

МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ
УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
"ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ"
АВТОМОБІЛЬНО-ДОРОЖНІЙ ІНСТИТУТ

Кафедра «Прикладна математика та інформатика»

**МЕТОДИЧНІ ВКАЗІВКИ
ДО ВИКОНАННЯ ПРАКТИЧНИХ РОБІТ
З ДИСЦИПЛІНИ " ЕКОНОМЕТРИКА"
(ДЛЯ СТУДЕНТІВ НАПРЯМУ 6.030502
"ЕКОНОМІЧНА КІБЕРНЕТИКА")**

22/18-2012-02

Горлівка – 2014

МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ
УКРАЇНИ
ДЕРЖАВНИЙ ВИЩИЙ НАВЧАЛЬНИЙ ЗАКЛАД
"ДОНЕЦЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ"
АВТОМОБІЛЬНО-ДОРОЖНІЙ ІНСТИТУТ

«ЗАТВЕРДЖУЮ»
Директор АДІ ДВНЗ «ДонНТУ»
М.М. Чальцев
06.01.2011

Кафедра «Прикладна математика та інформатика»

**МЕТОДИЧНІ ВКАЗІВКИ
ДО ВИКОНАННЯ ПРАКТИЧНИХ РОБІТ
З ДИСЦИПЛІНИ "ЕКОНОМЕТРИКА"
(ДЛЯ СТУДЕНТІВ НАПРЯМУ 6.030502
"ЕКОНОМІЧНА КІБЕРНЕТИКА")**

22/17-2012-02

«РЕКОМЕНДОВАНО»
Методична комісія факультету
«Економіка і Управління»
протокол №9 від 15.06.2012

«РЕКОМЕНДОВАНО»
Кафедра «Прикладна
математика
та інформатика»
протокол №8 від 27.03.2012

УДК

Методичні вказівки до виконання практичних робіт з дисципліни "Багатовимірний статистичний аналіз" (для студентів напряму підготовки 6.030502 "Економічна кібернетика") [Електронний ресурс] / укладачі В. Г. Хребет, Д. В. Фесенко.–Електрон. дані – Горлівка: ДВНЗ "ДонНТУ" АДІ, 2012. – 1 електрон. опт. диск (CD-R); 2 см. – Систем. вимоги: Pentium; 32 RAM; WINDOWS 98/2000/NT/XP;MS Word 2000. – Назва з титул. екрану.

Укладачі:	Хребет В. Г., к.ф.м.н., доц., Фесенко Д. В., асистент.
Відповідальний за випуск:	Хребет В. Г., к.ф.м.н., доц.
Рецензент:	Ніколаєнко В. Л., к.т.н., доц

©Державний вищий навчальний заклад
"Донецький національний технічний університет"
Автомобільно-Дорожній інститут, 2012.

ЗМІСТ

Вступ

Практична робота №1: Парна регресійна модель

Практична робота №2: Багатофакторна лінійна регресійна модель

Практична робота №3: Узагальнена лінійна модель

Практична робота №4: Системи регресійних рівнянь

Перелік рекомендованої літератури

ВСТУП

Курс багатовимірного статистичного аналізу є розділом прикладної статистики, який призначений для рішення наступних задач:

- 1) Вивчення структури та сили статистичного зв'язку між системами випадкових величин;
- 2) Заміна вихідного набору статистичних даних великої розмірності значно меншим масивом даних без істотної втрати інформації – зменшення розмірності;
- 3) Розбиття сукупності статистичних величин на однорідні в деякому відношенні групи – класифікація.

Аналіз та моделювання економічних систем, які є основними цілями економічної кібернетики неможливі без вирішення перелічених задач. Оскільки аналіз системи полягає у розкладанні її на складові частини та встановлення структури взаємодії між ними, а моделювання - це передусім максимально точне формулювання досліджуваного процесу або явища у якомога компактному вигляді методи багатовимірного аналізу є вкрай необхідними для підготовки фахівців із економічної кібернетики.

Дані методичні вказівки призначені для формування навичок розв'язання усіх проблем, які зустрічаються у багатовимірному статистичному аналізі: перша робота присвячена методам вимірювання кореляційних зв'язків між випадковими величинами, наступні три роботи – компонентний, факторний аналіз та шкалювання розв'язують задачу зниження розмірності та останні дві задачі – дискримінантний та кластерний аналізи формують навички автоматичної класифікації.

У зв'язку з тим, що більшість методів багатовимірного статистичного аналізу є досить трудомісткими вкрай необхідним є застосування спеціалізованих математичних пакетів, у якості якого була обрана популярна програма MathCad.

Розподіл робочого часу на виконання робіт

п/п	Робота	Витрати часу
	Парна регресія	
	Багатофакторна регресія	
	Узагальнена регресійна модель	
	Системи регресійних рівнянь	
	Всього	

Практичне заняття №1

Тема: «Парна регресійна модель»

План:

1. Метод найменших квадратів.
2. Припущення, покладені в основу парної лінійної моделі.
3. Властивості оцінок МНК.
4. Перевірка гіпотез на основі оцінок МНК.
5. Прогнозування.
6. Приведення нелінійних моделей к лінійному вигляду.

1. Метод найменших квадратів – один зі статистичних методів що дозволяє досліджувати залежність деякої функції Y від змінних X_1, X_2, \dots, X_k використовуючи дані спостережень змінних і функції. У багатьох економічних системах функція Y часто залежить лінійним образом тільки від однієї змінної:

$$y_i = a_0 + a_1 x_i$$

Однак у дійсності спостережувані значення функції рівні:

$$y_i = a_0 + a_1 x_i + u_i$$

Поява u_i викликана в основному такими причинами:

1. Функція у насправді може залежати від декількох змінних.
2. Можлива помилкова структура моделі (тобто залежність може бути нелінійною).
3. Помилками вимірювання.

Завдання полягає в тім, щоб на основі n пар значень (x_i, y_i) одержати уявлення про невідомі параметри регресії a_0 й a_1 . Відповідно до методу найменших квадратів (МНК) оцінками цих параметрів беруться такі числа \hat{a}_0 й \hat{a}_1 , які мінімізують цільову функцію:

$$z = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

де e_i – відхилення експериментальних даних від регресійної прямої;

$$\hat{y}_i = \hat{a}_0 + \hat{a}_1 x_i - \text{оцінені значення } a_0 + a_1 x_i.$$

Рішенням цієї задачі є

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

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

2. Для перевірки статистичних гіпотез про параметри моделі необхідно зробити припущення щодо того, що вносить імовірнісний характер у залежність – випадкових величинах u_i . По перше, ці величини очевидно складаються з великої кількості випадкових величин, кожна з яких набагато менше по абсолютній величині в порівнянні з $a_0 + a_1 x_i$ - зміною значення функції обумовленою зміною змінної. У зв'язку із цим і опираючись на центральну граничну теорему можна припустити, що:

I. Величини u_i мають нормальний закон розподілу.

Багато які з випадкових величин, що входять в u_i можуть діяти в протилежних напрямках, тому середнє значення їхньої суми дорівнює нулю або точніше:

II. Математичне очікування випадкових впливів дорівнює нулю.

$$E(u_i) = 0$$

Значення u_i у різні моменти часу і не залежать друг від друга. З курсу математичної статистики відомо, що необхідною й достатньою умовою незалежності нормально - розподілених величин є рівність нулю їх коваріації:

$$\text{III. } \text{cov}(u_i, u_j) = E((u_i - E(u_i))(u_j - E(u_j))) = E(u_i u_j) = \begin{cases} 0; i \neq j \\ \sigma_u^2, i = j \end{cases}$$

3. Використовуючи перераховані вище припущення, одержують кілька важливих результатів про характеристики оцінок МНК. В – перші оцінки МНК є лінійними функціями спостережуваних значень Y_i :

$$\hat{a}_1 = \sum_{i=1}^n \omega_i y_i$$

$$\hat{a}_0 = \sum_{i=1}^n c_i y_i$$

Де c_i, ω_i - постійні величини.

Величини \hat{a}_1, \hat{a}_0 є також лінійними функціями від випадкових збурювань u_i , а тому самі є випадковими, тобто міняються від вибірки до вибірки. Як і будь-які випадкові величини, вони мають свій закон розподілу, своє математичне очікування і свою дисперсію. Із припущень лінійної моделі випливає, що обидві оцінки розподілені за нормальним законом з параметрами:

$$E(\hat{a}_1) = a_1$$

$$E(\hat{a}_0) = a_0$$

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

$$D(\hat{a}_0) = \frac{\sigma_u^2 \sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}$$

Перші дві рівності означають, що оцінки \hat{a}_1, \hat{a}_0 є незміщеними. Якість всіх оцінок крім незміщеності характеризує також їхній розкид щодо оцінюваного параметра. Оцінка вважається тим ефективною, чим менше математичне очікування квадрата її відхилення щодо оцінюваної величини. Відповідно до теореми Маркова – Гауса оцінки МНК є найбільш ефективними в класі всіх лінійних незміщених оцінок, тобто мають найменшу дисперсію.

4. Для перевірки статистичних гіпотез необхідно спочатку одержати з наявних випадкових величин стандартні затабульовані величини.

Відомо, що величини:

$$t = \frac{(A_0 - a_0) \sqrt{n \cdot \sum (X_i - \bar{X})^2}}{\hat{\sigma}_u \sqrt{\sum X_i^2}} \quad (1.1)$$

$$t = \frac{(A_1 - a_1) \sqrt{\sum (X_i - \bar{X})^2}}{\hat{\sigma}_u} \quad (1.2)$$

мають розподіл Стюдента із числом ступенів вільності $n-k$, де n – кількість спостережень, а k – кількість оцінюваних параметрів (у випадку парної регресії їх 2). $\hat{\sigma}_u$ - незміщена оцінка дисперсії випадкових збурювань u :

$$\hat{\sigma}_u = \sqrt{\frac{\sum e_i^2}{n-k}}$$

Для перевірки гіпотез про значення параметрів необхідно підставити ті значення, що перевіряються, у відповідну формулу. Якщо виконується нерівність:

$$t \leq t_\varepsilon$$

де t_ε - критичне значення, то гіпотеза приймається на відповідному рівні значущості. Довірчі інтервали для параметрів регресії визначаються нерівностями:

$$\hat{a}_0 + t_{кр} \cdot \frac{\hat{\sigma}_u \cdot \sqrt{\sum X_i^2}}{\sqrt{n \sum (X_i - \bar{X})^2}} \geq a_0 \geq \hat{a}_0 - t_{кр} \cdot \frac{\hat{\sigma}_u \cdot \sqrt{\sum X_i^2}}{\sqrt{n \sum (X_i - \bar{X})^2}}$$

$$\hat{a}_1 + t_{кр} \cdot \frac{\hat{\sigma}_u}{\sqrt{\sum (X_i - \bar{X})^2}} \geq a_1 \geq \hat{a}_1 - t_{кр} \cdot \frac{\hat{\sigma}_u}{\sqrt{\sum (X_i - \bar{X})^2}}$$

5. Одне із застосувань регресійного аналізу полягає в тому, що він дозволяє прогнозувати невідомі значення функції при довільних значеннях аргументу. Існує два типи прогнозів:

1) Точковий прогноз – являє собою оцінку значення функції при деякому значенні змінної:

$$\hat{Y}_{np} = \hat{a}_0 + \hat{a}_1 X_{np}$$

2) Інтервальний прогноз – довірчий інтервал для значення функції при деякому значенні змінної. Відомо, що випадкова величина

$$t = \frac{\hat{Y}_{np} - Y_{np}}{\hat{\sigma}_u \sqrt{\frac{1}{n} + \frac{(X_{np} - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}}} \quad (1.3)$$

має розподіл Стюдента з $n-k$ ступенями вільності, тому з довірчою ймовірністю прогнозне значення функції буде знаходитися в інтервалі:

$$\hat{Y}_{np} \pm t_{\varepsilon/2} \hat{\sigma}_u \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2}} \quad (1.4)$$

де $t_{\varepsilon/2}$ - критичне значення розподілу Стюдента при $n-k$ ступенях вільності й $\varepsilon/2$ рівні значущості.

6. Відповідним перетворенням змінних деякі нелінійні моделі можна представити у лінійному вигляді. Взаємозв'язок між нелінійною та відповідною лінійною моделями показаний в таблиці 1.

Таблиця 1 – Перехід від нелінійної моделі до лінійної

Нелінійна модель	Перетворення змінних	Перетворення параметрів	
		пряме	зворотне
гіперболічна $y_i = \frac{a_1}{a_0 + x_i}$	$Y_i = \frac{1}{y_i}, \quad X_i = x_i$	$A_0 = \frac{a_0}{a_1}, A_1 = \frac{1}{a_1}$	$a_0 = \frac{A_0}{A_1}, \quad a_1 = \frac{1}{A_1}$
обернена квадратична $y_i = \frac{a_1}{(a_0 + x_i)^2}$	$Y_i = \frac{1}{\sqrt{y_i}}, \quad X_i = x_i$	$A_0 = \frac{a_0}{\sqrt{a_1}}, A_1 = \frac{1}{\sqrt{a_1}}$	$a_0 = \frac{A_0}{A_1}, a_1 = \frac{1}{A_1^2}$
експоненційна $y_i = a_0 \cdot e^{a_1 x_i}$	$Y_i = \ln(y_i), \quad X_i = x_i$	$A_0 = \ln(a_0), A_1 = a_1$	$a_0 = e^{A_0}, \quad a_1 = A_1$
дробно-лінійна $y_i = \frac{a_0 + a_1 x_i}{1 + x_i}$	$Y_i = y_i(1 + x_i),$ $X_i = x_i$	$A_0 = a_0, A_1 = a_1$	$a_0 = A_0, a_1 = A_1$
з постійною еластичністю $Y_i = a_0 \cdot X_i^a$	$Y_i = \ln(y_i),$ $X_i = \ln(x_i)$	$A_0 = \ln(a_0), A_1 = a$	$a_0 = e^{A_0}, \quad a_1 = A_1$

На основі статистичних даних про значення ціни X і попиту Y :

а) Оцінити параметри залежності між X і Y .

б) Перевірити гіпотезу H_0 про параметри моделі і побудувати довірчі інтервали для параметрів регресії.

в) Розв'язати прикладну задачу згідно варіанта.

Варіант 1

$X = (7 \ 5 \ 5 \ 11 \ 2 \ 10 \ 2 \ 17 \ 3 \ 18 \ 1 \ 7 \ 8 \ 18 \ 11)$

$Y = (33 \ 37 \ 44 \ 25 \ 42 \ 32 \ 43 \ 18 \ 46 \ 16 \ 48 \ 36 \ 35 \ 15 \ 27)$

Тип залежності: лінійна

Гіпотеза: $H_0 : a_1 = -1$

Задача: Обчислити імовірність того, що попит зникне при ціні 20.

Довірча імовірність $p=0.95$

Варіант 2

$X = (60 \ 44 \ 40 \ 59 \ 94 \ 39 \ 19 \ 57 \ 63 \ 5 \ 26 \ 89 \ 73 \ 39 \ 90)$

$Y = (16 \ 14 \ 17 \ 15 \ 7 \ 21 \ 30 \ 15 \ 12 \ 67 \ 38 \ 10 \ 9 \ 26 \ 18)$

Тип залежності: гіперболічна $Y_i = \frac{a_1}{a_0 + X_i}$

Гіпотеза $H_0 : a_1 = 1000$

Задача: Обчислити максимальний прибуток .

Довірча ймовірність $p=0.93$

Варіант 3

$X = (9 \ 4 \ 7 \ 5 \ 7 \ 6 \ 3 \ 2 \ 6 \ 9 \ 2 \ 1 \ 3 \ 6 \ 5)$

$Y = (16 \ 29 \ 20 \ 11 \ 17 \ 14 \ 35 \ 49 \ 15 \ 16 \ 46 \ 61 \ 40 \ 15 \ 26)$

Тип залежності: обернена квадратична $Y_i = \frac{a_1}{(a_0 + X_i)^2}$

Гіпотеза: $H_0 : a_1 = 10$

Задача: Обчислити максимально можливий прибуток

Довірча ймовірність $p=0.95$

Варіант 4

$X = (2 \ 9 \ 1 \ 7 \ 6 \ 3 \ 7 \ 9 \ 7 \ 4 \ 5 \ 4 \ 6 \ 2 \ 3)$

$Y = (391 \ 11 \ 487 \ 104 \ 312 \ 309 \ 109 \ 48 \ 81 \ 244 \ 156 \ 186 \ 140 \ 459 \ 305)$

Тип залежності: експоненційна $Y_i = a_0 \cdot e^{a_1 X_i}$

Гіпотеза $H_0 : a_1 = -0.5$

Задача: Побудувати довірчу область значень функції

Довірча ймовірність $p=0.97$

Варіант 5

$$X = (2 \ 9 \ 1 \ 7 \ 6 \ 3 \ 7 \ 9 \ 7 \ 4 \ 5 \ 4 \ 6 \ 1 \ 3)$$

$$Y = (89 \ 28 \ 104 \ 14 \ 36 \ 63 \ 34 \ 19 \ 27 \ 42 \ 32 \ 49 \ 21 \ 73 \ 80)$$

Тип залежності: дробно-лінійна $Y_i = \frac{a_0 + a_1 X_i}{1 + X_i}$

Гіпотеза $H_0 : a_1 = -1$

Задача: Знайдіть максимальний та мінімальний можливий попит

Довірча ймовірність $p=0.99$

Варіант 6

$$X = (1 \ 2 \ 6 \ 4 \ 8 \ 2 \ 7 \ 3 \ 1 \ 2 \ 9 \ 2 \ 1 \ 5 \ 6)$$

$$Y = (196 \ 183 \ 140 \ 161 \ 127 \ 172 \ 127 \ 164 \ 200 \ 173 \ 131 \ 175 \ 194 \ 151 \ 139)$$

Тип залежності: залежність з постійною еластичністю попиту $Y_i = a_0 \cdot X_i^{a_1}$

Гіпотеза: $H_0 : a_1 = -0.3$

Задача: Знайдіть ймовірність того, що при ціні 5 грн. попит становитиме не менше 300 шт.

Довірча ймовірність $p=0.92$

Варіант 7

$$X = (441 \ 75 \ 225 \ 549 \ 20 \ 269 \ 137 \ 94 \ 301 \ 452 \ 128 \ 462 \ 334 \ 511 \ 323)$$

$$Y = (467 \ 1247 \ 943 \ 265 \ 1402 \ 1043 \ 1279 \ 1400 \ 872 \ 506 \ 1328 \ 453 \ 752 \ 353 \ 835)$$

Тип залежності: лінійна

Гіпотеза: $H_0 : a_1 = -5$

Задача: Яку треба встановити ціну на продукцію з метою максимізації прибутку

Довірча ймовірність $p=0.9$

Варіант 8

$$X = (420 \ 89 \ 85 \ 416 \ 256 \ 579 \ 92 \ 493 \ 115 \ 490 \ 94 \ 439 \ 168 \ 409 \ 433)$$

$$Y = (1 \ 5 \ 6 \ 1 \ 2 \ 1 \ 5 \ 0 \ 3 \ 1 \ 5 \ 1 \ 3 \ 2 \ 1)$$

Тип залежності: гіперболічна $Y_i = \frac{a_1}{a_0 + X_i}$

Гіпотеза $H_0 : a_1 = 20$

Задача: Обчислить максимальний прибуток.

Варіант 9

$$X = (77 \ 15 \ 9 \ 47 \ 33 \ 61 \ 28 \ 44 \ 22 \ 52 \ 1 \ 77 \ 96 \ 90 \ 23)$$

$$Y = (14 \ 22 \ 22 \ 17 \ 18 \ 14 \ 19 \ 17 \ 19 \ 17 \ 25 \ 11 \ 11 \ 12 \ 21)$$

Тип залежності: гіперболічна $Y_i = \frac{a_1}{a_0 + X_i}$

Гіпотеза $H_0 : a_1 = 70$

Задача: Обчислить максимальний прибуток.

Порядок виконання роботи

Заносимо вихідні дані

Кількість спостережень $n := 15$ Кількість параметрів $k := 2$

$x := (1 \ 3 \ 4 \ 9 \ 9 \ 9 \ 10 \ 10 \ 11 \ 15 \ 16 \ 16 \ 17 \ 19 \ 19)$

$y := (103 \ 86 \ 74 \ 68 \ 75 \ 68 \ 71 \ 68 \ 77 \ 66 \ 65 \ 66 \ 64 \ 56 \ 65)$

Довірча імовірність $p := 0.92$ Рівень значущості $\epsilon := 1 - p$

$x := x^T$ $y := y^T$

Номер спостереження $i := 0, 1 .. n - 1$

У відповідності із своїм типом залежності за таблицею 1 перетворюємо змінні. Наприклад, для залежності з постійною еластичністю

$$Y_i := \ln(y_i) \quad X_i := \ln(x_i)$$

Обчислюємо параметри лінійної моделі

Середні значення змінних

$$Y_{sr} := \frac{\sum_i Y_i}{n} \quad X_{sr} := \frac{\sum_i X_i}{n}$$

Оцінка коефіцієнта нахилу

$$A_1 := \frac{\sum_i [(X_i - X_{sr}) \cdot (Y_i - Y_{sr})]}{\sum_i (X_i - X_{sr})^2}$$

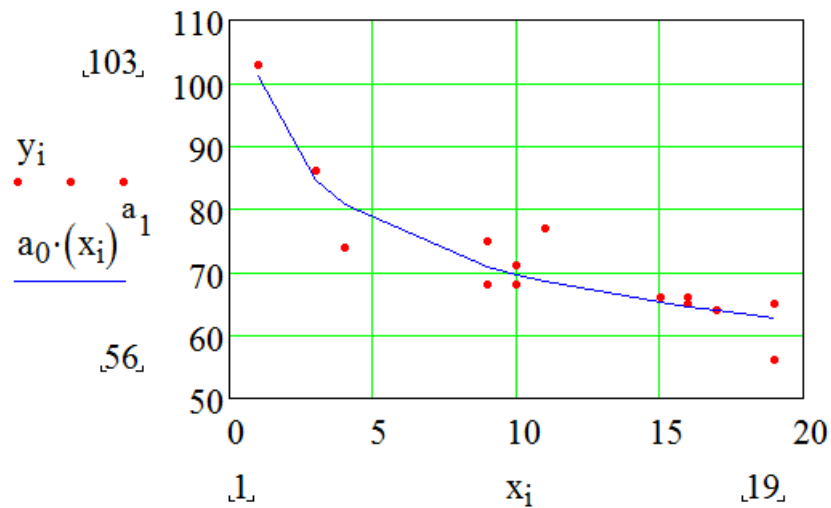
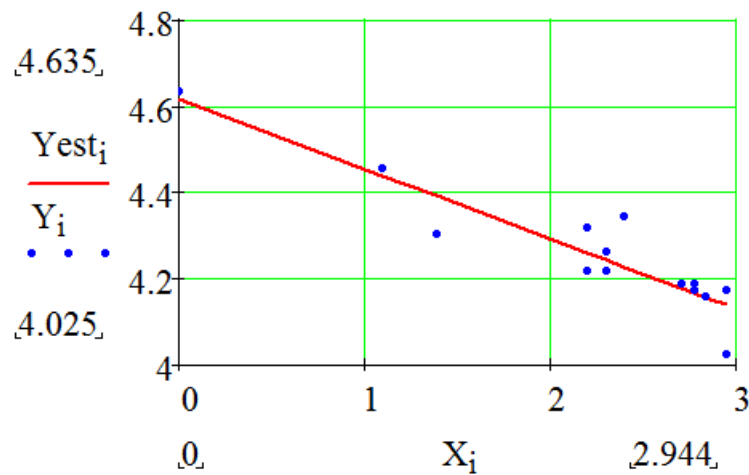
Оцінка вільного члену $A_0 := Y_{sr} - A_1 \cdot X_{sr}$

$$A_0 = 4.616 \quad A_1 = -0.162$$

Оцінені значення функції $Y_{est_i} := A_0 + A_1 \cdot X_i$

Параметри нелінійної залежності

$$a_0 := \exp(A_0) \quad a_1 := A_1$$



Помилки регресії $e := Y - Y_{est}$

Перевірка $\sum_i e_i = -0.000$ $\sum_i (X_i e_i) = -0$

Оцінка середньоквадратичного відхилення

$$\sigma := \sqrt{\frac{\sum_i (e_i)^2}{n - k}}$$

Для перевірки гіпотези відносно параметрів a_0 та a_1 користуються формулами відповідно (1.1) і (1.2). Наприклад, для перевірки гіпотези $H : a_0 = 90$ в моделі з постійною еластичністю обчислюють критерій

$$t_0 := \frac{(A_0 - \ln(90)) \cdot \sqrt{n \cdot \sum_i (X_i - X_{sr})^2}}{\sigma \cdot \sqrt{\sum_i (X_i)^2}}$$

$$t_0 = 2.582$$

Критичне значення t-статистики

$$tkr := qt\left(p + \frac{\varepsilon}{2}, n - k\right) = 1.899$$

Так як розрахункове значення t_0 виявилось вищим критичного гіпотезу H слід відхилити.

Будуємо довірчі інтервали для параметрів лінійної залежності.

$$A_0 - tkr \cdot \frac{\sigma \cdot \sqrt{\sum_i (X_i)^2}}{\sqrt{n \cdot \sum_i (X_i - X_{sr})^2}} = 4.531$$

$$A_0 + tkr \cdot \frac{\sigma \cdot \sqrt{\sum_i (X_i)^2}}{\sqrt{n \cdot \sum_i (X_i - X_{sr})^2}} = 4.701$$

$$A_1 - t_{kr} \cdot \frac{\sigma}{\sqrt{\sum_i (X_i - X_{sr})^2}} = -0.198$$

$$A_1 + t_{kr} \cdot \frac{\sigma}{\sqrt{\sum_i (X_i - X_{sr})^2}} = -0.125$$

Далі за формулами таблиці 1 обчислюємо границі довірчих інтервалів нелінійної моделі

a_0 знаходиться в межах від $\exp(4.531) = 93$ до $\exp(4.701) = 110$

a_1 знаходиться в межах від -0.198 до -0.125

Розв'язання прикладних задач

а) Знайти імовірність потрапляння прогнозного значення попиту у деякий інтервал можна за формулою (1.3). Наприклад, знайдемо імовірність того, що при ціні 10 попит становитиме більше 75. В позначеннях лінійної моделі

$$X_{пр} := \ln(5)$$

$$Y_{пр} := A_0 + A_1 \cdot X_{пр}$$

$$t := \frac{Y_{пр} - \ln(75)}{\sigma \cdot \sqrt{\frac{1}{n} + \frac{(X_{пр} - X_{sr})^2}{\sum_i (X_i - X_{sr})^2}}}$$

$$t = 2.005$$

Якщо попит перевищуватиме 75, то випадкова величина t , буде меншою 2.005. Імовірність цього дорівнює

$$pt(t, n - k) = 0.967$$

б) Знаходження максимального прибутку. Прибуток дорівнює добутку ціни на попит. В гіперболічній моделі прибуток f становить

$$f(x) := x \cdot \frac{500}{10 + x}$$

$$f'(x) = \frac{500(10 + x) - 500x}{(10 + x)^2}$$

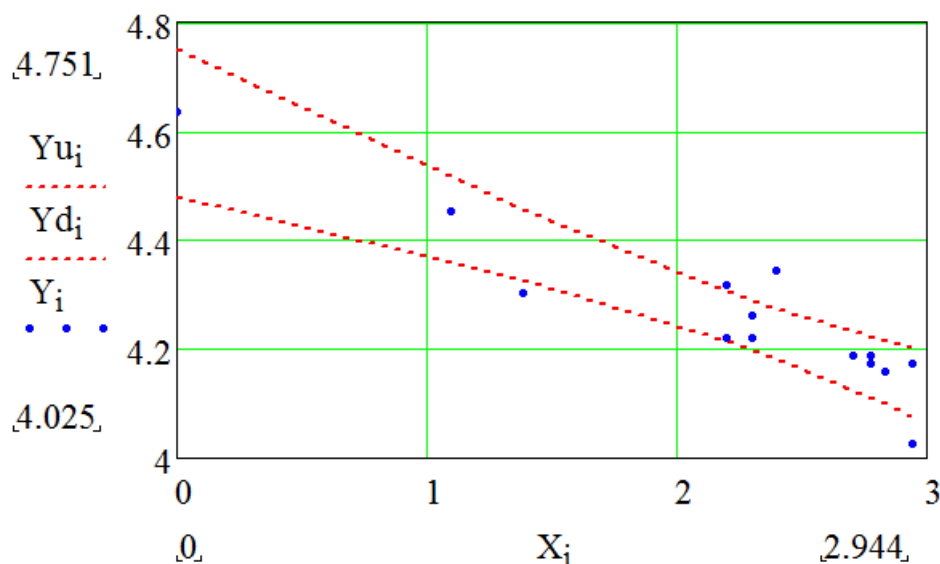
Похідна завжди позитивна, тому

$$\max f(x) = \lim_{x \rightarrow \infty} f(x) = 500$$

в) Для побудови довірчої області значень функції користуються формулою (1.4). Наприклад

Верхня межа довірчої області $Y_{u_i} := Y_{est_i} + t_{kr} \cdot \sigma \cdot \sqrt{\frac{1}{n} + \frac{(X_i - X_{sr})^2}{\sum_i (X_i - X_{sr})^2}}$

Нижня межа довірчої області $Y_{d_i} := Y_{est_i} - t_{kr} \cdot \sigma \cdot \sqrt{\frac{1}{n} + \frac{(X_i - X_{sr})^2}{\sum_i (X_i - X_{sr})^2}}$



Практична робота №2
Тема: «Багатофакторна лінійна регресійна модель»

План:

1. Постановка задачі загальної лінійної моделі. Припущення, покладені в основу багатофакторної лінійної моделі. Оцінювання параметрів багатофакторної лінійної моделі.
2. Властивості оцінок МНК багатофакторної моделі. Перевірка гіпотез у множинній регресії.
3. Поняття мультиколінеарності, її наслідки. Тестування на мультиколінеарність. Усунення мультиколінеарності.

1. Багатофакторна лінійна модель будується в припущенні, що досліджувана величина є функцією декількох змінних, що часто буває більш реалістично, ніж парна залежність. Вважають, що значення функції в i -м спостереженні визначається як:

$$Y_i = a_0 + a_1 X_{1i} + a_2 X_{2i} + \dots + a_k X_{ki} + u_i$$

Якщо ввести матричні позначення:

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, X = \begin{pmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{pmatrix}, a = \begin{pmatrix} a_0 \\ \vdots \\ a_k \end{pmatrix}, u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$$

те рівняння загальної моделі записується як:

$$Y = X \cdot a + u$$

Як і в парній лінійній моделі, щодо змінних і випадкових збурювань робляться деякі припущення:

I. Випадкові величини u_i розподілені незалежно друг від друга за нормальним законом з нульовими математичними очікуваннями і рівними дисперсіями, або в матричному вигляді:

$$E(u) = 0$$

$$\text{cov}(u) = \sigma^2 I_n$$

II. Ранг матриці X дорівнює $k+1$.

Перше припущення є багатомірним узагальненням передумов парної лінійної моделі. Воно може порушуватися в декількох найбільш типових випадках:

а) Наявність ненульової кореляції між випадковими збурюваннями. У цьому випадку коваріаційна матриця вектора u вже не буде діагональною. Такий вид порушень називають автокореляцією.

б) Дисперсії випадкових збурювань можуть бути не рівні між собою. У цьому випадку говорять про гетероскедастичність.

Порушення другого припущення спричинить неможливість використання МНК для оцінки параметрів моделі. Більше слабе порушення цієї умови, що полягає в сильних кореляціях між пояснюючими змінними, допускає використання МНК, однак його оцінки стають малоефективними. Таке явище одержало назву мультиколінеарність.

Відповідно з МНК як оцінку вектора a використовують такий вектор \hat{a} , що мінімізує критерій:

$$z = \sum_i \left(Y_i - \sum_k X_{ki} \hat{a}_k \right)^2$$

Рішенням такої задачі є:

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

2. Оцінки МНК множинної регресії успадковують властивості оцінок парної регресії, тобто також є найбільш ефективними серед всіх лінійних незміщених. Для перевірки гіпотез щодо параметрів регресії використовуються деякі результати про властивості оцінок регресійних коефіцієнтів:

- Коваріаційна матриця вектора оцінок:

$$\text{cov}(\hat{a}) = \sigma^2 (X'X)^{-1}$$

де σ^2 - дисперсія випадкових впливів:

- Дисперсія оцінки МНК значення функції при деяких значеннях незалежних змінних:

$$D(\hat{Y}_{np}) = \sigma^2 X_{np} \cdot (X'X)^{-1} X'_{np}$$

де X_{np} – вектор-рядок, що містить прогностні значення змінних.

• Сума квадратів помилок регресії має незалежний від вектора \hat{a} χ^2 розподіл з $n-k$ ступенями вільності:

$$e'e \sim \chi_{n-k}^2$$

де k – кількість параметрів.

Дисперсія різниці оціненого й зареєстрованого значення функції:

$$D(\hat{Y}_{np} - Y) = \sigma^2 \left(1 + X_{np} (X'X)^{-1} X'_{np} \right)$$

На основі цих результатів будують критерії для перевірки наступних гіпотез:

▪ Гіпотеза про значення деякого параметра регресії $H_0: a_i = c$. Гіпотеза приймається на відповідному рівні значимості, якщо величина

$$t = \frac{\hat{a}_i - c}{\sqrt{\hat{\sigma}^2 [(X'X)^{-1}]_{ii}}}$$

потрапляє в інтервал від $-t_{\alpha/2, n-k}$ до $t_{\alpha/2, n-k}$.

У випадку перевірки на значущість перевіряють гіпотезу $H_0: a_i = 0$

▪ Перевірка на значущість всієї регресії цілком $H_0: a_1 = a_2 \dots a_k = 0$. Гіпотеза приймається на рівні значимості α , якщо величина:

$$F = \frac{R^2 \cdot (n-k)}{(1-R^2) \cdot (k-1)}$$

виявляється не більше, ніж $F_{\alpha, k, n-k}$.

▪ Гіпотеза про прогнозне значення функції $H_0: Y_{np} = c$. Гіпотеза приймається на рівні значимості α , якщо величина

$$t = \frac{\hat{Y}_{np} - c}{\sqrt{\hat{\sigma}^2 X_{np} \cdot (X'X)^{-1} X'_{np}}}$$

потрапляє в інтервал від $-t_{\alpha/2, n-k}$ до $t_{\alpha/2, n-k}$.

▪ Гіпотеза про приналежність зареєстрованого значення функції тої ж генеральної сукупності, з якої були витягнуті інші значення $H_0: Y_{np} = a_0 + a_1 X_{1np} + \dots + a_k X_{knp} + \varepsilon_{np}$. Гіпотеза приймається на рівні значущості α , якщо величина

$$t = \frac{\hat{Y}_{np} - Y_{np}}{\sqrt{\hat{\sigma}^2 \left(1 + X_{np} (X'X)^{-1} X_{np}'\right)}}$$

потрапляє в інтервал від $-t_{\alpha/2, n-k}$ до $t_{\alpha/2, n-k}$.

3. Мультиколінеарність є однією з типових проблем, що зустрічаються при побудові багатofакторних моделей. Виділяють два види мультиколінеарності – повна й часткова. У першому випадку серед незалежних змінних присутні зв'язані точною лінійною залежністю. Тому матриця $X'X$ є виродженою і різних наборів коефіцієнтів, що мінімізують суму квадратів помилок виявляється нескінченно багато. Це не дозволяє оцінити вплив якої-небудь змінної або виконати прогноз по ній. В умовах часткової мультиколінеарності між змінними існують значні кореляційні зв'язки, які приводять до зменшення величини визначника матриці $X'X$ й, як наслідок, - до збільшення дисперсій вибіркового коефіцієнтів при відповідних змінних.

Основними ознаками мультиколінеарності є:

- Деякі або всі змінні виявляються незначущими, у той час як загальна якість моделі досить висока;
- Оцінки коефіцієнтів сильно змінюються при додаванні невеликої кількості спостережень;
- Деякі коефіцієнти виявляються несподівано великими або малими, або мають знаки протилежні очікуванім;
- Надмірно великі дисперсії деяких вибіркового коефіцієнтів.

Найкращим виходом для усунення мультиколінеарності є додавання нових спостережень, які дозволили б виділити вплив кожної зі змінних. У тому випадку, якщо це важко використовують два методи, які можуть і доповнювати один одного:

- Видалити деякі з мультиколінеарних змінних, якщо це не викличе серйозного зниження коефіцієнта детермінації. Це дозволить, принаймні, адекватно оцінити вплив тих змінних, що залишилися.

- Перейти до ортогональної системи змінних за допомогою, наприклад, методу головних компонентів. Недоліком цього методу є те, що не завжди вдається інтерпретувати отримані змінні.

Завдання

1. Оцінити параметри багатofакторної регресії, перевірити її на адекватність.

2. Перевірити наявність мультиколінеарності серед незалежних змінних та видалити мультиколінеарні змінні, якщо вони присутні. Знову оцінити адекватність моделі, побудувати довірчі інтервали для коефіцієнтів в моделях з мультиколінеарними змінними і без них та порівняти результати.

3. Побудувати довірчу область значень функції.

Приклад виконання роботи

Кількість спостережень $n := 10$ Кількість параметрів $k := 4$

Довірча імовірність $p := 0.95$

Рівень значущості $\gamma := 1 - p$

$$X := \begin{pmatrix} 3 & -24 & 3 & -24 & 16 & 26 & 13 & -4 & -4 & 4 \\ 10 & 29 & 13 & 23 & 5 & 3 & 14 & 10 & 18 & 8 \\ 13 & 7 & 16 & 0 & 23 & 29 & 27 & 7 & 14 & 12 \end{pmatrix}$$

$$Y := (66 \ 205 \ 112 \ 182 \ 102 \ 71 \ 155 \ 87 \ 165 \ 88)$$

$$X := X^T \quad Y := Y^T$$

Доповнимо матрицю X одиничним стовпцем та обчислимо вектор оцінок параметрів моделі

$$f(x, y) := 1$$

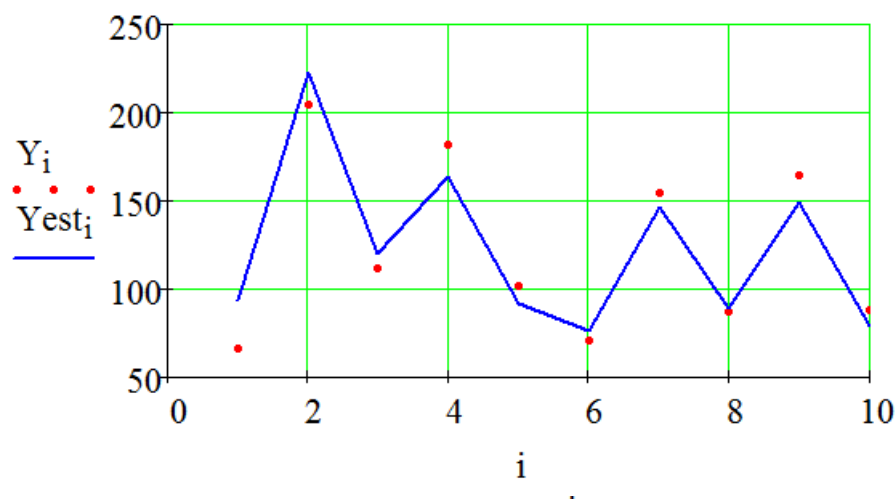
$$X := \text{augment}(\text{matrix}(10, 1, f), X)$$

$$\beta := (X^T \cdot X)^{-1} \cdot X^T \cdot Y$$

$$\beta = \begin{pmatrix} 3.227 \\ -6.456 \\ 0.238 \\ 8.292 \end{pmatrix}$$

Оціненні значення функції $Y_{est} := X \cdot \beta$

$i := 1, 2 \dots n$



Помилки регресії

$$e := Y - Y_{est}$$

Перевірка $\sum_i e_i = -0$ $X^T \cdot e = \begin{pmatrix} -0 \\ -0 \\ 0 \\ -0 \end{pmatrix}$

Оцінка середньоквадратичного відхилення помилок

$$\sigma := \sqrt{\frac{\sum_i (e_i)^2}{n - k}}$$

$$\sigma = 18.478$$

Оцінка коваріаційної матриці вектора параметрів регресії :

$$V := \sigma^2 \cdot (X^T \cdot X)^{-1}$$

$$Y_{sr} := \frac{\sum_i Y_i}{n}$$

Сума квадратів відхилень функції пояснена регресією:

$$RSS := \sum_i (Y_{est_i} - Y_{sr})^2$$

Загальна сума квадратів відхилень:

$$TSS := \sum_i (Y_i - Y_{sr})^2$$

Коефіцієнт детермінації:

$$R := \frac{RSS}{TSS}$$

$$R = 0.907$$

Розрахунковий F-критерій:

$$F := \frac{R \cdot (n - k)}{(1 - R) \cdot (k - 1)}$$

$$F = 19.506$$

Критичне значення F-розподілу:

$$F_{kr} := qF(p, k - 1, n - k)$$

$$F_{kr} = 4.757$$

Розрахункове значення перевищує критичне, тому модель є значущою. Обчислимо коефіцієнти кореляції між всіма парами незалежних змінних.

$$\text{corr}(X^{(2)}, X^{(3)}) = -0.891 \quad \text{corr}(X^{(2)}, X^{(4)}) = 0.916$$

$$\text{corr}(X^{(3)}, X^{(4)}) = -0.637$$

Видно, що тісно пов'язаними є перша змінна з другою і перша з третьою. Тому для усунення мультиколінеарності достатньо видалити першу змінну.

$$X := \text{augment}(X^{(1)}, X^{(3)}, X^{(4)})$$

$$b := (X^T \cdot X)^{-1} \cdot X^T \cdot Y$$

$$Y_{\text{est}} := X \cdot b$$

$$e := Y - Y_{\text{est}}$$

$$\sigma := \sqrt{\frac{\sum_i (e_i)^2}{n - k + 1}}$$

$$W := \sigma^2 \cdot (X^T \cdot X)^{-1}$$

Критичне значення t-розподілу:

$$\text{tkr} := \text{qt}\left(\frac{\gamma}{2}, n - k + 1\right)$$

Довірчі інтервали параметрів

$$\beta_0 \text{ знаходиться в інтервалі від } b_{1,1} - |\text{tkr}| \cdot \sqrt{W_{1,1}} = -49.653$$

$$\text{до } b_{1,1} + |\text{tkr}| \cdot \sqrt{W_{1,1}} = 62.03$$

$$\beta_2 \text{ знаходиться в інтервалі від } b_{2,1} - |\text{tkr}| \cdot \sqrt{W_{2,2}} = 4.578$$

$$\text{до } b_{2,1} + |\text{tkr}| \cdot \sqrt{W_{2,2}} = 9.145$$

$$\beta_3 \text{ знаходиться в інтервалі від } b_{3,1} - |\text{tkr}| \cdot \sqrt{W_{3,3}} = -0.247$$

$$\text{до } b_{3,1} + |\text{tkr}| \cdot \sqrt{W_{3,3}} = 3.741$$

Звуження довірчих інтервалів

$$\text{для константи } \frac{\left| \text{qt}\left(\frac{\gamma}{2}, n - k\right) \right| \cdot \sqrt{V_{1,1}}}{\left| \text{qt}\left(\frac{\gamma}{2}, n - k + 1\right) \right| \cdot \sqrt{W_{1,1}}} = 1.07$$

$$\begin{aligned} \text{для коефіцієнту при } X_2 & \frac{\left| \text{qt}\left(\frac{\gamma}{2}, n - k\right) \right| \cdot \sqrt{V_{3,3}}}{\left| \text{qt}\left(\frac{\gamma}{2}, n - k + 1\right) \right| \cdot \sqrt{W_{2,2}}} = 8.69 \\ \text{для коефіцієнту при } X_3 & \frac{\left| \text{qt}\left(\frac{\gamma}{2}, n - k\right) \right| \cdot \sqrt{V_{4,4}}}{\left| \text{qt}\left(\frac{\gamma}{2}, n - k + 1\right) \right| \cdot \sqrt{W_{3,3}}} = 9.8 \end{aligned}$$

Знову перевіримо модель на адекватність.

$$\text{RSS} := \sum_i (Y_{\text{est}i} - Y_{\text{sr}})^2$$

$$\text{TSS} := \sum_i (Y_i - Y_{\text{sr}})^2$$

$$R := \frac{\text{RSS}}{\text{TSS}}$$

$$R = 0.897$$

Видно, що коефіцієнт детермінації залишився майже незмінним, тому шляхом вилучення мультиколінеарної змінної вдалось збільшити точність оцінок без зменшення загальної якості моделі.

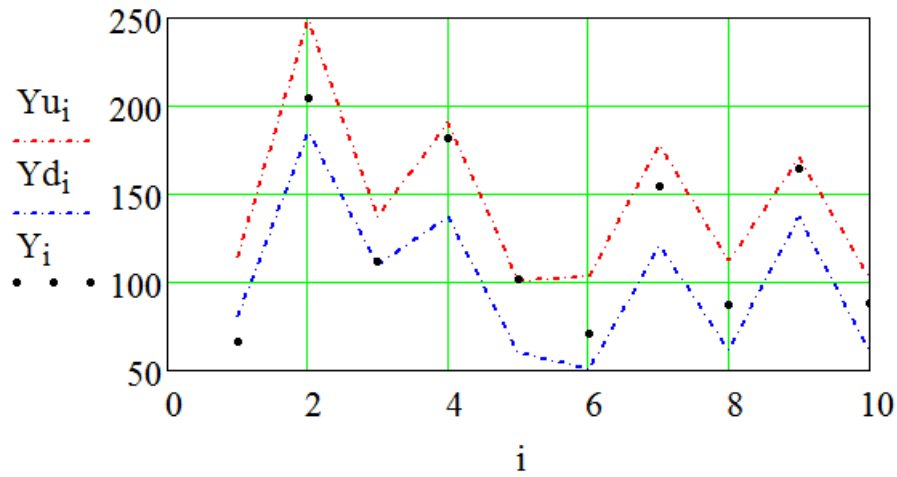
Оцінка дисперсій значень функції:

$$DY := \sigma^2 \cdot X \cdot (X^T \cdot X)^{-1} \cdot X^T$$

$$\text{Верхня межа довірчої області} \quad Y_{u_i} := Y_{\text{est}i} + |\text{tkr}| \cdot \sqrt{DY_{i,i}}$$

$$\text{Нижня межа довірчої області} \quad Y_{d_i} := Y_{\text{est}i} - |\text{tkr}| \cdot \sqrt{DY_{i,i}}$$

Видно, що немає точок, які виходять за межі довірчої області, тому всі спостереження генеровані однією сукупністю



Варіанти завдань

<p>Варіант 1</p> $X = \begin{pmatrix} 4 & -3 & 4 \\ 14 & -72 & 2 \\ 7 & 67 & 22 \\ 2 & 104 & 21 \\ 4 & 65 & 20 \\ 13 & -66 & 3 \\ 14 & -71 & 1 \\ 2 & 61 & 14 \\ 1 & 108 & 22 \\ 6 & 69 & 20 \end{pmatrix} \quad Y = \begin{pmatrix} 129 \\ 478 \\ -19 \\ -270 \\ -149 \\ 423 \\ 391 \\ -211 \\ -148 \\ -194 \end{pmatrix}$	<p>Варіант 2</p> $X = \begin{pmatrix} -65 & 15 & 7 \\ -24 & 2 & 10 \\ -123 & 29 & 17 \\ -86 & 18 & 11 \\ -52 & 11 & 11 \\ -100 & 26 & 13 \\ -57 & 7 & 15 \\ -13 & 2 & 3 \\ -51 & 5 & 16 \\ -42 & 7 & 9 \end{pmatrix} \quad Y = \begin{pmatrix} 148 \\ -20 \\ 258 \\ 216 \\ 111 \\ 228 \\ 13 \\ 41 \\ -33 \\ 62 \end{pmatrix}$
<p>Варіант 3</p> $X = \begin{pmatrix} 36 & 10 & 12 \\ 59 & 19 & 19 \\ 25 & 9 & 8 \\ 17 & 27 & 5 \\ 33 & 10 & 12 \\ 1 & 14 & 0 \\ 48 & 12 & 18 \\ 10 & 4 & 4 \\ 67 & 4 & 19 \\ 23 & 3 & 8 \end{pmatrix} \quad Y = \begin{pmatrix} 187 \\ 368 \\ 190 \\ 158 \\ 210 \\ 87 \\ 402 \\ 136 \\ 487 \\ 108 \end{pmatrix}$	<p>Варіант 4</p> $X = \begin{pmatrix} 11 & 14 & 27 \\ -31 & 28 & 26 \\ -66 & 29 & 12 \\ -6 & 15 & 20 \\ -61 & 27 & 8 \\ 27 & 10 & 29 \\ -16 & 17 & 18 \\ -27 & 10 & 4 \\ 10 & 9 & 18 \\ -51 & 28 & 11 \end{pmatrix} \quad Y = \begin{pmatrix} 50 \\ 360 \\ 436 \\ 100 \\ 433 \\ 22 \\ 155 \\ 95 \\ 5 \\ 305 \end{pmatrix}$

<p style="text-align: center;">Варіант 5</p> $X = \begin{pmatrix} 10 & 21 & 47 \\ 19 & 15 & 95 \\ 9 & 20 & 38 \\ 27 & 19 & 100 \\ 10 & 25 & 57 \\ 14 & 26 & 66 \\ 12 & 29 & 51 \\ 4 & 22 & 15 \\ 4 & 11 & 23 \\ 3 & 26 & 18 \end{pmatrix} \quad Y = \begin{pmatrix} -232 \\ -415 \\ -105 \\ -439 \\ -194 \\ -334 \\ -112 \\ -48 \\ -86 \\ 1 \end{pmatrix}$	<p style="text-align: center;">Варіант 6</p> $X = \begin{pmatrix} 19 & 2 & 60 \\ 3 & 30 & -49 \\ 15 & 19 & 22 \\ 21 & 27 & 30 \\ 11 & 27 & -10 \\ 29 & 14 & 92 \\ 14 & 23 & 9 \\ 16 & 22 & 17 \\ 24 & 11 & 70 \\ 22 & 25 & 45 \end{pmatrix} \quad Y = \begin{pmatrix} -63 \\ 77 \\ -1 \\ 14 \\ 24 \\ -25 \\ 16 \\ -5 \\ -51 \\ 17 \end{pmatrix}$
<p style="text-align: center;">Варіант 7</p> $X = \begin{pmatrix} 49 & 9 & 3 \\ -51 & 11 & 24 \\ 16 & 21 & 22 \\ 120 & 30 & 11 \\ 39 & 23 & 20 \\ -33 & 7 & 15 \\ -29 & 4 & 9 \\ -56 & 10 & 25 \\ 108 & 25 & 5 \\ 64 & 18 & 11 \end{pmatrix} \quad Y = \begin{pmatrix} 112 \\ 183 \\ 280 \\ 325 \\ 292 \\ 136 \\ 62 \\ 175 \\ 237 \\ 224 \end{pmatrix}$	<p style="text-align: center;">Варіант 8</p> $X = \begin{pmatrix} 25 & 15 & 24 \\ 37 & 19 & 12 \\ 51 & 25 & 25 \\ 28 & 14 & 5 \\ 9 & 4 & 24 \\ 47 & 27 & 23 \\ 25 & 12 & 17 \\ 36 & 19 & 28 \\ 65 & 30 & 13 \\ 28 & 14 & 17 \end{pmatrix} \quad Y = \begin{pmatrix} 191 \\ 436 \\ 382 \\ 335 \\ -24 \\ 403 \\ 131 \\ 348 \\ 577 \\ 139 \end{pmatrix}$
<p style="text-align: center;">Варіант 9</p> $X = \begin{pmatrix} 14 & 2 & -56 \\ 3 & 17 & 99 \\ 23 & 19 & 17 \\ 27 & 1 & -139 \\ 21 & 16 & 7 \\ 12 & 20 & 89 \\ 15 & 12 & 14 \\ 3 & 3 & 8 \\ 20 & 1 & -83 \\ 28 & 6 & -93 \end{pmatrix} \quad Y = \begin{pmatrix} -338 \\ 533 \\ 207 \\ -646 \\ 16 \\ 317 \\ -6 \\ 129 \\ -404 \\ -560 \end{pmatrix}$	<p style="text-align: center;">Варіант 10</p> $X = \begin{pmatrix} 11 & 20 & -44 \\ 17 & 16 & -59 \\ 25 & 4 & -105 \\ 7 & 3 & -32 \\ 1 & 13 & -4 \\ 19 & 26 & -76 \\ 29 & 18 & -105 \\ 24 & 18 & -96 \\ 12 & 8 & -47 \\ 29 & 15 & -116 \end{pmatrix} \quad Y = \begin{pmatrix} -135 \\ -145 \\ -105 \\ -39 \\ -59 \\ -231 \\ -175 \\ -203 \\ -80 \\ -195 \end{pmatrix}$

<p style="text-align: center;">Варіант 11</p> $X = \begin{pmatrix} 15 & 37 & 15 \\ 10 & 29 & 5 \\ 9 & 26 & 15 \\ 5 & 13 & 29 \\ 22 & 75 & 1 \\ 13 & 37 & 5 \\ 15 & 45 & 24 \\ 16 & 52 & 14 \\ 3 & 12 & 3 \\ 16 & 53 & 13 \end{pmatrix} \quad Y = \begin{pmatrix} 172 \\ 99 \\ 88 \\ -16 \\ 421 \\ 253 \\ 222 \\ 268 \\ 51 \\ 311 \end{pmatrix}$	<p style="text-align: center;">Варіант 12</p> $X = \begin{pmatrix} 4 & 8 & 5 \\ 0 & -0 & 2 \\ 10 & 24 & 2 \\ 18 & 45 & 9 \\ 11 & 24 & 7 \\ 10 & 8 & 23 \\ 4 & -1 & 11 \\ 8 & 20 & 4 \\ 0 & -23 & 23 \\ 0 & -28 & 29 \end{pmatrix} \quad Y = \begin{pmatrix} 47 \\ 14 \\ 66 \\ 175 \\ 90 \\ 120 \\ 9 \\ 85 \\ 29 \\ 39 \end{pmatrix}$
<p style="text-align: center;">Варіант 13</p> $X = \begin{pmatrix} 22 & 1 & 10 \\ 133 & 24 & 29 \\ 108 & 15 & 29 \\ 137 & 28 & 22 \\ 105 & 21 & 18 \\ 83 & 27 & 6 \\ 82 & 15 & 23 \\ 80 & 10 & 25 \\ 117 & 25 & 13 \\ 72 & 10 & 18 \end{pmatrix} \quad Y = \begin{pmatrix} 44 \\ 417 \\ 240 \\ 454 \\ 302 \\ 452 \\ 122 \\ 114 \\ 445 \\ 213 \end{pmatrix}$	<p style="text-align: center;">Варіант 14</p> $X = \begin{pmatrix} 15 & 13 & 13 \\ 20 & 26 & 24 \\ 108 & 2 & 27 \\ 42 & 6 & 16 \\ 59 & 2 & 17 \\ 57 & 14 & 24 \\ 62 & 1 & 19 \\ -52 & 24 & 6 \\ -20 & 14 & 5 \\ 27 & 25 & 25 \end{pmatrix} \quad Y = \begin{pmatrix} 60 \\ 125 \\ 195 \\ 112 \\ 113 \\ 144 \\ 121 \\ 30 \\ 27 \\ 99 \end{pmatrix}$
<p style="text-align: center;">Варіант 15</p> $X = \begin{pmatrix} 14 & -1 & 10 \\ 10 & 61 & 24 \\ 7 & 100 & 28 \\ 18 & -36 & 4 \\ 24 & 37 & 30 \\ 26 & -6 & 18 \\ 28 & -28 & 15 \\ 2 & 62 & 14 \\ 8 & 42 & 17 \\ 10 & -1 & 8 \end{pmatrix} \quad Y = \begin{pmatrix} 197 \\ 441 \\ 452 \\ 146 \\ 633 \\ 316 \\ 204 \\ 232 \\ 304 \\ 181 \end{pmatrix}$	<p style="text-align: center;">Варіант 16</p> $X = \begin{pmatrix} 21 & -57 & 2 \\ 2 & -5 & 2 \\ 25 & -80 & 20 \\ 21 & -65 & 24 \\ 28 & -72 & 2 \\ 7 & -18 & 21 \\ 19 & -46 & 18 \\ 9 & -28 & 3 \\ 21 & -67 & 11 \\ 24 & -76 & 2 \end{pmatrix} \quad Y = \begin{pmatrix} -479 \\ -96 \\ -681 \\ -659 \\ -722 \\ -228 \\ -451 \\ -242 \\ -690 \\ -650 \end{pmatrix}$

<p style="text-align: center;">Варіант 17</p> $X = \begin{pmatrix} -7 & 7 & 0 \\ 27 & 1 & 26 \\ 22 & 1 & 24 \\ -10 & 16 & 5 \\ -3 & 11 & 9 \\ 8 & 13 & 21 \\ -4 & 16 & 11 \\ 6 & 23 & 30 \\ 14 & 14 & 28 \\ 18 & 8 & 26 \end{pmatrix} \quad Y = \begin{pmatrix} 19 \\ 78 \\ 91 \\ 46 \\ 65 \\ 103 \\ 54 \\ 142 \\ 140 \\ 112 \end{pmatrix}$	<p style="text-align: center;">Варіант 18</p> $X = \begin{pmatrix} -8 & 3 & 17 \\ -19 & 9 & 3 \\ -18 & 9 & 23 \\ -9 & 4 & 16 \\ -49 & 25 & 13 \\ -36 & 18 & 21 \\ -15 & 8 & 13 \\ 1 & 0 & 17 \\ -48 & 24 & 19 \\ -12 & 6 & 15 \end{pmatrix} \quad Y = \begin{pmatrix} 89 \\ -88 \\ 38 \\ 45 \\ -122 \\ -95 \\ -48 \\ 175 \\ -202 \\ -11 \end{pmatrix}$
<p style="text-align: center;">Варіант 19</p> $X = \begin{pmatrix} 27 & 19 & 20 \\ 14 & 81 & 24 \\ 26 & 78 & 30 \\ 1 & 32 & 6 \\ 28 & -26 & 12 \\ 5 & 76 & 17 \\ 4 & 141 & 29 \\ 4 & 55 & 11 \\ 17 & -29 & 4 \\ 22 & -72 & 1 \end{pmatrix} \quad Y = \begin{pmatrix} -75 \\ -43 \\ -71 \\ 8 \\ -91 \\ -3 \\ -4 \\ 26 \\ -69 \\ -84 \end{pmatrix}$	<p style="text-align: center;">Варіант 20</p> $X = \begin{pmatrix} 29 & 47 & 18 \\ 8 & 47 & 16 \\ 20 & 72 & 25 \\ 29 & 55 & 17 \\ 8 & 0 & 0 \\ 6 & 79 & 25 \\ 5 & 11 & 4 \\ 22 & 34 & 11 \\ 17 & 6 & 2 \\ 30 & 1 & 0 \end{pmatrix} \quad Y = \begin{pmatrix} -5 \\ -41 \\ -50 \\ 37 \\ -2 \\ -52 \\ -31 \\ 32 \\ 34 \\ 110 \end{pmatrix}$

Практична робота №3
Тема: «Узагальнена лінійна модель»

План:

4. Оцінки узагальненої моделі.
5. Гетероскедастичність збурювань.
6. Автокореляція збурювань.

1. Одним з основних припущень загальної моделі є припущення про незалежність випадкових збурювань і сталість їхніх дисперсій. Оцінки МНК будуть найбільш ефективними тільки в тому випадку, якщо ці вимоги виконуються. В економетричній практиці це часто нереалістично, наприклад, дисперсія споживчих витрат родини збільшується зі збільшенням бюджету, а величина попиту на продукцію може залежати від попиту на цю же продукцію в попередньому періоді. Можна показати, що в обох зазначених випадках оцінювання за допомогою звичайного МНК приведе до серйозного збільшення дисперсій вибірових коефіцієнтів. Найкращі лінійні оцінки в цьому випадку одержують, перетворюючи вектор випадкових збурювань таким чином, щоб його коваріаційна матриця стала діагональною з рівними елементами на головній діагоналі, після чого використовують ту ж процедуру, що використовувалася при отриманні звичайних оцінок. У результаті одержують вираження для оцінок узагальненої моделі називані оцінками Ейткена:

$$b = (X_*^T X_*)^{-1} X_*^T Y_* = (X^T \Omega^{-1} X) X^T \Omega^{-1} Y$$

де X_* і Y_* являють собою перетворені значення змінних:

$$X_* = P^{-1} X \quad Y_* = P^{-1} Y$$

, а P - деяка позитивно визначена матриця така, що:

$$\Omega = P P^T$$

При перевірці гіпотез про параметри варто спочатку переходити до перетворених значень, наприклад, основою для виконання дисперсійного аналізу є вираз:

$$Y^T \Omega^{-1} Y = b^T X^T \Omega^{-1} Y + e^T \Omega^{-1} e$$

де $e = Y - Xb$.

Отримана модель є дуже загальною, але трудність її використання полягає в тому, що коваріаційна матриця вектора збурювань, як правило, невідома, а оцінити $n \cdot (n+1) / 2$ її невідомих елементів, використовуючи лише n спостережень неможливо. У силу цього застосовність оцінок Ейткена обмежена тим колом задач, у яких апріорно на елементи матриці Ω накладені обмеження що дозволяють зменшити кількість параметрів для оцінювання.

2. В умовах гетероскедастичності коваріаційна матриця випадкових збурювань виглядає як:

$$\Omega = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}$$

Тоді матриця перетворення P^{-1} дорівнює:

$$P^{-1} = \begin{pmatrix} 1/\sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & 1/\sqrt{\lambda_2} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & 1/\sqrt{\lambda_n} \end{pmatrix}$$

У випадку однофакторної моделі перетворені значення змінних визначаються як:

$$X_{*i} = \frac{X_i}{\sqrt{\lambda_i}}; Y_{*i} = \frac{Y_i}{\sqrt{\lambda_i}}$$

А найкращими оцінками будуть такі, які мінімізують вираз

$$z(b) = \sum_i \frac{1}{\lambda_i} (Y_i - b_0 - b_1 X_i)^2$$

, що являє собою зважену суму квадратів. Із цієї причини метод оцінювання параметрів моделі з гетероскедастичними залишками ще називають зваженим МНК.

Існує кілька розповсюджених процедур для тестування на гетероскедастичність: μ - критерій, тест Гольфреда-Квандта, тест Глейсера й ін. Тест Глейсера дозволяє, крім того, оцінити елементи матриці перетворення. У рамках цього тесту припускають, що дисперсія збурювань залежить від значень однієї або декількох пояснюючих змінних. Форму цього впливу звичайно специфікують у вигляді нескладних функцій: лінійної, гіперболічної й т.д. Вибір конкретної функції здійснюють на основі порівняння коефіцієнтів детермінації.

3. Автокореляційним процесом першого порядку називають таку схему генерування випадкових збурювань, при якій:

$$\varepsilon_i = \rho\varepsilon_{i-1} + \delta_i$$

де ρ - коефіцієнт автокореляції по модулі не перевищуючої одиниці;

δ_i - випадкові величини, розподілені незалежно друг від друга за нормальним законом з нульовими маточикунаннями й рівними дисперсіями σ_0^2 .

Можна показати, що в цьому випадку елементи коваріації матриці вектора ε визначаються як:

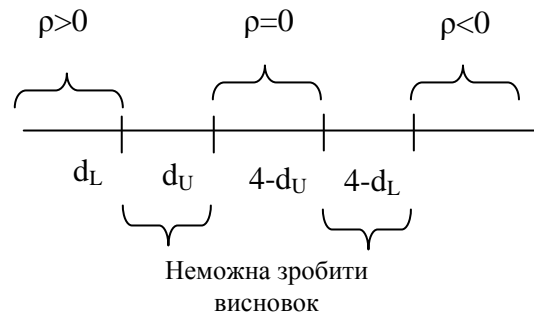
$$\text{cov}(\varepsilon_i, \varepsilon_{i-k}) = \frac{\sigma_0^2}{1-\rho^2} \rho^k$$

Основними негативними наслідками автокореляції є недооцінка дисперсій вибірових характеристик, причому вона може бути набагато більше істотної, чим у випадку з гетероскедастичністю.

Найпоширенішим методом тестування на автокореляцію є тест Дарбіна - Уотсона. Він заснований на обчисленні статистики:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2}$$

У випадку відсутності автокореляції ця величина повинна приймати значення приблизно рівні двом. Загальну схему тестування можна проілюструвати за допомогою рисунка:



де d_L і d_U – критичні значення статистики Дарбіна-Уотсона.

Оцінювання регресійної моделі з автокорельованими залишками здійснюють звичайно ітеративно за допомогою, наприклад, процедури Кохрейна-Орката:

- 1) Обчислюють оцінки МНК вектора A

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

- 2) Обчислюють помилки першої ітерації

$$e = Y - X\hat{A}$$

- 3) Обчислюють оцінку коефіцієнта автокореляції як оцінку МНК у залежності:

$$e_i = r \cdot e_{i-1} + \delta_i$$

- 4) Обчислюють оцінку УМНК вектора A :

$$\hat{A} = (X' \Sigma_e^{-1} X)^{-1} X' \Sigma_e^{-1} Y$$

де Σ_e^{-1} - оцінка ковариационної матриці вектора ε .

- 5) Переходять до кроку 2.

Процедуру закінчують при стабілізації значень r .

Недолік даної процедури полягає в можливості потрапити в локальний мінімум.

Завдання

1. Оцінити параметри регресійної моделі без врахування автокореляції.
2. За допомогою критерію Дарбіна-Уотсона зробити висновок про наявність автокореляції.

3. Оцінити параметри моделі методом Кохрена-Оркатта.

Варіанти завдань

Хід роботи

$$Y := (4 \ 1 \ 32 \ 20 \ 55 \ 40 \ 72 \ 51 \ 40 \ 44)$$

$$X := (1 \ 4 \ 12 \ 7 \ 16 \ 4 \ 14 \ 6 \ 2 \ 3)$$

$$i := 0, 1 \dots 9 \quad n := 10 \quad k := 2$$

$$a_1 := \frac{\sum_i [(X_{0,i} - \text{mean}(X))(Y_{0,i} - \text{mean}(Y))]}{\sum_i (X_{0,i} - \text{mean}(X))^2}$$

$$a_0 := \text{mean}(Y) - a_1 \text{mean}(X)$$

$$Y_{\text{est}0,i} := a_0 + a_1 X_{0,i}$$

$$e := Y - Y_{\text{est}}$$

Розрахункове значення d-статистики $d := \frac{\sum_{i=1}^9 (e_{0,i} - e_{0,i-1})^2}{\sum_{i=0}^9 (e_{0,i})^2}$

$$d = 0.226$$

При 5% рівні значущості і 10 спостереженнях критичні значення статистики Дарбіна-Уотсона

$$d_u := 1.32 \quad d_l := 0.88$$

Так як розрахункове значення статистики менше нижнього критичного між залишками є позитивна автокореляція.

$$j := 0, 1 \dots 9$$

$$f(i, j) := 1$$

$X := \text{augment}(\text{matrix}(10, 1, f), X)$

Ітерація 1

$X := X^T \quad Y := Y^T$

$e := e^T$

$$r := \frac{\sum_{i=1}^9 \left[\left(\begin{array}{c} \sum_{i=1}^9 e_i \\ e_i - \frac{\sum_{i=1}^9 e_i}{9} \end{array} \right) \cdot \left(\begin{array}{c} \sum_{i=1}^9 e_{i-1} \\ e_{i-1} - \frac{\sum_{i=1}^9 e_{i-1}}{9} \end{array} \right) \right]}{\sum_{i=1}^9 \left(\begin{array}{c} \sum_{i=1}^9 e_{i-1} \\ e_{i-1} - \frac{\sum_{i=1}^9 e_{i-1}}{9} \end{array} \right)^2}$$

Практична робота №4
Тема: Системи регресійних рівнянь

План:

1. Поняття системи регресійних рівнянь(СОУ).
2. Проблема ідентифікації. Методи ідентифікації.
3. Методи оцінювання систем регресійних рівнянь.

I. Поводження багатьох реальних економічних систем описується не одним єдиним рівнянням, а **декількома**, у які можуть входити тотожності, економічні залежності, зворотні зв'язки й т.д., наприклад, думаючи, що споживчі **витрати** C залежать від **доходів** Y і з огляду на те, що **доходи** являють собою споживчі й неспоживчі **витрати** **одержимо** модель:

$$\begin{cases} C_t = \alpha + \beta Y_t + \varepsilon_t \\ Y_t = C_t + Z_t \end{cases}$$

Всі **змінні** в такого роду моделях розбивають на **дві множини**: **ендогенні**(залежні, внутрішні) і **екзогенні**(незалежні, зовнішні). У розглянутій моделі ендогенними є C і Y , а **екзогенної** - Z .

Як і колись перед оцінюванням параметрів моделі роблять деякі припущення:

- Помилки **уважають** незалежними друг від друга з однаковою дисперсією σ^2 , і нульовими математичними **очікуваннями**;
- У **кожному** рівнянні екзогенні **змінні некорельовані** з відповідним випадковим **збурюванням**;

Якщо залишити в правій частині тільки випадкові помилки й позначити через B і Γ матриці коефіцієнтів при ендогенних і **екзогенних змінних**, то **СОР** можна **представити** в матричному **вигляді**:

$$B y_t + \Gamma x_t = u_t$$

Така форма **подання СОР** називається структурною формою. У розглянутому прикладі:

$$B = \begin{pmatrix} 1 & -\beta \\ -1 & 1 \end{pmatrix} \quad \Gamma = \begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix} \quad y_t = \begin{pmatrix} C_t \\ Y_t \end{pmatrix} \quad x_t = \begin{pmatrix} 1 \\ Z_t \end{pmatrix} \quad u_t = \begin{pmatrix} \varepsilon_t \\ 0 \end{pmatrix}$$

Другою можливою формою подання **СОР** є приведена форма, у якій у **кожному** рівнянні явно виражається одна з ендогенних **змінних**. Наведену форму можна **одержати** зі **структурної** шляхом множення її ліворуч на матрицю B^{-1} :

$$y_t = -B^{-1}\Gamma x_t + u_t = \Pi x_t + u_t$$

II. Можна показати, що в **СОР** ендогенні **змінні** й випадкові помилки **виявляються корельованими**. Це приводить до **того**, що оцінювання параметрів такої системи шляхом почергового застосування **МНК** до **кожного** з рівнянь у **загальному** випадку приводить до **зміщених** і неспроможних оцінок (тобто зсув неможливо усунути збільшенням **об'єму вибірки**).

Одним з можливих методів оцінювання **СОР** є т.зв. непрямий метод найменших квадратів (Indirect Least Squares, ILS). Він **полягає** в **первісній** оцінці **МНК** параметрів **приведеної** форми Π и в **подальшому** – **вираженні** матриць структурної форми B і Γ через матрицю Π . Оцінка матриці Π є **спроможною**, тому що вектора x_t і u_t є незалежними, однак виразити матриці B і Γ через матрицю Π **вдається** не завжди. При цьому можуть виникнути **три** випадки:

- Структурні коефіцієнти рівняння однозначно виражаються через параметри **приведеної** форми. У цьому випадку рівняння називають **ідентифікованим**;
- Структурні коефіцієнти рівняння **допускають** кілька різних **подань** через параметри **приведеної** форми. Такі рівняння називаю **сверхідентифікованими**;
- Структурні коефіцієнти рівняння неможливо виразити через параметри **наведеної** форми. Такі рівняння називають **неідентифікованими**.

Для того, щоб ідентифікувати яке-небудь рівняння можна накладати на **змінні структурної** форми додаткові обмеження. Існує кілька способів перевірки **ідентифікованості**:

$$1. \quad \rho(A\varphi) = G - 1$$

де A - матриця отримана шляхом об'єднання матриць B і Γ :

$$A = [B \quad \Gamma]$$

φ - матриця обмежень, у **стовпцях** якої **знаходяться** коефіцієнти при **змінних** що присутні в обмеженнях, наприклад, для обмежень $\beta_{11}=0$ і $\beta_{12}=\beta_{13}$:

$$\varphi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -1 \end{pmatrix}$$

G - Кількість ендогенних **змінних** у рівнянні.

2. Якщо як обмеження виступають тільки виключаючі обмеження, то кількість виключених з рівняння **екзогенних змінних** повинна бути не менше, ніж кількість включених у рівняння **ендогенних змінних зменшена на 1**.

Проілюструємо застосування правил ідентифікації до системи:

$$\begin{cases} \beta_{11}y_{1t} + \beta_{12}y_{2t} + \gamma_{11}x_{1t} + \gamma_{12}x_{2t} = u_{1t} \\ \beta_{21}y_{1t} + \beta_{22}y_{2t} + \gamma_{21}x_{1t} + \gamma_{22}x_{2t} = u_{2t} \end{cases}$$

при обмеженнях $\beta_{12} = 0$ $\gamma_{21} = 0$

Матриця обмежень для першого з рівнянь $\varphi = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$

и. $A\varphi = \begin{pmatrix} \beta_{12} \\ \beta_{22} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ Таким чином, $\rho(A\varphi) = 1 = G - 1 = 2 - 1$ і перше рівняння є **ідентифікованим**. Для другого рівняння:

$$\varphi = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix} \quad A\varphi = \begin{pmatrix} \gamma_{11} \\ \gamma_{21} \end{pmatrix} = \begin{pmatrix} \gamma_{11} \\ 0 \end{pmatrix} \quad \rho(A\varphi) = 1$$

і воно також є **ідентифікованим**. Значення **змінних** β_{11} і β_{22} дорівнюють одиниці відповідно до умови нормалізації, згідно з **якою** в **кожному** з регресійних рівнянь одна з ендогенних **змінних** входить із одиничним коефіцієнтом.

III. Всі методи оцінювання систем одночасних рівнянь можна розбити на **дві** групи: методи, у яких рівняння оцінюються поодиноці або цілком системою. До першої групи **відносять двокроковий МНК**, а до другого - **трикроковий МНК** і метод максимальної правдоподібності.

Найпоширенішим з них є **двокроковий МНК**, схему якого можна **навести** у вигляді послідовності:

- Запишемо рівняння, що цікавить нас, у вигляді:

$$y = Y_1\beta + X_1\gamma + u_1$$

де Y_1 – матриця значень **інших ендогенних змінних**, що **входять** у рівняння;

X_1 – матриця значень **екзогенних змінних**, що **входять** у рівняння.

- Побудові регресії **ендогенних змінних**, що **входять** у **певне** рівняння на всі екзогенні **змінні** системи:

$$\hat{Y}_1 = X(X'X)^{-1}X'Y_1$$

- Оцінці коефіцієнтів β і γ у регресії у на Y_1 і X_1 с заміною значень Y_1 на оцінені на попередньому кроці:

$$\begin{bmatrix} Y_1'X(X'X)^{-1}X'Y_1 & Y_1'X_1 \\ X_1'Y_1 & X_1'X_1 \end{bmatrix} \begin{bmatrix} b \\ c \end{bmatrix} = \begin{bmatrix} Y_1'X(X'X)^{-1}X'y \\ X_1'y \end{bmatrix}$$

При згаданих припущеннях про випадкові **збурювання** й екзогенні **змінні** можна показати, що оцінки **двокрокового МНК** є спроможними.

Приклад:

Умова: Задана система трьох регресійних рівнянь із трьома екзогенними змінними.

Потрібно:

1. Визначити, чи є дана система ідентифікованою.
2. В залежності від результату оцінити параметри тих рівнянь, які є ідентифікованими

Обмеження на коефіцієнти структурної форми:

$$\beta_{1,3} := 0 \quad \gamma_{1,1} := 0$$

$$\beta_{2,1} := 0 \quad \gamma_{2,3} := 0 \quad \gamma_{2,1} := \gamma_{2,2}$$

$$\gamma_{3,1} := \gamma_{3,2} \quad \gamma_{3,3} := 0$$

$$Y = \begin{pmatrix} 8.91 & 10.18 & 10.25 \\ 4.83 & 7.42 & 8.95 \\ 3.31 & 7.06 & 9.59 \\ 7.89 & 7.08 & 6.93 \\ 10.96 & 8.42 & 3.33 \\ -0.71 & 5.13 & 9.71 \\ -18.35 & -11.38 & -0.43 \\ -24.76 & -11.41 & 6.55 \\ -25.89 & -9.52 & 10.81 \\ -18.28 & -6.43 & 8.06 \end{pmatrix} \quad X = \begin{pmatrix} 0.62 & 2.48 & 2.11 \\ 1.64 & 1.29 & 2.93 \\ 3.55 & 0.37 & 7.23 \\ 4.1 & -0.99 & 4.02 \\ 4.13 & -4.57 & 5.36 \\ 9.09 & -3.9 & 10.05 \\ 10.26 & -6.13 & 2.18 \\ 13.11 & -4.95 & 5.01 \\ 17.88 & -7.27 & 9.43 \\ 16.62 & -8.97 & 8.57 \end{pmatrix}$$

1. Перевіримо виконання умов ідентифікації
Для першого рівняння система обмежень виглядатиме:

$$\phi := \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix} \quad A\phi := \begin{pmatrix} \beta_{1,3} & \gamma_{1,1} \\ \beta_{2,3} & \gamma_{2,1} \\ \beta_{3,3} & \gamma_{3,1} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \beta_{2,3} & \gamma_{2,1} \\ 1 & \gamma_{3,1} \end{pmatrix}$$

Ранг матриці $A\phi$ дорівнює $\rho(A\phi)=2=G-1=3-1=2$. Таким чином, перше рівня точно ідентифікованим. Його можна оцінити як за допомогою непрямого МНК, так і за допомогою двокрокового МНК.

Для другого рівняння:

$$\phi := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \quad A\phi := \begin{pmatrix} \beta_{1,1} & \gamma_{1,3} & \gamma_{1,1} - \gamma_{1,2} \\ \beta_{2,1} & \gamma_{2,3} & \gamma_{2,1} - \gamma_{2,2} \\ \beta_{3,1} & \gamma_{3,3} & \gamma_{3,1} - \gamma_{3,2} \end{pmatrix} = \begin{pmatrix} \beta_{1,1} & \gamma_{1,3} & -\gamma_{1,2} \\ 0 & 0 & 0 \\ \beta_{3,1} & 0 & \gamma_{3,1} - \gamma_{3,2} \end{pmatrix}$$

$\rho(A\phi)=2$. Друге рівняння є надідентифікованим. Його можна оцінити за допомогою двокрокового МНК

Для третього рівняння:

$$\phi := \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ -1 & 0 \\ 0 & 1 \end{pmatrix} \quad A\phi := \begin{pmatrix} \gamma_{1,1} - \gamma_{1,2} & \gamma_{1,3} \\ \gamma_{2,1} - \gamma_{2,2} & \gamma_{2,3} \\ \gamma_{3,1} - \gamma_{3,2} & \gamma_{3,3} \end{pmatrix} = \begin{pmatrix} -\gamma_{1,2} & \gamma_{1,3} \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$\rho(A\phi)=1 < G-1$. Третє рівняння не є ідентифікованим.

Оцінимо параметри першого рівняння за допомогою непрямого МНК

Оцінюємо параметри приведеної форми:

$$\pi_1 := \left((X)^T \cdot X \right)^{-1} \cdot (X)^T \cdot Y^{(1)} \quad \pi_2 := \left(X^T X \right)^{-1} X^T \cdot Y^{(2)} \quad \pi_3 := \left(X^T X \right)^{-1} X^T \cdot Y^{(3)}$$

$$\pi := \text{augment}(\pi_1, \pi_2, \pi_3)$$

$$\pi := \pi^T$$

$$\pi = \begin{pmatrix} -3.163 & -0.946 & 2.783 \\ -1.535 & 0.258 & 2.269 \\ 0.782 & 1.777 & 1.116 \end{pmatrix}$$

Рівняння $\beta\Pi=\Gamma$ для перших рядків матриць β та Γ виглядає як:

$$(\beta_{1,1} \ \beta_{1,2} \ \beta_{1,3}) \cdot \Pi = -(\gamma_{1,1} \ \gamma_{1,2} \ \gamma_{1,3})$$

або з урахуванням обмежень на коефіцієнти першого рядка структурної форми:

$$(1 \ \beta_{1,2} \ 0) \cdot \begin{pmatrix} \pi_{1,1} & \pi_{1,2} & \pi_{1,3} \\ \pi_{2,1} & \pi_{2,2} & \pi_{2,3} \\ \pi_{3,1} & \pi_{3,2} & \pi_{3,3} \end{pmatrix} = -(\gamma_{1,2} \ \gamma_{1,3})$$

Що дає систему із трьох рівнянь з трьома невідомими:

$$\pi_{1,1} + \beta_{1,2} \cdot \pi_{2,1} = 0$$

$$\pi_{1,2} + \beta_{1,2} \cdot \pi_{2,2} = -\gamma_{1,2}$$

$$\pi_{1,3} + \beta_{1,2} \cdot \pi_{2,3} = -\gamma_{1,3}$$

Ця система має єдиний розв'язок:

$$\beta_{1,2} := -\frac{\pi_{1,1}}{\pi_{2,1}}$$

$$\gamma_{1,2} := -(\pi_{1,2} + \beta_{1,2} \cdot \pi_{2,2})$$

$$\gamma_{1,3} := -(\pi_{1,3} + \beta_{1,2} \cdot \pi_{2,3})$$

$$\beta_{1,2} = -2.061 \quad \gamma_{1,2} = 1.478 \quad \gamma_{1,3} = 1.892$$

Оцінимо параметри другого рівняння за допомогою двокрокового МНК. В друге рівняння окрім другої ендогенної змінної входить ще і третя. Побудуємо регресію y_3 на всі екзогенні змінні.

$$y_3 := X \cdot \pi_3$$

Побудуємо регресію другої ендогенної змінної на екзогенні змінні, що входять до другого рівняння (першу та другу) та оцінені на попередньому кроці значення третьої ендогенної змінної:

$$X_3 := \text{augment}(y_3, X^{(1)}, X^{(2)})$$

$$\begin{pmatrix} \beta_{2,3} \\ \gamma_{2,1} \\ \gamma_{2,2} \end{pmatrix} := -(X_3^T X_3)^{-1} X_3^T Y^{(2)}$$

$$(\beta_{2,3} \ \gamma_{2,1} \ \gamma_{2,2}) = (-2.032 \ 3.124 \ 3.353)$$

Таким чином, перші два рівняння виглядають я

$$Y_{t,1} := 2.061 Y_{t,2} - 1.478 X_{t,2} - 1.892 X_{t,3}$$

$$Y_{t,2} := 2.032 Y_{t,3} - 3.124 X_{t,1} - 3.353 X_{t,2}$$

ЕЛЕКТРОННЕ НАВЧАЛЬНО-МЕТОДИЧНЕ ВИДАННЯ

Хребет Валерій Григорович
Фесенко Дмитро Володимирович

**МЕТОДИЧНІ ВКАЗІВКИ
ДО ВИКОНАННЯ ПРАКТИЧНИХ РОБІТ
З ДИСЦИПЛІНИ "ЕКОНОМЕТРИКА"
(ДЛЯ СТУДЕНТІВ НАПРЯМУ 6.030502
"ЕКОНОМІЧНА КІБЕРНЕТИКА" ТА
6.030601 "МЕНЕДЖМЕНТ ОРГАНІЗАЦІЙ")**

Підписано до випуску 2012 р. Гарнітура Times New.

Умов. друк. арк. 3,0. Зам. № 94

Державний вищий навчальний заклад

" Донецький національний технічний університет "

Автомобільно-Дорожній інститут

84646, м. Горлівка, вул. Кірова, 51

E-mail:

Редакційно – видавничий відділ

Свідоцтво про внесення до Державного реєстру видавців,
виготовників і розповсюджувачів видавничої продукції ДК № 2982 від
21.09.2007 р.