,0423	2,3596	150	1,2872	1,6551	1,9759	2,2641

Примітка. Критичні значення $t(\alpha; \nu)$ можна розрахувати у Microsoft Excel за допомогою функції СТЬЮДЕНТ.ОБР($1 - 0.5\alpha; \nu$) категорії «Статистичні». Наприклад, для розрахунку $t(0.05; 30)$ слід задати формулу '=СТЮДЕНТ.ОБР(1-0.025; 30)'.

В.5 z -перетворення Фішера

Табл. В.5. z -перетворення Фішера вибіркового коефіцієнта кореляції \hat{r}

\hat{r}	0,000	0,002	0,004	0,006	0,008	\hat{r}	0,000	0,002	0,004	0,006	0,008
0,00	0,0000	0,0020	0,0040	0,0060	0,0080	0,50	0,5493	0,5520	0,5547	0,5573	0,5600
0,01	0,0100	0,0120	0,0140	0,0160	0,0180	0,51	0,5627	0,5654	0,5682	0,5709	0,5736
0,02	0,0200	0,0220	0,0240	0,0260	0,0280	0,52	0,5763	0,5791	0,5818	0,5846	0,5874
0,03	0,0300	0,0320	0,0340	0,0360	0,0380	0,53	0,5901	0,5929	0,5957	0,5985	0,6013
0,04	0,0400	0,0420	0,0440	0,0460	0,0480	0,54	0,6042	0,6070	0,6098	0,6127	0,6155
0,05	0,0500	0,0520	0,0541	0,0561	0,0581	0,55	0,6184	0,6213	0,6241	0,6270	0,6299
0,06	0,0601	0,0621	0,0641	0,0661	0,0681	0,56	0,6328	0,6358	0,6387	0,6416	0,6446
0,07	0,0701	0,0721	0,0741	0,0761	0,0782	0,57	0,6475	0,6505	0,6535	0,6565	0,6595
0,08	0,0802	0,0822	0,0842	0,0862	0,0882	0,58	0,6625	0,6655	0,6685	0,6716	0,6746
0,09	0,0902	0,0923	0,0943	0,0963	0,0983	0,59	0,6777	0,6807	0,6838	0,6869	0,6900
0,10	0,1003	0,1024	0,1044	0,1064	0,1084	0,60	0,6931	0,6963	0,6994	0,7026	0,7057
0,11	0,1104	0,1125	0,1145	0,1165	0,1186	0,61	0,7089	0,7121	0,7153	0,7185	0,7218
0,12	0,1206	0,1226	0,1246	0,1267	0,1287	0,62	0,7250	0,7283	0,7315	0,7348	0,7381
0,13	0,1307	0,1328	0,1348	0,1368	0,1389	0,63	0,7414	0,7447	0,7481	0,7514	0,7548
0,14	0,1409	0,1430	0,1450	0,1471	0,1491	0,64	0,7582	0,7616	0,7650	0,7684	0,7718
0,15	0,1511	0,1532	0,1552	0,1573	0,1593	0,65	0,7753	0,7788	0,7823	0,7858	0,7893
0,16	0,1614	0,1634	0,1655	0,1676	0,1696	0,66	0,7928	0,7964	0,7999	0,8035	0,8071
0,17	0,1717	0,1737	0,1758	0,1779	0,1799	0,67	0,8107	0,8144	0,8180	0,8217	0,8254
0,18	0,1820	0,1841	0,1861	0,1882	0,1903	0,68	0,8291	0,8328	0,8366	0,8404	0,8441
0,19	0,1923	0,1944	0,1965	0,1986	0,2007	0,69	0,8480	0,8518	0,8556	0,8595	0,8634
0,20	0,2027	0,2048	0,2069	0,2090	0,2111	0,70	0,8673	0,8712	0,8752	0,8792	0,8832
0,21	0,2132	0,2153	0,2174	0,2195	0,2216	0,71	0,8872	0,8912	0,8953	0,8994	0,9035
0,22	0,2237	0,2258	0,2279	0,2300	0,2321	0,72	0,9076	0,9118	0,9160	0,9202	0,9245
0,23	0,2342	0,2363	0,2384	0,2405	0,2427	0,73	0,9287	0,9330	0,9373	0,9417	0,9461
0,24	0,2448	0,2469	0,2490	0,2512	0,2533	0,74	0,9505	0,9549	0,9594	0,9639	0,9684
0,25	0,2554	0,2575	0,2597	0,2618	0,2640	0,75	0,9730	0,9775	0,9822	0,9868	0,9915
0,26	0,2661	0,2683	0,2704	0,2726	0,2747	0,76	0,9962	1,0010	1,0058	1,0106	1,0154
0,27	0,2769	0,2790	0,2812	0,2833	0,2855	0,77	1,0203	1,0253	1,0302	1,0352	1,0403
0,28	0,2877	0,2899	0,2920	0,2942	0,2964	0,78	1,0454	1,0505	1,0557	1,0609	1,0661
0,29	0,2986	0,3008	0,3029	0,3051	0,3073	0,79	1,0714	1,0768	1,0822	1,0876	1,0931
0,30	0,3095	0,3117	0,3139	0,3161	0,3183	0,80	1,0986	1,1042	1,1098	1,1155	1,1212
0,31	0,3205	0,3228	0,3250	0,3272	0,3294	0,81	1,1270	1,1329	1,1388	1,1447	1,1507
0,32	0,3316	0,3339	0,3361	0,3383	0,3406	0,82	1,1568	1,1630	1,1692	1,1754	1,1817
0,33	0,3428	0,3451	0,3473	0,3496	0,3518	0,83	1,1881	1,1946	1,2011	1,2077	1,2144
0,34	0,3541	0,3564	0,3586	0,3609	0,3632	0,84	1,2212	1,2280	1,2349	1,2419	1,2490
0,35	0,3654	0,3677	0,3700	0,3723	0,3746	0,85	1,2562	1,2634	1,2707	1,2782	1,2857
0,36	0,3769	0,3792	0,3815	0,3838	0,3861	0,86	1,2933	1,3011	1,3089	1,3169	1,3249

Табл. В.5. z -перетворення Фішера вибіркового коефіцієнта кореляції \hat{r}

\hat{r}	0,000	0,002	0,004	0,006	0,008	\hat{r}	0,000	0,002	0,004	0,006	0,008
0,37	0,3884	0,3907	0,3931	0,3954	0,3977	0,87	1,3331	1,3414	1,3498	1,3583	1,3670
0,38	0,4001	0,4024	0,4047	0,4071	0,4094	0,88	1,3758	1,3847	1,3938	1,4030	1,4124
0,39	0,4118	0,4142	0,4165	0,4189	0,4213	0,89	1,4219	1,4316	1,4415	1,4516	1,4618
0,40	0,4236	0,4260	0,4284	0,4308	0,4332	0,90	1,4722	1,4828	1,4937	1,5047	1,5160
0,41	0,4356	0,4380	0,4404	0,4428	0,4453	0,91	1,5275	1,5393	1,5513	1,5636	1,5762
0,42	0,4477	0,4501	0,4526	0,4550	0,4574	0,92	1,5890	1,6022	1,6157	1,6296	1,6438
0,43	0,4599	0,4624	0,4648	0,4673	0,4698	0,93	1,6584	1,6734	1,6888	1,7047	1,7211
0,44	0,4722	0,4747	0,4772	0,4797	0,4822	0,94	1,7380	1,7555	1,7736	1,7923	1,8117
0,45	0,4847	0,4872	0,4897	0,4922	0,4948	0,95	1,8318	1,8527	1,8745	1,8972	1,9210
0,46	0,4973	0,4999	0,5024	0,5049	0,5075	0,96	1,9459	1,9721	1,9996	2,0287	2,0595
0,47	0,5101	0,5126	0,5152	0,5178	0,5204	0,97	2,0923	2,1273	2,1649	2,2054	2,2494
0,48	0,5230	0,5256	0,5282	0,5308	0,5334	0,98	2,2976	2,3507	2,4101	2,4774	2,5550
0,49	0,5361	0,5387	0,5413	0,5440	0,5466	0,99	2,6467	2,7587	2,9031	3,1063	3,4534

Примітки.

1. Наприклад, дано $z = 0.8752$. Для того, щоб знайти відповідне йому значення \hat{r} , потрібно скласти числа, що відповідають рядку і стовпчику таблиці, на перетині яких знаходиться значення 0.8752. В даному випадку маємо $\hat{r} = 0.70 + 0.004 = 0.704$.
2. Якщо $z < 0$, то і відповідне йому значення $r < 0$. Якщо $z = -0.8752$, то $\hat{r} = -0.704$.
3. У тих випадках, коли у таблиці немає у точності потрібного значення z , беруть два найближчих до нього (зверху і знизу) і знаходять відповідні їм значення \hat{r}_H і \hat{r}_B . Шукане значення \hat{r} буде належати інтервалу $[\hat{r}_H; \hat{r}_B]$.
4. Значення \hat{r} , відповідне z , можна знайти у Microsoft Excel, використавши функцію **ФИШЕРОБР**(z) категорії «Статистичні». Наприклад, якщо $z = 0.8752$, задаємо формулу '=ФИШЕРОБР(0,8752)'

В.6 Статистика Дарбіна — Уотсона

Табл. В.6. Значення статистик Дарбіна — Уотсона d_0 і d_u при рівні значущості $\alpha = 0.05$ (n — кількість спостережень, p — кількість пояснюючих змінних без урахування константи)

n	$p = 1$		$p = 2$		$p = 3$		$p = 4$		$p = 5$		$p = 6$		$p = 7$		$p = 8$		$p = 9$		$p = 10$	
	d_0	d_u	d_0	d_u	d_0	d_u	d_0	d_u	d_0	d_u	d_0	d_u	d_0	d_u	d_0	d_u	d_0	d_u	d_0	d_u
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	<u>2,29</u>	—	—	—	—	—	—	—	—	—	—	—	—	—	—
9	0,82	1,32	0,63	1,70	0,46	<u>2,13</u>	0,30	<u>2,59</u>	—	—	—	—	—	—	—	—	—	—	—	—
10	0,88	1,32	0,70	1,64	0,53	<u>2,02</u>	0,38	<u>2,41</u>	0,24	<u>2,82</u>	—	—	—	—	—	—	—	—	—	—
11	0,93	1,32	0,66	1,60	0,60	1,93	0,44	<u>2,28</u>	0,32	<u>2,65</u>	0,20	<u>3,01</u>	—	—	—	—	—	—	—	—
12	0,97	1,33	0,81	1,58	0,66	1,86	0,51	<u>2,18</u>	0,38	<u>2,51</u>	0,27	<u>2,83</u>	0,17	<u>3,15</u>	—	—	—	—	—	—
13	1,01	1,34	0,86	1,56	0,72	1,82	0,57	<u>2,09</u>	0,45	<u>2,39</u>	0,33	<u>2,69</u>	0,23	<u>2,99</u>	0,15	<u>3,27</u>	—	—	—	—
14	1,05	1,35	0,91	1,55	0,77	1,78	0,63	<u>2,03</u>	0,51	<u>2,30</u>	0,39	<u>2,57</u>	0,29	<u>2,85</u>	0,20	<u>3,11</u>	0,13	<u>3,36</u>	—	—
15	1,08	1,36	0,95	1,54	0,81	1,75	0,69	1,98	0,56	<u>2,22</u>	0,45	<u>2,47</u>	0,34	<u>2,73</u>	0,25	<u>2,98</u>	0,18	<u>3,22</u>	0,11	<u>3,44</u>
16	1,11	1,37	0,98	1,54	0,86	1,73	0,73	1,94	0,62	<u>2,16</u>	0,50	<u>2,39</u>	0,40	<u>2,62</u>	0,30	<u>2,86</u>	0,22	<u>3,09</u>	0,16	<u>3,30</u>
17	1,13	1,38	1,02	1,54	0,90	1,71	0,78	1,90	0,66	<u>2,10</u>	0,55	<u>2,32</u>	0,45	<u>2,54</u>	0,36	<u>2,76</u>	0,27	<u>2,98</u>	0,20	<u>3,18</u>
18	1,16	1,39	1,05	1,54	0,93	1,70	0,82	1,87	0,71	<u>2,06</u>	0,60	<u>2,26</u>	0,50	<u>2,46</u>	0,41	<u>2,67</u>	0,32	<u>2,87</u>	0,24	<u>3,07</u>
19	1,18	1,40	1,07	1,54	0,97	1,69	0,86	1,85	0,75	<u>2,02</u>	0,65	<u>2,21</u>	0,55	<u>2,40</u>	0,46	<u>2,59</u>	0,37	<u>2,78</u>	0,29	<u>2,97</u>
20	1,20	1,41	1,10	1,54	1,00	1,68	0,98	1,83	0,79	1,99	0,69	<u>2,16</u>	0,60	<u>2,34</u>	0,50	<u>2,52</u>	0,42	<u>2,70</u>	0,34	<u>2,89</u>
21	1,22	1,42	1,13	1,54	1,03	1,67	0,93	1,81	0,83	1,96	0,73	<u>2,12</u>	0,64	<u>2,29</u>	0,55	<u>2,46</u>	0,46	<u>2,63</u>	0,38	<u>2,81</u>
22	1,24	1,43	1,15	1,54	1,05	1,66	0,96	1,80	0,86	1,94	0,77	<u>2,09</u>	0,68	<u>2,25</u>	0,59	<u>2,41</u>	0,50	<u>2,57</u>	0,42	<u>2,73</u>
23	1,26	1,44	1,17	1,54	1,08	1,66	0,99	1,79	0,90	1,92	0,80	<u>2,06</u>	0,72	<u>2,21</u>	0,63	<u>2,36</u>	0,55	<u>2,51</u>	0,47	<u>2,67</u>
24	1,27	1,45	1,19	1,55	1,10	1,66	1,01	1,78	0,93	1,90	0,84	<u>2,04</u>	0,75	<u>2,17</u>	0,67	<u>2,32</u>	0,58	<u>2,46</u>	0,51	<u>2,61</u>
25	1,29	1,45	1,21	1,55	1,12	1,65	1,04	1,77	0,95	1,89	0,87	<u>2,01</u>	0,78	<u>2,14</u>	0,70	<u>2,28</u>	0,62	<u>2,42</u>	0,54	<u>2,56</u>
26	1,30	1,46	1,22	1,55	1,14	1,65	1,06	1,76	0,98	1,87	0,90	1,99	0,82	<u>2,12</u>	0,74	<u>2,25</u>	0,66	<u>2,38</u>	0,58	<u>2,51</u>
27	1,32	1,47	1,24	1,56	1,16	1,65	1,08	1,75	1,00	1,86	0,93	1,97	0,85	<u>2,09</u>	0,77	<u>2,22</u>	0,69	<u>2,34</u>	0,62	<u>2,47</u>
28	1,33	1,48	1,26	1,56	1,18	1,65	1,10	1,75	1,03	1,85	0,95	1,96	0,87	<u>2,07</u>	0,80	<u>2,19</u>	0,72	<u>2,31</u>	0,65	<u>2,43</u>
29	1,34	1,48	1,27	1,56	1,20	1,65	1,12	1,74	1,05	1,84	0,98	1,94	0,90	<u>2,05</u>	0,83	<u>2,16</u>	0,75	<u>2,28</u>	0,68	<u>2,40</u>
30	1,35	1,49	1,28	1,57	1,21	1,65	1,14	1,74	1,07	1,83	1,00	1,93	0,93	<u>2,03</u>	0,85	<u>2,14</u>	0,78	<u>2,25</u>	0,71	<u>2,36</u>
31	1,36	1,50	1,30	1,57	1,23	1,65	1,16	1,74	1,09	1,83	1,02	1,92	0,95	<u>2,02</u>	0,88	<u>2,12</u>	0,81	<u>2,23</u>	0,74	<u>2,33</u>
32	1,37	1,50	1,31	1,57	1,24	1,65	1,18	1,73	1,11	1,82	1,04	1,91	0,97	<u>2,00</u>	0,90	<u>2,10</u>	0,84	<u>2,20</u>	0,77	<u>2,31</u>
33	1,38	1,51	1,32	1,58	1,26	1,65	1,19	1,73	1,13	1,81	1,06	1,90	0,99	1,99	0,93	<u>2,09</u>	0,86	<u>2,18</u>	0,80	<u>2,28</u>
34	1,39	1,51	1,33	1,58	1,27	1,65	1,03	1,73	1,14	1,81	1,08	1,89	1,02	1,98	0,95	<u>2,07</u>	0,89	<u>2,16</u>	0,82	<u>2,26</u>
35	1,40	1,52	1,34	1,58	1,28	1,65	1,22	1,73	1,16	1,80	1,10	1,88	1,03	1,97	0,97	<u>2,05</u>	0,91	<u>2,14</u>	0,85	<u>2,24</u>
36	1,41	1,53	1,35	1,59	1,30	1,65	1,24	1,72	1,18	1,80	1,11	1,88	1,05	1,96	0,99	<u>2,04</u>	0,93	<u>2,13</u>	0,87	<u>2,22</u>
37	1,42	1,53	1,36	1,59	1,31	1,66	1,25	1,72	1,19	1,80	1,13	1,87	1,07	1,95	1,01	<u>2,03</u>	0,95	<u>2,11</u>	0,89	<u>2,20</u>
38	1,43	1,54	1,37	1,59	1,32	1,66	1,26	1,72	1,20	1,79	1,15	1,86	1,09	1,94	1,03	<u>2,02</u>	0,97	<u>2,10</u>	0,91	<u>2,18</u>
39	1,44	1,54	1,38	1,60	1,33	1,66	1,27	1,72	1,22	1,79	1,16	1,86	1,10	1,93	1,05	<u>2,01</u>	0,99	<u>2,09</u>	0,93	<u>2,16</u>
40	1,44	1,54	1,39	1,60	1,34	1,66	1,29	1,72	1,23	1,79	1,18	1,85	1,12	1,92	1,06	2,00	1,01	<u>2,07</u>	0,95	<u>2,15</u>
45	1,48	1,57	1,43	1,62	1,38	1,67	1,34	1,72	1,29	1,78	1,24	1,84	1,19	1,90	1,14	1,96	1,09	<u>2,02</u>	1,04	<u>2,09</u>
50	1,50	1,59	1,46	1,63	1,42	1,67	1,38	1,72	1,34	1,77	1,29	1,82	1,25	1,88	1,20	1,93	1,16	1,99	1,11	<u>2,04</u>
55	1,53	1,60	1,49	1,64	1,45	1,68	1,41	1,72	1,37	1,77	1,33	1,81	1,29	1,86	1,25	1,91	1,21	1,96	1,17	<u>2,01</u>
60	1,55	1,62	1,51	1,65	1,48	1,69	1,44	1,73	1,41	1,77	1,37	1,81	1,34	1,85	1,30	1,89	1,26	1,94	1,22	1,98
65	1,57	1,63	1,54	1,66	1,50	1,70	1,47	1,73	1,44	1,77	1,40	1,81	1,37	1,84	1,34	1,88	1,30	1,92	1,27	1,96
70	1,58	1,64	1,55	1,67	1,53	1,70	1,49	1,74	1,46	1,77	1,43	1,80	1,40	1,84	1,37	1,87	1,34	1,91	1,31	1,95
75	1,60	1,65	1,57	1,68	1,54	1,71	1,52	1,74	1,49	1,77	1,46	1,80	1,43	1,83	1,40	1,87	1,37	1,90	1,34	1,94
80	1,61	1,66	1,59	1,69	1,56	1,72	1,53	1,74	1,51	1,77	1,48	1,80	1,45	1,83	1,43	1,86	1,40	1,89	1,37	1,93
85	1,62	1,67	1,60	1,70	1,58	1,72	1,55	1,75	1,53	1,77	1,50	1,80	1,47	1,83	1,45	1,86	1,42	1,89	1,40	1,92
90	1,64	1,68	1,61	1,70	1,59	1,73	1,57	1,75	1,54	1,78	1,52	1,80	1,49	1,83	1,47	1,85	1,45	1,88	1,42	1,91
95	1,65	1,69	1,62	1,71	1,60	1,73	1,58	1,76	1,56	1,78	1,54	1,80	1,51	1,83	1,49	1,85	1,47	1,88	1,44	1,90
100	1,65	1,69	1,63	1,72	1,61	1,74	1,59	1,76	1,57	1,78	1,55	1,80	1,53	1,83	1,51	1,85	1,48	1,87	1,46	1,90
150	1,72	1,75	1,71	1,72	1,69	1,77	1,68	1,79	1,67	1,80	1,65	1,82	1,64	1,83	1,62	1,85	1,61	1,86	1,59	1,88
200	1,76	1,78	1,75	1,79	1,74	1,80	1,73	1,81	1,72	1,82	1,71	1,83	1,70	1,84	1,69	1,85	1,68	1,86	1,67	1,87

Примітка. Підкреслені значення d_u означають неможливість отримання оцінки відсутності автокореляції залишків за допомогою даного критерію при відповідних n і p .

ДЛЯ НОТАТОК

Навчальне видання

А. Т. Яровий, Є. М. Страхов

ЕКОНОМЕТРІЯ

Підписано до друку 16.01.2017 р.
Формат 60×84/16. Папір офсетний.
Ум. друк. арк. 7,56.
Наклад 100 прим.

ФОП «Сухачов»

65023, м. Одеса, вул. Садова, 14
Свідоцтво про внесення до Державного реєстру
суб'єктів видавничої справи ДК №4362 від 13.07.2012 р.
Тел.: (098) 457-75-00, (048) 706-78-70, (067) 798-05-70, (094) 931-98-70
E-mail: osvita.odessa@gmail.com, www.svitknigi.com

Видавництво «Освіта України»

04136, м. Київ, вул. Маршала Гречка, 13, оф. 218
Свідоцтво про внесення до Державного реєстру
суб'єктів видавничої справи ДК №1957 від 23.04.2009 р.
Тел./факс (044) 507-04-52, (044) 228-81-29, (063) 237-59-92
E-mail: osvita2005@mail.ru, www.rambook.ru

Видавництво «Освіта України» запрошує авторів до співпраці
з випуску видань, що стосуються питань управління,
модернізації, інноваційних процесів, технологій,
методичних і методологічних аспектів освіти та навчального процесу
у вищих навчальних закладах.

Надаємо всі види видавничих та поліграфічних послуг.

