Міністерство освіти і науки України
Центральноукраїнський національний технічний університет

Економічний факультет

Кафедра економіки, підприємництва та готельно-ресторанної справи

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

до виконання лабораторних робіт
з дисципліни «МОДЕЛЮВАННЯ В УПРАВЛІННІ СОЦІАЛЬНО-ЕКОНОМІЧНИМИ СИСТЕМАМИ»

для здобувачів освітньо-кваліфікаційного рівня «магістр»
зі спеціальностей 051 «Економіка»
(освітньо-професійні програми «Економіка підприємства», «Економіка агробізнесу та агротрейдінг») та 076 «Підприємництво та торгівля» (освітньо-професійна програма «Економіка та організація бізнесу»)
усіх форм навчання

Кропивницький
2024

Міністерство освіти і науки України
Центральноукраїнський національний технічний університет

Економічний факультет

Кафедра економіки, підприємництва та готельно-ресторанної справи

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

до виконання практичних робіт
з дисципліни «МОДЕЛЮВАННЯ В УПРАВЛІННІ СОЦІАЛЬНО-ЕКОНОМІЧНИМИ СИСТЕМАМИ»

для здобувачів освітньо-кваліфікаційного рівня «магістр»
зі спеціальностей 051 «Економіка»
(освітньо-професійні програми «Економіка підприємства», «Економіка агробізнесу та агротрейдінг») та 076 «Підприємництво та торгівля» (освітньо-професійна програма «Економіка та організація бізнесу»)
усіх форм навчання

Затверджено
на засіданні кафедри економіки, підприємництва та готельно-ресторанної справи
Протокол № 8 від 09.01.2024

Кропивницький
2024

МОДЕЛЮВАННЯ В УПРАВЛІННІ СОЦІАЛЬНО-ЕКОНОМІЧНИМИ СИСТЕМАМИ. Методичні вказівки до виконання практичних робіт для здобувачів освітньо-кваліфікаційного рівня «магістр» зі спеціальностей 051 «Економіка» (освітньо-професійні програми «Економіка підприємства», «Економіка агробізнесу та агротрейдінг») та 076 «Підприємництво та торгівля» (освітньо-професійна програма «Економіка та організація бізнесу») усіх форм навчання /Уклад. Б.В. Дмитришин; Центральноукраїн. нац. техн. ун-т. Кропивницький : ЦНТУ, 2024. 68 с.

Укладач: Дмитришин Б.В., канд.екон.наук, доц.

Рецензент: Зайченко В.В., докт. екон. наук, проф.

ЗМІСТ

ВСТУП

Практичні заняття з курсу “Моделювання в управлінні соціально-економічними системами” обов’язкові для студентів I курсу здобуття освітньо-кваліфікаційного рівня «магістр» зі спеціальностей 051 «Економіка» (освітньо-професійні програми «Економіка підприємства», «Економіка агробізнесу та агротрейдінг») та 076 «Підприємництво та торгівля» (освітньо-професійна програма «Економіка та організація бізнесу») усіх форм навчання.

Метою даного практикуму, є набуття навичок роботи з пакетами прикладних комп’ютерних програм, зокрема Excel, Mathcad (опційно) та STATISTICA (опційно), під час побудови та аналізу економетричних моделей, їх розв’язків, а також оволодіння процесом постановки задач, пов’язаних з соціально-економічними системами і процесами.

Вказівки до кожної практичної роботи містить тему, конкретну мету й певні теоретичні відомості. Теорію до практичної роботи необхідно самостійно опрацювати й відповісти на контрольні питання. На заняттях у комп’ютерному класі бажано виконати всі процедури зазначені в основних завданнях по економіко-математичному моделюванню і захистити виконану практичну роботу. Базу даних для економетричних моделей студенти формують за таблицями завдань, поданих в кожній роботі, згідно порядку виконання і свого варіанту.

В основних завданнях, викладених після теоретичних відомостей, зазначено, які дані таблиць використовуються в практичній роботі, спосіб вибору варіанту, за яким формується база даних, вид моделі і пакети прикладних програм, які найбільш доцільно використати.

При необхідності номер варіанту по роботі студент може узгоджувати з викладачем.

По кожній виконаній здобувачем практичній роботі складається звіт (файл формату Microsoft Word .doc, .docx, .rtf або посилання на гугл-документ / документ OpenOffice), в якому подається номер лабораторної роботи, тема, мета, вибрана база даних, математична модель, алгоритм виконання лабораторної роботи у реалізованому вигляді, а також аналіз розв’язку моделі та висновки.

Після виконання практичної роботи студентом проводиться її захист. На захисті викладач задає питання з теорії, процесу виконання на комп’ютері, розглядається сутність використаних понять та видів економетричних моделей та методи їх розв’язку й аналізу.

Типові контрольні питання розміщені у кінці кожної практичної роботи. В кінці методичних вказівок також наведено посилання на літературні джерела. Все це необхідно для самостійного опрацювання з відповідного матеріалу.

Звіт про виконання практичної роботи, розрахунки й аналіз повинні зберігатися до кінця семестру й складання підсумкового контролю з дисципліни. Слід пам’ятати, що всі роботи, без винятку, кожен студент повинен виконати самостійно і здати викладачу збірку звітів з цих робіт під час одержання заліку.

В комп’ютерному класі проводяться консультації і є графік додаткових занять, на яких студенти відробляють пропущені заняття, або виконують завдання, що не встигли виконати на основному занятті.

Практична робота №1

Тема: Побудова та аналіз одночинникової економетричної моделі.

Мета: На запропонованій базі даних побудувати й оцінити одночинникову економетричну модель, а також провести аналіз її достовірності

Короткі теоретичні відомості

При дослідженні різноманітних економічних явищ і процесів у вигляді економетричної моделі зв'язків між економічними показниками, часто можна виявити такий показник, який здійснює найсуттєвіший вплив на результативну ознаку і є найбільш важливим.

Кількісний зв'язок між змінною у, що характеризує результативну ознаку і незалежною змінною x, що характеризує найбільш важливий чинник, дає одночинникова економетрична модель. Загальний вигляд такої моделі:

$$y = f(x, e), \tag{1.1}$$

де е – стохастична складова моделі.

Аналітична форма економетричної моделі залежить від економетричної сутності зв'язків.

В економічній практиці найбільш поширеними є такі форми аналітичних залежностей:

$$y = b_0 + b_1 x + e; \tag{1.2}$$

$$y = b_0 \exp(b_1 x) + e; \tag{1.3}$$

$$y = b_0 x^{b_1} + e; \tag{1.4}$$

$$y = b_0 + \frac{b_1}{x} + e; \tag{1.5}$$

де b_1, b_0 - невідомі параметри моделі.

Можна помітити, що за допомогою елементарних перетворень нелінійні форми залежності можна привести до лінійних не враховуючи стохастичної складової моделі.

Наприклад, логарифмуванням залежності (1.3) і (1.4) без стохастичних частин, можна привести до вигляду:

$$\ln y = \ln b_0 + b_1 x, \tag{1.6}$$

$$\ln y = \ln b_0 + b_1 \ln x. \tag{1.7}$$

Після чого, шляхом заміни змінних, легко отримати явні лінійні аналітичні залежності.

Нехай лінійна одночинникова економетрична модель має вигляд:

$$y = b_0 + b_1 x + e, \tag{1.8}$$

Стохастична складова e має нульове математичне сподівання $(M(e) = 0)$ і постійну дисперсію ($\sigma_e^2 = const$).

В цьому випадку невідомі параметри моделі (1.8) можна оцінити на основі звичайного методу найменших квадратів (1 МНК).

З курсу математики відомо, що 1 МНК використовує принцип мінімізації квадратів залишків стохастичної складової моделі. Застосовуючи необхідну умову мінімізації функції двох змінних, отримуємо систему нормальних рівнянь:

$$\begin{cases} n\breve{b}_0 + \sum\limits_{i=1}^{n} x_i \cdot \breve{b}_1 = \sum\limits_{i=1}^{n} y_i \\ \sum\limits_{i=1}^{n} x_i \cdot \breve{b}_0 + \sum\limits_{i=1}^{n} x_i^2 \cdot \breve{b}_1 = \sum\limits_{i=1}^{n} x_i y_i \end{cases}, \tag{1.9}$$

де $\sum_{i=-1}^{n} x_i$, $\sum_{i=-1}^{n} y_i$, $\sum_{i=-1}^{n} x_i^2$, $\sum_{i=-1}^{n} x_i y_i$ - адитивні величини, які можна розрахувати на основі бази вихідних даних, n – кількість статистичних спостережень.

Розв'язок системи рівнянь (1.10) дає можливість одержати оцінки невідомих параметрів моделі (1.8) $\hat{b}_0, \hat{b}_1$:

$$\hat{b}_1 = \frac{\sum_{i=1}^{n} (y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{n} (x_i - \overline{x})^2}, \tag{1.10}$$

$$\hat{b}_0 = \overline{y} - \hat{b}_1 x. \tag{1.11}$$

Для визначення параметра $\hat{b}_1$ можна використати і інші вирази:

$$\hat{b}_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \overline{x} \cdot \overline{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\overline{x})^2} = \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)}; \tag{1.12}$$

де $\operatorname{cov}(x, y)$ – коваріація між змінними x і y;

$\operatorname{var}(x)$ – дисперсія чинника x.

За допомогою матричних перетворень параметри моделі можна дістати наступним чином:

$$B=(X^T X)^{-1} X^T Y$$

де Y – матриця-стовбець значень y_i;

X^T – транспонована матриця $X=[X_0, X_1]$

Стовбець X_0 складається з одних одиниць

Лінійна економетрична модель матиме вигляд:

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x, \tag{1.13}$$

де символ “кутик” над y, b_0, b_1 означає, що їхні значення отримують в результаті розрахунків, тобто вони є оцінками того реального значення, яке можна встановити в процесі статистичного спостереження.

Достовірність побудованої моделі (1.13) можна перевірити, знаючи дисперсії залишків та результативної ознаки:

$$\sigma_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n - m}, \tag{1.14}$$

$$\sigma_y^2 = \frac{\sum_{i=1}^{n} (y_i - \overline{y})^2}{n - 1}, \tag{1.15}$$

де $e_i = y_i - \hat{y}_i$, m – кількість невідомих параметрів моделі (у даному випадку $m = 2$).

Коефіцієнти детермінації і кореляції визначають за виразами:

$$R^2 = \frac{\sigma_y^2 - \sigma_e^2}{\sigma_y^2}; \tag{1.16}$$

$$R = \sqrt{R^2}. \tag{1.17}$$

Стандартну помилку кожного параметра моделі (1.13) знаходять за виразом:

$$S_{\hat{b}_{j-1}} = \sigma\sqrt{c_{jj}}, \tag{1.18}$$

де c_{jj} – відповідний діагональний елемент матриці помилок C – матриці, оберненої до матриці системи нормальних рівнянь (1.10):

$$C = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}, \tag{1.19}$$

Враховуючи (1.19) та (1.18), маємо вирази для стандартної та відносної γ оцінок параметрів моделі:

$$S_{b_0} = \sqrt{\sigma_e^2 \cdot c_{11}}, \tag{1.20}$$

$$S_{b_1} = \sqrt{\sigma_e^2 \cdot c_{22}}, \tag{1.21}$$

$$\gamma_{b_0} = \frac{S_{b_0}}{b_0} \cdot 100\%, \tag{1.22}$$

$$\gamma_{b_1} = \frac{S_{b_1}}{b_1} \cdot 100\%, \tag{1.23}$$

За оціненим параметром $\hat{b}_1 = \frac{d\hat{y}}{dx}$ можна визначити коефіцієнт еластичності:

$$E_{y/x} = \frac{d\hat{y}}{dx} \cdot \frac{\bar{x}}{\bar{y}} = \hat{b}_1 \cdot \frac{\bar{x}}{\bar{y}}, \tag{1.24}$$

який показує, на скільки відсотків у середньому зміниться результат, якщо чинник зміниться на один відсоток.

Для обробки одновимірних залежностей в Excel є декілька вбудованих функцій.

Функція КОРРЕЛ повертає коефіцієнт парної кореляції між x та y, функція НАКЛОН – коефіцієнт регресії b_1, функція ОТРЕЗОК – вільний член лінійної моделі b_0, функція ПРЕДСКАЗ обчислює розрахункове значення за лінійною моделлю $Y_p=b_0+b_1*x$ для вказаного значення x. У всіх цих функціях треба вказувати посилання на діапазони даних X та Y, причому для функції ПРЕДСКАЗ ці посилання мають бути абсолютними ($), щоб функція правильно працювала при копіюванні формули. Першим аргументом функції ПРЕДСКАЗ є посилання на поточне значення *x* (це посилання не має бути абсолютним, інакше при копіюванні функція буде повертати одне і те саме значення). Функцію ПРЕДСКАЗ можна застосовувати також як функцію діапазону, для чого необхідно попередньо виділити діапазон результату, в якості першого аргументу функції задати не одне значення *x*, а весь діапазон *X*, введення завершити комбінацією Ctrl+Shift+OK.

За допомогою функції ЛИНЕЙН можна докладно аналізувати як одномірні, так і багатомірні залежності, що знадобиться в подальшому.

Як для будь-якої функції діапазону тут треба попередньо виділити діапазон виводу результатів, який буде завжди мати 5 рядків, а число стовбців має дорівнювати числу параметрів моделі (зараз у нас три параметри b0, b1). Не знімаючи виділення, набираємо знак рівності (=), викликаємо функцію ЛИНЕЙН і заповнюємо її поля вводу: Известные значения Y – стовбець Y, Известные значения X – X0, X1, Конст – 1 (відповідає логічному значенню ІСТИНА), Статистика – 1 (те саме). Введення завершуємо комбінацією Ctrl+Shift+OK.

В перших двох рядках діапазону результату функції ЛИНЕЙН виводяться коефіцієнти регресії і стандартні похибки цих коефіцієнтів в зворотному порядку. В інших рядках перших двох стовбців наводяться наступні характеристики: коефіцієнт детермінації R2, стандартна похибка Se, дисперсійне відношення Фішера F, залишкове число степенів свободи dfE, суми квадратів розрахункових значень і залишків моделі.

Завдання

За даними витрат на просування деякого товару (x, тис. дол. США) та розміром отриманого прибутку (y, тис. дол. США) за шістнадцять місяців поданими в таблиці 1.1, побудувати одночинникову економетричну модель. Оцінити параметри й провести аналіз.

Таблиця 1.1*

№ п/п	x_i	y_i
1	48+N/10	1590+N/10
2	55+N/10	1760+N/10
3	61+N/10	1975+N/10
4	67+N/10	2205+N/10
5	78+N/10	2460+N/10
6	88+N/10	2725+N/10
7	91+N/10	3035+N/10
8	114+N/10	3320+N/10
9	122+N/10	3485+N/10
10	147+N/10	3575+N/10
11	163+N/10	3785+N/10
12	200+N/10	4025+N/10
13	240+N/10	4150+N/10
14	275+N/10	4275+N/10
15	300+N/10	4540+N/10
16	330+N/10	4730+N/10

*Примітка: варіант формується за допомогою Таблиці 1.1, в якій N – це дві останні цифри номера Вашої залікової книжки.

Порядок виконання

1. За своїм варіантом сформувати вихідну базу даних та побудувати таблицю статистичних показників.
2. Скласти систему нормальних рівнянь (1.9) і розв'язати її відносно невідомих коефіцієнтів.
3. Порівняти одержані розв'язки з оцінками, проведеними за виразами (1.10), (1.11) і (1.12).
4. Отримати оцінки параметрів моделі за допомогою матричних розрахунків
5. Отримати аналітичний вираз економетричної моделі (1.13) і дати графічну інтерпретацію моделі та сукупності бази даних. За виразами (1.14) і (1.15) обчислити дисперсії залишків та результативної ознаки.
6. Оцінити за виразом (1.16) ступінь значущості і характер вимірюваного зв'язку, а за (1.17) тісноту зв'язку між змінними моделі.
7. За виразом (1.19) побудувати матрицю помилок.
8. Визначити стандартні та відносні помилки параметрів моделі за виразами (1.20), (1.22), а також обчислити відносні їх розходження за виразами (1.22), (1.23).
9. Визначити коефіцієнт еластичності (1.24) залежності між змінними в отриманій економетричній моделі.
10. Проаналізувати отримані результати і зробити висновки.
11. Перевірити значення отриманих параметрів моделі та її якісних характеристик за допомогою вбудованих функцій КОРРЕЛ, НАКЛОН, ОТРЕЗОК, ЛИНЕЙН.
12. Отримати прогнозні значення Y для x=350+N/10 та 375+N/10 за допомогою функції ПРЕДСКАЗ.

Приклад виконання

1. Нехай база статистичних даних, що характеризує зміну певного економічного показника за дванадцять місяців, представлена двома першими колонками таблиці 1.2

За незалежну змінну вибираємо номер місяця, за залежну – значення економічного показника. В таблиці 1.2 наведені розрахунки показників x_i^2, $x_i\ y_i$, $\hat{y}_i$, $e_i = y_i - \hat{y}$, e_i^2, $(y_i - \bar{y})^2$ які будуть використані в наступних пунктах алгоритму.

Таблиця 1.2

x_i	y_i	x_i^2	$x_i\ y_i$	$\hat{y}_i$	$e_i = y_i - \hat{y}$	e_i^2	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1	2	3	4	5	6	7	8	9
1	5,23	1	5,23	5,81	-0,58	0,34	-3,07	9,42
2	6,37	4	12,74	6,26	0,11	1,01	-1,93	3,72
3	7,45	9	22,35	6,71	0,74	0,55	-0,85	0,72
4	7,17	16	28,68	7,16	0,01	0	-1,13	1,28
5	8,01	25	40,05	7,61	0,40	0,16	-0,29	0,08
6	8,50	36	51,00	8,06	0,44	0,19	0,2	0,04
7	8,07	49	56,49	8,51	-0,44	0,19	-0,23	0,05
8	7,66	64	61,28	8,96	-1,3	1,69	-0,64	0,41
9	9,75	81	87,75	9,41	0,34	0,12	1,45	2,10
10	9,87	100	98,70	9,86	0,01	0	1,57	2,46
11	10,54	121	115,94	10,31	0,230	0,05	2,24	5,02
12	11,00	144	132,00	10,76	0,24	0,06	2,7	7,29
Σ=78	99,62	650	712,21		0,2	3,36	~0	32,59

2. Вважаємо, що між чинником x та результативною змінною y існує лінійна залежність. Для складення системи нормальних рівнянь (1.10) доповнимо таблицю рядком Σ, де будемо проставляти значення сум чисел, що стоять у відповідних стовпчиках.

Використавши значення рядка Σ отримаємо систему рівнянь:

$$\begin{cases} 12\hat{b}_0 + 78\hat{b}_1 = 99{,}62 \\ 78\hat{b}_0 + 650\hat{b}_1 = 712{,}21 \end{cases} \Rightarrow \begin{cases} \hat{b}_0 + 6{,}5\hat{b}_1 = 8{,}30 \\ \hat{b}_0 + 8{,}33\hat{b}_1 = 9{,}13 \end{cases}$$

Віднімемо від другого рівняння перше, знаходимо:

$\hat{b}_1 = \dfrac{9{,}13 - 8{,}30}{8{,}33 - 6{,}5} = 0{,}454$, з першого рівняння:

$\hat{b}_0 = 8{,}3 - 6{,}5 \cdot 0{,}454 = 5{,}35$

При обчисленні параметра $\hat{b}_1$ доцільно збільшити точність, оскільки вплив його на точність моделі значний.

3. Коефіцієнт $\hat{b}_1$ можна обчислити і іншим шляхом, наприклад, з виразу:

$$\hat{b}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\,\bar{y}}{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2}.$$

Знаходимо середні значення змінних:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{12} \cdot 78 = 6{,}5;$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{12} \cdot 99{,}62 = 8{,}3.$$

$$\text{Звідси: } \hat{b}_1 = \frac{\frac{1}{12}\cdot 712{,}21 - 6{,}5\cdot 8{,}3}{\frac{1}{12}\cdot 650 - 6{,}5^2} = \frac{5{,}361}{11{,}9167} = 0{,}450\,;$$

$$\hat{b}_0 = \bar{y} - b_1\bar{x} = 8{,}3 - 0{,}45\cdot 6{,}5 = 5{,}37$$

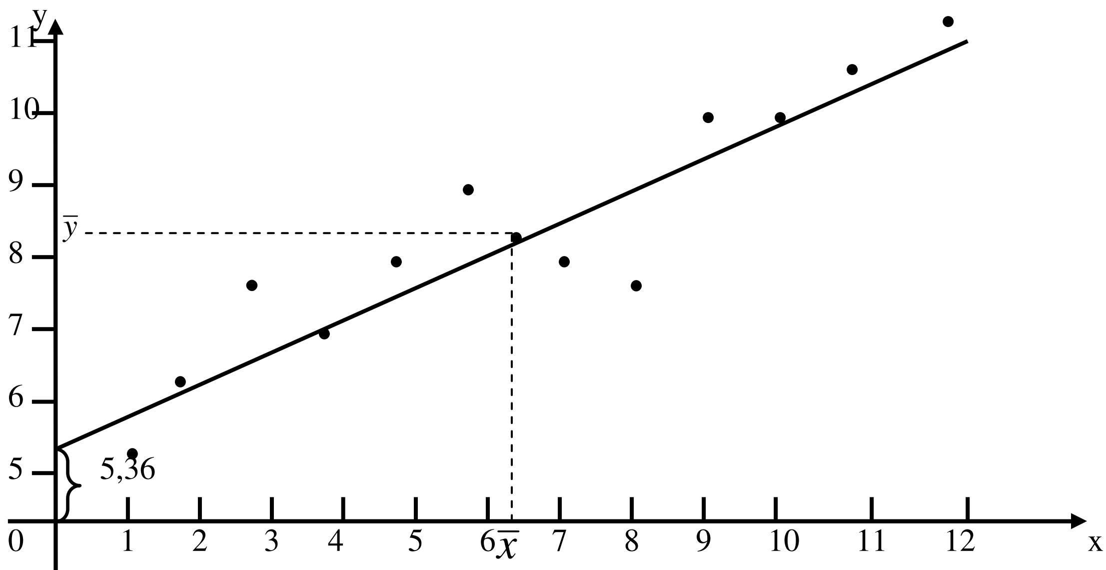
4. За допомогою матричних розрахунків отримуємо

$$B=(X^TX)^{-1}X^TY=(0{,}452;5{,}362)^T$$

5. За виразом (1.13) аналітична формула одночинникової лінійної моделі має вигляд:

$$\hat{y} = 5{,}36 + 0{,}45x\,.$$

Щоб отримати графічну інтерпретацію моделі, побудуємо в системі координат x0y пряму, що відповідає отриманому рівнянню. Нанесемо також на тій же координатній площині точки з координатами $(x_i; y_i)$, взятими з бази даних.



При побудові теоретичної прямої використаємо ту обставину, що їй належать точки (0;5,36) та $(\bar{x}; \bar{y})$ =(6,5; 8,3).

Як видно з рисунка теоретична лінія загалом відображає основні тенденції взаємозв’язку між змінними x та y.

6. Для обчислення дисперсії залишків, використаємо значення елементів стовпчика (6), які обчислені за виразом

$$e_i = y_i - \hat{y}_i$$

з використанням відповідно стовпчиків 2 і 5. Далі обчислюємо e_i^2. Після цього знаходимо дисперсію залишків:

$$\sigma_e^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-m} = \frac{3{,}36}{12-2} = 0{,}336$$

Знаходимо також відхилення залежної змінної від її середнього значення та його квадрат за виразами:

$$y_i - \bar{y} \quad та \quad (y_i - \bar{y})^2$$

Результати обчислень заносимо у колонки 8 та 9 таблиці 1.2.

Знаходимо дисперсію результативної змінної:

$$\sigma_y^2 = \frac{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}{n-1} = \frac{32,59}{12-1} = 2,96\,.$$

7. Використавши знайдені значення σ_e^2 і σ_y^2 визначаємо коефіцієнт детермінації:

$$R^2 = \frac{\sigma_y^2 - \sigma_e^2}{\sigma_y^2} = \frac{2,96 - 0,336}{2,96} = 0,886\,.$$

За отриманими значеннями коефіцієнта детермінації можна зробити висновок про значущість зв'язку змінних x і y.

Визначаємо коефіцієнт кореляції:

$$R = \sqrt{R^2} = \sqrt{0,886} = 0,931\,.$$

Рівень отриманого значення свідчить про велику тісноту зв'язку між змінними x та y.

8. Матрицю помилок знаходимо за формулою:

$$C = \begin{pmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{pmatrix}^{-1} = \begin{pmatrix} 12 & 78 \\ 78 & 650 \end{pmatrix}^{-1}.$$

Визначник цієї матриці дорівнює :

$$\Delta = \begin{vmatrix} 12 & 78 \\ 78 & 650 \end{vmatrix} = 650 \times 12 - 78^2 = 1716\,.$$

Звідси :

$$C = \begin{pmatrix} 12 & 78 \\ 78 & 650 \end{pmatrix}^{-1} = \frac{1}{1716}\begin{pmatrix} 650 & -78 \\ -78 & 12 \end{pmatrix} = \begin{pmatrix} \frac{325}{858} & -\frac{1}{22} \\ -\frac{1}{22} & \frac{1}{143} \end{pmatrix}.$$

9. Знаходимо стандартну $S_{\hat{b}_0}$ та відносну $\gamma_{\hat{b}_0}$ помилки оцінювання параметра $\hat{b}_0$:

$$S_{\hat{b}_0} = \sqrt{\sigma_e^2 \cdot c_{11}} = \sqrt{0,336 \cdot \frac{325}{858}} = 0,357\,;$$

$$\gamma_{\hat{b}_0} = \frac{S_{\hat{b}_0}}{\hat{b}_0} \cdot 100\% = \frac{0,357}{5,36} \cdot 100\% = 6,66\%\,.$$

Для оцінки параметра $\hat{b}_1$, маємо:

$$S_{\hat{b}_1} = \sqrt{\sigma_e^2 \cdot c_{22}} = \sqrt{0,336 \cdot \frac{1}{143}} = 0,048\,;$$

$$\gamma_{\hat{b}_1} == \frac{S_{\hat{b}_1}}{\hat{b}_1} \cdot 100\% = \frac{0,048}{0,45} \cdot 100\% = 10,67\%$$

10. Знаходимо коефіцієнт еластичності:

$$E_{y/x} = \hat{b}_1 \cdot \frac{\bar{x}}{\bar{y}} = 0,45 \cdot \frac{6,5}{8,3} = 0,352$$

11. За допомогою вбудованої функції КОРРЕЛ знаходимо коефіцієнт кореляції між залежною та незалежною змінними.

R=0,947,

що досить близько до значення отриманого через коефіцієнт детермінації.

Функція НАКЛОН видає значення коефіцієнта b1=0,452, що цілком відповідає отриманим іншими методиками значенням.

За допомогою функції ОТРЕЗОК розраховуємо значення вільного члена. b0=5,362.

Вихідний діапазон функції ЛИНЕЙН має наступний вигляд:

0,452308	5,361667
0,048441	0,356514
0,897104	0,579267
87,18586	10
29,25526	3,355505

В першому рядку в оберненому порядку виведені параметри моделі, в другому – їхні стандартні похибки, що практично збігаються з отриманими в п.9.

12. За допомогою функції ПРЕДСКАЗ розрахуємо прогнозне значення залежної змінної для x=13 та 14.

Для x=13 y=11,242,

для x=14 y=11,694.

13. Враховуючи наведені обчислення та оцінки, можна зробити наступні висновки:

13.1. Обчислення значень b_0 та b_1 за різними методиками дають помилку у третьому знаку після коми, що свідчить про достатню точність обчислень.

13.2. Аналітичний вигляд моделі, як це свідчить з порівняння графічної інтерпретації моделі та бази даних (експериментальних точок), в основному відображає основну тенденцію взаємозв'язку між змінними x та y, а саме: збільшення значення x призводить до збільшення значення y.

13.3. Невелике значення дисперсії залишків відхилень між теоретичною оцінкою та експериментальними значеннями, та їх невідчутна залежність від рівня значень x та y свідчить про доцільність використання методу 1МНК.

13.4. Рівні значень коефіцієнтів детермінації та кореляції свідчать про значущість моделі та тісний лінійний зв'язок чинника і економічного показника.

13.5. Рівні значень стандартних та відносних помилок оцінювання параметрів моделі не є занадто великим, а це говорить про невелику зміщеність оцінок.

13.6. Отримане значення коефіцієнта еластичності вказує на те, що при збільшенні або зменшення значення чинника на 1% значення економічного показника відповідно збільшується або зменшується у середньому на 0,352%.

Контрольні запитання

1. Що собою являє проста економетрична модель?
2. Які форми зв'язків між економетричними показниками найбільш поширенні на практиці?
3. За допомогою яких математичних перетворень нелінійні форми зв'язку можна привести до лінійних?
4. У чому сутність методу найменших квадратів?
5. Як отримати і розв'язати систему нормальних рівнянь?
6. Як оцінити дисперсії залишків і результативної ознаки?
7. Що таке коефіцієнт детермінації? Що він визначає?
8. Що таке коефіцієнт кореляції і що він характеризує?
9. Як побудувати матрицю помилок і оцінити стандартні помилки оцінювання параметрів моделі?
10. Як знайти стандартні помилки і відносні розходження оцінок параметрів економетричної моделі?
11. Економетричний зміст коефіцієнта еластичності. Як обчислити його значення?
12. Як Ви оцінюєте можливості та доцільність використання електронних таблиць Excel при виконанні цієї лабораторної роботи?

13. Як Ви оцінюєте можливості та доцільність використання системи Mathcad при виконанні даної лабораторної роботи?

14. Які переваги має система Statistica при складанні економетричних моделей та оцінювання їхніх параметрів?

15. Порівняйте традиційний та комп’ютерний методи складання та аналізу економетричної моделі.

16. Як оцінити параметри нелінійної однофакторної моделі?

17. Як оцінити параметри лінійної моделі в матричній формі?

18. Для чого призначені та як використовуються вбудовані функції Excel КОРРЕЛ, НАКЛОН, ОТРЕЗОК, ПРЕДСКАЗ?

19. З чого складається діапазон виводу функції ЛИНЕЙН?

Практична робота №2

Тема: Побудова та аналіз багаточинникової економетричної моделі.

Мета: Оцінити параметри багаточинникової економетричної моделі методом 1 МНК й дати їм економічне тлумачення.

Короткі теоретичні відомості

Для кількісного опису зв'язку між результативною ознакою і кількома чинниками необхідно побудувати економетричну модель, що базується на регресійному аналізі.

У загальному випадку ця модель має вигляд

$$\mathbf{y} = \mathbf{f}\,(\mathbf{x_1}, \mathbf{x_2}, \mathbf{x_3}, \ldots, \mathbf{x_k}, \mathbf{e}), \tag{2.1}$$

де $\mathbf{y}$ – вектор змінної, що характеризує результативну ознаку; $\mathbf{x}_j, \boldsymbol{j} = \overline{\mathbf{1}, \boldsymbol{k}}$ – вектор незалежних змінних, що характеризують чинники моделі; $\boldsymbol{e}$ – стохастична складова.

Залежну змінну $\mathbf{y}$ називають також пояснювальною, ендогенною змінною; незалежні змінні $\mathbf{x}_j$ – пояснюючими, продетермінованими, екзогенними змінними.

У найпростішому випадку економетрична модель є лінійною, аналітична форма якої записується у вигляді:

$$\mathbf{y} = \boldsymbol{b_0} + \boldsymbol{b_1}\mathbf{x_1} + \boldsymbol{b_2}\mathbf{x_2} + \ldots + \boldsymbol{b_k}\mathbf{x}_{\boldsymbol{k}} + \boldsymbol{e}, \tag{2.2}$$

де b_p, $p = \overline{0, k}$ – невідомі параметри моделі.

В матричній формі лінійна багаточинникова економетрична модель має вигляд:

$$Y = X \cdot B + \boldsymbol{e} \tag{2.3}$$

де Y – вектор ендогенних змінних; X – матриця екзогенних змінних; B – вектор оцінок параметрів моделі; $\boldsymbol{e}$ – вектор залишків.

Параметри моделі (2.2) можна оцінити на основі методу 1 МНК. Але при цьому необхідно дотримуватися таких передумов (гіпотез):

- математичне сподівання залишків дорівнює нулю, тобто

$$M(\boldsymbol{e}) = 0; \tag{2.4}$$

- значення вектора залишків незалежні між собою і мають постійну дисперсію:

$$M(\boldsymbol{e} \cdot \boldsymbol{e}^T) = \sigma_e^2 \cdot E \tag{2.5}$$

де $\boldsymbol{e}^T$ – транспонований вектор залишків, E – одинична матриця;

- незалежні змінні моделі не зв'язані з залишками, тобто

$$M(X^T \cdot \boldsymbol{e}) = 0; \tag{2.6}$$

де X^T – транспонована матриця незалежних змінних.

Незалежні змінні створюють лінійно-незалежну систему векторів:

$$\operatorname{cov}ar\left(X_i X_j\right) = 0 \qquad i \neq j \qquad i = \overline{1, k} \tag{2.7}$$

$$\operatorname{cov}ar\left(X_i X_j\right) \neq 0 \qquad i = j \qquad j = \overline{1, k}$$

Оператор оцінювання параметрів багаточинникової економетричної моделі на основі 1 МНК дорівнює:

$$\hat{B} = \left(X^T X\right)^{-1} X^T Y \tag{2.8}$$

Можна довести, що оцінки *B*, які можна отримати на основі оператора оцінювання (2.8), мінімізують суму квадратів залишків ($\sum e^2{}_i \Rightarrow \min$). При цьому значення вектора $\hat{B}$ є розв'язком системи нормальних рівнянь:

$$\left(X^T X\right)\hat{B} = X^T Y \tag{2.9}$$

де $(X^T X)^{-1}$ – матриця моментів. Числа, що стоять на головній діагоналі матриці, характеризують величину дисперсій незалежних змінних чинників, а інші елементи відповідають взаємним коваріаціям.

Оцінка параметрів багаточинникової економетричної моделі повинна бути незміщеною і виконуватися рівність:

$$M\left(\hat{B}\right) = B \tag{2.10}$$

У випадку якщо рівність (2.10) не виконується, знаходять різницю $M\left(\hat{B}\right) - B = Q$ – зміщення оцінки.

Оцінка параметрів моделі повинна бути обґрунтованою, тобто при заданій малій величині $\varepsilon > 0$ справедливе співвідношення:

$$\lim_{n \to \infty} P\left\{\left|\hat{B} - B\right| < \varepsilon\right\} = 1 \tag{2.11}$$

Елементами оператора оцінювання

$$\hat{B} = \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \\ \ldots \\ \hat{b}_n \end{pmatrix} \tag{2.12}$$

є оцінки невідомих параметрів моделі:

$$\hat{Y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2 + \ldots + \hat{b}_i x_i + \ldots \hat{b}_n x_n \tag{2.13}$$

Вектор залишків знаходять як різницю векторів:

$$\boldsymbol{e} = Y - \hat{Y}, \tag{2.14}$$

елементи якого визначають за виразом**:**

$$e_i = y_i - \hat{y}_i \tag{2.15}$$

Дисперсії залишків e_i і залежної змінної y_i можна в матричній формі розрахувати за виразами:

$$\sigma_e{}^2 = \frac{1}{n-m}\,(e^T \cdot e) \tag{2.16}$$

Матрицю коваріацій оцінок параметрів моделі в матричній формі визначають за виразом:

$$\mathrm{var}(\hat{B}) = \sigma_e{}^2 \left(X^T X\right)^{-1} \tag{2.17}$$

Діагональні елементи цієї матриці характеризують дисперсії оцінок параметрів моделі:

$$\sigma_{b_0}{}^2, \sigma_{b_1}{}^2, \sigma_{b_2}{}^2, \ldots \sigma_{b_m}^2$$

Всі інші елементи матриці (2.18) визначають рівень коваріації між оцінками параметрів моделі.

Стандартні помилки оцінок параметрів знаходять за виразами:

$$S_{b_0} = \sqrt{\sigma_{b_0}{}^2}\,;\; S_{b_1} = \sqrt{\sigma_{b_1}{}^2}\,, \ldots S_{b_n} = \sqrt{\sigma_{b_n}{}^2} \tag{2.18}$$

а їхні відносні відхилення – за виразами:

$$\gamma_0 = \frac{S_{b_0}}{\hat{b}_0} \times 100\%\,;\; \gamma_1 = \frac{S_{b_1}}{\hat{b}_1} \times 100\% \ldots \gamma_n = \frac{S_{bi}}{\hat{b}_n} \times 100\% \tag{2.19}$$

В середовищі Microsoft Excel для матричних розрахунків наявні такі вбудовані функції:

ТРАНСП(масив) – повертає транспоновану матрицю

МУМНОЖ(масив1;масив2) – повертає добуток двох матриць

МОБР(масив) – розраховує обернену матрицю

Нагадаємо, що ці функції є функціями діапазонів, тому перед їхнім введенням необхідно виділити діапазон клітинок, що співпадає за розмірністю з очікуваною матрицею-результатом, а завершувати введення функції діапазону треба комбінацією клавіш Ctrl+Shift+Пробіл (або Ctrl+Shift+Ok).

Завдання

Дані, що характеризують прибуток підприємства, для шістнадцяти періодів наведені у наступній таблиці 2.1

Таблиця 2.1

№ п/п	Прибуток, млн грн	Матеріальні витрати, млн грн	Витрати на оплату праці, млн грн	Відрахування на соціальні заходи, *100 тис. грн	Амортизація, млн грн
1	$19{,}4+\frac{N}{10}$	$159+\frac{N}{10}$	$26+\frac{N}{10}$	$205+\frac{N}{10}$	$32+\frac{N}{10}$
2	$16{,}8+\frac{N}{10}$	$34+\frac{N}{10}$	$28+\frac{N}{10}$	$46+\frac{N}{10}$	$59+\frac{N}{10}$
3	$18{,}0+\frac{N}{10}$	$253+\frac{N}{10}$	$31+\frac{N}{10}$	$246+\frac{N}{10}$	$30+\frac{N}{10}$
4	$19{,}8+\frac{N}{10}$	$263+\frac{N}{10}$	$40+\frac{N}{10}$	$344+\frac{N}{10}$	$43+\frac{N}{10}$
5	$19{,}2+\frac{N}{10}$	$216+\frac{N}{10}$	$26+\frac{N}{10}$	$216+\frac{N}{10}$	$39+\frac{N}{10}$
6	$17{,}2+\frac{N}{10}$	$216+\frac{N}{10}$	$30+\frac{N}{10}$	$269+\frac{N}{10}$	$32+\frac{N}{10}$
7	$25{,}0+\frac{N}{10}$	$168+\frac{N}{10}$	$29+\frac{N}{10}$	$73+\frac{N}{10}$	$42+\frac{N}{10}$

8	$15{,}2+\frac{N}{10}$	$35+\frac{N}{10}$	$26+\frac{N}{10}$	$42+\frac{N}{10}$	$21+\frac{N}{10}$
9	$13{,}8+\frac{N}{10}$	$52+\frac{N}{10}$	$24+\frac{N}{10}$	$49+\frac{N}{10}$	$20+\frac{N}{10}$
10	$27{,}0+\frac{N}{10}$	$342+\frac{N}{10}$	$31+\frac{N}{10}$	$302+\frac{N}{10}$	$137+\frac{N}{10}$
11	$19{,}4+\frac{N}{10}$	$178+\frac{N}{10}$	$30+\frac{N}{10}$	$319+\frac{N}{10}$	$73+\frac{N}{10}$
12	$21{,}4+\frac{N}{10}$	$240+\frac{N}{10}$	$32+\frac{N}{10}$	$330+\frac{N}{10}$	$25+\frac{N}{10}$
13	$24{,}2+\frac{N}{10}$	$336+\frac{N}{10}$	$40+\frac{N}{10}$	$451+\frac{N}{10}$	$59+\frac{N}{10}$
14	$19{,}4+\frac{N}{10}$	$172+\frac{N}{10}$	$28+\frac{N}{10}$	$226+\frac{N}{10}$	$82+\frac{N}{10}$
15	$14{,}0+\frac{N}{10}$	$59+\frac{N}{10}$	$29+\frac{N}{10}$	$60+\frac{N}{10}$	$13+\frac{N}{10}$
16	$14{,}4+\frac{N}{10}$	$28+\frac{N}{10}$	$26+\frac{N}{10}$	$30+\frac{N}{10}$	$9+\frac{N}{10}$

Дослідити залежність між прибутком підприємства та іншими чинниками, що наведені у таблиці 2.1, і провести аналіз побудованої багатофакторної економетричної моделі. Залежність між результуючою ознакою та чинниками вважати лінійною. При проведенні розрахунків використати систему Excel або Mathcad. Сформувати базу даних згідно свого варіанту:в таблиці 2.1 за N прийняти дві останні цифрі номера своєї залікової книжки.

Порядок виконання

1. Подати базу вихідних даних у вигляді матриці незалежних змінних X , де елементами першого стовпця є одиниці, а також вектора Y залежної змінної.

2. Знайти добутки матриці $X^T \cdot X$ і $X^T \cdot Y$ у виразі для оператора оцінювання (2.8).

3. Знайти обернену матрицю $\left(X^T \cdot X\right)^{-1}$, а також сам оператор оцінювання за виразом (2.8).

4. Прийняти елементи матриці $\vec{B}$ за оціночні невідомі коефіцієнти економетричної моделі.

5. Побудувати економетричну модель для бази даних на основі виразу (2.13).

6. Визначити розрахункові значення залежної змінної $\hat{Y}$, підставивши в модель (2.13) значення незалежних змінних.

7. Отримати вектор залишків (2.14), елементи якого знайти за виразом (2.15).

8. Розрахувати дисперсії залишків і залежної змінної за виразом (2.16) і (2.17).

9. Побудувати матрицю коваріацій оцінок параметрів за виразом (2.17) та знайти дисперсні оцінки параметрів моделі як діагональні елементи матриці.

10. Знайти стандартні помилки (2.18) та їхні відносні відхилення (2.19) оцінювання невідомих параметрів моделі.

11. Знайти параметри моделі, їхні стандартні похибки та інші якісні характеристики моделі за допомогою функції ЛИНЕЙН. Порівняти отримані результати з результатами попередніх розрахунків.

12. Дати змістовне економічне тлумачення параметрів моделі.

Приклад виконання

Нехай відомі дані про збитковість свинини по деяким підприємствам за рік (таблиця 2.2):

Таблиця 2.2

№ підпри-ємства	Середньо-добовий приріст, X_1 грн.	Затрати на 1 ц. приросту X_2 люд.-год.	Затрати на 1 голову, X_3 грн.	Затрати на 1 ц. приросту, X_4 грн.	Собівартість 1 ц., X_5 грн.	Рівень збитковості, Y, %
1	140	96,5	210	393,31	239,57	36,4
2	180	53,8	175	233,21	170,38	8,7
3	122	99,9	199	403,39	290,18	50,6
4	110	130,9	198	429,05	223,86	29,2
5	159	108,3	273	456,17	217,07	45,0
6	79	179,7	172	531,96	327,12	64,2
7	194	72,5	196	316,63	200,74	25,6
8	86	132,1	204	591,22	282,19	57,3
9	111	121,6	200	499,58	302,78	47,2
10	112	126,0	205	594,5	289,73	31,2
11	159	201,8	153	311,38	211,81	68,1
12	54	109,0	317	1019,06	367,02	33,9
13	136	243,6	204	399,6	233,87	44,6
14	85	80,9	190	466,38	280,67	38,1
15	228	103,3	246	349,97	265,82	40,2
16	129	53,9	256	404	230,92	40,6

1. Виходячи з бази даних, вважаємо рівень збитковості залежним фактором, інші змінні – чинниками. Таким чином матриця незалежних змінних X та вектор залежної змінної Y мають вигляд:

$$X = \begin{pmatrix} 1 & 140 & 96,5 & 210 & 393,31 & 239,57 \\ 1 & 180 & 53,8 & 175 & 233,21 & 170,38 \\ 1 & 122 & 99,9 & 199 & 403,39 & 290,18 \\ 1 & 110 & 130,9 & 198 & 429,05 & 223,86 \\ 1 & 159 & 108,3 & 273 & 456,17 & 217,07 \\ 1 & 79 & 179,7 & 172 & 531,96 & 327,12 \\ 1 & 194 & 72,5 & 196 & 316,63 & 200,74 \\ 1 & 86 & 132,1 & 204 & 591,22 & 282,19 \\ 1 & 111 & 121,6 & 200 & 499,58 & 302,78 \\ 1 & 112 & 126 & 205 & 594,5 & 289,73 \\ 1 & 159 & 201,8 & 153 & 311,38 & 211,81 \\ 1 & 54 & 109 & 317 & 1019,06 & 367,02 \\ 1 & 136 & 243,6 & 204 & 399,6 & 233,87 \\ 1 & 82 & 80,9 & 190 & 466,38 & 280,67 \\ 1 & 228 & 103,3 & 246 & 349,97 & 265,82 \\ 1 & 129 & 53,9 & 256 & 404 & 230,92 \end{pmatrix} \qquad Y = \begin{pmatrix} 36,4 \\ 8,7 \\ 50,6 \\ 29,2 \\ 45,0 \\ 64,2 \\ 25,6 \\ 57,3 \\ 47,2 \\ 31,2 \\ 68,1 \\ 33,9 \\ 44,6 \\ 38,1 \\ 40,2 \\ 40,6 \end{pmatrix}$$

2. Використовуючи програму "Матриці" (програмний продукт Mathcad) або функцій MS Excel *ТРАНСП* (транспонування матриці) і *МУМНОЖ* (множення матриць) виконаємо над сформованими об'єктами операції: (X^TX) та (X^TY), маємо:

$$X^{\mathrm{T}} \cdot X = \begin{pmatrix} 16 & 2.081\times10^3 & 1.834\times10^3 & 3.398\times10^3 & 7.399\times10^3 & 4.134\times10^3 \\ 2.081\times10^3 & 3.024\times10^5 & 2.298\times10^5 & 4.386\times10^5 & 8.714\times10^5 & 5.126\times10^5 \\ 1.834\times10^3 & 2.298\times10^5 & 2.429\times10^5 & 3.847\times10^5 & 8.707\times10^5 & 4.834\times10^5 \\ 3.398\times10^3 & 4.386\times10^5 & 3.847\times10^5 & 7.474\times10^5 & 1.639\times10^6 & 8.889\times10^5 \\ 7.399\times10^3 & 8.714\times10^5 & 8.707\times10^5 & 1.639\times10^6 & 3.898\times10^6 & 2.025\times10^6 \\ 4.134\times10^3 & 5.126\times10^5 & 4.834\times10^5 & 8.889\times10^5 & 2.025\times10^6 & 1.107\times10^6 \end{pmatrix}$$

$$X^{\mathrm{T}} \cdot Y = \begin{pmatrix} 660.9 \\ 8.315\times10^4 \\ 8.609\times10^4 \\ 1.387\times10^5 \\ 3.099\times10^5 \\ 1.749\times10^5 \end{pmatrix}$$

3. Знаходимо також $(X^TX)^{-1}$ (функція MS Excel *МОБР*):

$$\left(X^{\mathrm{T}} \cdot X\right)^{-1} = \begin{pmatrix} 8.525 & -0.015 & -5.623\times10^{-3} & -0.01 & 2.174\times10^{-3} & -0.018 \\ -0.015 & 1.184\times10^{-4} & -7.974\times10^{-6} & -7.527\times10^{-5} & 3.261\times10^{-5} & 3.986\times10^{-6} \\ -5.623\times10^{-3} & -7.974\times10^{-6} & 3.17\times10^{-5} & 2.656\times10^{-5} & -6.037\times10^{-6} & 7.654\times10^{-8} \\ -0.01 & -7.527\times10^{-5} & 2.656\times10^{-5} & 1.346\times10^{-4} & -4.06\times10^{-5} & 2.696\times10^{-5} \\ 2.174\times10^{-3} & 3.261\times10^{-5} & -6.037\times10^{-6} & -4.06\times10^{-5} & 2.012\times10^{-5} & -2.468\times10^{-5} \\ -0.018 & 3.986\times10^{-6} & 7.654\times10^{-8} & 2.696\times10^{-5} & -2.468\times10^{-5} & 9.103\times10^{-5} \end{pmatrix}$$

4.Для отримання значень невідомих параметрів моделі використовуємо вираз для оператора оцінювання:

$$\hat{B} = \left(X^T \cdot X\right)^{-1} \cdot (X^T \cdot Y)$$

Маємо:

$$\hat{B} = \begin{pmatrix} -6.753 \\ -0.108 \\ 0.175 \\ 0.11 \\ -0.082 \\ 0.215 \end{pmatrix}$$

Отже, для даної моделі: $\hat{b}_0 = -6{,}753$; $\hat{b}_1 = -0{,}108$; $\hat{b}_2 = 0{,}175$; $\hat{b}_3 = 0{,}11$; $\hat{b}_4 = -0{,}082$; $\hat{b}_5 = 0{,}215$;

5. Таким чином багаточинникова економетрична, модель має вигляд:

$$Y = -6{,}753 - 0{,}108X_1 + 0{,}175X_2 + 0{,}11X_3 - 0{,}082X_4 + 0{,}215X_5$$

6. Розрахункові (теоретичні) значення залежної змінної визначаємо за формулою:

$$\hat{Y} = X \cdot B$$

При цьому маємо:

$$\hat{Y} = X \cdot \hat{B} = \begin{pmatrix} 37.568 \\ 20.215 \\ 48.935 \\ 39.147 \\ 34.558 \\ 61.919 \\ 24.009 \\ 41.816 \\ 48.835 \\ 39.447 \\ 48.475 \\ 36.765 \\ 61.381 \\ 41.643 \\ 42.587 \\ 33.6 \end{pmatrix}$$

7. Отримаємо вектор залишків, елементи якого знайдемо за виразом:

$$e_i = y_i - \hat{y}_i$$

$$e = \begin{pmatrix} -1.168 \\ -11.515 \\ 1.665 \\ -9.947 \\ 10.442 \\ 2.281 \\ 1.591 \\ 15.484 \\ -1.635 \\ -8.247 \\ 19.625 \\ -2.865 \\ -16.781 \\ -3.543 \\ -2.387 \\ 7 \end{pmatrix}$$

8. Розраховуємо дисперсію залишків і залежної змінної за виразами:

$$\sigma_e^2 = \frac{1}{n-m} \cdot e^T \cdot e$$

$$\sigma_e^2 = 140{,}508$$

9. Побудуємо матрицю коваріацій оцінок параметрів за виразом:

$$Var\left(\hat{B}\right) = \sigma_e^2 \left(X^T X\right)^{-1}$$

тобто отримуємо

$$Var(B) = \begin{pmatrix} 8{,}525 & -0{,}015 & -5.623\cdot10^{-3} & -0{,}01 & 2.174\cdot10^{-3} & -0{,}018 \\ -0{,}015 & 1{,}184\cdot10^{-4} & -7{,}974\cdot10^{-6} & -7{,}527\cdot10^{-5} & 3{,}261\cdot10^{-5} & 3{,}986\cdot10^{-6} \\ -5{,}623\cdot10^{-3} & -7{,}974\cdot10^{-6} & 3{,}17\cdot10^{-5} & 2{,}656\cdot10^{-5} & -6{,}037\cdot10^{-6} & 7{,}654\cdot10^{-8} \\ -0{,}01 & -7{,}527\cdot10^{-5} & 2{,}656\cdot10^{-5} & 1{,}346\cdot10^{-4} & -4{,}06\cdot10^{-5} & 2{,}696\cdot10^{-5} \\ 2{,}174\cdot10^{-3} & 3{,}261\cdot10^{-5} & -6{,}037\cdot10^{-6} & -4{,}06\cdot10^{-5} & 2{,}012\cdot10^{-5} & -2{,}468\cdot10^{-5} \\ -0{,}018 & 3{,}986\cdot10^{-6} & 7{,}654\cdot10^{-8} & 2{,}696\cdot10^{-5} & -2{,}468\cdot10^{-5} & 9{,}103\cdot10^{-5} \end{pmatrix}$$

З матриці коваріаційних оцінок знаходимо дисперсійні оцінки параметрів моделі як діагональні елементи цієї матриці:

$\sigma_{\hat{b}_0}^2 = 8{,}525$; $\sigma_{\hat{b}_1}^2 = 1{,}184 \cdot 10^{-4}$; $\sigma_{\hat{b}_2}^2 = 3{,}17 \cdot 10^{-5}$; $\sigma_{\hat{b}_3}^2 = 1{,}346 \cdot 10^{-4}$;

$\sigma_{\hat{b}_4}^2 = 2{,}012 \cdot 10^{-5}$; $\sigma_{\hat{b}_5}^2 = 9{,}103 \cdot 10^{-5}$.

10. Знаходимо стандартні помилки та їхні відносні відхилення оцінок параметрів.

$S_{\hat{b}_0} = \sqrt{\sigma_{\hat{b}_0}^2} = 2{,}92$; $S_{\hat{b}_1} = \sqrt{\sigma_{\hat{b}_1}^2} = 1{,}088 \cdot 10^{-2}$; $S_{\hat{b}_2} = \sqrt{\sigma_{\hat{b}_2}^2} = 5{,}63 \cdot 10^{-2}$;

$S_{\hat{b}_3} = \sqrt{\sigma_{\hat{b}_3}^2} = 1{,}16 \cdot 10^{-2}$; $S_{\hat{b}_4} = \sqrt{\sigma_{\hat{b}_4}^2} = 4{,}485 \cdot 10^{-2}$; $S_{\hat{b}_5} = \sqrt{\sigma_{\hat{b}_5}^2} = 9{,}541 \cdot 10^{-2}$.

$$\gamma_0 = \frac{S_{\hat{b}_0}}{|\hat{b}_0|} \cdot 100\% = \frac{2,92}{6,753} \cdot 100\% = 43,24\%$$

$$\gamma_1 = \frac{S_{\hat{b}_1}}{|\hat{b}_1|} \cdot 100\% = \frac{1,088 \cdot 10^{-2}}{0,108} \cdot 100\% = 10,07\%$$

$$\gamma_2 = \frac{S_{\hat{b}_2}}{|\hat{b}_2|} \cdot 100\% = \frac{5,63 \cdot 10^{-2}}{0,175} \cdot 100\% = 32,1\%$$

$$\gamma_3 = \frac{S_{\hat{b}_3}}{|\hat{b}_3|} \cdot 100\% = \frac{1,16 \cdot 10^{-2}}{0,11} \cdot 100\% = 10,54\%$$

$$\gamma_4 = \frac{S_{\hat{b}_4}}{|\hat{b}_4|} \cdot 100\% = \frac{4,485 \cdot 10^{-2}}{0,082} \cdot 100\% = 54,7\%$$

$$\gamma_5 = \frac{S_{\hat{b}_5}}{|\hat{b}_5|} \cdot 100\% = \frac{9,541 \cdot 10^{-2}}{0,215} \cdot 100\% = 44,38\%$$

11. Розраховуємо параметри моделі за допомогою функції ЛИНЕЙН. Для цього спочатку виділяємо діапазон клітинок 6 (5 факторів + вільний член) на 5 (стандартна висота діапазону виводу функції.

В якості параметрів обираємо по черзі:

- стовбець Y;
- п'ять стовбців із значеннями незалежних змінних;
- константа=1;
- статистика=1.

Діапазон виводу в даному прикладі матиме наступний вигляд:

0,215681	-0,08274	0,111053	0,174951	-0,10863	-6,51657
0,112986	0,053356	0,137399	0,066609	0,130168	34,72531
0,573976	11,84423	#Н/Д	#Н/Д	#Н/Д	#Н/Д
2,69457	10	#Н/Д	#Н/Д	#Н/Д	#Н/Д
1890,051	1402,859	#Н/Д	#Н/Д	#Н/Д	#Н/Д

Як бачимо, оцінки параметрів, що знаходяться в першому рядку діапазону виводу (в зворотному порядку) майже співпадають з розрахованими вище вручну.

12. Висновки

1. Отримані значення параметрів моделі мають такий економічний зміст:

- при збільшені середньодобових приросту, та затрат на 1 ц. приросту на одиницю, ми досягнемо зменшення рівня збитковості відповідно на 0,108 та 0,082 відсотків
- при збільшені затрат на 1 ц. приросту у людиногодинах, затрат на 1 голову у гривнях та собівартості на одиницю, ми досягнемо збільшення рівня збитковості відповідно на 0,175; 0,11; 0,215 відсотків.

2. Велике значення стандартної помилки при визначені коефіцієнта b_0 свідчить про недоліки у специфікацій моделі, на це вказує і велике значення стандартної помилки при визначені коефіцієнта b_4. Мале абсолютне значення коефіцієнта b_4, та велике значення

стандартної помилки при його визначені свідчать при недоцільності включення цього фактора, а саме затрат на 1 голову, до економічної моделі.

Контрольні запитання

1. Що таке багаточинникова економетрична модель?
2. Що таке ендогенні та екзогенні змінні в економетричній моделі?
3. Загальний вигляд лінійної багаточинникової економетричної моделі в аналітичній та матричній формах?
4. Яких передумов необхідно дотримуватись, щоб параметри моделі можна оцінити було на основі методу 1 МНК?
5. Що можна визначити оператором оцінювання?
6. Як за допомогою системи Mathcad визначити елементи матриці моментів?
7. Як обчислити дисперсії залишків і залежної змінної?
8. Що таке матриця коваріації і що визначають її коефіцієнти?
9. Як обчислити стандартні помилки та відносні відхилення оцінювання невідомих параметрів моделі?
10. У чому , з точки зору економіки , сутність параметрів моделі?
11. Як ви оцінюєте можливості електронних таблиць Excel при виконанні цієї роботи?
12. Як ви визначити дисперсії оцінок параметрів моделі?
13. Як знайти параметри багатофакторної моделі за допомогою функції ЛИНЕЙН?
14. Які функції для матричних розрахунків наявні в Excel?

Практична робота №3

Тема: Дисперсійний аналіз економетричної моделі.

Мета: На основі дисперсійного аналізу економетричної моделі реалізувати алгоритм покрокової регресії та перевірити гіпотезу про суттєвість зв'язку між залежною та незалежними змінними.

Короткі теоретичні відомості

Для економетричної моделі існує відповідність між оцінками її параметрів та коефіцієнтом кореляції, що характеризує щільність зв'язку між змінними. Для простої економетричної моделі її можна записати так:

$$\hat{a} = r\,\frac{\sigma_Y}{\sigma_X},$$

де r –коефіцієнт парної кореляції; σ_Y, σ_X – середньоквадратичні відхилення відповідно залежної і незалежної змінної.

Це співвідношення було покладено в основу алгоритму визначення альтернативної оцінки параметрів моделі за методом 1МНК. Алгоритм має назву покрокової регресії і наступні кроки:

Етап 1. Стандартизація (нормалізація) всіх змінних моделі:

$$Y_i^* = \frac{Y_i - \overline{Y}}{\sigma_Y};\ .\ X_{ij}^* = \frac{X_{ij} - \overline{X}_j}{\sigma_{X_j}}$$

Етап 2. Визначення вектора r_{YX_j}, елементи якого розраховуються таким чином:

$$r_{YX_j} = \frac{1}{n}(Y^{*T}\ X_j^*)^T\ ;\ \ j = \overline{1,m}.$$

Етап 3. Із усіх елементів вектора r_{YXj} вибирається той, якому відповідає $\max\left\{\left|r_{YXj}\right|\right\}$. Це означає, що незалежна змінна X_j найтісніше зв'язана з залежною змінною Y. Будується економетрична модель:

$$\hat{Y}^* = \hat{\beta}_j X_j^*\ .$$

Етап 4. Серед інших елементів вектора r_{YXj} знову вибирається $\max\left\{\left|r_{YXj}\right|\right\}$. Якщо даному коефіцієнту кореляції відповідає X_{j+1}^*, то саме цей чинник вводиться в побудовану раніше економетричну модель; в результаті дістанемо:

$$Y^* = \hat{\beta}_j X_j^* + \hat{\beta}_{j+1} X_{j+1}^* \quad \text{і т.д.}$$

Процес продовжується до тих пір, поки всі незалежні змінні поступово не будуть включені в модель. Якщо є обмеження, яке вказує на недоцільність розширення економетричної моделі за рахунок змінних, що залишилися, то процес розрахунку закінчується раніше. Таким обмеженням може бути співвідношення між коефіцієнтом кореляції чи детермінації, виправленими й невиправленими на число ступеней свободи.

Таким чином етап 4 включатиме декілька кроків залежно від того, скільки чинників буде вводитися в модель.

Система нормальних рівнянь у даному алгоритмі:

$$\begin{cases} r_{YX_1}=\hat{\beta}_1+\hat{\beta}_2 r_{X_1X_2}+\hat{\beta}_3 r_{X_1X_3}+...+\hat{\beta}_m r_{X_1X_m} \\ r_{YX_2}=\hat{\beta}_1 r_{X_1X_2}+\hat{\beta}_2+\hat{\beta}_3 r_{X_2X_3}+...+\hat{\beta}_m r_{X_2X_m} \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ r_{YX_m}=\hat{\beta}_1 r_{X_1X_m}+\hat{\beta}_2 r_{X_2X_m}+\hat{\beta}_3 r_{X_3X_m}+...+\hat{\beta}_m \end{cases}$$

Позначимо елементи r_{YXj} через вектор r_{YX}, а елементи матриці r_{XkXj} – через матрицю r, тоді система рівнянь у матричному вигляді матиме такий вигляд:

$$r\vec{\beta} = r_{YX}.$$

Звідси $\vec{\beta} = r^{-1} r_{YX}$, тобто отримаємо альтернативний оператор оцінювання параметрів моделі за методом 1МНК.

Матриця r є матрицею парних коефіцієнтів між чинниками моделі. На головній діагоналі цієї матриці завжди стоятимуть одиниці, так як ці елементи характеризують зв'язок кожного чинника з тим же самим чинником. Крім того кореляційна матриця r симетрична відносно головної діагоналі.

Знайти матрицю можна використовуючи функцію КОРРЕЛ або за виразом

$$r = \frac{1}{n}(X_j^{*T}\ X_j^{*})$$

Оскільки оцінки параметрів моделі $\vec{\beta}$ відносяться до стандартизованих змінних, то щоб перейти до оцінок параметрів моделі, в якій змінні мають свій початковий вимір, необхідно:

$$\hat{a}_j=\hat{\beta}_j\frac{\sigma_Y}{\sigma_{X_j}},\ j=\overline{1,m};\quad \hat{a}_0=\bar{y}-\sum_j \hat{a}_j\bar{x}_j.$$

Якщо оцінка параметрів моделі отримана на основі покрокової регресії, то для визначення коефіцієнта детермінації можна використати такі співвідношення:

$$R^2=1-\frac{|r|}{R_{11}};\ R^2=\hat{\beta}_1 r_{YX_1}+\hat{\beta}_2 r_{YX_2}+\hat{\beta}_3 r_{YX_3}+\ ...\ +\hat{\beta}_m r_{YX_m}$$

де $|r|$ – визначник матриці r; R_{11} – алгебраїчне доповнення першого елемента матриці r.

Гіпотеза про наявність чи відсутність зв'язку між залежною і незалежною змінними може бути перевірена на основі F – критерію: $F = \dfrac{\sigma_Y^2-\sigma_U^2}{\sigma_U^2} = \dfrac{\sigma_P^2}{\sigma_U^2}$.

Фактичне значення F – критерію порівнюється з табличним при ступенях свободи $n-m$ і $m-1$ і вибраному рівні довіри. Якщо $F_{факт} > F_{табл}$, то гіпотеза про суттєвість зв'язку між залежною і незалежними змінними економетричної моделі підтверджується, в протилежному випадку – відкидається.

Альтернативна формула розрахунку F – критерію через коефіцієнт детермінації:

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

Значущість оцінок параметрів моделі можна визначити на основі t – критерію Стьюдента: $t_j = \dfrac{|a_j|}{\sqrt{\sigma_U^2 c_{jj}}} = \dfrac{|a_j|}{S_{a_j}}$.

Значення критерію t_j порівнюється з табличним при вибраному рівні значущості і

$n-m$ ступенями свободи. Якщо $t_{\text{факт}} > t_{\text{табл}}$, то відповідний параметр економетричної моделі є достовірним.

Завдання

Використавши базу даних практичної роботи №2 для свого варіанту провести дослідження за порядком, наведеним нижче. При виконанні роботи використати електронні таблиці Excel.

Порядок виконання роботи

1. Згідно свого варіанту вибрати базу даних і подати її за допомогою електронних таблиць Excel. Специфікувати модель в лінійній формі.
2. Знайти середні значення (функція СРЗНАЧ) та стандартні відхилення (функція СТАНДОТКЛОНП) для всіх змінних моделі, які розмістити в двох останніх рядках таблиці бази даних.
3. Провести нормалізацію змінних за формулами з етапу 1 теоретичних відомостей або використовуючи функцію НОРМАЛИЗАЦИЯ. Отримані значення розмістити у вигляді таблиці.
4. Знайти коефіцієнти парної кореляції між нормалізованими змінними *X* та *Y*. Значення коефіцієнтів розташувати у вигляді вектора. При цьому використовувати формулу з етапу 2 теоретичних відомостей або команду КОРРЕЛ.
5. Охарактеризувати тісноту зв'язку між залежною змінною і кожною з незалежних. Вибрати з елементів матриці r_{yx} найбільший за модулем та побудувати однофакторну економетричну модель (див. етап 3 теоретичних відомостей).
6. З тих елементів матриці r_{yx}, що залишилися вибираємо наступний за рівнем значення елемент (по модулю) та будуємо двофакторну економетричну модель: отримуємо рівняння регресії з двома незалежними змінними. Процедуру слід повторювати до отримання повного рівняння регресії, аналізуючи на кожному кроці ступінь визначеності моделі з допомогою коефіцієнта детермінації.
7. Провести аналіз отриманих результатів за допомогою *F*-критерію Фішера та *t*-критерію Стьюдента.
8. Виконати перехід до економетричної моделі, в якій змінні виражені в абсолютних значеннях.
9. Зробити висновки.

Приклад виконання

На основі методу покрокової регресії побудувати економетричну модель, яка характеризує збитковість свинини по деяким підприємствам. Перевірити достовірність моделі та її параметрів, використовуючи елементи дисперсійного аналізу. Дати змістовне тлумачення параметрів моделі. Вихідні дані наведені в табл. 3.1. Вони аналогічні до бази вихідних даних з практичної роботи №2.

Таблиця 3.1

№ підпр.	Середньо-добовий приріст, X_1 грн.	Затрати на 1 ц. приросту X_2 люд.-год.	Затрати на 1 голову, X_3 грн.	Затрати на 1 ц. приросту, X_4 грн.	Собівартість 1 ц., X_5 грн.	Рівень збитковості, *Y*, %
1	140	96,5	210	393,31	239,57	36,4
2	180	53,8	175	233,21	170,38	8,7
3	122	99,9	199	403,39	290,18	50,6
4	110	130,9	198	429,05	223,86	29,2

5	159	108,3	273	456,17	217,07	45,0
6	79	179,7	172	531,96	327,12	64,2
7	194	72,5	196	316,63	200,74	25,6
8	86	132,1	204	591,22	282,19	57,3
9	111	121,6	200	499,58	302,78	47,2
10	112	126,0	205	594,5	289,73	31,2
11	159	201,8	153	311,38	211,81	68,1
12	54	109,0	317	1019,06	367,02	33,9
13	136	243,6	204	399,6	233,87	44,6
14	85	80,9	190	466,38	280,67	38,1
15	228	103,3	246	349,97	265,82	40,2
16	129	53,9	256	404	230,92	40,6

Розв'язання

1. Специфікуємо модель в лінійній формі:

$$Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + a_4 X_4 + a_5 X_5 + u \ ;$$

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X_1 + \hat{a}_2 X_2 + \hat{a}_3 X_3 + \hat{a}_4 X_4 + \hat{a}_5 X_5 .$$

Оскільки оцінка параметрів моделі за методом 1МНК виконуватиметься на основі покрокової регресії, то має бути побудована економетрична модель виду:

$$Y^* = \hat{\beta}_1 X_1^* + \hat{\beta}_2 X_2^* + \hat{\beta}_3 X_3^* + \hat{\beta}_4 X_4^* + \hat{\beta}_5 X_5^* ,$$

де $Y^* = \dfrac{Y_i - \overline{Y}}{\sigma_Y}$, $X_{ji}^* = \dfrac{X_{ji} - \overline{X}_j}{\sigma_{X_j}}$ – нормалізовані змінні.

2. Середні арифметичні та стандартні (середньоквадратичні) відхилення розміщуємо в двох рядках після вихідних даних, користуючись функціями СРЗНАЧ та СТАНДОТКЛОНП.

Таблиця 3.2

№ підприємства	X_1	X_2	X_3	X_4	X_5	Y
1	140	96,5	210	393,31	239,57	36,4
2	180	53,8	175	233,21	170,38	8,7
3	122	99,9	199	403,39	290,18	50,6
4	110	130,9	198	429,05	223,86	29,2
5	159	108,3	273	456,17	217,07	45,0
6	79	179,7	172	531,96	327,12	64,2
7	194	72,5	196	316,63	200,74	25,6
8	86	132,1	204	591,22	282,19	57,3
9	111	121,6	200	499,58	302,78	47,2
10	112	126,0	205	594,5	289,73	31,2
11	159	201,8	153	311,38	211,81	68,1
12	54	109,0	317	1019,06	367,02	33,9
13	136	243,6	204	399,6	233,87	44,6
14	85	80,9	190	466,38	280,67	38,1
15	228	103,3	246	349,97	265,82	40,2
16	129	53,9	256	404	230,92	40,6
СРЗНАЧ	*130,25*	*119,613*	*212,375*	*462,463*	*258,358*	*41,306*
СТАНДОТКЛОНП	*49,658*	*14,346*	*44,3741*	*49,9146*	*40,0903*	*172,515*

3. Нормалізуємо змінні моделі. Розрахунки здійснюємо або з використанням функції

НОРМАЛИЗАЦИЯ, або за формулами нормалізації. Результати представляємо у вигляді таблиці.

Таблиця 3.3

Y*	X1*	X2*	X3*	X4*	X5*
-0,34200	0,21972	-0,46304	-0,05924	-0,40085	-0,37835
-2,27285	1,12115	-1,31850	-0,93227	-1,32889	-1,77169
0,64783	-0,18592	-0,39492	-0,33362	-0,34242	0,64082
-0,84388	-0,45635	0,22614	-0,35857	-0,19368	-0,69472
0,25748	0,64790	-0,22664	1,51221	-0,03648	-0,83145
1,59583	-1,15495	1,20381	-1,00710	0,40285	1,38472
-1,09482	1,43665	-0,94386	-0,40845	-0,84534	-1,16030
1,11486	-0,99720	0,25018	-0,20890	0,74635	0,47992
0,41083	-0,43381	0,03982	-0,30868	0,21515	0,89456
-0,70447	-0,41128	0,12797	-0,18396	0,76537	0,63176
1,86768	0,64790	1,64656	-1,48103	-0,87577	-0,93738
-0,51626	-1,71834	-0,21261	2,60973	3,22660	2,18822
0,22959	0,12958	2,48399	-0,20890	-0,36439	-0,49314
-0,22349	-1,01974	-0,77558	-0,55811	0,02270	0,44931
-0,07711	2,20286	-0,32681	0,83873	-0,65208	0,15027
-0,04923	-0,02817	-1,31650	1,08817	-0,33889	-0,55254

4. Формуємо вектор парних коефіцієнтів кореляції між залежною і незалежними змінними $\{ r_{YX_j} \}$. Знайти коефіцієнти кореляції можна користуючись функцією КОРРЕЛ або за виразом:

$$r_{YX_j} = \frac{1}{n}(Y^{*T} X_j^*)^T$$

В нашому випадку:

$$r_{yx} = \begin{pmatrix} 0.27636 \\ 0.61427 \\ -0.18312 \\ 0.10726 \\ 0.36267 \end{pmatrix}$$

5. Враховуючи, що

$$\max\{r_{YX_j}\} = r_{YX_2} = 0{,}61427,$$

то на першому етапі треба побудувати економетричну модель виду:

$$Y^* = \hat{\beta}_2 X_2^* \cdot$$

Вираз для знаходження параметру $\hat{\beta}_2$ має вигляд:

$$r_{YX_2} = \hat{\beta}_2 r_{X_2X_2};$$

$$\hat{\beta}_2 = \frac{r_{YX_2}}{r_{X_2X_2}} = \frac{0{,}61427}{1};$$

$$\hat{\beta}_2 = 0{,}61427$$

Запишемо модель:

$$Y^* = 0{,}61427\ X_2^*$$

Визначеність моделі на цьому кроці характеризує коефіцієнт детермінації:

$$R^2 = \hat{\beta}_2 r_{YX_2} = 0{,}61427 \cdot 0{,}61427 = 0{,}377$$

Бачимо, що модель визначена недостатньо.

6. На другому кроці методу покрокової регресії включимо в економетричну модель чинник X_5^*, так як він найбільш тісно пов'язаний із залежною змінною Y серед тих чинників, які ще не включені в модель. В результаті остання набуде такого вигляду:

$$Y^* = \hat{\beta}_2 X_2^* + \hat{\beta}_5 X_5^* \cdot$$

Знайти невідомі параметри моделі $\hat{\beta}$ можна або розв'язавши систему рівнянь

$$\begin{cases} r_{YX_2} = \hat{\beta}_2 + \hat{\beta}_5 r_{X_2X_5} \\ r_{YX_5} = \hat{\beta}_2 r_{X_2X_5} + \hat{\beta}_5 \end{cases}$$

або у матричному вигляді за виразом

$$\hat{\beta} = r^{-1} r_{YX}\,.$$

Зауважимо, що на кожному наступному кроці методу системи рівнянь будуть мати все більшу розмірність (на останньому – 5 рівнянь з 5 невідомими). Більш доцільно буде застосування виразу $\hat{\beta} = r^{-1} r_{YX}$, так як електронні таблиці Excel дозволяють без особливих складностей знайти обернені матриці та добуток двох матриць. В той же час при виконанні розрахунків без використання ПК більш легшим варіантом є складання та розв'язання систем рівнянь (загальний вигляд системи див. етап 4 з теоретичних відомостей).

Щоб знайти вектор оцінок параметрів $\hat{\beta}$ у матричному вигляді, спочатку знайдемо кореляційну матрицю r.

У матричному вигляді $r = \frac{1}{n}(X_j^{*T}\ X_j^*)$ або скориставшись функцією КОРРЕЛ.

$$r = \begin{vmatrix} 1 & -0{,}18501 & -0{,}12147 & -0{,}74317 & -0{,}70912 \\ -0{,}18501 & 1 & -0{,}29603 & 0{,}076374 & 0{,}149016 \\ -0{,}12147 & -0{,}29603 & 1 & 0{,}614561 & 0{,}344676 \\ -0{,}74317 & 0{,}076374 & 0{,}614561 & 1 & 0{,}824059 \\ -0{,}70912 & 0{,}149016 & 0{,}344676 & 0{,}824059 & 1 \end{vmatrix}$$

Залишаємо у матриці r ті рядки і стовпці, які відповідають чинникам X_2^* та X_5^*. У векторі r_{YX} теж залишаємо лише коефіцієнти, які характеризують зв'язок з цими ж чинниками.

Маємо:

$$r_{(2,5)} = \begin{vmatrix} 1 & 0{,}149016 \\ 0{,}149016 & 1 \end{vmatrix}$$

$$r_{(2,5)}^{-1} = \begin{vmatrix} 1{,}02271 & -0{,}1524 \\ -0{,}1524 & 1{,}02271 \end{vmatrix}$$

$$r_{YX_{2,5}} = \begin{vmatrix} 0{,}614271 \\ 0{,}362666 \end{vmatrix}$$

$$\hat{\beta} = r^{-1} r_{YX} = \begin{vmatrix} 0{,}572951 \\ 0{,}277287 \end{vmatrix}$$

Економетрична модель має вигляд:

$$Y^* = 0,57295 X_2^* + 0,27729 X_5^* \cdot$$

Коефіцієнт детермінації:

$$R^2 = \hat{\beta}_2 r_{YX_2} + \hat{\beta}_5 r_{YX_5} = 0,57295 \cdot 0.61427 + 0.27729 \cdot 0.36267 = 0,4525$$

Модель залишається недостатньо визначеною.

Вводимо в модель чинник X_1^*

$$r_{YX_{1,2,5}} = \begin{vmatrix} -0,27636 \\ 0,614271 \\ 0,362666 \end{vmatrix} \qquad r_{(1,2,5)} = \begin{vmatrix} 1 & -0,18501 & -0,70912 \\ -0,18501 & 1 & 0,149016 \\ -0,70912 & 0,149016 & 1 \end{vmatrix}$$

$$r_{(1,2,5)}^{-1} = \begin{vmatrix} 2,037862 & 0,165364 & 1,420447 \\ 0,165364 & 1,036129 & -0,03714 \\ 1,420447 & -0,03714 & 2,012801 \end{vmatrix}$$

$$\hat{\beta} = \begin{vmatrix} 0,053548 \\ 0,577295 \\ 0,314612 \end{vmatrix}$$

$$Y^* = 0,05355 X_1^* + 0,5773 X_2^* + 0,31461 X_5^*$$

$R^2 =$ 0,453917

Вводимо в модель чинник X_3^*

$$r_{YX_{1,2,3,5}} = \begin{vmatrix} -0,27636 \\ 0,614271 \\ -0,18312 \\ 0,362666 \end{vmatrix} \qquad r_{(1,2,3,5)} = \begin{matrix} 1 & -0,18501 & -0,12147 & -0,70912 \\ -0,18501 & 1 & -0,29603 & 0,149016 \\ -0,12147 & -0,29603 & 1 & 0,344676 \\ -0,70912 & 0,149016 & 0,344676 & 1 \end{matrix}$$

$$r_{(1,2,3,5)}^{-1} = \begin{vmatrix} 2,088286 & 0,07668 & -0,26113 & 1,559424 \\ 0,07668 & 1,192104 & 0,459267 & -0,28157 \\ -0,26113 & 0,459267 & 1,35231 & -0,71972 \\ 1,559424 & -0,28157 & -0,71972 & 2,395847 \end{vmatrix}$$

$$\hat{\beta} = \begin{vmatrix} 0,083358 \\ 0,524866 \\ -0,15438 \\ 0,396775 \end{vmatrix}$$

$$Y^* = 0,08336 X_1^* + 0,52487 X_2^* - 0,15438 X_3^* + 0,39678 X_5^*$$

$R^2 =$ 0,47154

Вводимо в модель останній чинник X_4^*

$$r_{YX} = \begin{vmatrix} -0,27636 \\ 0,614271 \\ -0,18312 \\ 0,107259 \\ 0,362666 \end{vmatrix} \qquad r = \begin{vmatrix} 1 & -0,18501 & -0,12147 & -0,74317 & -0,70912 \\ -0,18501 & 1 & -0,29603 & 0,076374 & 0,149016 \\ -0,12147 & -0,29603 & 1 & 0,614561 & 0,344676 \\ -0,74317 & 0,076374 & 0,614561 & 1 & 0,824059 \\ -0,70912 & 0,149016 & 0,344676 & 0,824059 & 1 \end{vmatrix}$$

$$r^{-1} = \begin{vmatrix} 3,804998 & -0,26657 & -2,16359 & 4,072691 & 0,127521 \\ -0,26657 & 1,260734 & 0,839653 & -0,81431 & 0,004736 \\ -2,16359 & 0,839653 & 3,460607 & -4,51335 & 0,867111 \\ 4,072691 & -0,81431 & -4,51335 & 9,661965 & -3,39702 \\ 0,127521 & 0,004736 & 0,867111 & -3,39702 & 3,590193 \end{vmatrix}$$

$$Y^* = -0,336 X_1^* + 0,6087 X_2^* + 0,31035 X_3^* - 0,99488 X_4^* + 0,74656 X_5^*$$

$$\hat{\beta}= \begin{vmatrix} -0{,}336 \\ 0{,}608714 \\ 0{,}310353 \\ -0{,}99488 \\ 0{,}746561 \end{vmatrix}$$

$$R^2 = 0{,}573982$$

Отримане значення коефіцієнта детермінації свідчить про те, що варіація рівня збитковості на 57,4% визначається варіацією введених в модель чинників (факторів).

Визначаємо коефіцієнт множинної кореляції:

$$R = \sqrt{R^2} = \sqrt{0{,}574} \approx 0{,}757.$$

Коефіцієнт кореляції характеризує досить тісний зв'язок факторів із рівнем збитковості.

7. Оцінимо достовірність моделі та її параметрів на основі критеріїв Фішера та Стьюдента.

F-критерій Фішера:

$$F = \frac{R^2/(m-1)}{(1-R^2)/(n-m)} = \frac{0{,}574/(6-1)}{(1-0{,}574)/(16-6)} = 2{,}69$$

Зауваження: за m в формулі критерію Фішера необхідно приймати загальну кількість змінних у моделі, а не кількість чинників.

Знайдене фактичне значення F-критерію порівнюємо з табличним значенням F^* при ступенях вільності (свободи) n-m і m-1 і рівні довіри α.

При ступенях свободи $df_1 = m-1 = 6-1 = 5$ і $df_2 = n-m = 16-6 = 10$; рівні довіри $\alpha = 0{,}05$ (імовірність правильних висновків становить $p = 1 - 0{,}05 = 0{,}95$ або 95%) $F_{табл} = 3{,}33$. Оскільки $F_{факт} < F_{табл}$, то гіпотеза про наявність і суттєвість лінійного зв'язку, який вимірюється на основі економетричної моделі, відхиляється. Це означає, що економетрична модель є недостовірною, тому перевірка значущості оцінок параметрів моделі за критерієм Стьюдента є недоцільною.

8. Виконаємо перехід до економетричної моделі, в якій змінні виражені в абсолютних значеннях:

$$\hat{a}_1 = \hat{\beta}_1 \frac{\sigma_Y}{\sigma_{X1}} = -0{,}336 \cdot \frac{14{,}346}{44{,}374} = -0{,}1086\ ;$$

$$\hat{a}_2 = 0{,}175\ ;\ \hat{a}_3 = 0{,}11106\ ;\ \hat{a}_5 = -0{,}08273\ ;\ \hat{a}_5 = 0{,}21568\ ;$$

$$\hat{a}_0 = \overline{Y} - \hat{a}_1 \overline{X}_1 - \hat{a}_2 \overline{X}_2 - \hat{a}_3 \overline{X}_3 - \hat{a}_4 \overline{X}_4 - \hat{a}_5 \overline{X}_5 = -6{,}51907$$

$$\hat{Y} = -6{,}51907 - 0{,}1086 X_1 + 0{,}175 X_2 + 0{,}11106 X_3 - 0{,}08273 X_4 + 0{,}21568 X_5 .$$

Аналогічну економетричну модель було отримано в попередній роботі. Дещо відрізняється вільний член (-6,519 проти -6,753), проте ця відмінність є невеликою і виникла через похибки в обчисленнях внаслідок заокруглень.

9. Наведемо розраховані економетричні моделі і дамо змістовне тлумачення параметрів цих моделей.

1) $Y^* = -0{,}336 X_1^* + 0{,}6087 X_2^* + 0{,}31035 X_3^* - 0{,}99488 X_4^* + 0{,}74656 X_5^*$

2) $\hat{Y} = -6{,}51907 - 0{,}1086 X_1 + 0{,}175 X_2 + 0{,}11106 X_3 - 0{,}08273 X_4 + 0{,}21568 X_5$

Перш за все звернемо увагу на відсутність вільного члена в першій економетричній моделі. Це пов'язано з тим, що всі змінні нормалізовані і мають одну й ту саму одиницю виміру. Параметри першого рівняння характеризують граничну зміну залежної змінної, якщо незалежна збільшиться на величину свого середньоквадратичного відхилення σ_{X_j}. Так, якщо X_1^* збільшиться на σ_{X_1}, то Y^* зменшиться на $0{,}336\sigma_Y$ при незмінній величині всіх інших факторів; якщо X_2^* збільшиться на σ_{X_2}, то Y^* збільшиться на $0{,}6087\sigma_Y$ і т.д. Враховуючи, що всі змінні мають одну й ту саму величину виміру, параметри першої економетричної моделі характеризують порівняльну силу впливу незалежних змінних на залежну. В нашому випадку найсильніше на рівень збитковості впливають затрати на 1ц приросту і собівартість 1ц свинини.

В другій економетричній моделі, коли кожний економічний показник має свою початкову одиницю виміру, є вільний член. Його рівень залежить від початку відрахунку змінних, а також від одиниць виміру кожної змінної моделі. Параметри цієї моделі вже були охарактеризовані в попередній роботі.

Контрольні запитання

1. Сутність алгоритму покрокової регресії.
2. Кореляційна матриця: її побудова та властивості.
3. Характеристика парних коефіцієнтів кореляції.
4. Відмінність парних коефіцієнтів кореляції від множинного.
5. Процедура нормалізації змінних.
6. Економетричний зміст нормалізованих змінних.
7. Матрична форма розрахунку оцінок параметрів моделі в нормалізованих змінних.
8. Процедура переходу до оцінок параметрів моделі, в яких змінні мають свій початковий вимір.
9. Критерій Фішера. Процедура його застосування.
10. Критерій Стьюдента та процедура його застосування.
11. Економетричний зміст оцінок параметрів моделі в нормалізованих змінних.

Практична робота №4

Тема: Дослідження наявності мультиколінеарності бази даних.

Мета: На основі алгоритму Феррара – Глобера дослідити базу даних чинників економетричної моделі на мультиколінеарність.

Короткі теоретичні відомості

В економічній практиці при побудові економетричної моделі часто можна зустрітися з проблемою взаємозв'язку між чинниками (пояснюючими змінними) моделі. Якщо цей взаємозв'язок суттєвий, то оцінка параметрів моделі може мати велику похибку. Таке явище називають мультиколінеарність.

Якщо параметри моделі оцінені методом 1МНК, то мультиколінеарність приводить до зміщення отриманих оцінок, а отже до неможливості з достатньою точністю зробити конкретні висновки про результати взаємозв'язку між екзогенними та ендогенними змінними.

Виявити мультиколінеарність можна за такими ознаками:

1. Якщо серед парних коефіцієнтів кореляції чинників є такі, рівень яких наближається або дорівнює множинному коефіцієнту кореляції.

2. За визначником (детермінантом) $|r|$ матриці коефіцієнтів парної кореляції, числові значення якого знаходяться на множині: $|r| \in [0;1]$. При чому, якщо $|r| = 0$, то існує повна мультиколінеарність, а при $|r| = 1$ – мультиколінеарність відсутня. Оскільки на числове значення $|r|$ впливає дисперсія чинників, то цей показник можна вважати точковою мірою тісноти мультиколінеарності: чим ближче $|r|$ до нуля, тим певніше можна стверджувати, що між чинниками існує мультиколінеарність.

3. Одержане мале значення параметру b_k при високому рівні частинного коефіцієнта кореляції R_k^2 і коли при цьому значення F-критерію суттєво відрізняється від нуля, це свідчить про наявність мультиколінеарності.

4. Якщо коефіцієнт частинної детермінації R_k^2, що розрахований для регресивних залежностей між однією незалежною змінною (чинником) та іншими, має значення, яке близьке до 1, то це може свідчити про наявність мультиколінеарності.

5. Включення нового чинника при наявності мультиколініеарності в вихідний базі економетричної моделі суттєво змінює оцінку параметрів моделі при незначному підвищенні (або зниженні) коефіцієнтів кореляції чи детермінації. При цьому цей чинник знаходиться в лінійній залежності від інших чинників, введених в покрокову регресію раніше.

Разом з тим слід зазначити, що жодна з цих ознак не визначає чіткої межі між суттєвою і несуттєвою мультиколінеарністю.

Найбільш повне дослідження явище мультиколінеарності можна здійснити на основі алгоритму Феррера-Глобера. Алгоритм містить в собі наступні критерії:

- критерій Спірмера (χ^2-критерій), на основі якого перевіряється мультиколінеарність всієї сукупності чинників моделі;
- критерій Фішера (F-критерій) для перевірки наявності мультиколінеарності кожного чинника з сукупністю інших чинників моделі;

- критерій Стьюдента (t-критерій), для визначення наявності мультиколінеарності кожної пари чинників.

Конкретні висновки щодо мультиколінеарності можна зробити тільки при порівнянні фактичних (емпіричних) значень цих критеріїв з їхніми табличними (критичними) значеннями.

Розглянемо сутність алгоритму Феррара-Глобера та процедуру визначення за ним мультиколінеарності.

Передусім розглянемо сукупність вихідних значень чинників як елементи векторів $x_1, x_2 ... x_k$.

Елементи стандартних (нормативних) векторів розраховуються за виразом:

$$x_{ik}^* = \frac{x_{ik} - \overline{x_k}}{\sqrt{n\sigma_{x_k}^2}}, \tag{4.1}$$

де n – кількість спостережень, елементів векторів ($i = \overline{i,n}$);

m – число незалежних змінних ($k = \overline{i,m}$);

$\bar{x}_k$ – середня арифметична k-ї незалежної змінної (чинника)

$$\overline{x_1} = \frac{\sum_{i=1}^{n} x_{i1}}{n}; \ \overline{x_2} = \frac{\sum_{i=1}^{n} x_{i2}}{n}; \ ... \ \overline{x_k} = \frac{\sum_{i=1}^{n} x_{ik}}{n}; \tag{4.2}$$

$\sigma_{x_k}^2$ – дисперсія k-ї незалежної змінної (чинника):

$$\sigma_{x_1}^2 = \frac{\sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2}{n}; \ \sigma_{x_2}^2 = \frac{\sum_{i=1}^{n}(x_{i2} - \bar{x}_2)^2}{n}; \ ... \ \sigma_{x_3}^2 = \frac{\sum_{i=1}^{n}(x_{ik} - \bar{x}_k)^2}{n}. \tag{4.3}$$

Знаючи дисперсію можна обчислити знаменник для стандартизації кожного чинника: для $x_1 : \sqrt{\mathrm{n}\sigma_{\mathrm{x}_1}}$; для $x_2 : \sqrt{\mathrm{n}\sigma_{\mathrm{x}_2}}$; ... для $x_k : \sqrt{\mathrm{n}\sigma_{\mathrm{x}_k}}$.

Кореляційну матрицю моделі (матрицю моментів стандартизованої системи нормативних рівнянь) можна отримати з добутку матриць:

$$r = \mathrm{X}^{*\mathrm{T}}\mathrm{X}^*, \tag{4.4}$$

де X^* – матриця стандартизованих значень чинників:

$$X^* = \begin{pmatrix} x_{11}^* & x_{12}^* & ... & x_{1n}^* \\ x_{21}^* & x_{22}^* & ... & x_{2n}^* \\ ... & ... & ... & ... \\ x_{n1}^* & x_{n2}^* & ... & x_{nn}^* \end{pmatrix}; \tag{4.5}$$

$\mathrm{X}^{*\mathrm{T}}$ – транспонована матриця X^*:

$$X^{*T} = \begin{pmatrix} x_{11}^* & x_{21}^* & ... & x_{n1}^* \\ x_{12}^* & x_{22}^* & ... & x_{n2}^* \\ ... & ... & ... & ... \\ x_{1n}^* & x_{2n}^* & ... & x_{nn}^* \end{pmatrix}. \tag{4.6}$$

Кожен елемент кореляційної матриці r характеризує тісноту зв'язку однієї незалежної змінної (чинника) з іншою. Діагональні елементи характеризують тісноту кожної незалежної змінної з цією самою змінною і дорівнюють одиниці. Матриця r має загальний вигляд:

$$r = \begin{pmatrix} r_{X_1X_1} & r_{X_1X_2} & r_{X_1X_3} & \cdots & r_{X_1X_m} \\ r_{X_2X_1} & r_{X_2X_2} & r_{X_2X_3} & \cdots & r_{X_2X_m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{X_mX_1} & r_{X_mX_2} & r_{X_mX_3} & \cdots & r_{X_mX_m} \end{pmatrix}, \tag{4.7}$$

Визначник кореляційної матриці дорівнює:

$$D = |r| = \begin{bmatrix} 1 & r_{X_1X_2} & r_{X_1X_3} & \cdots & r_{X_1X_m} \\ r_{X_2X_1} & 1 & r_{X_2X_3} & \cdots & r_{X_2X_m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{X_mX_1} & r_{X_mX_2} & r_{X_mX_3} & \cdots & 1 \end{bmatrix}. \tag{4.8.}$$

Фактичне значення критерію Спірмена (χ^2-критерію) можна обчислити врахувавши значення визначника кореляційної матриці (4.8):

$$\chi^2 = -\left[n - 1 - \frac{1}{6}(2m+5)\right] \ln D,$$

де D – визначник кореляційної матриці (4.7).

Значення критерію χ^2 обчислене за виразом (4.9), порівнюється з табличним значенням $\chi^2_{таб}$, при $\frac{1}{2}m(m-1)$ ступенях вільності і рівні значущості α (0,01 або 0,05). Якщо $\chi^2 < \chi^2_{таб}$, то вихідна база значень чинників економетричної моделі не має мультиколініарності.

На основі кореляційної матриці (4.7) та визначника можна побудувати матрицю, обернену до неї:

$$C = r^{-1} = \left(X^{*T} \cdot X^{*}\right)^{-1}, \tag{4.10}$$

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ c_{m1} & c_{m2} & \cdots & c_{mm} \end{pmatrix}. \tag{4.11}$$

Використовуючи діагональні елементи матриці С, можна розрахувати значення F-критеріїв:

$$F_1 = (c_{11} - 1)\frac{n-m}{m-1}$$

$$F_2 = (c_{22} - 1)\frac{n-m}{m-1}$$

....

$$F_m = (c_{mm} - 1)\frac{n-m}{m-1}. \tag{4.12}$$

При рівні значущості α і ступенях вільності $\gamma = n - m$ і $\gamma_2 = m - 1$ знаходять табличне (критичне) значення критерію Фішера $F_{таб}$

Якщо виконується умова:

$$F_{1\ \text{факт.}} < F_{\text{таб}}$$
$$F_{2\ \text{факт.}} < F_{\text{таб}}$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$F_{m\ \text{факт.}} < F_{\text{таб}}, \tag{4.13}$$

то жоден чинник не мультиколінеарний з іншими чинниками. В протилежному випадку колінеарність наявна.

Для визначення наявності попарної мультиколінеарності розраховують часткові коефіцієнти кореляції, використавши елементи матриці С:

$$r_{12} = \frac{-c_{12}}{\sqrt{c_{11}c_{22}}}; \quad r_{13} = \frac{-c_{13}}{\sqrt{c_{11}c_{33}}};\ldots\ r_{kj} = \frac{-c_{kj}}{\sqrt{c_{kk}c_{jj}}}, \tag{4.14}$$

де c_{kj} – елемент матриці С, що знаходиться на перетині k-того рядку і j-того стовпця; $k = \overline{1,m}$; $j = \overline{1,m}$; одночасно $k \neq j$, c_{kk}, c_{jj} – діагональні елементи матриці С.

Часткові коефіцієнти кореляції визначають тісноту зв'язку між двома чинниками за умови, що інші не впливають на цей зв'язок.

Якщо порівняти часткові коефіцієнти кореляції з парними, то можна помітити, що вони значно менші парних. Це підтверджує той факт, що на основі парних коефіцієнтів не можна робити остаточний висновок про наявність чи відсутність мультиколінеарності. Тому необхідно на основі часткових коефіцієнтів кореляції матриці X, визначити значення критеріїв Стьюдента (t-критеріїв):

$$t_{12} = \frac{r_{12}\sqrt{n-m}}{\sqrt{1-r_{12}^2}};\ t_{13} = \frac{r_{13}\sqrt{n-m}}{\sqrt{1-r_{13}^2}};\ t_{23} = \frac{r_{23}\sqrt{n-m}}{\sqrt{1-r_{23}^2}};\ \text{або}$$

$$t_{kj} = \frac{r_{kj}\sqrt{n-m}}{\sqrt{1-r_{kj}^2}} \tag{4.15}$$

де $k = \overline{1,m}$; $j = \overline{1,m}$

Аналогічно до розглянутих критеріїв фактичні значення критеріїв t_{kj} порівнюється з критичними (табличними) при n-m ступенях вільності і рівні значущості α. Якщо $t_{kj\text{факт.}} > t_{\text{таб.}}$, то між незалежними змінними (чинниками) x_k і x_j існує мультиколінеарність, тобто ця пара змінних тісно пов'язана певною залежністю.

Завдання

Використовуючи базу даних лабораторної роботи №2 та результати отримані в лабораторній роботі №3,дослідити базу даних на наявність мультиколінарності за допомогою алгоритму Феррара-Глобера. При цьому рекомендується притримуватися наведеного нижче порядку виконання роботи.

Порядок виконання роботи

1. Згідно свого варіанту вибрати базу вихідних даних і ввести її в ПЕОМ.

2. Обчислити для кожного чинника за формулами:

(4.2), (4.3) середню арифметичну дисперсію, а також знаменник виразу (4.1) При цьому доцільно використовувати електронні таблиці Excel та програму STATISTICA.

Результати обчислень подати у вигляді таблиці.

3. Враховуючи обчисленні значення в п. 2, за виразом (4.1) знайти елементи матриці X^* стандартизованих значень чинників.

4. За добутком матриць X^{*T} і X^* знайти всі елементи кореляційної матриці r (вираз (4.4)), і проаналізувати їх.

5. Визначити детермінант D кореляційної матриці r.

6. За виразом (4.9) знайти фактичне значення критерію Спірмена (χ^2-критерію) порівняти його з критичним (табличним) значенням для ступеню вільності $\frac{1}{2}$m(m-1) і рівня значущості α. Зробити висновок щодо наявності мультиколінеальності в сукупності чинників економетричної моделі.

7. Знайти матрицю С, обернену до матриці r за виразом (4.10).

8. Використовуючи діагональні елементи матриці С (4.11), розрахувати фактичні значення критеріїв Фішера за виразами (4.12).

9. Перевіривши умову (4.13), зробити висновок щодо коленіарності кожного чинника моделі з іншими чинниками.

10. Використовуючи значення елементів матриці С (4.11), за виразами (4.14) розрахувати часткові коефіцієнти кореляції.

11. Порівняти часткові коефіцієнти кореляції з парними, вираз(4.7) і зробити висновок.

12. Використовуючи значення часткових коефіцієнтів кореляції, обчислити фактичні значення критеріїв Студента(t-критеріїв) за виразами (4.15).

13. Для числа ступенів вільності n-m і рівні значущості α визначити критичне (табличне) значення t-критеріїв і порівняти його з фактичними значеннями, визначенними в п. 12. Зробити висновок щодо наявності мультиколінеарності і економічної доцільності виключення певного чинника.

Приклад виконання

1.Нехай маємо базу даних, сформовану у практичній роботі №2 згідно свого варіанту.

2. Обчислюємо для кожного чинника середнє значення та дисперсію, нормалізуємо змінні, використовуючи електронні таблиці Excel. Результат формуємо в таблиці на окремому аркуші поряд з базою даних, при цьому використовуємо результати попередньої роботи.

3. Знаходимо кореляційну матрицю за формулою:

$$r = \frac{1}{n} X^{*T} * X^*,$$

де X^* матриця щойно отриманих нормалізованих змінних.

Для множення матриць використаємо пакет прикладних програм Mathcad.

4. Аналізуючи елементи кореляційної матриці, можна зробити висновок, що існує значна кореляція між першою змінною та четвертою, першою та п’ятою, третьою та четвертою, четвертою та п’ятою.

Досить слабо пов’язана з іншими друга змінна. Попередній аналіз свідчить про досить тісний зв’язок четвертої змінної з іншими.

5. Обчислюємо визначник кореляційної матриці:

$$D = |r| = 0{,}037$$

6. Знаходимо значення критерію Спірмена за виразом (4.9):

$$\chi^2_{\delta\grave{a}\hat{e}\grave{o}} = -(15 - \frac{1}{6}(2 \cdot 5 + 5)) \ln D = -12{,}5 \cdot \ln 0{,}037 = 41{,}21$$

Табличне значення χ^2-критерію для ступенів вільності $\frac{1}{2}m(m-1) = \frac{1}{2}5(5-1) = 10$ і рівня значущості $\alpha = 0{,}05$ дорівнює $\chi^2_{таб} = 18{,}31$.

Оскільки $\chi^2_{факт} > \chi^2_{таб}$ доходимо до висновку, що в масиві змінних існує мультиколінеарність.

7. Знаходимо матрицю C, обернену до матриці r:

$$C = r^{-1} = \begin{pmatrix} 3.805 & -0.267 & -2.164 & 4.073 & 0.127 \\ -0.267 & 1.261 & 0.84 & -0.814 & 4.758*10^{-3} \\ -2.167 & 0.84 & 3.461 & -4.513 & 0.867 \\ 4.073 & -0.814 & -4.513 & 9.663 & -3.397 \\ 0.127 & 4.758*10^{-3} & 0.867 & -3.397 & 3.59 \end{pmatrix}$$

8.Визначаємо фактичні значення критеріїв Фішера, використовуючи діагональні елементи матриці C.

$$F_1=(c_{11}-1)\frac{n-m}{m-1}=(3{,}805-1)\frac{16-5}{5-1}=7{,}714$$

$$F_2=(c_{22}-1)\frac{n-m}{m-1}=(1{,}261-1)\frac{16-5}{5-1}=0{,}718$$

$$F_3=(c_{33}-1)\frac{n-m}{m-1}=(3{,}461-1)\frac{16-5}{5-1}=6{,}768$$

$$F_4=(c_{44}-1)\frac{n-m}{m-1}=(9{,}663-1)\frac{16-5}{5-1}=23{,}823$$

$$F_5=(c_{55}-1)\frac{n-m}{m-1}=(3{,}59-1)\frac{16-5}{5-1}=7{,}123$$

9. Знаходимо критичне (табличне) значення критерію $F_{таб} = 3{,}36$. Можно бачити що для рівня значущості α =0,05 (95% довіри) і ступенів вільності n-m=11 та m-1=4 для отриманих в п. 8 результатів лише друга змінна, для якої $F_r < F_{таб}$ слабо пов'язана з іншими. Для всіх інших змінних $F_{факт} > F_{таб}$, що свідчить про високу мультиколінеарність між кожною з цих змінних та іншими. Особливо великий цей показник для четвертої змінної.

10. Знаходимо частинні коефіцієнти кореляції, які характеризують тісноту зв'язку між двома змінами, за умови, що інші на цей зв'язок не впливають. При обчисленнях використовуємо елементи матриці C:

$$r_{12} = -\frac{c_{12}}{\sqrt{c_{11}} \cdot \sqrt{c_{22}}} = -\frac{-0.267}{\sqrt{3.805} \cdot \sqrt{1.261}} = 0{,}122$$

$$r_{13} = -\frac{c_{13}}{\sqrt{c_{11}} \cdot \sqrt{c_{33}}} = -\frac{-2.164}{\sqrt{3.805} \cdot \sqrt{3.461}} = 0{,}596$$

$$r_{14} = -\frac{c_{14}}{\sqrt{c_{11} \cdot \sqrt{c_{44}}}} = -\frac{4.073}{\sqrt{3.805 \cdot \sqrt{9.663}}} = -0{,}672$$

$$r_{15} = -\frac{c_{15}}{\sqrt{c_{11}}\cdot\sqrt{c_{55}}} = -\frac{0.127}{\sqrt{3.805}\cdot\sqrt{3.59}} = -0{,}034$$

$$r_{23} = -\frac{c_{23}}{\sqrt{c_{22}\cdot\sqrt{c_{22}}}} = -\frac{0.84}{\sqrt{1.261}\cdot\sqrt{3.461}} = -0{,}402$$

$$r_{24} = -\frac{c_{24}}{\sqrt{c_{22}}\cdot\sqrt{c_{44}}} = -\frac{-0{,}814}{\sqrt{1{,}261}\cdot\sqrt{9{,}663}} = 0{,}233$$

$$r_{25} = -\frac{c_{25}}{\sqrt{c_{22}}\cdot\sqrt{c_{55}}} = -\frac{4{,}758^{-3}}{\sqrt{1{,}261}\cdot\sqrt{3{,}59}} = 2{,}237\cdot 10^{-3}$$

$$r_{34} = -\frac{c_{34}}{\sqrt{c_{33}}\cdot\sqrt{c_{44}}} = -\frac{-4.513}{\sqrt{3.461}\cdot\sqrt{9.663}} = 0{,}78$$

$$r_{35} = -\frac{c_{35}}{\sqrt{c_{33}}\cdot\sqrt{c_{55}}} = -\frac{0.867}{\sqrt{3.461}\cdot\sqrt{3.59}} = -0{,}246$$

$$r_{45} = -\frac{c_{45}}{\sqrt{c_{44}}\cdot\sqrt{c_{55}}} = -\frac{-3.397}{\sqrt{9.663}\cdot\sqrt{3.59}} = 0{,}577$$

11. Порівнюючи частинні коефіцієнти кореляції з парними, робимо висновок про те, що перші як правило, менші за другі. Частинні коефіцієнти кореляції дозволяють набагато точніше фіксувати зв’язок між парою змінних.

12. Використовуючи одержані значення частинних коефіцієнтів кореляції, визначаємо критерії Стьюдента:

$$t_{12} = \frac{r_{12}\sqrt{n-m}}{\sqrt{1-r_{12}^2}} = \frac{0{,}122\cdot\sqrt{11}}{\sqrt{1-(0{,}122)^2}} = \frac{0{,}122\cdot 3{,}317}{\sqrt{0{,}985}} = 0{,}408$$

$$t_{13} = \frac{r_{13}\sqrt{n-m}}{\sqrt{1-r_{13}^2}} = \frac{0{,}596\cdot\sqrt{11}}{\sqrt{1-0{,}596^2}} = 2{,}462$$

$$t_{14} = \frac{r_{14}\sqrt{n-m}}{\sqrt{1-r_{14}^2}} = \frac{0{,}672\cdot\sqrt{11}}{\sqrt{1-0{,}672^2}} = 3{,}01$$

$$t_{15} = \frac{r_{15}\sqrt{n-m}}{\sqrt{1-r_{15}^2}} = \frac{0{,}034\cdot\sqrt{11}}{\sqrt{1-0{,}034^2}} = 0{,}112$$

$$t_{23} = \frac{r_{23}\sqrt{n-m}}{\sqrt{1-r_{23}^2}} = \frac{0{,}402\cdot\sqrt{11}}{\sqrt{1-0{,}402^2}} = 1{,}456$$

$$t_{24} = \frac{r_{24}\sqrt{n-m}}{\sqrt{1-r_{24}^2}} = \frac{0{,}233\cdot\sqrt{11}}{\sqrt{1-0{,}233^2}} = 0{,}844$$

$$t_{25} = \frac{r_{25}\sqrt{n-m}}{\sqrt{1-r_{25}^2}} = \frac{2{,}237\cdot 10^{-3}\cdot\sqrt{11}}{\sqrt{1-2{,}237^2\cdot 10^{-6}}} = 0{,}742\cdot 10^{-2}$$

$$t_{34} = \frac{r_{34}\sqrt{n-m}}{\sqrt{1-r_{34}^2}} = \frac{0{,}78\cdot\sqrt{11}}{\sqrt{1-0{,}78^2}} = 4.133$$

$$t_{35} = \frac{r_{35}\sqrt{n-m}}{\sqrt{1-r_{35}^2}} = \frac{0,246\sqrt{11}}{\sqrt{1-0,246^2}} = 0,826$$

$$t_{45} = \frac{r_{35}\sqrt{n-m}}{\sqrt{1-r_{35}^2}} = \frac{0,557\sqrt{11}}{\sqrt{1-0,577^2}} = 2,583$$

13. Визначаємо критичне значення t – критерію при n – m ступенях вільності і рівня значущості $\alpha = 0,05$. Обчислення показують, що фактичні значення t–критерію t_{13}, t_{14}, t_{34}, t_{45}, перевищують табличне значення ($t_{таб} = 1,796$) причому $t_{34} > t_{14} > t_{45} > t_{13} > t_{таб}$.

Це говорить про доцільність виключення з моделі четвертої змінної і подання моделі,

враховуючи результати лабораторної роботи №3, у вигляді:

$$У = -44,675+0,175X_2+0,215X_5-0,108X_1$$

Контрольні запитання

1. В чому полягає мультиколінеарність змінних?
2. Які ознаки мультиколінеарності?
3. Як впливає мультиколінеарність на якість економетричної моделі?
4. Як проводиться дослідження на мультиколінеарність за допомогою критерію Спірмена?
5. Що визначається за допомогою χ^2-критерію?
6. Як використовується критерій Фішера при дослідженнях на мультиколінеарність?
7. Що визначає F-критерій в дослідженнях на мультиколінеарність?
8. Як використовується критерій Стьюдента у дослідженнях на мультиколінеарність?
9. Що визначається за допомогою t-критерію?
10. Як визначається матриця помилок С?
11. Що таке частинні коефіцієнти кореляції?
12. Як знаходяться частинні коефіцієнти кореляції?
13. У чому сутність алгоритму Феррара-Глобера?

Практична робота №5

Тема: Виробничі функції Кобба - Дугласа.

Мета: Оволодіти обробкою даних за загальною степеневою моделлю і моделлю Кобба – Дугласа.

Короткі теоретичні відомості

В 1927 р. економіст Пол Дуглас виявив, що на діаграмі логарифмів капітальних затрат (K), обсягів випуску (Y) та затрат праці (L) відстані від точок графіку ln Y до точок графіків ln K і ln L складають постійну пропорцію.

Математик Чарльз Кобб показав, що таку особливість має степенева залежність $Y=AK^bL^c$, якщо показники степеня b і c в сумі дорівнюють одиниці (b+c=1).

Отже, отримуємо відому функцію Кобба – Дугласа:

$$Y=AK^bL^{1-b}. \tag{5.1}$$

Параметри b і c в степеневій залежності є коефіцієнтами еластичності випуску за затратами капіталу і праці. Іншими словами, загальна степенева функція описує залежності з постійною еластичністю.

Коефіцієнт еластичності за капіталом $E_{Y/K}$ показує, на скільки відсотків збільшиться обсяг випуску при збільшенні затрат капіталу на 1%. Аналогічно визначається коефіцієнт еластичності за працею $E_{Y/L}$.

Можна також показати, що параметри b і c в степеневій залежності характеризують прогнозовані долі виробничих факторів, зокрема b є доля прибутку на одиницю затрат капіталу, а c – на одиницю затрат праці.

Після логарифмування модель стає лінійною відносно параметрів і може бути оцінена методом найменших квадратів. За додаткової умови (b+c=1) модель перетворюється до вигляду:

$$\ln\frac{Y}{L} = a + b \cdot \ln\frac{Y}{K} \tag{5.2}$$

Завдання

За даними обсягу випуску Y, затрат капіталу K та праці L (у відсотках до 1899 р.), наведеними у таблиці 5.1, розрахувати параметри загальної степеневої моделі та виробничої функції Кобба - Дугласа. Провести якісний аналіз отриманих моделей.

Таблиця 5.1*

Рік, T	Y	K	L
1900	101+N/10	107+N/10	105+N/10
1901	112+N/10	114+N/10	110+N/10
1902	122+N/10	122+N/10	118+N/10
1903	124+N/10	131+N/10	129+N/10
1904	122+N/10	138+N/10	116+N/10
1905	143+N/10	149+N/10	125+N/10
1906	152+N/10	163+N/10	133+N/10
1907	151+N/10	176+N/10	138+N/10
1908	126+N/10	185+N/10	121+N/10
1909	155+N/10	198+N/10	140+N/10
1910	159+N/10	208+N/10	144+N/10
1911	153+N/10	216+N/10	145+N/10
1912	177+N/10	226+N/10	152+N/10
1913	184+N/10	236+N/10	154+N/10
1914	169+N/10	244+N/10	149+N/10
1915	189+N/10	266+N/10	154+N/10
1916	225+N/10	298+N/10	182+N/10

1917	227+N/10	335+N/10	196+N/10
1918	223+N/10	366+N/10	200+N/10
1919	218+N/10	387+N/10	193+N/10
1920	231+N/10	407+N/10	193+N/10
1921	179+N/10	417+N/10	147+N/10
1922	240+N/10	431+N/10	161+N/10

*Примітка: варіант формується за допомогою Таблиці 5.1, в якій N – це дві останні цифри номера Вашої залікової книжки.

Порядок виконання

1. За своїм варіантом сформувати вихідну базу даних та побудувати таблицю статистичних показників.
2. Розрахувати значення логарифмів ln Y, ln K та ln L.
3. За допомогою вбудованої функції Excel ЛИНЕЙН обчислити параметри загальної степеневої моделі $Y=AK^bL^c$, а також значення критерію Стьюдента для кожного з параметрів.
4. Розрахувати значення логарифмів ln (K/L) та ln (Y/L), а на їх основі за допомогою формули (5.2) – значення параметрів виробничої функції Кобба – Дугласа.
5. Обчислити критерії Стьюдента для функції Кобба – Дугласа.

Приклад виконання

1. Нехай база статистичних даних, що характеризує залежність обсягу виробництва від затрат капіталу та праці, має вигляд, представлений у таблиці 5.2

Таблиця 5.2

Рік, T	Y	K	L
1900	101	107	105
1901	112	114	110
1902	122	122	118
1903	124	131	129
1904	122	138	116
1905	143	149	125
1906	152	163	133
1907	151	176	138
1908	126	185	121
1909	155	198	140
1910	159	208	144
1911	153	216	145
1912	177	226	152
1913	184	236	154
1914	169	244	149
1915	189	266	154
1916	225	298	182
1917	227	335	196
1918	223	366	200
1919	218	387	193
1920	231	407	193
1921	179	417	147
1922	240	431	161

2. За допомогою функції Excel LN знаходимо значення логарифмів пояснювальної та пояснюючих змінних. Результат наведено у таблиці 5.3

Таблиця 5.3

Рік, T	ln Y	ln K	ln L
1900	4,615121	4,672829	4,65396
1901	4,718499	4,736198	4,70048
1902	4,804021	4,804021	4,770685
1903	4,820282	4,875197	4,859812
1904	4,804021	4,927254	4,75359
1905	4,962845	5,003946	4,828314
1906	5,023881	5,09375	4,890349
1907	5,01728	5,170484	4,927254
1908	4,836282	5,220356	4,795791
1909	5,043425	5,288267	4,941642
1910	5,068904	5,337538	4,969813
1911	5,030438	5,375278	4,976734
1912	5,17615	5,420535	5,023881
1913	5,214936	5,463832	5,036953
1914	5,129899	5,497168	5,003946
1915	5,241747	5,583496	5,036953
1916	5,4161	5,697093	5,204007
1917	5,42495	5,814131	5,278115
1918	5,407172	5,902633	5,298317
1919	5,384495	5,958425	5,26269
1920	5,442418	6,008813	5,26269
1921	5,187386	6,033086	4,990433
1922	5,480639	6,066108	5,081404

3. За допомогою функції ЛИНЕЙН обчислюємо параметри загальної степеневої моделі. Діапазон виводу функції має такий вигляд:

c	b	a
0,778451	**0,247149**	**-0,11101**
0,151219	0,064823	0,469783
0,948437	0,060816	#Н/Д
183,936	20	#Н/Д
1,360598	0,073971	#Н/Д

tc	tb	ta
5,147846	3,812683	-0,23629

Нижче блоку виводу функції ЛИНЕЙН доданий рядок статистик Стьюдента (відношення значення параметру до його помилки).

Отримали b=0,247 і c=0,778 – близькі значення до 0,25 і 0,75.

В сумі b+c=1,0256 дуже близько до одиниці, вільний член моделі не значимий за критерієм Стьюдента (-0,24) і може бути прирівняний до нуля. Таким чином, вихідні дані добре описуються моделлю ln Y = 0,25 ln K + 0,75 ln L, звідки випливає помічена Дугласом пропорційність відрізків між графіками логарифмів Y K і L.

4. Приймаємо зв'язок b+c=1 між коефіцієнтами еластичностей випуску за затратами праці і капіталу і приводимо загальну степеневу модель ln Y = a + b*ln K + c*ln L до вигляду ln (Y/L) = a + b ln (K/L). Розраховані значення логарифмів ln (K/L) та ln (Y/L) наведені у таблиці 5.4.

Таблиця 5.4

Рік, T	ln (K/L)	ln (Y/L)
1900	0,018868	-0,03884
1901	0,035718	0,018019
1902	0,033336	0,033336
1903	0,015385	-0,03953
1904	0,173663	0,050431
1905	0,175633	0,134531
1906	0,203401	0,133531
1907	0,24323	0,090026
1908	0,424565	0,040491
1909	0,346625	0,101783
1910	0,367725	0,099091
1911	0,398545	0,053704
1912	0,396654	0,152269
1913	0,426879	0,177983
1914	0,493222	0,125952
1915	0,546544	0,204794
1916	0,493087	0,212094
1917	0,536016	0,146835
1918	0,604316	0,108854
1919	0,695735	0,121805
1920	0,746123	0,179728
1921	1,042654	0,196953
1922	0,984704	0,399235

За допомогою функції ЛИНЕЙН розраховуємо параметри виробничої функції Кобба – Дугласа:

b	a
0,259218	**0,011554**
0,044439	0,021993
0,618363	0,059451
34,0261	21
0,120264	0,074224

Функцією ЛИНЕЙН оцінені параметри моделі Кобба – Дугласа: ln (Y/L) = 0,012+0,259ln(K/L).

5. Обчислюємо критерії Стьюдента для функції Кобба – Дугласа:

tb	ta
5,83319	0,525337

Контрольні запитання

1. Що собою являє загальна степенева модель?
2. В чому відмінність між загальною степеневою моделлю та виробничою функцією Кобба – Дугласа?
3. За допомогою яких математичних перетворень нелінійні форми зв’язку можна привести до лінійних?
4. Що показують коефіцієнти еластичності за капіталом та працею?
5. Як за допомогою критерію Стьюдента оцінити достовірність оцінок параметрів моделі?
6. Які елементи містить діапазон виводу функції ЛИНЕЙН?

Практична робота №6

Тема: Двокроковий метод найменших квадратів.
Мета: Знайти 1МНК і 2МНК оцінки параметрів економетричної моделі Л. Клейна.

Короткі теоретичні відомості

Економетричні моделі складаються у вигляді системи декількох регресійних рівнянь, тотожностей і нерівностей. Модель у такій формі називається структурною та відображає основні закони економіки. Для прикладу розглянемо першу економетричну модель Л. Клейна для описання економіки США в період між двома світовими війнами (1921 – 1941). Модель Клейна складається з трьох регресійних рівнянь і трьох тотожностей:

$$C=a_0+a_1*P+a_2*PLAG+a_3*(W+WP)+u;$$
$$I=b_0+b_1*P+b_2*PLAG+b_3*KLAG+v;$$
$$W=c_0+c_1*X+c_2*XLAG+c_3*t+w; \qquad (6.1)$$
$$X=C+I+G;$$
$$P=X-W-T;$$
$$K=KLAG+I.$$

Ці шість рівностей визначають шість результативних ознак, які називаються *ендогенними* (внутришньосистемними) змінними:

C – (consumption) – обсяг споживання;
I – (investment) – інвестиції;
W – (private wage bill) – особистий дохід в приватному секторі;
X – (private production) – обсяг виробництва приватних фірм;
P – (profits) – прибуток;
K – (capital stock) – акціонерний капітал.

Серед визначаючих змінних є ті самі ендогенні змінні і ендогенні змінні з запізненням на один період (тобто дані попереднього року): XLAG, PLAG, KLAG. В науковій літературі зазвичай такі змінні позначають X_{t-1}, P_{t-1}, K_{t-1}, інші змінні мають індекс t: C_t , I_t , W_t , X_t , P_t , K_t . Далі серед визначаючих змінних є так звані *екзогенні* (позасистемні) змінні, значення яких задаються ззовні:

t – (year) – час;
WP – (government wage bill) – заробітна плата в державному секторі;
G – (government demand) – державні закупки;
T – (taxes) – податки.

Лагові ендогенні й екзогенні змінні називаються *наперед визначеними*, їх значення відомі до теперішнього моменту часу.

Члени u, v, w описують випадкові збурення в кожному рівнянні.

Зазначимо, що в системі *одночасних рівнянь* деякі змінні в одних рівняннях виступають в ролі результативних ознак, а в інших – в якості пояснюючих змінних (ендогенні пояснюючі змінні). Це призводить до деяких ускладнень в оцінюванні параметрів моделі. Так, оцінювання звичайним МНК кожного регресійного рівняння окремо приводить до наступних результатів:

$$C=16{,}24+0{,}193*P+0{,}090*PLAG+0{,}796*(W+WP)+u;$$
$$I=10{,}13+0{,}480*P+0{,}333*PLAG-0{,}110*KLAG+v;$$
$$W=1{,}497+0{,}439*X+0{,}146*XLAG+0{,}130*t+w.$$

Оскільки всі змінні виражені в одних одиницях (млрд. долл. США за курсом 1931 р.), то коефіцієнти регресії при P та PLAG можна трактувати як норми відрахувань прибутку в споживання й інвестиції. Але тоді вийшло, що відрахування в споживання складають 19% прибутку поточного року та тільки 9% прибутку минулого року; аналогічно, в інвестиції відраховується 48% прибутку поточного року та 33,3% прибутку минулого року. Це зовсім не відповідає дійсності. Тільки невелика частина прибутку поточного року може бути освоєна в цьому ж році і тільки тому, що фінансовий рік не співпадає з календарним.

Звичайні МНК-оцінки параметрів моделі виявилися зміщеними.

Причина настільки великого зміщення полягає в порушенні першої (основної) передумови регресійного аналізу, згідно з якою випадкові збурення відносяться тільки до результативних ознак і не мають корелювати з пояснюючими змінними. Однак в системі одночасних рівнянь через систему тотожностей випадкові помилки адитивно переносяться на ендогенні пояснюючі змінні. Так, з першого рівняння випливає, що випадкова помилка u додається до C, далі згідно першій тотожності X=C+I+G ця помилка переноситься на X і, нарешті, згідно другій тотожності P=X-W-T помилка u переноситься на P – пояснюючу змінну першого рівняння: u → C → X → P. Випадкова помилка другого рівняння v також тісно корелює з P – пояснюючою змінною другого рівняння: v →I →X →P. З третім рівнянням все в порядку і його параметри оцінені без зміщення.

Коли правильно визначена причина хвороби, знаходяться способи її лікування. Одним з таких способів є заміна в моделі *пояснюючої* ендогенної змінної на так звану *інструментальну змінну*, яка тісно корелює зі змінною, що замінюється, але не корелює з випадковим членом. Інструментальні змінні CR, IR, WR складаються у вигляді лінійних комбінацій наперед визначених змінних, на які випадкова помилка не впливає.

Перенесемо всі ендогенні змінні в ліву частину рівностей і запишемо структурну модель в такій матричній формі:

$$AY=BZ+E,$$

де Y^T=(C, I, W, X, P, K) – матриця значень ендогенних змінних;

Z^T=(X0, XLAG, PLAG, KLAG, WP, G, T, t) – матриця значень наперед визначених змінних; X0=1;

E^T=(u, v, w, 0, 0, 0) – випадкові члени кожної рівності;

A – матриця коефіцієнтів перед ендогенними змінними [6x6];

B – коефіцієнти перед наперед визначеними змінними [6x8].

Виразимо всі ендогенні змінні через наперед визначені, для чого помножимо записану вище матричну рівність на A^{-1} і отримаємо модель в так званій *приведеній формі*:

$$\begin{aligned} &CR=a0+a1*PLAG+a2*XLAG+a3*KLAG+a4*WP+a5*G+a6*T+a7*t; \\ &IR=b0+b1*PLAG+b2*XLAG+b3*KLAG+b4*WP+b5*G+b6*T+b7*t; \\ &WR=c0+c1*PLAG+c2*XLAG+c3*KLAG+c4*WP+c5*G+c6*T+c7*t. \end{aligned} \tag{6.2}$$

Тут через CR, IR, WR позначені розрахункові значення відповідних ендогенних змінних. Розрахункові значення інших ендогенних змінних визначаються системою тотожностей:

$$\begin{aligned} &XR=CR+IR+G; \\ &PR=XR-WR-T; \\ &KR=KLAG+IR. \end{aligned} \tag{6.3}$$

Розрахункові значення тісно пов’язані з відповідними результативними ознаками C, I, W, X, P, K (з коефіцієнтами детермінації 0,938; 0,847; 0,950; 0,918; 0,826; 0,977) і виражені тільки через наперед визначені змінні. Тому вони є найкращими інструментальними змінними.

Процедура оцінювання тепер природнім чином розпадається на два етапи (процедура 2 МНК). На першому етапі оцінюються (звичайним МНК) регресійні рівняння моделі в приведеній формі і обчислюються розрахункові значення всіх ендогенних змінних. На другому кроці оцінюється кожне рівняння структурної моделі з попередньою заміною ендогенних пояснюючих змінних на відповідні розрахункові значення.

Для моделі Клейна на другому кроці 2 МНК отримані рівняння:

$$\begin{aligned} &C=16,55+0,017*PR+0,216*PLAG+0,810*(WR+WP)+u; \\ &I=20,28+0,150*PR+0,616*PLAG-0,160*KLAG+v; \\ &W=1,500+0,439*XR+0,147*XLAG+0,130*t+w. \end{aligned}$$

Третє рівняння практично не змінилося, так і має бути. Характер же відрахувань прибутку на споживання та інвестиції докорінно змінився – основні відрахування тепер припадають на прибуток минулого року. Цікаво, що в сумі всі відрахування склали 100%

(0,017+0,216+0,150+0,616=1).

Істотне коригування параметрів структурної моделі відбулося в основному за рахунок пояснюючої ендогенної змінної P на свою інструментальну змінну PR (див. рис. 6.1). Ці поправки, взагалі, незначні. Дивно навіть, що такі малі поправки в даних змогли запобігти настільки великим зміщенням в оцінках параметрів моделі.

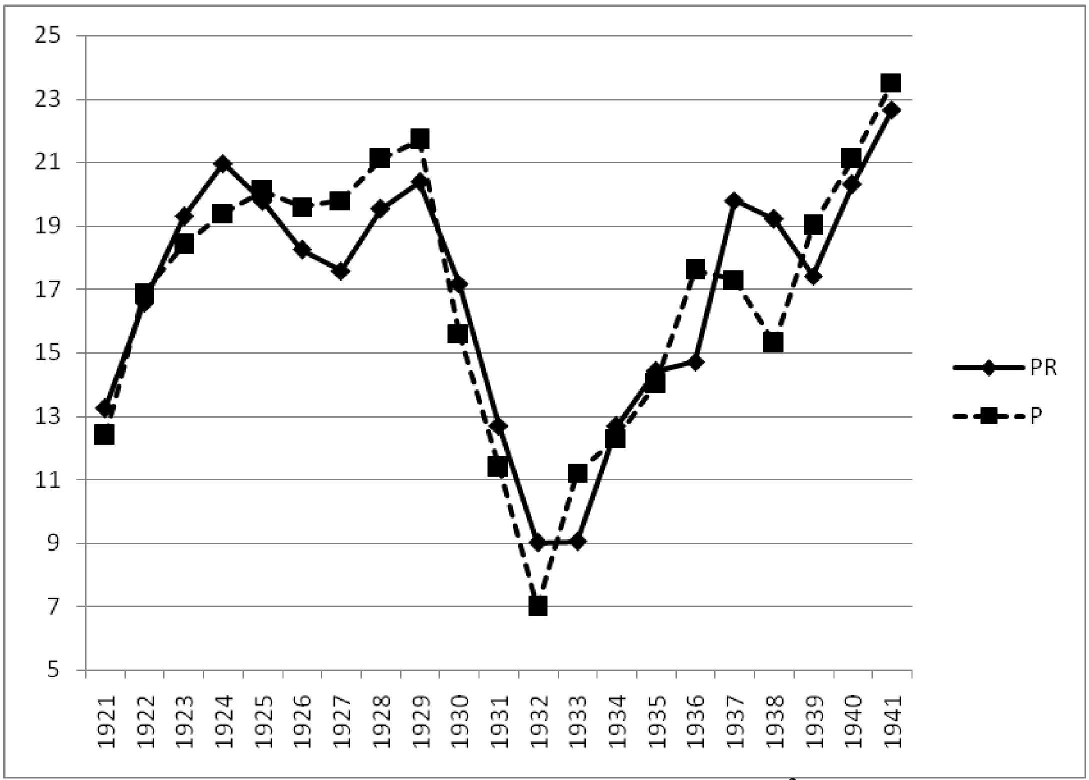


Рис. 6.1. Графік динаміки змінних P та PR (R^2=0,8261)

Завдання

Має місце наступна вибірка статистичних даних (таблиця 6.1), де

C – (consumption) – обсяг споживання;

I – (investment) – інвестиції;

W – (private wage bill) – особистий дохід в приватному секторі;

X – (private production) – обсяг виробництва приватних фірм;

P – (profits) – прибуток;

K – (capital stock) – акціонерний капітал;

t – (year) – час;

WP – (government wage bill) – заробітна плата в державному секторі;

G – (government demand) – державні закупки;

T – (taxes) – податки.

Таблиця 6.1*

Рік	C	I	W	P	X	K	WP	G	T	t
1921	41,9+N/10	-0,2	25,5+N/10	12,4	45,6+N/10	182,6+N/10	2,7	3,9	7,7	-10
1922	45+N/10	1,9	29,3+N/10	16,9	50,1+N/10	184,5+N/10	2,9	3,2	3,9	-9
1923	49,2+N/10	5,2	34,1+N/10	18,4	57,2+N/10	189,7+N/10	2,9	2,8	4,7	-8
1924	50,6+N/10	3	33,9+N/10	19,4	57,1+N/10	192,7+N/10	3,1	3,5	3,8	-7
1925	52,6+N/10	5,1	35,4+N/10	20,1	61+N/10	197,8+N/10	3,2	3,3	5,5	-6
1926	55,1+N/10	5,6	37,4+N/10	19,6	64+N/10	203,4+N/10	3,3	3,3	7	-5
1927	56,2+N/10	4,2	37,9+N/10	19,8	64,4+N/10	207,6+N/10	3,6	4	6,7	-4
1928	57,3+N/10	3	39,2+N/10	21,1	64,5+N/10	210,6+N/10	3,7	4,2	4,2	-3

1929	57,8+N/10	5,1	41,3+N/10	21,7	67+N/10	215,7+N/10	4	4,1	4	-2
1930	55+N/10	1	37,9+N/10	15,6	61,2+N/10	216,7+N/10	4,2	5,2	7,7	-1
1931	50,9+N/10	-3,4	34,5+N/10	11,4	53,4+N/10	213,3+N/10	4,8	5,9	7,5	0
1932	45,6+N/10	-6,2	29+N/10	7	44,3+N/10	207,1+N/10	5,3	4,9	8,3	1
1933	46,5+N/10	-5,1	28,5+N/10	11,2	45,1+N/10	202+N/10	5,6	3,7	5,4	2
1934	48,7+N/10	-3	30,6+N/10	12,3	49,7+N/10	199+N/10	6	4	6,8	3
1935	51,3+N/10	-1,3	33,2+N/10	14	54,4+N/10	197,7+N/10	6,1	4,4	7,2	4
1936	57,7+N/10	2,1	36,8+N/10	17,6	62,7+N/10	199,8+N/10	7,4	2,9	8,3	5
1937	58,7+N/10	2	41+N/10	17,3	65+N/10	201,8+N/10	6,7	4,3	6,7	6
1938	57,5+N/10	-1,9	38,2+N/10	15,3	60,9+N/10	199,9+N/10	7,7	5,3	7,4	7
1939	61,6+N/10	1,3	41,6+N/10	19	69,5+N/10	201,2+N/10	7,8	6,6	8,9	8
1940	65+N/10	3,3	45+N/10	21,1	75,7+N/10	204,5+N/10	8	7,4	9,6	9
1941	69,7+N/10	4,9	53,3+N/10	23,5	88,4+N/10	209,4+N/10	8,5	13,8	11,6	10

*Примітка: варіант формується за допомогою Таблиці 6.1, в якій N – це дві останні цифри номера Вашої залікової книжки.

За даною вибіркою знайти 1МНК та 2МНК оцінки параметрів економетричної моделі Клейна, порівняти їх за зробити висновки щодо випадків застосування та результатів розрахунків за двокроковим методом найменших квадратів.

Порядок виконання

1. За своїм варіантом сформувати вихідну базу даних та побудувати таблицю статистичних показників.
2. Розрахувати значення лагових змінних XLAG, PLAG, KLAG.
3. За допомогою системи рівнянь (6.2) знайти параметри моделей розрахунку інструментальних змінних.
4. Обчислити значення інструментальних змінних CR, IR, WR, а також KR, XR, PR, використовуючи формулу (6.3).
5. Побудувати порівняльні графіки ендогенних та відповідних інструментальних змінних. Навести на графіках коефіцієнти детермінації.
6. Функцією ЛИНЕЙН оцінити кожне рівняння структурної моделі Клейна, замінюючи в них ендогенні пояснюючі змінні відповідними інструментальними.
7. Розрахувати параметри моделі Клейна звичайним 1МНК та порівняти результати з отриманими в попередньому пункті. Зробити висновки щодо ефективності використання двокрокового МНК.

Приклад виконання

1. Нехай база статистичних даних, що характеризує залежність обсягу виробництва від затрат капіталу та праці, має вигляд, представлений у таблиці 6.2

Таблиця 6.2

Рік	C	I	W	P	X	K	WP	G	T	t
1921	41,9	-0,2	25,5	12,4	45,6	182,6	2,7	3,9	7,7	-10
1922	45	1,9	29,3	16,9	50,1	184,5	2,9	3,2	3,9	-9
1923	49,2	5,2	34,1	18,4	57,2	189,7	2,9	2,8	4,7	-8
1924	50,6	3	33,9	19,4	57,1	192,7	3,1	3,5	3,8	-7
1925	52,6	5,1	35,4	20,1	61	197,8	3,2	3,3	5,5	-6
1926	55,1	5,6	37,4	19,6	64	203,4	3,3	3,3	7	-5
1927	56,2	4,2	37,9	19,8	64,4	207,6	3,6	4	6,7	-4
1928	57,3	3	39,2	21,1	64,5	210,6	3,7	4,2	4,2	-3
1929	57,8	5,1	41,3	21,7	67	215,7	4	4,1	4	-2
1930	55	1	37,9	15,6	61,2	216,7	4,2	5,2	7,7	-1
1931	50,9	-3,4	34,5	11,4	53,4	213,3	4,8	5,9	7,5	0

1932	45,6	-6,2	29	7	44,3	207,1	5,3	4,9	8,3	1
1933	46,5	-5,1	28,5	11,2	45,1	202	5,6	3,7	5,4	2
1934	48,7	-3	30,6	12,3	49,7	199	6	4	6,8	3
1935	51,3	-1,3	33,2	14	54,4	197,7	6,1	4,4	7,2	4
1936	57,7	2,1	36,8	17,6	62,7	199,8	7,4	2,9	8,3	5
1937	58,7	2	41	17,3	65	201,8	6,7	4,3	6,7	6
1938	57,5	-1,9	38,2	15,3	60,9	199,9	7,7	5,3	7,4	7
1939	61,6	1,3	41,6	19	69,5	201,2	7,8	6,6	8,9	8
1940	65	3,3	45	21,1	75,7	204,5	8	7,4	9,6	9
1941	69,7	4,9	53,3	23,5	88,4	209,4	8,5	13,8	11,6	10

2. Значення лагових змінних формуються з початкової бази даних шляхом зсуву значень відповідних показників на один рік вперед. Наприклад лаговий показник для 1930 року буде дорівнювати значенню відповідного показника у 1929 році. Отримані значення наведені в таблиці 6.3

Таблиця 6.3

Рік	PLAG	XLAG	KLAG
1921	12,7	44,9	182,8
1922	12,4	45,6	182,6
1923	16,9	50,1	184,5
1924	18,4	57,2	189,7
1925	19,4	57,1	192,7
1926	20,1	61	197,8
1927	19,6	64	203,4
1928	19,8	64,4	207,6
1929	21,1	64,5	210,6
1930	21,7	67	215,7
1931	15,6	61,2	216,7
1932	11,4	53,4	213,3
1933	7	44,3	207,1
1934	11,2	45,1	202
1935	12,3	49,7	199
1936	14	54,4	197,7
1937	17,6	62,7	199,8
1938	17,3	65	201,8
1939	15,3	60,9	199,9
1940	19	69,5	201,2
1941	21,1	75,7	204,5

3. Функцією ЛИНЕЙН обчислимо параметри моделі в приведеній формі, де в полі *Известные_значения_у* – послідовно задані стовбці C; I; W; в полі *Известные_значения_х* – стовбці наперед визначених змінних (PLAG, XLAG, KLAG, WP, G, T, t); в полі *Константа* – 1 (константа вимагається); в полі *Статистика* – 0 (статистика не вимагається). Без статистики вивід функції ЛИНЕЙН займає один рядок (8 комірок). Нагадуємо, що коефіцієнти регресії функцією ЛИНЕЙН виводяться в зворотному порядку. Для зручності наступних розрахунків нижче вони представлені у правильному порядку.

	t	T	G	WP	KLAG	XLAG	PLAG	Const
CR:	0,701	-0,366	0,205	0,193	-0,147	0,230	0,748	58,302
IR:	0,332	-0,162	0,100	-0,717	-0,193	-0,113	0,926	35,518
WR:	0,714	-0,604	0,866	-0,444	-0,123	0,095	0,872	43,436

	Const	PLAG	XLAG	KLAG	WP	G	T	t
CR:	58,302	0,748	0,230	-0,147	0,193	0,205	-0,366	0,701
IR:	35,518	0,926	-0,113	-0,193	-0,717	0,100	-0,162	0,332
WR:	43,436	0,872	0,095	-0,123	-0,444	0,866	-0,604	0,714

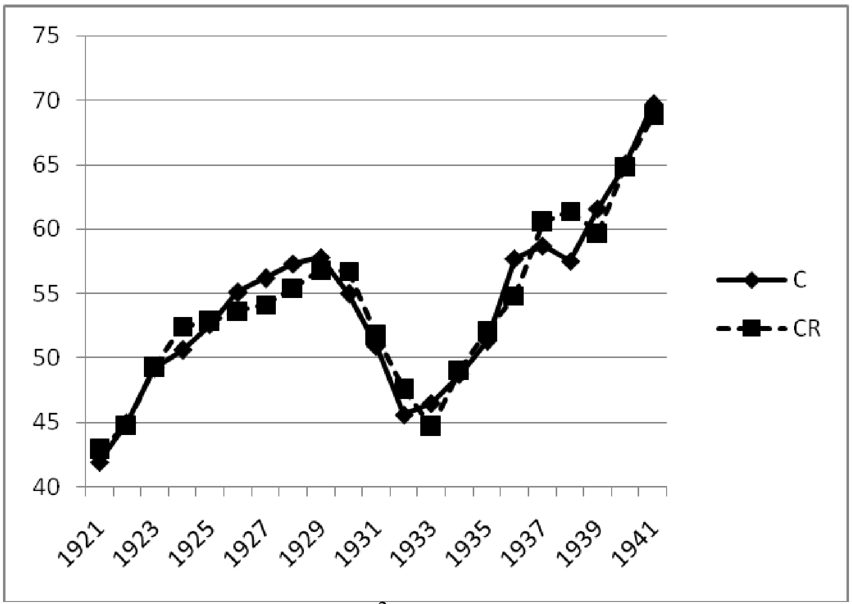
4. Розрахункові значення CR, IR, WR обчислені за формулою: *Розрахункові Значення у = Const+СУММПРОИЗВ(Параметри; Відомі значення х)* і наведені в таблиці 6.4.

Розрахункові значення інших показників отримані за допомогою тотожностей: KR=KLAG+IR, XR=CR+IR+G, WRS=WR+WP, PR=XR-WR-T.

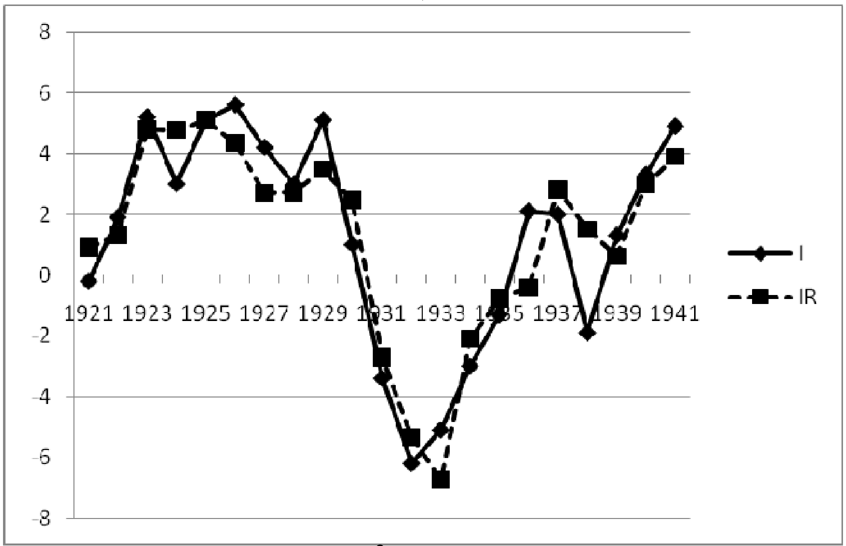
Таблиця 6.4

Рік	CR	IR	WR	KR	XR	XLAG	t	WRS	PR	PLAG	KLAG
1921	42,838	0,923	26,706	183,723	47,662	44,9	-10	29,406	13,256	12,7	182,8
1922	44,790	1,337	28,850	183,937	49,327	45,6	-9	31,750	16,577	12,4	182,6
1923	49,240	4,795	32,853	189,295	56,835	50,1	-8	35,753	19,282	16,9	184,5
1924	52,446	4,787	35,973	194,487	60,733	57,2	-7	39,073	20,960	18,4	189,7
1925	52,789	5,113	35,936	197,813	61,202	57,1	-6	39,136	19,767	19,4	192,7
1926	53,634	4,358	36,054	202,158	61,292	61	-5	39,354	18,239	20,1	197,8
1927	54,142	2,714	36,583	206,114	60,856	64	-4	40,183	17,573	19,6	203,4
1928	55,444	2,730	38,632	210,330	62,374	64,4	-3	42,332	19,542	19,8	207,6
1929	56,812	3,484	40,021	214,084	64,396	64,5	-2	44,021	20,375	21,1	210,6
1930	56,700	2,478	39,498	218,178	64,378	67	-1	43,698	17,180	21,7	215,7
1931	51,690	-2,707	34,678	213,993	54,883	61,2	0	39,478	12,705	15,6	216,7
1932	47,552	-5,320	29,832	207,980	47,132	53,4	1	35,132	9,000	11,4	213,3
1933	44,649	-6,712	27,184	200,388	41,638	44,3	2	32,784	9,054	7	207,1
1934	49,050	-2,080	31,499	199,920	50,970	45,1	3	37,499	12,671	11,2	202
1935	52,027	-0,766	34,040	198,234	55,661	49,7	4	40,140	14,421	12,3	199
1936	54,813	-0,399	34,303	197,301	57,315	54,4	5	41,703	14,712	14	197,7
1937	60,546	2,829	41,178	202,629	67,675	62,7	6	47,878	19,796	17,6	199,8
1938	61,401	1,509	41,603	203,309	68,210	65	7	49,303	19,207	17,3	201,8
1939	59,678	0,632	40,591	200,532	66,911	60,9	8	48,391	17,420	15,3	199,9
1940	64,882	2,996	45,372	204,196	75,278	69,5	9	53,372	20,306	19	201,2
1941	68,774	3,899	52,216	208,399	86,473	75,7	10	60,716	22,657	21,1	204,5

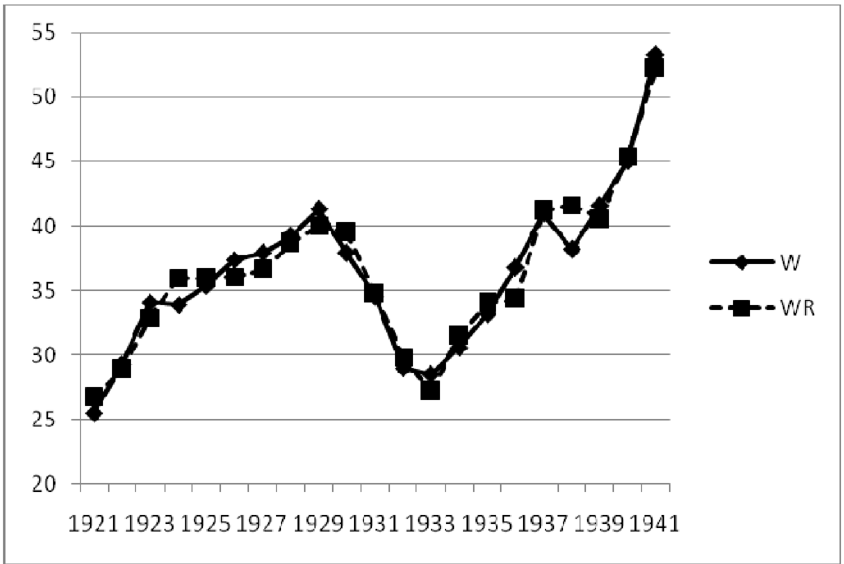
5. Будуємо в Excel графіки попарного порівняння ендогенних і відповідних їм інструментальних змінних. Для цього обираємо тип діаграми График. Для нашого прикладу результат наведено нижче. Під кожним рисунком вказаний коефіцієнт детермінації (можна знайти за допомогою функції ЛИНЕЙН: перша цифра в третьому рядку виводу значень цієї функції).



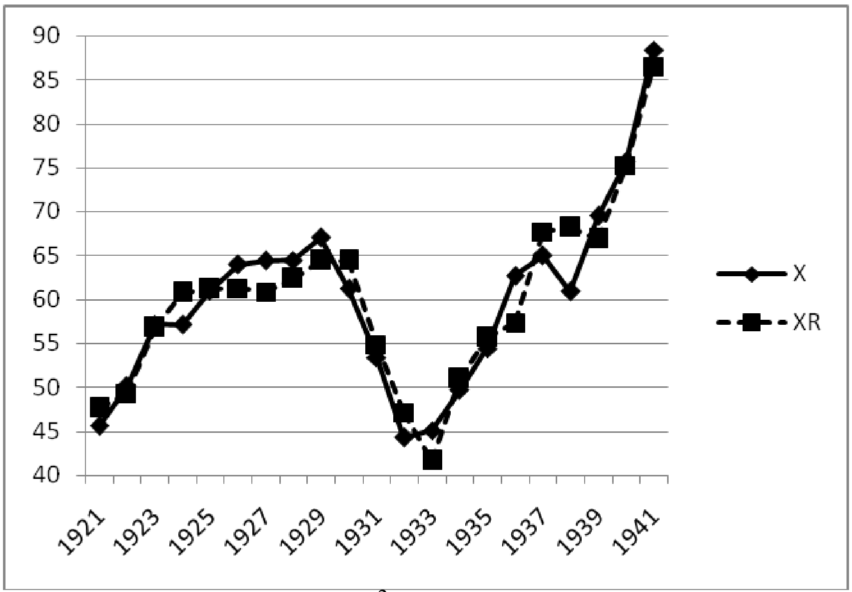
R^2=0,9383



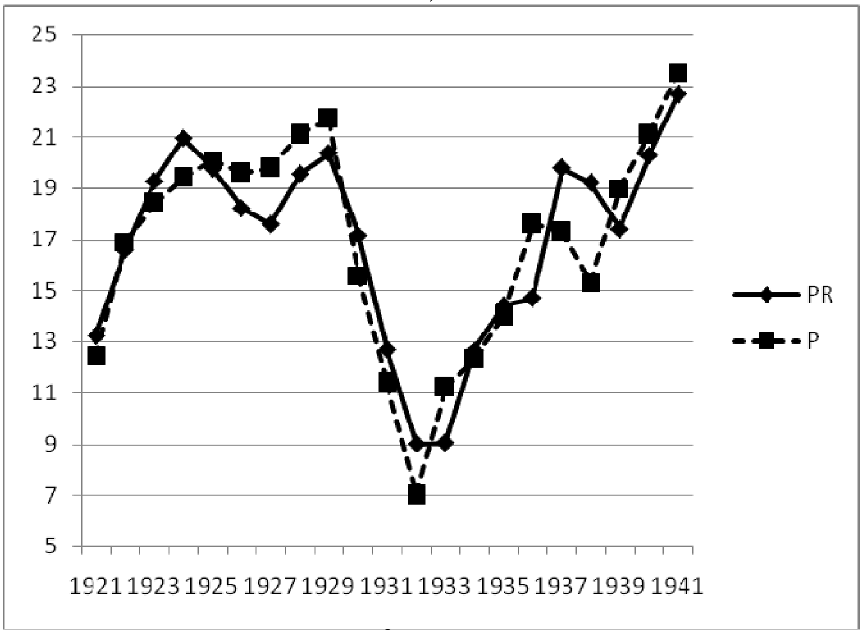
R^2=0,8469



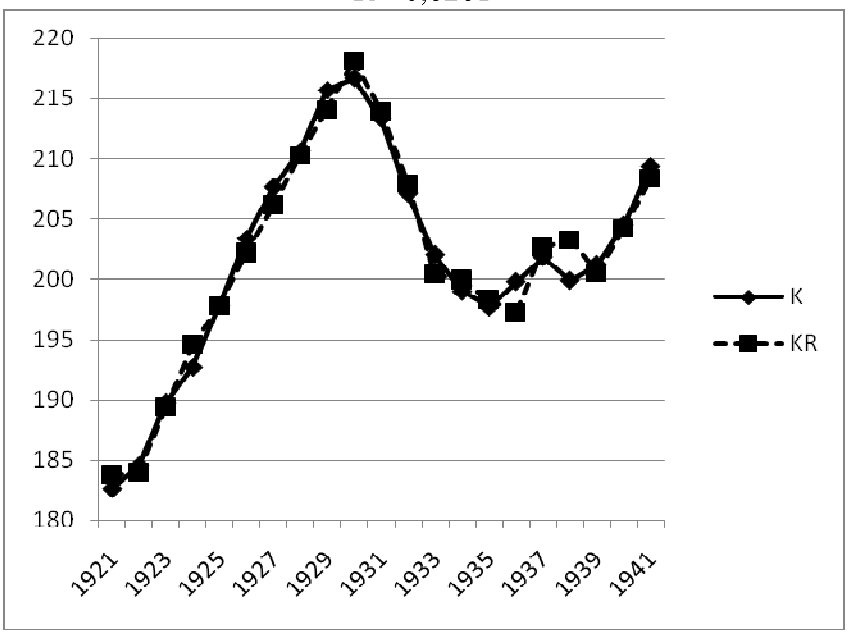
R^2=0,9497



R^2=0,9179



R^2=0,8261



R^2=0,9773

Як видно з графіків, інструментальні змінні дуже тісно корелюють з відповідними ендогенними і тому можуть з високою мірою достовірності використовуватися замість них в розрахунках.

6 – 7. Переходимо до другого кроку 2МНК процедури. Функцією ЛИНЕЙН треба оцінити кожне рівняння структурної моделі, замінюючи в них ендогенні пояснюючі змінні відповідними розрахунковими значеннями.

Оскільки функція ЛИНЕЙН вимагає, щоб в таблиці стовпці аргументів розташовувались поруч, після стовпця XR вставлені стовпці XLAG, t, WRS=WR+WP і тільки після цього слідують стовпці PR,PLAG, KLAG. Тепер, задаючи в різних комбінаціях сусідні стовпці змінних, можна функцією ЛИНЕЙН оцінювати кожне рівняння моделі в структурній формі.

Для порівняння результатів розрахунку зі звичайною процедурою 1МНК нижче складена таблиця, аналогічна попередній з вихідними (не розрахунковими) значеннями ендогенних змінних.

Результати другого кроку 2МНК процедури представлені нижче.

C=a0+a1*P+a2*PLAG+a3*WS+u

	PLAG	P	WS	Const
1МНК	0,09	0,193	0,796	16,24
2МНК	0,216	0,017	0,81	16,55

I=b0+b1*P+b2*PLAG+b3*KLAG+v

	KLAG	PLAG	P	Const
1МНК	-0,11	0,333	0,48	10,13
2МНК	-0,16	0,616	0,15	20,28

W=c0+c1*X+C2*XLAG+C3*t+w

	t	XLAG	X	Const
1МНК	0,13	0,146	0,439	1,497
2МНК	0,13	0,147	0,439	1,5

Блоки розрахунків по обох методиках (1МНК та 2МНК) розташовані поруч, звідки виписуємо результат в таблицю 6.5.

Таблиця 6.5

1МНК	2МНК
C=16,24+0,193*P+0,090*PLAG+0,796*WS	C=16,55+0,017*P+0,216*PLAG+0,810*WS
I=10,13+0,48*P+0,333*PLAG-0,11*KLAG	I=20,28+0,15*P+0,616*PLAG-0,16*KLAG
W=1,497+0,439*X+0,146*KLAG+0,13*t	W=1,500+0,439*X+0,147*XLAG+0,13*t

На відміну від стандартних, 2МНК-оцінки відповідають сенсу задачі.

Зверніть увагу, як істотно змінились оцінки параметрів перших двох рівнянь, що виявилось наслідком на вигляд цілком незначних поправок в пояснюючих ендогенних змінних (див. графіки значень, що спостерігаються, і розрахункових значень).

Контрольні запитання

1. В яких випадках застосовується двокроковий метод найменших квадратів?
2. Що являють собою наперед визначені змінні?
3. Дайте економічну інтерпретацію кожній рівності економетричної моделі Клейна.
4. Що таке інструментальні змінні і для чого вони застосовуються?
5. Які розрахунки відбуваються на кожному з двох кроків процедури 2МНК?
6. Чим викликана на різниця в оцінках параметрів моделі Клейна за методиками 1МНК та 2МНК?

Практична робота №7

Тема: Гетероскедастичність.

Мета: Ознайомитися з ефектом порушення припущення Гаусса – Маркова про рівноточність спостережень $M(\varepsilon i^2)=\sigma\varepsilon^2$. Оволодіти методом зважених найменших квадратів для отримання стійких оцінок параметрів моделі при порушенні припущення про рівноточність спостережень.

Короткі теоретичні відомості

Зазвичай припускається, що всі спостереження однаково надійні, що ймовірності випадкових похибок однакові для всіх спостережень. Ця властивість називається гомоскедастичністю (в перекладі – «однаковий розкид»). Якщо ж різні спостереження мають різну мінливість, то така властивість називається гетероскедастичністю (в перекладі – «неоднаковий розкид»).

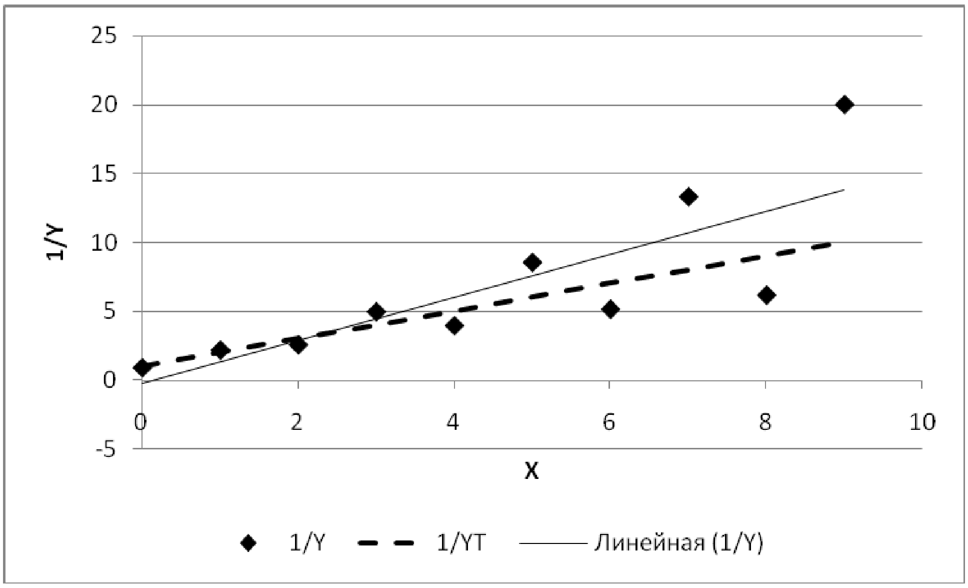
Це припущення не використовувалося при доказі незміщеності МНК-оцінок параметрів моделі, тому вважається, що за відсутності гомоскедастичності невірно розраховуються тільки стандартні похибки коефіцієнтів регресії, статистики Стьюдента та довірчі межі на розрахункові значення. Однак теоретична незміщеність ще не означає, що коефіцієнти регресії оцінюються правильно, так як МНК не робить розмежувань між спостереженнями, надаючи їм однакову вагу, незалежно від того, чи є конкретне спостереження надійним, чи ні.

Іноді появу гетероскедастичності можна передбачити завчасно. Так, при функціональних перетвореннях залежної змінної (результативної ознаки Y) нелінійним чином перетворюється вся система випадкових похибок, що відносяться до різних спостережень, що може призвести до небажаних наслідків.

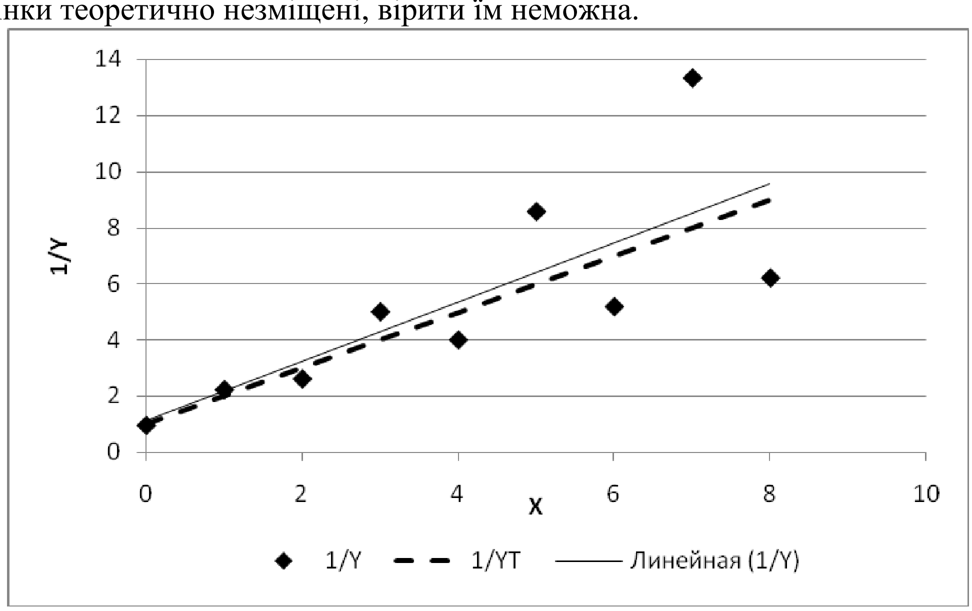
Розглянемо умовний приклад. Нехай результативна ознака породжується нелінійною залежністю YT=1/(1+X), де X=0, 1, 2,…9. Нам відомі значення Y з невеликими похибками e=±0,05, знаки яких чергуються (Y=YT+e). Чи можливо за цими даними методом найменших квадратів відновити (з достатнім ступенем точності) вихідну залежність? Форму зв’язку часто вдається вгадати. Якщо на вихідному графіку емпіричні точки не групуються навколо деякої прямої, то виконують пробні функціональні перетворення змінних (частіше за все застосовується логарифмування або перехід до обернених величин).

В нашому прикладі в координатах (X, 1/Y) емпіричні точки явно групуються навколо прямої, тому приймаємо в цих координатах лінійну модель 1/Y=b0+b1*X, звідки отримуємо правильну форму зв’язку Y=1/(b0+b1*X). Залишилось визначити оцінки коефіцієнтів регресії, які у вихідній залежності дорівнюють одиниці β0=β1=1. У функціональних масштабах (X, 1/Y) графік вихідної залежності зображено пунктиром. З наведеного рисунку ясно видно, що після переходу до обернених величин 1/Y з’явився ефект «збовтування»: розкид точок тепер збільшується зі збільшенням X, причому остання точка навіть схожа на викид. І як всякий викид, ця точка перетягнула до себе всю лінію регресії. Рівняння лінійного тренду стало таким:

1/Yp=–0,23+1,56*X. Отримані оцінки b0=–0,23 і b1=1,56 зовсім не схожі на очікувані одиничні значення β0=β1=1.



Спробуємо видалити останню сумнівну точку (див. новий графік). Виявляється, після вибракування найбільш сумнівної точки роль викиду взяло на себе передостаннє спостереження. Положення лінії регресії помітно змінилось. Результати звичайного МНК виявились нестійкими – при послідовних вибракуваннях лінія регресії «виляє хвостом». І хоча МНК-оцінки теоретично незміщені, вірити їм неможна.



Ідея методу зважених найменших квадратів полягає у введенні вагової функції, яка надає велику вагу надійним спостереженням і меншу – малонадійним.

В нашому випадку вагова функція знаходиться з наступних міркувань. В нелінійній моделі Y=1/(b0+b1*X)+u всі спостереження були рівноточними. Для можливості оцінки параметрів моделі стандартною процедурою МНК вимагається, щоб модель була лінійною відносно параметрів; тому переходимо до обернених величин 1/Y=b0+b1*X+e. Зазначимо, що тепер випадковий член (е) відноситься не до вихідної змінної Y, а до перетвореної величини 1/Y. Однак, згідно з основної передумови регресійного аналізу, випадковий член має відноситися саме до вихідної залежної змінної Y і ні до якої іншої. Питання: чи можна перетворити модель так, щоб вона була лінійною відносно параметрів і в той же час була вирішена відносно Y? Для цього помножимо обидві частини перетвореного рівняння на Y2, в результаті отримаємо: $Y=b_0*Y^2+b_1*X*Y^2+u$. В цій формі моделі задоволені одразу обидві умови – модель вирішена відносно Y і лінійна відносно параметрів b_0,b_1. Оцінюємо параметри цієї моделі за всіма 10-ма спостереженнями $Y=0{,}963*Y^2+0{,}868*X*Y^2$ та за 9-ма спостереженнями (без сумнівної останньої точки) $Y=0{,}963*Y^2+0{,}866*X*Y^2$. Як бачимо, тепер результати виявились стійкими до вибракування декількох сумнівних спостережень. Отримані оцінки коефіцієнтів регресії дуже близькі до очікуваних значень вихідної залежності $\beta0=\beta1=1$.

Завдання

Дано змінну X, теоретичне значення Yт та похибку e (табл. 7.1). Знайти оцінки параметрів моделі залежності 1/Y від X звичайним та зваженим методами найменших квадратів.

Таблиця 7.1*

№	X	YT	e
1	0	1	N
2	1	0,5	-N
3	2	0,333333	N
4	3	0,25	-N
5	4	0,2	N
6	5	0,166667	-N
7	6	0,142857	N
8	7	0,125	-N
9	8	0,111111	N
10	9	0,1	-N

*Примітка: варіант формується за допомогою Таблиці 7.1, в якій N – розмір похибки за варіантом. Варіант призначається викладачем.

Варіант 1. N=0,03
Варіант 2. N=0,04
Варіант 3. N=0,06
Варіант 4. N=0,07

Порядок виконання

1. За своїм варіантом сформувати вихідну базу даних та побудувати таблицю статистичних показників.
2. Обчислити параметри моделі залежності вихідних змінних звичайним МНК.
3. Ввести в модель ваговий коефіцієнт Y^2 та провести розрахунок методом зважених найменших квадратів..
4. Прибрати з таблиці останній рядок спостережень та повторити розрахунки п.3 за дев'ятьма рядками даних. Порівняти отримані результати.

Приклад виконання

1. Нехай база даних має вигляд, представлений у таблиці 7.2

Таблиця 7.2

№	X	YT	e
1	0	1	0,05
2	1	0,5	-0,05
3	2	0,333333	0,05
4	3	0,25	-0,05
5	4	0,2	0,05
6	5	0,166667	-0,05
7	6	0,142857	0,05
8	7	0,125	-0,05
9	8	0,111111	0,05
10	9	0,1	-0,05

2. Сформуємо масив даних для розрахунку моделі звичайним методом найменших квадратів. Результат наведено в таблиці 7.3.

Таблиця 7.3

№	Y	1/Y	1/YT
1	1,05	0,952381	1
2	0,45	2,222222	2
3	0,383333	2,608696	3
4	0,2	5	4
5	0,25	4	5
6	0,116667	8,571429	6
7	0,192857	5,185185	7
8	0,075	13,33333	8
9	0,161111	6,206897	9
10	0,05	20	10

В третьому стовбці таблиці розрахована величина, обернена до Y, яка буде використовуватись в моделі в якості залежної змінної. Для порівняння в четвертому стовбці розрахований аналогічний показник без впливу похибки.

За даними X, 1/Y функцією ЛИНЕЙН отримані незадовільні оцінки параметрів і статистичні характеристики (перший блок виводу), які вже обговорені вище.

ЛИНЕЙН	b	a
	1,564069	**-0,2303**
	0,398858	2,129321
	0,657785	3,62281
	15,37711	8
	201,8209	104,998
Стьюдент	3,921366	-0,10816

3. Вводимо в модель ваговий коефіцієнт Y^2. Вихідна база набуває вигляду, представленого у перших двох стовбцях табл. 7.4.

Таблиця 7.4

YY	YYX	1/Yp(10)	1/Yp(9)
1,1025	0	0,963138	0,963282
0,2025	0,2025	1,830857	1,828863
0,146944	0,293889	2,698576	2,694444
0,04	0,12	3,566295	3,560024
0,0625	0,25	4,434014	4,425605
0,013611	0,068056	5,301734	5,291186
0,037194	0,223163	6,169453	6,156766
0,005625	0,039375	7,037172	7,022347
0,025957	0,207654	7,904891	7,887927
0,0025	0,0225	8,77261	

На основі перших двох стовбців цієї таблиці, а також стовбця Y функцією ЛИНЕЙН розраховуємо параметри моделі $Y=b_0*Y^2+b_1*X*Y^2$:

b	a
0,867719	**0,963138**
0,087484	0,042529
0,989065	0,047337
361,7994	8
1,621405	0,017926

9,918658	22,64688

В останньому рядку наведені значення критеріїв Стьюдента для параметрів моделі.

4. Повторюємо розрахунки функцією ЛИНЕЙН, не беручи до уваги останній рядок таблиці. Отримаємо наступний результат:

b	a
0,865581	**0,963282**
0,09152	0,044454
0,98953	0,049478
330,8038	7
1,619694	0,017137
9,457881	21,6694

Як бачимо, значення параметрів майже не змінюються при зменшенні кількості спостережень, що свідчить про відсутність явища гетероскедастичності.

Розрахункові значення 1/Y за десятьма та дев'ятьма рядками вихідних даних представлено в останніх двох стовбцях таблиці 7.3.

Контрольні запитання

1. Що являє собою властивість гетероскедастичності та в чому вона проявляється?
2. Як поводить себе графік тренду залежності при зменшенні кількості спостережень за наявності гомо- та гетероскедастичності?
3. В який спосіб можна позбавитись гетероскедастичності?
4. В чому полягає ідея методу зважених найменших квадратів?
5. В який спосіб обернену залежність можна привести до вигляду лінійної?

Практична робота №8

Тема: Моделі з лаговими змінними.

Мета: Скласти модель з розподіленими лагами для опису витрат на житло в залежності від рівня доходів і відносних цін поточного і декількох попередніх періодів; перетворити цю модель в авторегресійну методом Койка, оцінити її параметри і зробити висновки відносно коротко та довгострокового впливу пояснюючих змінних.

Короткі теоретичні відомості

Моделі, що враховують лагові ендогенні змінні, прийнято називати авторегресійними; якщо ж вони включають тільки лагові екзогенні змінні, їх іменують як моделі з розподіленими лагами. Цей вид моделей застосовується для опису економічних процесів, що мають довгостроковий характер.

Розглянемо авторегресійну модель для опису витрат на житло Y в залежності від доходу X та відносної ціни P. До цієї моделі призвели наступні міркування.

Очевидно, що попит на деякий товар визначається не тільки поточним доходом, але також доходом попередніх періодів. За даними витрат на житло в США (дані наведено в додатку) можна скласти наступні моделі:

a) ln Yp=-1,51+1,18*lnX-0,35*lnP; R^2=0,991
(tb) (0,8 (21,0) (1,1)

a1) ln Yp=0,86+1,09*lnXLAG-0,73*lnPLAG R^2=0,994
(tb) (0,5) (19,4) (2,2)

a2) ln Yp=2,31+1,02*lnXLAG2-0,95*lnPLAG2 R^2=0,996
(tb) (1,4) (20,6) (3,3)

Тут LAG і LAG2 означає зсув назад на один і два періоди.

Зазначимо, що між рівняннями a, a1, a2 практично немає особливих відмінностей; значення еластичностей затрат за доходом майже однакові; всі рівняння значимі в цілому і мають майже однакові коефіцієнти детермінації. Немає ніяких підстав для надання переваги одній з цих моделей. Це не випадково, адже лагові змінні зазвичай тісно корелюють: R(X,XLAG)=0,997; R(P,PLAG)=0,951. Тому коефіцієнти регресії в будь-якому з цих рівнянь фактично відбивають сумарні ефекти за декілька часових періодів.

Правильна модель має враховувати ефекти кожного періоду окремо, таким чином, ми приходимо до моделі з розподіленими лагами. Побудуємо послідовно моделі з врахуванням екзогенних лагових змінних за попередні періоди:

a) ln Yp=-1,51+1,18*lnX-0,35*lnP;
(tb) (0,8) (21,0) (1,1)

b) ln Yp=0,17+0,22*lnX+1,00*lnP+0,88*lnXLAG-1,60*lnPLAG
(tb) (0,1) (0,8) (2,8) (2,9) (4,0)

c)lnYp=2,14+0,31*lnX-0,16*lnP+0,18*lnXLAG-0,65*lnPLAG+0,55*lnXLAG2-1,44*lnPLAG2
(tb) (1,2) (1,2) (0,3) (0,4) (0,7) (1,6) (2,5)

Всі три моделі (a, b, c) значимі в цілому і мають приблизно однаковий коефіцієнт детермінації R^2=0,991 ÷ 0,997. Однак зі збільшенням числа членів коефіцієнти регресії стають все менш значимими і все більш нестабільними. Цей неприємний ефект пояснюється мультиколінеарністю – тісними кореляційними зв'язками між лаговими змінними.

Для того щоб скоротити число пояснюючих змінних і тим самим позбавитися (або хоча б зменшити) ефекту мультиколінеарності, Л. Койк запропонував прийняти гіпотезу про експоненціальне зменшення коефіцієнтів регресії при послідовних лагових змінних – коефіцієнти регресії (ваги) мають зменшуватися в геометричній прогресії.

Запишемо модель

lnY=a+(b*lnX+c*lnP)+(b1*lnXLAG+c1*lnPLAG)+(b2*lnXLAG2+c2*lnPLAG2)+…+e

для попереднього періоду

lnYLAG=a+(b*lnXLAG+c*lnPLAG)+(b1*lnXLAG2+c1*lnPLAG2)+…+eLAG

і врахуємо, що послідовні ваги зменшуються в геометричній прогресії:
b1=b*q; b2=b1*q; b3=b2*q; …
c1=c*q; c2=c1*q; c3=c2*q; …
Помножимо другу рівність на q і віднімемо його від першого.
В результаті отримаємо економічну авторегресійну модель:

$$lnY=a1+b*lnX+c*lnP+q*lnYLAG+u \qquad (8.1)$$

де a1=a*(1-q); u=e-q*eLAG (u – новий випадковий член).
Розраховуємо параметри цієї авторегресійної моделі:
lnYp=0,50+0,15*lnX-0,16*lnP+0,845*lnYLAG; R2=0,9996
(tb) (1,3) (3,1) (2,4) (22,4)

Тепер ми можемо оцінити як коротко так і довгострокові ефекти. В короткостроковому аспекті (для поточного періоду) значення YLAG необхідно розглядати як фіксоване, тоді еластичності витрат за доходом і ціною будуть дорівнювати коефіцієнтам регресії: b=0,15 і c=-0,16. В довгостроковому аспекті, коли X, P, Y приймаються рівними своїм рівноважним значенням (тоді YLAG=Y), з'ясовується, що довгостроковий вплив X на Y дорівнює величині b/(1-q)=0,15/(1-0,845)=0,96, а довгостроковий вплив P на Y дорівнює c/(1-q)=-0,16/(1-0,845)=-1,02. Ці цифри близькі до значимих коефіцієнтів регресії a, a1, a2.

Завдання

За даними витрат на житлові послуги (Y, млрд. дол. США), рівня особистого доходу у розпорядженні (X, млрд. дол.. США) та індексу реальних цін на житло в США (P, для 1972 р. P=100) за 24 роки поданими в таблиці 8.1, побудувати п'ять типів моделей з лаговими змінними (a, a1, a2, b, c) та авторегресійну модель Койка (8.1).

Таблиця 8.1*

Рік	Y	X	P
1960	64+N	489,7+N	104,5
1961	67+N	503,8+N	105,1
1962	70,7+N	524,9+N	105
1963	74+N	542,3+N	104,8
1964	77,4+N	580,8+N	104,5
1965	81,6+N	616,3+N	104
1966	85,3+N	646,8+N	102,6
1967	89,1+N	673,5+N	102,2
1968	93,5+N	701,3+N	100,9
1969	98,4+N	722,5+N	100
1970	102+N	751,6+N	99,6
1971	106,4+N	779,2+N	100
1972	112,5+N	810,3+N	100
1973	118,2+N	865,3+N	99,1
1974	124,2+N	858,4+N	95,1
1975	128,3+N	875,8+N	93,3
1976	134,9+N	906,8+N	93,7
1977	141,3+N	942,9+N	94,5
1978	148,5+N	988,8+N	94,7
1979	154,8+N	1015,5+N	93,8
1980	159,8+N	1021,6+N	93
1981	164,8+N	1049,3+N	94,2
1982	167,5+N	1058,3+N	96,7
1983	171,3+N	1095,4+N	99,2

*Примітка: варіант формується за допомогою Таблиці 8.1, в якій N – це дві останні цифри номера Вашої залікової книжки.

Порядок виконання

1. За своїм варіантом сформувати вихідну базу даних та побудувати таблицю статистичних показників.

2. Розрахувати значення логарифмів ln Y, ln X, ln P.

3. Сформувати стовбці з лаговими змінними з запізненням на 1 та 2 роки: ln YLAG, ln XLAG, ln PLAG, ln XLAG2, ln PLAG2.

4. Отримати оцінки параметрів моделей a, a1, a2. Розрахувати коефіцієнти детермінації.

5. Отримати оцінки параметрів моделей b та c.

6. Розрахувати параметри авторегресійної моделі (8.1). Обчислити для неї коефіцієнт детермінації.

Приклад виконання

1. Нехай база статистичних даних, що характеризує зміну витрат на житлові послуги в залежності від особистого доходу у розпорядженні та індексу цін, представлена в таблиці 8.2

Таблиця 8.2

Рік	Y	X	P
1960	64	489,7	104,5
1961	67	503,8	105,1
1962	70,7	524,9	105
1963	74	542,3	104,8
1964	77,4	580,8	104,5
1965	81,6	616,3	104
1966	85,3	646,8	102,6
1967	89,1	673,5	102,2
1968	93,5	701,3	100,9
1969	98,4	722,5	100
1970	102	751,6	99,6
1971	106,4	779,2	100
1972	112,5	810,3	100
1973	118,2	865,3	99,1
1974	124,2	858,4	95,1
1975	128,3	875,8	93,3
1976	134,9	906,8	93,7
1977	141,3	942,9	94,5
1978	148,5	988,8	94,7
1979	154,8	1015,5	93,8
1980	159,8	1021,6	93
1981	164,8	1049,3	94,2
1982	167,5	1058,3	96,7
1983	171,3	1095,4	99,2

2. За допомогою функції Excel LN розраховуємо значення логарифмів всіх трьох змінних. Результат заносимо до таблиці 8.3.

Таблиця 8.3

Рік	lnY	lnX	lnP
1960	4,15888	6,19379	4,64919
1961	4,20469	6,22218	4,65491
1962	4,25845	6,26321	4,65396
1963	4,30407	6,29582	4,65205
1964	4,34899	6,36441	4,64919

1965	4,40183	6,42373	4,64439
1966	4,44617	6,47204	4,63084
1967	4,48976	6,51249	4,62693
1968	4,53796	6,55294	4,61413
1969	4,58904	6,58272	4,60517
1970	4,62497	6,6222	4,60116
1971	4,66721	6,65827	4,60517
1972	4,72295	6,6974	4,60517
1973	4,77238	6,76308	4,59613
1974	4,82189	6,75507	4,55493
1975	4,85437	6,77514	4,53582
1976	4,90453	6,80992	4,5401
1977	4,95089	6,84896	4,5486
1978	5,00058	6,89649	4,55071
1979	5,04213	6,92314	4,54116
1980	5,07392	6,92913	4,5326
1981	5,10473	6,95588	4,54542
1982	5,12098	6,96442	4,57161
1983	5,14342	6,99887	4,59714

3. Зсуваючи значення логарифмів, отримані в п.2, на один чи два роки вперед, отримаємо значення лагових змінних, наведених в таблиці 8.4.

Таблиця 8.4

lnYLAG	lnXLAG	lnPLAG	lnXLAG2	lnPLAG2
4,1092	6,1732	4,6492	6,1327	4,6501
4,15888	6,19379	4,64919	6,1732	4,6492
4,20469	6,22218	4,65491	6,19379	4,64919
4,25845	6,26321	4,65396	6,22218	4,65491
4,30407	6,29582	4,65205	6,26321	4,65396
4,34899	6,36441	4,64919	6,29582	4,65205
4,40183	6,42373	4,64439	6,36441	4,64919
4,44617	6,47204	4,63084	6,42373	4,64439
4,48976	6,51249	4,62693	6,47204	4,63084
4,53796	6,55294	4,61413	6,51249	4,62693
4,58904	6,58272	4,60517	6,55294	4,61413
4,62497	6,6222	4,60116	6,58272	4,60517
4,66721	6,65827	4,60517	6,6222	4,60116
4,72295	6,6974	4,60517	6,65827	4,60517
4,77238	6,76308	4,59613	6,6974	4,60517
4,82189	6,75507	4,55493	6,76308	4,59613
4,85437	6,77514	4,53582	6,75507	4,55493
4,90453	6,80992	4,5401	6,77514	4,53582
4,95089	6,84896	4,5486	6,80992	4,5401
5,00058	6,89649	4,55071	6,84896	4,5486
5,04213	6,92314	4,54116	6,89649	4,55071
5,07392	6,92913	4,5326	6,92314	4,54116
5,10473	6,95588	4,54542	6,92913	4,5326
5,12098	6,96442	4,57161	6,95588	4,54542

4. За допомогою функції ЛИНЕЙН розраховуємо параметри моделей a, a1 та a2. Коефіцієнт детермінації знаходиться в діапазоні виводу функції ЛИНЕЙН у третьому рядку ліворуч. Нижче діапазону виводу розраховані значення критеріїв Стьюдента для параметрів моделі.

lnP	lnX	Const
-0,352	**1,17632**	**-1,5095**
0,32851	0,05608	1,84876
0,99085	0,03156	#Н/Д
1137,62	21	#Н/Д
2,26637	0,02092	#Н/Д
tb		
-1,0715	20,9747	-0,8165

lnPLAG	lnXLAG	Const
-0,7291	**1,08689**	**0,85706**
0,32632	0,05613	1,8507
0,99389	0,02579	#Н/Д
1708,52	21	#Н/Д
2,27332	0,01397	#Н/Д
tb		
-2,2344	19,3636	0,4631

lnPLAG2	lnXLAG2	Const
-0,9466	**1,02453**	**2,30796**
0,28992	0,04962	1,64389
0,9955	0,02214	#Н/Д
2321,71	21	#Н/Д
2,27699	0,0103	#Н/Д
tb		
-3,2651	20,6475	1,40397

Як видно з діапазону виводу, коефіцієнти детермінації складають відповідно: для моделі a – 0,99085; для моделі a1 – 0,99389; для моделі a2 – 0,9955.

Маємо

a) ln Yp=-1,51+1,18*lnX-0,35*lnP; R^2=0,991

(tb) (0,8 (21,0) (1,1)

a1) ln Yp=0,86+1,09*lnXLAG-0,73*lnPLAG R^2=0,994

(tb) (0,5) (19,4) (2,2)

a2) ln Yp=2,31+1,02*lnXLAG2-0,95*lnPLAG2 R^2=0,996

(tb) (1,4) (20,6) (3,3)

5. Далі за допомогою функції ЛИНЕЙН знаходимо значення параметрів моделей b та c:

lnPLAG	lnXLAG	lnP	lnX	Const
-1,6041	**0,8816**	**0,99998**	**0,22072**	**0,17474**
0,40004	0,29962	0,35929	0,29277	1,62757
0,99604	0,02182	#Н/Д	#Н/Д	#Н/Д
1196,16	19	#Н/Д	#Н/Д	#Н/Д
2,27824	0,00905	#Н/Д	#Н/Д	#Н/Д
tb				
-4,0098	2,94235	2,78319	0,75391	0,10736

lnPLAG2	lnXLAG2	lnPLAG	lnXLAG	lnP	lnX	Const
-1,4369	**0,55322**	**0,65013**	**0,17794**	**-0,1575**	**0,31397**	**2,13496**
0,57262	0,34191	0,95747	0,46653	0,55917	0,2662	1,82551
0,99714	0,01961	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
988,377	17	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д
2,28075	0,00654	#Н/Д	#Н/Д	#Н/Д	#Н/Д	#Н/Д

tb

-2,5093	1,61802	0,67901	0,38141	-0,2816	1,17945	1,16952

Як бачимо, значення критеріїв Стьюдента для більшості параметрів є незадовільними. Це відображає вплив мультиколінеарності між вихідними та відповідними лаговими змінними.

Отримаємо такі моделі:

b) ln Yp=0,17+0,22*lnX+1,00*lnP+0,88*lnXLAG-1,60*lnPLAG

(tb) (0,1) (0,8) (2,8) (2,9) (4,0)

c)lnYp=2,14+0,31*lnX-0,16*lnP+0,18*lnXLAG-0,65*lnPLAG+0,55*lnXLAG2-1,44*lnPLAG2

(tb) (1,2) (1,2) (0,3) (0,4) (0,7) (1,6) (2,5)

6. Наприкінці розраховуємо параметри авторегресійної моделі (8.1).

c	b	q	a1
-0,1579	0,14847	0,84513	0,50189
0,06657	0,04732	0,03779	0,3822
0,99965	0,00634	#Н/Д	#Н/Д
18952,3	20	#Н/Д	#Н/Д
2,28648	0,0008	#Н/Д	#Н/Д
tb			
-2,3721	3,13757	22,3643	1,31316

З діапазону виводу функції ЛИНЕЙН дістаємо коефіцієнт детермінації 0,99965

lnYp=0,50+0,15*lnX-0,16*lnP+0,845*lnYLAG; R2=0,9996

(tb) (1,3) (3,1) (2,4) (22,4)

Контрольні запитання

1. Дайте визначення поняття «лагові змінні».
2. Що представляє собою авторегресійна модель? В чому її відмінність від звичайної моделі з розподіленими лагами?
3. Чим викликаний ефект мультиколінеарності в моделях з розподіленими лагами (b та c)?
4. З чого складається діапазон виводу функції ЛИНЕЙН? Де в ньому знаходиться коефіцієнт детермінації?
5. Яким чином та з якою метою розраховуються критерії Стьюдента для параметрів моделі?

Рекомендована література

Базова

1. Пономаренко В.С., Малярець Л.М. Багатовимірний аналіз соціально-економічних систем : навчальний посібник. Харків : Вид. ХНЕУ, 2019. 384 с.

2. Григорків В.С. Моделювання економіки: підручник. Чернівці: Чернівецький нац. ун-т ім. Ю. Федьковича, 2019. 360 с.

3. Корхін А.С., Турчанінова І.Ю. Моделювання економіки: навч. пос. Д.: ДВНЗ «НГУ», 2016. 104 с.

4. Наконечний С.І., Терещенко Т.О., Романюк Т.П. Економетрія: Підручник Видання 2-ге, доповнене та перероблене. К.: КНЕУ, 2000. 296с.

Допоміжна

1. Бакурова А.В., Діденко А.В., Пшегорлинська О.А. Комплексний підхід до управління життєздатністю підприємств на основі гармонійного аналізу та системно-динамічного моделювання. Інтелект XXI, 2018, № 3, С. 31-35.

2. Вітлінський В. В. Моделювання економіки: Навч. посібник. К.: КНЕУ, 2007. 408 с.

3. Корольов О.А., Рязанцева В.В. Практикум з економетрії: завдання з практичними рекомендаціями, алгоритмами та прикладом їх наскрізного виконання. Ч.1. Регресійний аналіз: Навчальний посібник. К.: Видавництво Європейського університету, 2002. 250с.

4. Моделі, методи і засоби управління соціально-економічними об'єктами : монографія / О. О. Арсірій, Т. Л. Будорацька, М. Г. Глава [та ін.]. Одеса : ОНПУ, 2016. 225 с.

5. Applicable Analysis and Discrete Mathematics. – Belgrade. – ISSN 1452 -8630 Volume 13, Number 2, October 2019

6. Shemayev V. Cognitive approach to modeling reflexive control in socio-economic systems. Information and Security, 2007. № 22. P. 28-37. URL : http://it4sec.org/system/files/22.04_Shemayev.pdf

7. Wei Shouke, et al. System dynamics simulation model for assessing socio-economic impacts of different levels of environmental flow allocation in the Weihe River Basin, China. European Journal of Operational Research, 2012. 221.1. P. 248-262. URL: https://www.sciencedirect.com/science/article/abs/pii/S037722171200207X

8. Weidlich W. Stability and Cyclicityin Social Systems / W. Weidlich // Behavioral Science. 1988. 33. P. 241-256.

Для нотаток

………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………
………………………………………………………………………………………………………