Міністерство освіти і науки України

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ
УНІВЕРСИТЕТ
БУДІВНИЦТВА ТА АРХІТЕКТУРИ

О.О. Аршава

Методичні вказівки до практичних занять

з дисципліни «Економіко-математичні методи та моделі (економетрика)» для студентів спеціальності 051 «Економіка»

Харків 2018

Міністерство освіти і науки України

**ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
БУДІВНИЦТВА ТА АРХІТЕКТУРИ**

Спеціальність 051

О.О. Аршава

Методичні вказівки до практичних занять з дисципліни «Економіко-математичні методи та моделі (економетрика)» для студентів спеціальності 051 «Економіка»

Затверджено на засіданні кафедри
вищої математики
Протокол № 1 від 30.08.2018 р.

Харків 2018

Методичні вказівки до практичних занять з курсу «Економіко-математичні методи та моделі (економетрика)» для студентів спеціальності 051 «Економіка» / Укладач О. О. Аршава. – Харків: ХНУБА, 2018. – 51 с.

Рецензент А.П. Харченко

Кафедра вищої математики

ВСТУП

Основна мета методичних вказівок — надати допомогу студентам у вивченні теоретичного матеріалу з дисципліни «Економіко-математичні методи та моделі (економетрика)», оволодіти навичками та прийомами розв'язування задач та прикладів.

Представлений матеріал відповідає змісту та послідовності викладання його на лекційних заняттях. Методичні вказівки містять: програму кожного модуля, навчальний матеріал для проведення практичних занять, питання для самоперевірки, зміст теоретичного матеріалу, що потрібно вивчити перед кожним практичним заняттям, бібліографічний список.

Методичні вказівки доцільно використовувати як для роботи в аудиторії, так і для індивідуальної самостійної роботи.

ПРОГРАМА МОДУЛЯ №1
Багатофакторна класична лінійна модель. Метод МНК. Мультиколінеарність.

Тема 1. Парна лінійна економетрична модель.
Тема 2. Множинна лінійна регресія.
Тема 3. Мультиколінерність.

ПРАКТИЧНЕ ЗАНЯТТЯ 1

Парна лінійна регресія

Для виконання завдання студент повинен знати:
1 Мету і зміст запропонованого завдання, порядок її виконання.
2 Означення функціональної статистичної та кореляційної залежності.
3 Формули оцінок параметрів лінійної парної регресії.
4 Формули надійних інтервалів параметрів узагальненої регресійної моделі і надійних інтервалів прогнозу.
5 F - критерій перевірки гіпотези адекватності прийнятої моделі статистичним даним.
6 Формули оцінки коефіцієнта кореляції та детермінації.
7 Формулу оцінки коефіцієнта еластичності.

Студент повинен вміти користуватися пакетом *Excel*:
1 Будувати таблиці.
2 Складати і копіювати формули.
3 Користуватися статистичними і математичними функціями.
4 Будувати графіки і вміти їх друкувати

5 Перевіряти значущість коефіцієнта кореляції.

Студент повинен підготувати алгоритм розв'язування даного завдання з використанням електронних таблиць Excel.

Завдання

1 На основі статистичних даних ознаки y_i і фактора x_i знайти:

а) оцінки параметрів лінії регресії $\hat{y} = b_0 + b_1 x$;

б) оцінку коефіцієнта кореляції;

в) оцінку коефіцієнта детермінації.

2 Використовуючи критерій Фішера з надійністю 0,95, оцінити адекватність прийнятої економічної моделі статистичним даним.

3 Використовуючи t - статистику з надійністю 0,95, оцінити значущість оцінки коефіцієнта кореляції та параметрів лінії регресії.

4 Якщо модель адекватна статистичним даним, то знайти:

а) надійні інтервали параметрів узагальненої регресійної моделі з вірогідністю 0,95;

б) прогноз показника та його надійні інтервали;

в) коефіцієнт еластичності для прогнозу.

5 Побудувати графіки статистичних даних, лінії регресії та її надійної зони, коефіцієнта еластичності.

6 На основі одержаної економетричної моделі зробити висновки.

Числовий приклад побудови та дослідження простої економетричної моделі

Приклад. На основі статистичних даних по дев'яти металобазах (таблиця 1.1) побудувати економетричну модель, яка характеризує залежність між витратами обігу, грн., та вантажообігом, грн.

Таблиця 1.1

Таблиця вихідних даних

№з/п	Витрати обігу	Вантажообіг
1	2,7	15,6
2	3	15,3
3	2,8	14,9
4	2,9	15,1
5	2,6	16,1
6	2,5	16,7
7	2,8	15,4
8	2,6	17,1
9	2,5	16,8

У таблиці 1.2 наведено вихідні дані та їх елементарні перетворення для побудови моделі.

Таблиця 1.2

№ з/п	Y	X	X^2	XY	$Y-\bar{Y}$	$X-\bar{X}$	$(Y-\bar{Y})^2$	$(X-\bar{X})^2$	$(X-\bar{X})(Y-\bar{Y})$
1	2,7	15,6	243,36	42,12	-0,011	-0,289	0,0001	0,0835	0,0032
2	3	15,3	234,09	45,9	0,289	-0,589	0,0835	0,3468	-0,1701
3	2,8	14,9	222,01	41,72	0,089	-0,989	0,0079	0,9779	-0,0879
4	2,9	15,1	228,01	43,79	0,189	-0,789	0,0357	0,6223	-0,1490
5	2,6	16,1	259,21	41,86	-0,111	0,211	0,0123	0,0446	-0,0235
6	2,5	16,7	278,89	41,75	-0,211	0,811	0,0446	0,6579	-0,1712
7	2,8	15,4	237,16	43,12	0,089	-0,489	0,0079	0,2390	-0,0435
8	2,6	17,1	292,41	44,46	-0,111	1,211	0,0123	1,4668	-0,1346
9	2,5	16,8	282,24	42	-0,211	0,911	0,0446	0,8301	-0,1923
Сума	24,4	143	2277,38	386,72			0,2489	5,2689	-0,9689
Середнє	2,711111	15,88889	253,042	42,96889					

Розв'язання.

1 Обчислимо середні значення початкових даних, використовуючи таблицю 1.2:

$\bar{X}$=(15,5+15,3+14,9+15,1+16,1+16,7+15,4+17,1+16,8)/9=15,889

$\bar{Y}$=(2,7+3+2,8+2,9+2,6+2,5+2,8+2,6+2,5)/9=2,711

$\overline{XY}$=(42,12+45,9+41,72+43,79+41,86+41,75+43,12+44,46+42)/9=42,969

$\overline{X^2}$=(243,36+234,09+222,01+228,01+259,21+……+292,41+282,24)/9=253,042

Ідентифікуємо змінні:

Y - витрати (залежна змінна), X - вантажообіг (незалежна змінна).

Специфікуємо економетричну модель у лінійній формі: $Y = b_0 + b_1 X + \varepsilon$, де b_0, b_1 – параметри моделі; ε – стохастична складова (збурення).

Оцінимо параметри моделі $\hat{Y} = \hat{b}_0 + \hat{b}_1 X$ методом МНК за формулами:

$$\hat{b}_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}, \qquad \hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x} \cdot$$

Одержимо:

$$\hat{b}_1 = \frac{42,969 - 2,711 \cdot 15,889}{253,042 - 15,889^2} = -0,184, \qquad \hat{b}_0 = 2,711 - (-0,184) \cdot 15,889 = 5,633.$$

2 Обчислимо дисперсії залежної змінної та збурення.

Таблиця 1.3

Дисперсія залежної змінної та збурення

$\hat{Y}$	$e=Y-\hat{Y}$	e^2	$(Y-\hat{Y})^2$
2,7642	-0,06	0,004	0,004
2,8194	0,18	0,033	0,033
2,8930	-0,09	0,009	0,009
2,8562	0,04	0,002	0,002
2,6723	-0,07	0,005	0,005
2,5620	-0,06	0,004	0,004
2,8010	0,00	0,000	0,000
2,4884	0,11	0,012	0,012
2,5436	-0,04	0,002	0,002
Сума	$-18 \cdot 10^{-16}$	0,071	0,071

Для оцінки дисперсії збурення використаємо формулу $\hat{\sigma}^2 = \dfrac{\sum e_i^2}{n-2}$:

$$\hat{\sigma}^2 = \frac{0,071}{9-2} = 0,0101$$

Оцінимо дисперсії для параметрів b_0 та b_1:

$$\hat{\sigma}_{b_0}^2 = \frac{0,0101 \cdot 2277,4}{9 \cdot 5,27} = 0,485, \qquad \hat{\sigma}_{b_1}^2 = \frac{0,0101}{5,27} = 0,00192.$$

3 Перевіримо гіпотези відносно коефіцієнтів лінійного рівняння регресії, використовуючи *t*- статистику.

Для перевірки гіпотези

$$H_0 : \hat{b}_1 = b_1^0,$$
$$H_1 : \hat{b}_1 \neq b_1^0,$$

скористуємось статистикою

$$t = \frac{\hat{b}_1 - b_1^0}{\hat{\sigma}_{\hat{b}_1}}.$$

У нашому випадку перевіримо гіпотезу:

$$\begin{matrix} H_0 : \hat{b}_1 = 0 \\ H_1 : \hat{b}_1 \neq 0 \end{matrix}, \qquad t_{розр} = \left| \frac{\hat{b}_1 - b_1^0}{\hat{\sigma}_{\hat{b}_1}} \right| = 4,199.$$

Критичне значення $t_{кр}$ при рівні значущості $\alpha = 0,05$ з числом ступенів свободи $\upsilon = n - 2 = 7$ буде таким: $t_{кр} = 2,365$.

Бачимо, що $t_{розр} > t_{кр}$, а це означає, що приймається гіпотеза $H_1 : \hat{b}_1 \neq 0$.

4 Побудуємо довірчі інтервальні оцінки коефіцієнтів рівняння регресії, використовуючи формули:

$$\hat{b}_0 - t_{\frac{\alpha}{2},n-2} \cdot \hat{\sigma}_{\hat{b}_0} < b_0 < \hat{b}_0 - t_{\frac{\alpha}{2},n-2} \cdot \hat{\sigma}_{\hat{b}_0},$$
$$\hat{b}_1 - t_{\frac{\alpha}{2},n-2} \cdot \hat{\sigma}_{\hat{b}_1} < b_1 < \hat{b}_1 - t_{\frac{\alpha}{2},n-2} \cdot \hat{\sigma}_{\hat{b}_1},$$

Довірчій інтервал для b_0:

$$\hat{b}_0 - t_{\frac{\alpha}{2},n-2} \cdot \hat{\sigma}_{\hat{b}_0} = 5,633 - 2,365 \cdot 0,697 = 3,986,$$
$$\hat{b}_0 + t_{\frac{\alpha}{2},n-2} \cdot \hat{\sigma}_{\hat{b}_0} = 5,633 + 2,365 \cdot 0,697 = 7,280,$$

або

$$(3,99 < b_0 < 7,28).$$

Довірчий інтервал для b_1:

$$\hat{b}_1 - t_{\frac{\alpha}{2},n-2} \cdot \hat{\sigma}_{\hat{b}_0} = (-0,184) - 2,365 \cdot 0,044 = -0,287,$$
$$\hat{b}_1 + t_{\frac{\alpha}{2},n-2} \cdot \hat{\sigma}_{\hat{b}_0} = (-0,184) + 2,365 \cdot 0,044 = -0,080,$$

або

$$(-0,29 < b_1 < -0,08).$$

5 Виконаємо перевірку моделі на адекватність в цілому.

Спочатку обчислимо коефіцієнт детермінації:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \overline{y})^2} = 1 - \frac{0,071}{0,2489} = 0,716.$$

Далі обчислимо фактичне значення параметра критерію Фішера F і порівняємо його з табличним $F_{\text{табл}}$.

$$F = \frac{R^2}{1-R^2} \cdot \frac{n-2}{1} = \frac{0,716}{1-0,716} \cdot \frac{9-2}{1} = 17,635.$$

Табличне значення параметра Фішера $F_{\text{табл}}$ при рівні значущості $\alpha = 0,05$ та числах ступенів свободи $\upsilon_1 = 1$ та $\upsilon_2 = 9-2 = 7$ буде таким $F_{\text{табл}}$ = 5,59.

Оскільки $F_{\text{табл}} < F$, робимо висновок, що гіпотеза про значущість зв'язку незалежної та залежної змінних приймається і, отже, модель є статистично значущою.

6 Обчислимо прогнозові значення.

Спочатку знайдемо точковий прогноз: $\hat{y}_{n+1} = \hat{b}_0 + \hat{b}_1 x_{n+1}$, де $x_{n+1} = x_{10} = 17,4$.

$$\hat{y}_{10} = 5,633 + (-0,184) \cdot 17,4 = 2,43.$$

Далі знайдемо інтервальний прогноз, використовуючи формулу

$$\left(\hat{y}_{n+1} - \cdot t_{\frac{\alpha}{2},n-2} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{\left(x_{n+1} - \overline{x}\right)^2}{\sum \left(x_i - \overline{x}\right)^2}}; \hat{y}_{n+1} + \cdot t_{\frac{\alpha}{2},n-2} \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{\left(x_{n+1} - \overline{x}\right)^2}{\sum \left(x_i - \overline{x}\right)^2}} \right),$$

$$\left(2,43 - 2,365 \cdot 0,1 \sqrt{1 + \frac{1}{9} + \frac{(15,889 - 17,4)^2}{5,27}}; 2,43 + 2,365 \cdot 0,1 \sqrt{1 + \frac{1}{9} + \frac{(15,889 - 17,4)^2}{5,27}} \right),$$

або

$$(2,138 < y_{10} < 2,729).$$

7 Обчислимо коефіцієнт еластичності:

$$\overline{\text{E}} = b_1 \cdot \frac{\overline{x}}{\overline{y}} = -0,184 \cdot \frac{15,889}{2,711} = -1,078.$$

Теорія Множинна лінійна регресія. [19], с. 14 –26.

Питання для самоперевірки

1 Поясніть сутність методу найменших квадратів для оцінок параметрів парних економетричних моделей.

2 Назвіть припущення методу найменших квадратів.

3 Які причини присутності випадкового збурення у регресійних моделях?

4 Яка модель зветься гетероскедаксичною?

5 Які властивості мають оцінки параметрів моделі?

6 Чому треба перевіряти статистичну значущість оцінок параметрів моделі?

7 Поясніть сутність критерію Стьюдента.

8 Як знаходяться довірчі інтервали для оцінок параметрів?

9 Що таке адекватна модель?
10 Назвіть методи визначення адекватності.
11 Поясніть сутність критерію Фішера.
12 Що таке коефіцієнт еластичності?

ПРАКТИЧНЕ ЗАНЯТТЯ 2-3

Множинна лінійна регресія

Для виконання завдання студент повинен знати:
1 Мету і зміст даної роботи, порядок її проведення.
2 Як знайти коваріаційну матрицю збурень.
3 Формулу обчислення визначників 2-го та 3-го порядків.
4 Матричну форму системи нормальних рівнянь.
5 Матричну формулу оцінки параметрів системи регресії.
6 Формулу коефіцієнтів множинної кореляції.
7 Формулу *t*-статистики значущості параметрів регресії.
8 Формулу прогнозу та його довірчого інтервалу.
9 Частинні коефіцієнти еластичності та їх економічну інтерпретацію.

Студент повинен вміти користуватися пакетом Excel:
1 Знаходити транспоновану обернену матрицю та добуток матриць;
2 Використовуючи матричні операції, знаходити розв'язок системи нормальних рівнянь;
3 Користуватися вбудованими статистичними функціями: "СРЗНАЧ" - середнє значення; зміщене середнє квадратичне відхилення;
4 Використовуючи категорію "Математика", знаходити довірчі інтервали прогнозу;
5 На основі розрахунків робити висновки.

Студент повинен підготувати алгоритм розв'язання задачі з використанням електронних таблиць Excel.

Завдання

Економічний показник Y залежить від двох факторів, на основі статистичних даних потрібно:

1 Використовуючи сервіс: "математика/матриц" знайти оцінки параметрів лінійної регресії. Результат отриманих оцінок перевірити, використовуючи сервіс: "статистика/лінейн".

2 Використовуючи F-критерій з надійністю p=0,9, перевірити статистичну значущість коефіцієнта детермінації.

3 Якщо математична модель із заданою надійністю адекватна експериментальним даним, то, використовуючи t-статистику з надійністю p = 0,95, оцінити значущість параметрів регресії.

4 Знайти значення прогнозу показника для заданих значень факторів, його довірчий інтервал із надійністю p=0,95.

5 Знайти частинні коефіцієнти еластичності для точки прогнозу.

6 На основі отриманих розрахунків зробити економічний аналіз.

Числовий приклад побудови та аналізу багатофакторної лінійної економетричної моделі

Приклад. Оцінити параметри економетричної моделі, що характеризує залежність між витратами, вантажообігом та фондомісткістю.

Вихідні дані в умовних одиницях наведено в таблиці 2.1.

Таблиця 2.1

Таблиця вихідних даних

№з/п	Витрати y	Вантажообіг x_1	Фондомісткість x_2
1	2,72	15,6	106,3
2	3,04	13,5	128,5
3	2,84	15,3	118
4	2,89	14,9	121,2
5	2,58	15,1	120
6	2,64	16,1	118,4
7	2,52	16,7	108,4
8	2,75	15,4	110
9	2,63	17,1	105,9

Розв'язання.

1 Запишемо економетричну модель

$$y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon$$

$$\hat{y} = \hat{b}_0 + \hat{b}_1 x_1 + \hat{b}_2 x_2$$

де y, $\hat{y}$ – відповідно фактичні та розрахункові значення витрат за моделлю, x_1 – вантажообіг, x_2 – фондомісткість, ε – стахостична складова, b_0, b_1, b_2 – оцінки параметрів моделі.

2 Побудуємо оператор оцінювання параметрів моделі за методом МНК, використовуючи формулу (2.6) (див. [19]), де

$$\hat{b} = \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{pmatrix}, \qquad X = \begin{pmatrix} 1 & 15,6 & 106,3 \\ 1 & 13,5 & 128,5 \\ 1 & 15,3 & 118 \\ 1 & 14,9 & 121,2 \\ 1 & 15,1 & 120 \\ 1 & 16,1 & 118,4 \\ 1 & 16,7 & 108,4 \\ 1 & 15,4 & 110 \\ 1 & 17,1 & 105,9 \end{pmatrix}, \qquad Y = \begin{pmatrix} 2,72 \\ 3,04 \\ 2,84 \\ 2,89 \\ 2,58 \\ 2,64 \\ 2,52 \\ 2,75 \\ 2,63 \end{pmatrix},$$

$$X' = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 15,6 & 13,5 & 15,3 & 14,9 & 15,1 & 16,1 & 16,7 & 15,4 & 17,1 \\ 106,3 & 128,5 & 118 & 121,2 & 120 & 118,4 & 108,4 & 110 & 105,9 \end{pmatrix}$$

X' – матриця, транспонована до матриці X.

Матриця X, крім векторів незалежних змінних, містить вектор одиниць. Він дописується в цій матриці тоді, коли економетрична модель має вільний член. Не дописуючи такого вектора одиниць, вільний член можна обчислити, скориставшись рівністю: $\hat{b}_0 = \bar{y} - \hat{b}_1\bar{x}_1 - \hat{b}_2\bar{x}_2$, де $\bar{y}$ – середнє значення залежної змінної.

Згідно з оператором оцінювання (2.6) (див. [19]) знайдемо:

$$(X'X) = \begin{pmatrix} 9 & 139,7 & 1036,7 \\ 139,7 & 2177,4 & 16037,7 \\ 1036,7 & 16037,7 & 119909,3 \end{pmatrix}, \qquad (X'X)^{-1} = \begin{pmatrix} 293,14 & -9,444 & -1,271 \\ -9,444 & 0,335 & 0,037 \\ -1,271 & 0,037 & 0,006 \end{pmatrix},$$

$$(X'Y) = \begin{pmatrix} 24,61 \\ 380,85 \\ 2841,5 \end{pmatrix}, \qquad \hat{b} = (X'X)^{-1}X'Y = \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} = \begin{pmatrix} 5,024 \\ -0,137 \\ -0,001 \end{pmatrix}.$$

3 Тоді економетрична модель має такий вигляд: $\hat{y} = 5,024 - 0,137x_1 - 0,001x_2$.

4 Отже, коли за всіх однакових умов незалежна змінна x_1 (вантажообіг) збільшується (зменшується) на одиницю, то залежна змінна $\hat{y}$ (оцінка витрат) при цьому зменшується (збільшується) на 0,137 одиниць. Якщо за інших незмінних умов незалежна змінна x_2 (фондомісткість) збільшується (зменшується) на одиницю, то залежна змінна $\hat{y}$ (оцінка витрат) при цьому зменшується (збільшується) на 0,001 одиниць.

5 Відповідно до отриманих оцінок параметрів моделі отримаємо обчислені значення $\hat{y}$, які приведені у таблиці 2.2.

Таблиця 2.2

Обчислення значення $\hat{y}$

№з/п	Витрати y	Витрати $\hat{y} = \hat{b}_0 + \hat{b}_1x_1 + \hat{b}_2x_2$	Збурення $e = y - \hat{y}$
1	2,72	2,74	-0,02
2	3,04	2,99	0,05
3	2,84	2,76	0,08
4	2,89	2,81	0,08
5	2,58	2,79	-0,21
6	2,64	2,65	-0,01
7	2,52	2,58	-0,06
8	2,75	2,76	-0,01
9	2,63	2,53	0,1

6 Обчислимо незміщену оцінку дисперсії збурень за формулою $\hat{\sigma}^2 = \dfrac{e'e}{n-p-1}$, де e'- транспонований стовпчик збурень e.

$$e' = \begin{pmatrix} -0,02 & 0,05 & 0,08 & 0,08 & -0,21 & -0,01 & -0,06 & -0,01 & 0,1 \end{pmatrix}$$

$$e'e = 0,07103$$

$$\hat{\sigma}^2 = \frac{e'e}{n-p-1} = \frac{0,07103}{9-2-1} = 0,01184.$$

7 Побудуємо коваріаційну матрицю $\hat{K}_b$ (проведемо її оцінку) за формулою $\hat{K}_b = \hat{\sigma}^2 (X'X)^{-1}$:

$$\hat{K}_b = \hat{\sigma}^2 (X'X)^{-1} = 0,01184 \begin{pmatrix} 293,14 & -9,444 & -1,271 \\ -9,444 & 0,335 & 0,037 \\ -1,271 & 0,037 & 0,006 \end{pmatrix} = \begin{pmatrix} 3,4705 & -0,111806 & -0,015051 \\ -0,111806 & 0,003968 & 0,000436 \\ -0,015051 & 0,000436 & 0,000072 \end{pmatrix}$$

Обчислимо стандартні помилки оцінок параметрів моделі за формулою

$$\hat{\sigma}_{\hat{b}_i} = \sqrt{\hat{\sigma}^2_{\hat{b}_i}} = \hat{\sigma}\sqrt{(X'X)^{-1}_{ii}}:$$

$$\hat{\sigma}_{\hat{b}_0} = \sqrt{\hat{\sigma}^2_{\hat{b}_0}} = \sqrt{3,4705} = 1,8629,$$

$$\hat{\sigma}_{\hat{b}_{01}} = \sqrt{\hat{\sigma}^2_{\hat{b}_1}} = \sqrt{0,003968} = 0,063,$$

$$\hat{\sigma}_{\hat{b}_2} = \sqrt{\hat{\sigma}^2_{\hat{b}_2}} = \sqrt{0,000072} = 0,0085.$$

9 Проведемо перевірку статистичної значущості коефіцієнтів рівняння регресії за формулою $t_{i\text{ф}} = \dfrac{\hat{b}_i}{\hat{\sigma}_{\hat{b}_i}}$:

$$t_{0\text{ф}} = \frac{\hat{b}_0}{\hat{\sigma}_{\hat{b}_0}} = \frac{5,024}{1,8629} = 2,6966$$

$$t_{1\text{ф}} = \frac{\hat{b}_1}{\hat{\sigma}_{\hat{b}_1}} = \frac{-0,137}{0,063} = -2,1751$$

$$t_{2\text{ф}} = \frac{\hat{b}_2}{\hat{\sigma}_{\hat{b}_2}} = \frac{-0,001}{0,0085} = -0,1663$$

10 Для обраних параметрів економетричної моделі (при рівні значущості $\alpha = 0,05$ та $\upsilon = n-p-1 = 9-2-1 = 6$) знайдемо критичне значення параметра розподілу Стьюдента $t_{\text{кр}} = 2,4469$.

Як бачимо, $|t_{0\text{ф}}| > t_{\text{кр}}$, $|t_{1\text{ф}}| < t_{\text{кр}}$ (але близьке до $t_{\text{кр}}$), $|t_{2\text{ф}}| < t_{\text{кр}}$. Отже, на практиці це означає, що оцінка параметра $\hat{b}_2$ не є статистично значущою, отже, його треба відкинути. А це означає, що з моделі треба вилучити фактор x_2, що з економічної точки зору може бути невиправданим. Тоді треба повернутися до початку побудови моделі та перевірити її.

11 Знайдемо довірчий інтервал, який з надійністю $(1-\alpha)$ накриває невідоме значення параметра b_1 рівняння регресії, за формулою

$$\hat{b}_i - t_{\alpha/2, n-p-1} \cdot \hat{\sigma}_{\hat{b}_i} < b_i < \hat{b}_i + t_{\alpha/2, n-p-1} \cdot \hat{\sigma}_{\hat{b}_i}:$$

$$-0,137-2,4469\cdot 0,063<b_1<-0,137+2,4469\cdot 0,063.$$

12 Перевіримо якість рівняння в цілому. Для цього обчислимо коефіцієнт детермінації за формулою (2.15) (див. [19]), виконавши попередні розрахунки в таблиці 2.3.

Таблиця 2.3

Коефіцієнт детермінації

№з/п	y_i	$\widehat{y}_i=\widehat{b}_0+\widehat{b}_1x_1+\widehat{b}_2x_2$	$(y_i-\widehat{y}_i)^2$	$(y_i-\overline{y})^2$
1	2,72	2,74	0,000266	0,000209
2	3,04	2,99	0,002233	0,093364
3	2,84	2,76	0,006252	0,011142
4	2,89	2,81	0,006206	0,024198
5	2,58	2,79	0,042234	0,023853
6	2,64	2,65	0,000116	0,00892
7	2,52	2,58	0,003925	0,045986
8	2,75	2,76	0,000072	0,00242
9	2,63	2,53	0,009728	0,010909
Сума			0,071	0,219

Остаточно маємо:

$$R^2=1-\frac{\sum\left(y_i-\widehat{y}_i\right)^2}{\sum\left(y_i-\overline{y}\right)^2}=1-\frac{0,071}{0,219}=0,675.$$

Обчислимо скорегований коефіцієнт детермінації за формулою

$$\overline{R}^2=1-(1-R^2)\cdot\frac{n-1}{n-p-1}=1-(1-0,675)\cdot\frac{9-1}{9-2-1}=0,567.$$

13 Проведемо аналіз значущості коефіцієнта детермінації. Обчислимо значення параметра Фішера за формулою

$$F_{\text{ф}}=\frac{R^2}{1-R^2}\cdot\frac{n-p-1}{p}=\frac{0,675}{1-0,675}\cdot\frac{9-2-1}{2}=6,242.$$

Для перевірки гіпотези $H_0:\ b_1=b_2=\ldots=b_p=0$ знайдемо за таблицею розподілу Фішера з $\upsilon_1=p=2$, $\upsilon_2=n-p-1=9-2-1=6$ ступенями свободи та рівнем значущості $\alpha=0,05$ критичне значення $F_{\text{кр}}=5,14$. Бачимо, що обчислене значення параметра Фішера $F_{\text{ф}}$ більше за критичне $F_{\text{кр}}$, тобто $F_{\text{ф}}>F_{\text{кр}}$. Отже, нульова гіпотеза $H_0:\ b_1=b_2=\ldots=b_p=0$ відхиляється, а це означає сукупний вплив змінних x_1, x_2 на залежну змінну істотний.

14 Використаємо побудовану регресійну модель для прогнозу залежної змінної. Задамо прогнозовані значення незалежних змінних $x'_{10}=(1;x_{10;1};x_{10;2})=(1;17,5;112)$. Тоді прогнозовані значення знайдемо за формулою:

$$\widehat{y}_{10}=\widehat{b}_0+\widehat{b}_1x_1+\widehat{b}_2x_2=5,024-0,137\cdot 17,5-0,001\cdot 112=2,468.$$

Для побудови довірчого інтервалу для індивідуального значення використаємо формулу

$$\widehat{y}_{n+1}-t_{\alpha/2,n-p-1}\cdot\widehat{\sigma}\left[1+x'_{n+1}(X'X)^{-1}x_{n+1}\right]^{1/2}<y_{n+1}<\widehat{y}_{n+1}+t_{\alpha/2,n-p-1}\cdot\widehat{\sigma}\left[1+x'_{n+1}(X'X)^{-1}x_{n+1}\right]^{1/2},$$

де $x'_{n+1}=x'_{10}=(1;17,5;112)$

$$t_{\alpha/2,n-p-1}\cdot\hat{\sigma}\left[1+x'_{n+1}(X'X)^{-1}x_{n+1}\right]^{1/2}=2,4469\cdot 0,1088\left[1+(1\quad 17,5\quad 112)\begin{pmatrix}293,14 & -9,444 & -1,271\\ -9,444 & 0,335 & 0,037\\ -1,271 & 0,037 & 0,006\end{pmatrix}\begin{pmatrix}1\\17,5\\112\end{pmatrix}\right]^{\frac{1}{2}}=0,378$$

$$2,468-0,378<y_{10}<2,468+0,378 \qquad \text{або} \qquad 2,0896<y_{10}<2,8463\,.$$

15 Обчислимо середні коефіцієнти еластичності за формулою $\bar{Э}_{yx_j}=\hat{b}_j\dfrac{\bar{X}_j}{\bar{Y}}$:

$$\bar{X}_1=\frac{1}{n}\sum_{i=1}^{9}x_{i1}=\frac{1}{9}(15,6+13,5+.....+17,1)=15,52$$

$$\bar{X}_2=\frac{1}{n}\sum_{i=1}^{9}x_{i2}=\frac{1}{9}(106,3+128,5+.....+105,9)=115,89$$

$$\bar{Y}=\frac{1}{n}\sum_{i=1}^{9}y_i=\frac{1}{9}(2,72+3,04+.....+2,63)=2,73$$

$$\bar{Э}_{yx_1}=\hat{b}_1\frac{\bar{X}_1}{\bar{Y}}=-0,137\cdot\frac{15,52}{2,73}=-0,778$$

$$\bar{Э}_{yx_2}=\hat{b}_2\frac{\bar{X}_2}{\bar{Y}}=-0,001\cdot\frac{115,189}{2,73}=-0,059\,.$$

Отже, бачимо, що при збільшенні фактора x_1 на один відсоток результативна ознака y у середньому зменшиться на 0,78%, а при такому ж збільшенні фактора x_2 результативна ознака y зменшиться на 0,06%.

Теорія Мультиколінерність. [19], с. 26 – 31.

Питання для самоперевірки

1 Що таке класична лінійна економетрична модель?
2 Яку матрицю називають матрицею факторів?
3 Охарактеризуйте поняття гомоскедастичності.
4 Як одержати оцінки параметрів?
5 Які властивості мають оцінки параметрів?
6 Сформулюйте гіпотезу про значущість коефіцієнтів рівняння регресії.
7 Для чого застосовують коефіцієнт детермінації?
8 Як перевірити гіпотезу про адекватність лінійної моделі?
9 Сформулюйте задачу точкового прогнозування.
10 Сформулюйте задачу інтервального прогнозування.

ПРАКТИЧНЕ ЗАНЯТТЯ 4

Мультиколінеарність

Для виконання завдання студент повинен знати:
1 Мету і зміст даної роботи, порядок її проведення.
2 Як знайти кореляційну матрицю.
3 Алгоритм методу Феррара-Глобера перевірки мультиколінеарності.

4 Формулу обчислення визначників 2-го та 3-го порядків.

5 Матричну форму системи нормальних рівнянь.

6 Матричну формулу оцінки параметрів системи регресії.

7 Формулу частинних коефіцієнтів кореляції.

8 Формулу коефіцієнтів множинної кореляції.

9 Формулу *t*-статистики значущості параметрів регресії.

10 Формулу прогнозу та його довірчого інтервалу.

11 Частинні коефіцієнти еластичності та їх економічну інтерпретацію.

Студент повинен вміти користуватися пакетом Excel:

1 Знаходити транспоновану обернену матрицю та добуток матриць.

2 Використовуючи матричні операції, знаходити розв'язок системи нормальних рівнянь.

3 Користуватися вбудованими статистичними функціями: "СРЗНАЧ" - середнє значення; зміщене середнє квадратичне відхилення.

4 Використовуючи категорію "Математика", знаходити довірчі інтервали прогнозу.

5 На основі розрахунків робити висновки.

Студент повинен підготувати алгоритм розв'язання задачі з використанням електронних таблиць Excel.

Числовий приклад на дослідження наявності мультиколінеарності

Приклад. Економічний показник *Y* залежить від трьох факторів, на основі статистичних даних потрібно:

1 Використовуючи χ^2-критерій, з надійністю p = 0,95 оцінити наявність загальної мультиколінеарності.

2 Якщо існує загальна мультиколінеарність, то, використовуючи *F* та *t*-статистику з надійністю p=0,95, виявити пари факторів, між якими існує мультиколінеарність. Якщо такі пари факторів існують, то один із факторів цієї пари виключити із розгляду.

3 Використовуючи сервіс: "математика/матриц" знайти оцінки параметрів лінійної регресії. Результат отриманих оцінок перевірити, використовуючи сервіс: "статистика/лінейн".

4 Використовуючи *F*-критерій з надійністю p=0,95, перевірити статистичну значущість коефіцієнта детермінації.

5 Якщо математична модель із заданою надійністю адекватна експериментальним даним, то, використовуючи *t*-статистику з надійністю p = 0,95, оцінити значущість параметрів регресії.

6 Знайти значення прогнозу показника для заданих значень факторів, його довірчий інтервал із надійністю p =0,95.

7 Знайти частинні коефіцієнти еластичності для точки прогнозу.

8 На основі отриманих розрахунків зробити економічний аналіз.

Розв'язання. Досліджуємо наявність мультиколінеарності в масиві X_1, X_2, X_3, використовуючи алгоритм Феррара-Глобера.

Крок 1 Стандартизуємо початкові змінні:

№ п/п	X_1	X_2	X_3
1	2,37	10,27	6,4
2	4,77	11,07	7,88
3	6,24	13,08	8,5
4	8,7	14,52	8,86
5	10,79	16,28	10,51
6	13,6	17,29	10,53
7	16,31	19,04	11,74
8	16,4	20,45	13,96
9	21,25	21,94	13,86
10	22,87	22,55	14,6
11	25,15	22,56	14,24
12	28,27	24,79	16,59
13	28,7	24,82	15,03
14	30,0	25,11	15,34
15	32,25	26,11	15,84
16	33,85	27,58	17,3

Стандартизуємо змінні за формулою

$$x_{ij}^* = \frac{x_{ij} - \overline{X}_j}{\sqrt{\sigma_{x_j}^2 \cdot n}}, \quad j - \overline{1,p}, \quad i = \overline{1,n},$$

де n – число спостережень, i – номер значення j–тої змінної,

$$\overline{X}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}, \quad \sigma_{x_j}^2 = \overline{X_j^2} - \overline{X_j}^2 .$$

$$\overline{X_1} = \frac{2,37+4,77+6,24+10,79+13,6+16,4+21,25+22,87+22,15+28,27+28,7+30+32,25+33,85}{16} = 18,845 .$$

Середні значення для X_2 та X_3 обчислюються аналогічно:

$$X_2 = 19,841, \quad X_3 = 12,574 .$$

Дисперсія $\sigma_{x_j}^2$ знаходиться за формулою

$$\sigma_{x_j}^2 = \frac{\sum_{i=1}^{n}(x_{ij} - \overline{X}_j)^2}{n} .$$

$\sigma_{x_1}^2 =$ ((2,37-18,845)^2+(4,77-18,845)^2+(6,24-18,845)^2+ ... +(33,85-18,845)^2)/16 =107,54.

Дисперсії для X_2 та X_3 знаходяться аналогічно:

$$\sigma_{x_2}^2 = 107,54, \ \sigma_{x_3}^2 = 107,54.$$

Далі обчислюємо стандартизовані (нормалізовані) значення змінних:

$$x_{11}^* = \frac{2,37 - 18,845}{\sqrt{107,54*16}} = -0,397, \quad x_{12}^* = \frac{4,77 - 18,845}{\sqrt{107,54*16}} = -0,339.$$

Таблиця стандартизованих змінних приймає наступний вигляд:

X_1^*	X_2^*	X_3^*
-0,397	-0,388	-0,916
-0,339	-0,352	-0,807
-0,304	-0,261	-0,762
-0,245	-0,196	-0,735
-0,194	-0,116	-0,614
-0,126	-0,07	-0,612
-0,061	0,009	-0,523
-0,059	0,073	-0,36
0,058	0,14	-0,367
0,097	0,168	-0,313
0,152	0,168	-0,339
0,227	0,269	-0,166
0,238	0,27	-0,281
0,269	0,283	-0,258
0,323	0,329	-0,221
0,362	0,395	-0,114

Крок 2 Знаходження кореляційної матриці R :

$$R = X^{*\prime} X^* = \begin{pmatrix} -0,397 & -0,339 & \cdots & 0,362 \\ -0,388 & -0,352 & \cdots & 0,395 \\ -0,916 & -0,807 & \cdots & -0,114 \end{pmatrix} \begin{pmatrix} -0,397 & -0,388 & -0,916 \\ -0,339 & -0,352 & -0,807 \\ \vdots & \vdots & \vdots \\ 0,362 & 0,395 & -0,114 \end{pmatrix},$$

де $X^{*\prime}$ – матриця транспонованих пояснювальних змінних.

Отримаємо

$$R = \begin{pmatrix} 0,938 & 0,928 & 0,909 \\ 0,928 & 0,937 & 0,926 \\ 0,909 & 0,926 & 0,938 \end{pmatrix}.$$

Обчислимо визначник кореляційної матриці :

$$det\, R = |R| = 0,00036.$$

Крок 3 Обчислимо значення критерію χ^2:

$$\chi^2 = -\left[n - 1 - \frac{1}{6}(2 \cdot m + 5)\right] \ln R = -\left[16 - 1 - \frac{1}{6}(2 \cdot 3 + 5)\right] \ln 0,00036 = 104,41 \ .$$

При ступені свободи $\upsilon = \frac{1}{2} p(p-1) = \frac{1}{2} \cdot 3 \cdot (3-1) = 3$ і рівні значущості $\alpha = 0,05$ критерій $\chi^2_{\alpha,\upsilon} = 7,815$. Оскільки $\chi^2_{\alpha,\upsilon} > \chi^2$, доходимо висновку, що в масиві змінних існує мультиколінеарність.

Крок 4 Знайдемо матрицю, обернену до матриці R.

$$C = R^{-1} = (X^{*\prime} X^*)^{-1};$$

$$C = \frac{1}{\det R}\begin{pmatrix} B_{11} & B_{12} & B_{13} \\ B_{21} & B_{22} & B_{23} \\ B_{31} & B_{23} & B_{33} \end{pmatrix}.$$

де B_{ij} – алгебраїчні доповнення відповідних елементів матриці $(X^{*\prime}X^{*})^{-1}$.

$$\begin{pmatrix} 58,528 & -76,77 & 19,087 \\ -76,97 & 145,23 & -69,026 \\ 19,087 & -69,03 & 50,749 \end{pmatrix}$$

Крок 5 Використовуючи діагональні елементи матриці *C*, обчислимо значення *F* – критерія:

$$F_1 = (c_{11} - 1)\frac{n-m}{m-1} = (58,53 - 1)\frac{16-3}{3-1} = 191,73,$$

$$F_2 = (c_{22} - 1)\frac{n-m}{m-1} = (145,23 - 1)\frac{16-3}{3-1} = 480,78,$$

$$F_3 = (c_{33} - 1)\frac{n-m}{m-1} = (50,75 - 1)\frac{16-3}{3-1} = 165,83.$$

Для рівня значущості $\alpha = 0,05$ і ступенів свободи $\upsilon_1 = p = 2$ і $\upsilon_2 = n - p - 1 = 12$ критичне (табличне) значення критерію - $F = 19,41$.

Оскільки $F_1 > F_{\alpha,\upsilon_1,\upsilon_2}$, $F_2 > F_{\alpha,\upsilon_1,\upsilon_2}$, $F_3 > F_{\alpha,\upsilon_1,\upsilon_2}$,

то кожна з пояснювальних змінних мультиколінеарна з двома іншими. Щоб визначити наявність попарної мультиколінеарності, продовжимо дослідження .

Крок 6 Обчислимо частинні коефіцієнти кореляції r_{ij} , скориставшись елементами матриці *C*:

$$r_{12} = \frac{-c_{12}}{\sqrt{c_{11}\cdot c_{22}}} = \frac{76,77}{\sqrt{58,53\cdot 145,23}} = 0,83,$$

$$r_{13} = \frac{-c_{13}}{\sqrt{c_{11}\cdot c_{33}}} = \frac{-19,09}{\sqrt{58,53\cdot 50,75}} = -0,35 ,$$

$$r_{23} = \frac{-c_{23}}{\sqrt{c_{22}\cdot c_{33}}} = \frac{69,03}{\sqrt{145,23\cdot 50,75}} = 0,80.$$

Порівнявши частинні коефіцієнти кореляції з парними, якими є відповідно елементи кореляційної матриці *R*: r_{12}=0,928, r_{13}=0,909, r_{23}=0,926, можна помітити, що вони близькі за величиною (крім r_{13}), отже зробити висновки про наявність певної мультиколінеарності поки що не можливо.

Крок 7 Визначимо значення *t*-критерію на основі частинних коефіцієнтів кореляції:

$$t_{12} = \frac{r_{12}\sqrt{n-m}}{\sqrt{1-r_{12}^2}} = \frac{0,83\sqrt{16-3}}{\sqrt{1-0,83^2}} = 5,42,$$

$$t_{13} = \frac{r_{13}\sqrt{n-m}}{\sqrt{1-r_{13}^2}} = \frac{0,35\sqrt{16-3}}{\sqrt{1-0,35^2}} = 1,35,$$

$$t_{23} = \frac{r_{23}\sqrt{n-m}}{\sqrt{1-r_{23}^2}} = \frac{0,8\sqrt{16-3}}{\sqrt{1-0,8^2}} = 4,88.$$

Табличне значення t-критерію при n-p-1 = 12 ступенях свободи і рівні значущості α=0,05 дорівнює 2,179. Тільки числове значення t_{13} менше за табличне значення. Отже, робимо висновок, що X_2 колінеарне з X_1 та X_3. Тому далі фактор X_2 потрібно виключити із розгляду і оцінити параметри економетричної моделі, що характеризує залежність Y від змінних X_1, X_3 .

Теорія Узагальнена лінійна регресійна модель. [19], с. 31 – 33.

Питання для самоперевірки

1 Що означає термін “мультиколінеарність”?
2 У чому полягає різниця між повною і неповною мультиколінеарністю?
3 Які основні наслідки мультиколінеарності?
4 Які ознаки мультиколінеарності?
5 Що таке частинний коефіцієнт кореляції?
6 Як знайти частинний коефіцієнт кореляції?
7 Які критерії включає алгоритм Феррара - Глобера?
8 Із яких кроків складається алгоритм Феррара - Глобера?
9 Які засоби вилучення мультиколінеарності?
10 У чому полягає сутність методу головних компонентів?

ПРОГРАМА МОДУЛЯ № 2

Узагальнена лінійна регресійна модель. Гетероскедастичність. Автокореляція. Моделі часових рядів

Тема 4. Виробнича регресія. Узагальнена лінійна регресійна модель.
Тема 5. Гетероскедастичність.
Тема 6. Автокореляція.
Тема 7. Моделі часових рядів.

ПРАКТИЧНЕ ЗАНЯТТЯ 5

Виробнича регресія. Узагальнена лінійна регресійна модель

Для виконання роботи студент повинен знати:

1) Мету і зміст роботи, порядок її виконання.

2) Метод приведення нелінійної регресії до лінійної регресії.

3) Яким чином скласти систему нормальних рівнянь для оцінки параметрів приведеної лінійної регресії.

4) Методику розв’язання системи лінійних рівнянь.

5) Яким чином знайти оцінку частинного коефіцієнта еластичності та його економічну інтерпретацію.

6) Метод знаходження оберненої матриці.

7) Правило знаходження добутку матриць.

8) Яким чином обчислюється оцінка прогнозу і його надійний інтервал.

9) Критерій Фішера оцінки адекватності прийнятої моделі експериментальним даним.

Студент повинен вміти користуватися пакетом Excel:

1) будувати таблиці;

2) складати і копіювати формули;

3) працювати з базою даних;

4) розв'язувати систему лінійних рівнянь;

5) користуватися статистичними та математичними вбудованими функціями;

6) будувати графіки.

Студент повинен підготувати: алгоритм розв'язання даної задачі з використанням електронних таблиць Excel.

Числовий приклад побудови та дослідження виробничої регресії

Приклад. Нехай виробнича регресія має вигляд: $Y = a_0 x_1^{a_1} x_2^{a_2}$, де Y – обсяг випущеної продукції; x_1 – працезатрати; x_2 – основні засоби розглянутої галузі.

На основі статистичних даних потрібно:

1) Знайти оцінки параметрів виробничої регресії, використовуючи МНК.

2) З надійністю p=0,95 встановити адекватність прийнятої математичної моделі статистичним даним.

3) Якщо модель адекватна статистичним даним, то знайти частинні коефіцієнти еластичності.

4) Знайти значення прогнозу і його надійний інтервал.

5) Побудувати ізокванту для $Y = y_2$.

6) Побудувати криву граничної продуктивності праці.

7) Використовуючи розрахунки, зробити висновки.

x_1	50	60	70	80	90	100	110	120	130	140	150	160
x_2	90	100	110	120	130	140	150	160	170	180	190	200
y	152	172	192	213	232	253	275	293	314	334	354	
№	1	2	3	4	5	6	7	8	9	10	11	12

Розв'язання.

1 Ввести задані X_1, X_2, Y в початкову таблицю (B2:D12).

2 Привести до лінійної форми. Для цього:

а) прологарифмувати задані X_1, X_2, Y та ввести їх у початкову таблицю: у комірку E2: "=LN(B2)", у комірку F2: "=LN(C2)", у комірку G2: "=LN(D2)". Скопіювати ці формули у кожному відповідному стовпчику до рядка 12.

б) побудувати транспоновану матрицю Z':

- виділити діапазон (I16:S18) – 3 рядки, 11 стовпчиків;
- кнопка "Вставка функції";
- категорія "Ссылки и массивы", функція "ТРАНСП";
- у віконці "Массив" натиснути справа червону стрілку;
- виділити діапазон (B16:D26), закрити віконце (натиснути червону стрілку у віконці);
- натиснути (Ctrl + Shift + Enter) – отримаємо транспоновану матрицю.

3 Знайти добуток матриць $(Z'xZ)$:

а) виділити діапазон (G23:I25);

б) кнопка "Вставка функции", категорія "Математические", функція "МУМНОЖ";

в) у віконце “массив 1” ввести діапазон (I16:S18), у віконце “массив 2” ввести діапазон (B16:D26);

г) натиснути (Ctrl + Shift + Enter) – отримаємо добуток матриць.

4 Обчислити обернену матрицю:

- виділити діапазон (G29:I31);

- кнопка “Вставка функции”, категорія “Математические”, функція “МОБР” (далі, як у пункті 2б).

5 У діапазоні D42 – D44 міститься значення a_1, a_2, a_0 – виписати до зошиту.

6 У комірку H42 ввести формулу: “=D44*F42*G42” і скопіювати її до комірки H53.

7 У комірці L56 – значення коефіцієнта детермінації (виписати), у комірці L57 – обчислене значення критерію Фішера. Порівняти його зі значенням у комірці K58. Зробити висновки.

8 У комірці G55 введена формула обчислення $\hat{Y}_{\text{прогн.}}$. Ретельно вивчити усі складові цієї формули та виписати свої значення $x_{1пр.}$, $x_{2пр.}$, $\hat{Y}_{\text{пр.}}$

9 У комірках G57 – G59 приведені прогнозові значення для лінеарізованої моделі. Розібратися, бути готовим давати пояснення.

10 У діапазоні D73 – E83 приведені дані для побудови “Изольвенти”. Побудувати її за своїми даним.

11 У діапазоні G90 – H101 приведені дані для побудови ГПп. Побудувати її для своїх даних.

Висновки

1 Оскільки для k_1= 2 і k_2 = 11-3=8 F(2; 8; 0,95) =4,46, тобто $F_{\text{роз.}}$> F(2; 8; 0,95), то з надійністю P = 0,95 можна вважати, що прийнята математична модель адекватна експериментальним даним та її можна застосувати для аналізу господарської діяльності підприємства.

2 Параметри a_1 = 0,219 і a_2 = 0,813 є частинним коефіцієнтами еластичності, тобто зміна фактора X_1 (працезатрати) на 1% при незмінному факторі X_2 (основні засоби) викликає зміну обсягу випуску продукції на 0,219%, аналогічно зміна фактора X_2 на 1% при незмінному факторі X_1 викликає зміну обсягу продукції на 0,813%.

3 Темпи приросту показника виражаються лінійно через темпи приросту факторів:

$$E_y = 0,213E_{x1} + 0,83E_{x2},$$

де E, E_{x1}, E_{x2} - темпи приросту показника і факторів відповідно.

4 Для факторів X_1 = 160, X_2 = 200 визначена оцінка прогнозу Y_p = 374,79 і з надійністю p = 0,95 вона буде належати інтервалу (374,71; 374,87).

5 Оскільки сумарний коефіцієнт еластичності $A = a_1 + a_2$ = 1,032, то при збільшенні обсягу основних засобів та обсягу працезатрат у k раз, обсяг випуску продукції збільшиться у $k^{1,032}$ разів.

Теорія Гетероскедастичність. [19], с. 33 – 38.

Питання для самоперевірки

1. Чим відрізняється узагальнена модель від класичної?
2. Які параметри невідомі в узагальненій регресійній моделі?
3. До яких наслідків приведе застосування МНК до узагальненої моделі?
4. У чому полягає сутність узагальненого методу найменших квадратів?
5. Чи будуть найкращими оцінки УМНК?
6. Чи можна оцінити матрицю *Ко* без додаткових припущень відносно її структури?

ПРАКТИЧНЕ ЗАНЯТТЯ 6

Гетероскедастичність

Приклад.

1. Дослідити наявність гетероскедастичності у загальній економетричній моделі за допомогою параметричного тесту Гольдфельда – Кванта;

2. Оцінити параметри моделі за допомогою узагальненого методу найменших квадратів (методу Ейткена).

3. Здійснити порівняльний аналіз кількісних характеристик взаємозв'язку, здобутих методом найменших квадратів і узагальненим методом Ейткена.

Вихідні дані наведено в таблиці 6.1.

Таблиця 6.1

Вихідні дані

Місяць	Попит, грн.	Дохід на одну особу, грн.	Ціна одиниці товару, грн.	Ціна одиниці взаємозамінних товарів, грн.
1-й	50,00	200,00	22,00	2,00
2-й	52,00	210,00	21,00	3,00
3-й	51,00	250,00	21,50	2,50
4-й	55,00	220,00	19,00	4,00
5-й	56,00	225,00	18,00	5,00
6-й	54,00	220,00	18,55	4,50
7-й	60,00	230,00	16,00	6,00
8-й	58,00	228,00	17,00	9,00
9-й	62,00	240,00	15,00	10,00
10-й	64,00	250,00	12,00	11,00
11-й	65,00	255,00	11,00	11,00
12-й	67,00	260,00	11,00	12,00

Розв'язання.

1. Дослідимо наявність гетероскедастичності у загальній економетричній моделі за допомогою метода Гольфельда-Квандта.

1.1. Ідентифікуємо змінні.

Y – попит (залежна змінна);

X_1 – дохід на одну людину (незалежна, або пояснююча змінна);

X_2 – ціна одиниці (незалежна, або пояснююча змінна);

X_3 – ціна одиниці взаємозамінних товарів (незалежна, або пояснююча змінна).

Звідси дістаємо загальний вигляд моделі:

$$Y = f(X_1, X_2, X_3, u),$$

де u – стохастична складова.

1.2.Специфікуємо економетричну модель у лінійній формі. Дістанемо лінійну модель

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + u \text{ (теоретична модель);}$$

$$\hat{Y} = \hat{a}_0 + \hat{a}_1 X_1 + \hat{a}_2 X_2 + \hat{a}_3 X_3, \qquad u = Y - \hat{Y}.$$

1.3. Установимо наявність гетероскедастичності, застосовуючи тест Гольдфельда– Квандта. Розглядувану сукупність спостережень впорядкуємо за X_1 від меншого до більшого. Відшукаємо c спостережень, які містяться всередині сукупності

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

Зауважимо, що ми висунули таку гіпотезу: дохід може викликати гетероскедастичність залишків.

У цьому прикладі n = 12, $c \approx$ 3. Отже, якщо ми відкинемо 3 елементи із середини, то в сукупності залишиться 9 елементів, які не діляться на 2 без остачі. Тому потрібно відкинути 3 елементи з середини та один елемент на початку сукупності й дістати дві сукупності : n_1, n_2 = 4.

1.3.1. Знайдемо економетричну модель для сукупності n_1 = 4.

Оцінимо кількісно параметри моделі, скориставшись стандартною програмою "Лінійн":

	0,2370	2,3333
	0,0392	8,3821
	0,9481	0,5092
лін1:	36,5714	2,0000
	9,4815	$\boxed{0,5185}$

Таким чином, на основі першої сукупності дістанемо таку економетричну модель:

$$\hat{Y} = 2,33 + 0,24X_1.$$

1.3.2. Побудуємо економетричну модель для сукупності n_2 = 4.

Оцінимо кількісно параметри моделі, скориставшись також стандартною програмою "Лінійн":

	0,2400	4,2000
	0,0302	7,6004
	0,9692	0,4472
лін2:	63,0000	2,0000
	12,6000	$\boxed{0,4000}$

Отже, на основі другої сукупності дістанемо таку економетричну модель:

$$\hat{Y} = 4{,}2 + 0{,}24X_1.$$

1.3.3. На підставі наведених щойно даних знайдемо суму квадратів залишків (ці дані можна дістати із стандартних програм "Лін1" та "Лін2", що виділені в наведених розрахунках вище):

$$S_1 = u'_1 u_1 = \sum\left(Y_1 - \hat{Y}_1\right)^2;$$

$$S_2 = u'_2 u_2 = \sum\left(Y_2 - \hat{Y}_2\right)^2;$$

$$S_1 = 0{,}5185;$$

$$S_2 = 0{,}4000.$$

1.3.4. Знайдемо критерій

$$R^* = \frac{S_2}{S} \quad R^* = \frac{0{,}4}{0{,}5185} = 0{,}77$$

Порівняємо це значення з табличним значенням критерію Фішера (F –критерію), коли маємо

$$\frac{n - c - 2m}{2} = 2$$

ступенів свободи та α = 0,5. Значення критерію R^* менше за табличне значення критерію Фішера (0,77 < 19), а, отже, у масиві змінних гетероскедастичності немає.

Висновок. В економетричній моделі для вихідних даних виявлено змішану гетероскедастичність, а тому для оцінювання параметрів моделі не можна застосовувати метод найменших квадратів.

2. Оцінимо параметри моделі за допомогою УМНК.

Якщо при економетричному моделюванні для певних вихідних даних буде виявлено явище гетероскедастичності, то параметри моделі слід оцінювати за узагальненим методом найменших квадратів (методом Ейткена). Оператор оцінювання цим методом можна записати у вигляді

$$\hat{A} = \left(X' S^{-1} X\right)^{-1} X' S^{-1} Y.$$

У матриці S^{-1} використаємо умову

$$\lambda_i = \hat{u}_i^2.$$

Запишемо матрицю пояснювальних змінних

$$X = \begin{bmatrix} 1{,}00 & 200{,}00 & 22{,}00 & 2{,}00 \\ 1{,}00 & 210{,}00 & 21{,}00 & 3{,}00 \\ 1{,}00 & 205{,}00 & 21{,}50 & 2{,}50 \\ 1{,}00 & 220{,}00 & 19{,}00 & 4{,}00 \\ 1{,}00 & 225{,}00 & 18{,}00 & 5{,}00 \\ 1{,}00 & 220{,}00 & 18{,}50 & 4{,}50 \\ 1{,}00 & 230{,}00 & 16{,}00 & 6{,}00 \\ 1{,}00 & 228{,}00 & 17{,}00 & 9{,}00 \\ 1{,}00 & 240{,}00 & 15{,}00 & 10{,}00 \\ 1{,}00 & 250{,}00 & 12{,}00 & 11{,}00 \\ 1{,}00 & 255{,}00 & 11{,}00 & 11{,}00 \\ 1{,}00 & 260{,}00 & 11{,}00 & 12{,}00 \end{bmatrix}.$$

Знайдемо матрицю, транспоновану до матриці пояснювальних змінних:

$$X' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 200 & 210 & 205 & 220 & 225 & 220 & 230 & 228 & 240 & 250 & 255 & 260 \\ 22 & 21 & 21{,}5 & 19 & 18 & 18{,}5 & 16 & 17 & 15 & 12 & 11 & 11 \\ 2 & 3 & 2{,}5 & 4 & 5 & 4{,}5 & 6 & 9 & 10 & 11 & 11 & 12 \end{bmatrix}$$

Знайдемо матрицю S^{-1}:

$$S^{-1} = \begin{bmatrix} 1{,}048 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0{,}999 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0{,}952 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0{,}816 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0{,}860 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0{,}773 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0{,}472 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}691 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}408 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}349 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}349 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}294 \end{bmatrix}$$

Обчислимо діагональні елементи матриці S^{-1} згідно з третьою моделлю залишків за тестом Глейсера

$$\left(\left|\hat{u}\right|\right)^2 = \left(\hat{a}_0 + \hat{a}_1 X_3\right)^2, \text{ або } \left(\left|\hat{u}\right|\right)^2 = \left(1{,}11997 - 0{,}04813 X_3\right)^2 .$$

Нарешті запишемо добуток останньої матриці та матриці S^{-1} :

Перемножимо матрицю, транспоновану до матриці пояснюючих змінних, та матрицю S^{-1}:

$$X'S^{-1} = \begin{bmatrix} 1{,}0 & 1{,}0 & 1{,}0 & 0{,}8 & 0{,}9 & 0{,}8 & 0{,}5 & 0{,}7 & 0{,}4 & 0{,}3 & 0{,}3 & 0{,}3 \\ 209{,}6 & 204{,}9 & 199{,}9 & 179{,}5 & 189{,}2 & 174{,}0 & 107{,}5 & 158{,}9 & 97{,}9 & 87{,}2 & 88{,}9 & 76{,}5 \\ 23{,}1 & 21{,}5 & 20{,}0 & 15{,}1 & 16{,}3 & 13{,}9 & 8{,}0 & 11{,}1 & 6{,}1 & 4{,}2 & 3{,}8 & 3{,}2 \\ 2{,}1 & 2{,}5 & 2{,}9 & 3{,}7 & 3{,}4 & 3{,}9 & 4{,}2 & 4{,}1 & 4{,}1 & 3{,}8 & 3{,}8 & 3{,}5 \end{bmatrix}.$$

Матриця залежної замінної Y (попит на товар) має такий вигляд:

$$Y = \begin{bmatrix} 50{,}00 \\ 51{,}00 \\ 52{,}00 \\ 55{,}00 \\ 54{,}00 \\ 56{,}00 \\ 58{,}00 \\ 60{,}00 \\ 65{,}00 \\ 64{,}00 \\ 65{,}00 \\ 67{,}00 \end{bmatrix}.$$

Перемножимо матрицю $X'S^{-1}$ та вектор Y попиту на товар

$$X'S^{-1}Y = \begin{bmatrix} 446{,}321 \\ 99490{,}1 \\ 8022{,}9 \\ 2463{,}87 \end{bmatrix}.$$

Перемножимо матрицю $X'S^{-1}$ на матрицю X незалежних змінних

$$X'S^{-1}X = \begin{bmatrix} 8{,}01047 & 1773{,}97581 & 146{,}33183 & 42{,}0992 \\ 1773{,}97581 & 395075{,}267 & 31965{,}61685 & 9720{,}730565 \\ 146{,}331183 & 31965{,}61685 & 2762{,}153715 & 689{,}47742 \\ 42{,}0992 & 9720{,}730565 & 689{,}40742 & 299{,}235335 \end{bmatrix}.$$

Знайдемо матрицю, обернену до попередньої

$$\left(X'S^{-1}X\right)^{-1} = \begin{bmatrix} 2797{,}45912 & -9{,}05130 & -43{,}79182 & 1{,}35283 \\ -9{,}05130 & 0{,}03007 & 0{,}13528 & -0{,}01515 \\ -43{,}79182 & 0{,}13528 & 0{,}73864 & 0{,}06469 \\ 1{,}35283 & 0{,}01515 & 0{,}06469 & 0{,}15613 \end{bmatrix}.$$

Перемноживши попередню матрицю та матрицю $X'S^{-1}Y$, дістанемо матрицю оцінок параметрів економетричної моделі за допомогою узагальненого методу найменших квадратів:

$$A = \begin{bmatrix} 45{,}8773 \\ 0{,}10393 \\ -0{,}77676 \\ 0{,}1928 \end{bmatrix}.$$

Економічна модель попиту на товар запишеться таким чином:

$$\hat{Y} = 45{,}88 + 0{,}1X_1 - 0{,}77X_2 + 0{,}2X_3 .$$

3. Здійснимо порівняльний аналіз кількісних характеристик взаємозв'язку, здобутих методом найменших квадратів і узагальненим методом Ейткена.

Середня ефективність чинників, здобутих методом найменших квадратів і узагальненим методом Ейткена, буде однакова:

$$\mu_1 = 0{,}253; \mu_2 = 3{,}436; \mu_3 = 8{,}675.$$

Середня ефективність чинників подається у вигляді

$$\mu_i = \frac{\overline{Y}}{\overline{X}_j}.$$

Висновок. Якщо дохід на одну особу зміниться на одиницю, а решта чинників (показників) будуть сталими, то попит у середньому збільшиться на 0,25 одиниці; якщо ціна одиниці товару зміниться на одиницю, а решта показників будуть сталими, то попит у середньому збільшиться на 3,44 одиниці. Якщо ціна одиниці взаємозамінних товарів зміниться на одиницю, а інші змінні будуть сталими, то попит у середньому збільшиться на 8,68 одиниці.

Теорія Автокореляція. [19], с. 38 – 45.

Питання для самоперевірки

1 Дайте означення гомоскедастичності і гетероскедастичності.

2 Як впливає явище гетероскедастичності на оцінку параметрів моделі?

3 Назвіть методи визначення гетероскедастичності.

4 Як застосовується тест Голфелда - Квандта для визначення гетероскедастичності?

5 У чому полягають методи формування матриці K_0.

6 Запишіть оператор оцінювання параметрів моделі з гетероскедастичними збуреннями за методом Ейткена.

ПРАКТИЧНЕ ЗАНЯТТЯ 7

Автокореляція

Приклад побудови економетричної моделі з автокорельованими залишками

Приклад. Згідно з даними таблиці 7.1 побудувати економетричну модель продуктивності праці у випадку, якщо залишки, здобуті 1МНК, є автокорельованими.

Таблиця 7.1

Вихідні дані

Місяць	Продуктивність праці, гр.од/ людино-год.	Фондомісткість продукції, гр. од.	Коефіцієнт плинності робочої сили, %	Рівень втрат робочого часу, %	Стаж, років
1-й	52	72	13,0	2,7	5,0
2-й	53	74	12,5	2,8	5,5
3-й	50	72	12,0	3,0	5,0
4-й	51	73	11,0	3,2	6,0
5-й	54	70	10,1	3,2	7,0
6-й	55	67	9,0	3,3	8,0
7-й	57	6	8,5	3,4	10,0
8-й	52	62	8,2	3,6	10,0
9-й	60	72	8,0	3,7	10,5
10-й	60	72	5,5	3,7	11,0
11-й	62	74	5,0	3,4	13,0
12-й	64	75	4,7	4,0	10,0
13-й	65	76	4,6	4,2	12,0
14-й	67	80	4,0	4,3	13,0
15-й	67	82	4,1	4,7	14,0
16-й	62	84	4,2	4,8	14,5
17-й	60	84	4,5	4,8	15,5

1. Дослідити залишки на наявність автокореляції.

2.Оцінити параметри моделі методом Ейткена та методом перетворення вихідної інформації.

3.Дослідити статистичну значущість моделі та оцінок її параметрів.

4.Зробити порівняльний аналіз кількісних характеристик взаємозв'язку, здобутих методом 1МНК і методом Ейткена.

Розв'язання.

1.Дослідимо залишки, здобуті згідно з 1МНК, на наявність автокореляції, обчисливши критерій Дарбіна – Уотсона.

Знайдемо оцінку критерію Дарбіна – Уотсона:

$$DW = \frac{\sum_{t=2}^{10}(u_t - u_{t-1})^2}{\sum_{t=1}^{10} u_t^2} = \frac{110{,}721}{53{,}454} = 2{,}07$$

.

Порівняємо значення критерію DW з табличними при α=0,05 і n = 17. Критичні значення критерію DW у цьому разі такі:

DW_1=0,48 – нижня межа;

DW_2=1,85 – верхня межа.

Оскільки $DW_{факт} > DW_2$, то при α=0,05 можна стверджувати що залишки u_t не є автокорельованими.

2.Скориставшись методом Ейткена, оцінимо параметри економетричної моделі з автокорельованими залишками. Оператор оцінювання можна записати таким чином:

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

або

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

де S^{-1} – матриця, обернена до матриці S; V^{-1} – матриця, обернена до матриці V.

2.1. Формування матриці S^{-1}.

$$S^{-1} = \frac{1}{1-\rho^2}\begin{pmatrix} 1 & -\rho & 0 & 0 & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & 0 & 0 & \dots & 0 \\ 0 & -\rho & 1+\rho^2 & -\rho & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}.$$

де $\rho \approx r_{скор} \approx 0{,}36$.

Запишемо матрицю S^{-1} для вихідних даних:

$$S^{-1} = \begin{pmatrix} 1{,}145776 & -0{,}40869 & \dots & 0 & 0 \\ -0{,}40869 & 1{,}291552 & \dots & 0 & 0 \\ 0 & -0{,}40869 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & -0{,}40869 & 0 \\ 0 & 0 & \dots & 1{,}291552 & -0{,}40869 \\ 0 & 0 & \dots & -0{,}40869 & 1{,}145776 \end{pmatrix}$$

Звідси:

$$X'S^{-1} = \begin{pmatrix} 0{,}737089 & 0{,}474178 & \dots & 0{,}474178 & 0{,}737089 \\ 52{,}25306 & 36{,}72395 & \dots & 40{,}64836 & 61{,}91551 \\ 9{,}786506 & 5{,}92723 & \dots & 1{,}909812 & 3{,}439508 \\ 1{,}949273 & 1{,}286831 & \dots & 2{,}316925 & 3{,}538029 \\ 3{,}481103 & 3{,}016668 & \dots & 6{,}671243 & 11{,}83357 \end{pmatrix};$$

$$X'S^{-1}X = \begin{pmatrix} 8{,}586855 & 636{,}5822 & 65{,}72254 & 31{,}75024 & 86 \\ 636{,}5822 & 47569{,}29 & 4778{,}06 & 2379{,}979 & 6492{,}853 \\ 65{,}72254 & 4778{,}06 & 597{,}6239 & 225{,}6698 & 564{,}7057 \\ 31{,}75024 & 2379{,}979 & 225{,}6698 & 121{,}7652 & 336{,}5092 \\ 86 & 6492{,}853 & 564{,}7057 & 336{,}5092 & 977{,}0056 \end{pmatrix};$$

$$\left(X'S^{-1}X\right)^{-1} = \begin{pmatrix} 35{,}22012 & -0{,}17412 & -1{,}3524 & -2{,}49053 & -0{,}4291 \\ -0{,}17412 & 0{,}00483 & -0{,}00422 & -0{,}03267 & -0{,}00308 \\ -1{,}13524 & -0{,}00422 & 0{,}070437 & 0{,}141314 & 0{,}0386 \\ -2{,}49053 & -0{,}032267 & 0{,}141314 & 1{,}124469 & -0{,}03266 \\ -0{,}4291 & -0{,}00308 & 0{,}0386 & 0{,}03266 & 0{,}0482 \end{pmatrix}.$$

2.2. Оцінимо параметри моделі на основі оператора Ейткена:

$$\hat{A} = \left(X'S^{-1}X\right)^{-1} X'S^{-1}Y.$$

$$X'S^{-1}Y = \begin{pmatrix} 501{,}5681 \\ 37422{,}77 \\ 3690{,}201 \\ 1882{,}427 \\ 5175{,}631 \end{pmatrix}; \qquad \hat{A} = \begin{pmatrix} 50{,}86868 \\ 0{,}404484 \\ -1{,}6715 \\ -2{,}4793 \\ -0{,}04822 \end{pmatrix}.$$

$$Y = \begin{pmatrix} 52 \\ 53 \\ 50 \\ 51 \\ 54 \\ 55 \\ 57 \\ 52 \\ 60 \\ 60 \\ 62 \\ 64 \\ 65 \\ 67 \end{pmatrix}; \quad \hat{Y} = \begin{pmatrix} 51{,}32675 \\ 52{,}69943 \\ 52{,}25446 \\ 53{,}78636 \\ 54{,}02904 \\ 54{,}35809 \\ 54{,}84946 \\ 52{,}83263 \\ 56{,}93973 \\ 61{,}09437 \\ 63{,}38643 \\ 62{,}94946 \\ 62{,}92878 \\ 65{,}25346 \end{pmatrix}; \quad u = \begin{pmatrix} 0{,}453266 \\ 0{,}090345 \\ 5{,}082595 \\ 7{,}763817 \\ 0{,}000843 \\ 0{,}412052 \\ 4{,}62482 \\ 0{,}693287 \\ 9{,}365269 \\ 1{,}197645 \\ 1{,}922189 \\ 1{,}103644 \\ 4{,}289945 \\ 3{,}050391 \end{pmatrix};$$

$$D_u^2 = 4{,}82151; \quad D_y^2 = 34{,}01471; \quad R^2 = 0{,}85825; \quad R = 0{,}926419; \quad F = 18{,}16435;$$

$$\operatorname{cov}ar\left(\hat{A}\right)=\begin{pmatrix} 169{,}8141 & -0{,}83952 & -5{,}47259 & -12{,}0811 & -2{,}68889 \\ -0{,}83952 & 0{,}023288 & -0{,}02035 & -0{,}1575 & -0{,}01485 \\ -5{,}47359 & -0{,}02035 & 0{,}339613 & 0{,}681346 & 0{,}18611 \\ -12{,}0081 & -0{,}1575 & 0{,}681346 & 5{,}421639 & -0{,}15746 \\ -2{,}06889 & -0{,}01485 & 0{,}18611 & -0{,}15746 & 0{,}232394 \end{pmatrix};$$

$$S\left(\hat{A}\right)=\begin{pmatrix} 13{,}03128 \\ 0{,}152603 \\ 0{,}582763 \\ 2{,}328441 \\ 0{,}482073 \end{pmatrix}; \qquad tj=\begin{pmatrix} 3{,}903585 \\ 2{,}650565 \\ -2{,}86824 \\ -1{,}06479 \\ -0{,}010003 \end{pmatrix}.$$

Економетричну модель продуктивності праці подамо у такому вигляді:

$$y = 50{,}869 + 0{,}404x_1 - 1{,}672x_2 - 2{,}479x_3 - 0{,}048x_4. \qquad (7.1)$$

3. Дослідимо статистичну значущість моделі та оцінок її параметрів.

Коефіцієнт детермінації R^2=0,858. Він показує, що варіація продуктивності праці на 85,8% визначається досліджуваними чинниками.

Коефіцієнт кореляції R=0,926. На підставі цього коефіцієнта можна стверджувати, що зв'язок між продуктивністю праці та досліджуваними чинниками – фондомісткістю, коефіцієнтом плинності, рівнем втрат робочого часу і стажем – тісний.

Критерій Фішера F = 18,16. Табличне його значення рівня значущості α = 0,05, коли є m – 1= 4, n – m =12 ступенів свободи, дорівнює 5,78. Оскільки $F_{факт}$ >$F_{табл.}$,то гіпотеза про статистичну значущість зв'язку на основі економетричної моделі (7.1) підтверджується.

Перевіримо достовірність кожної оцінки параметрів моделі, зокрема на основі t – критеріїв. Вони дорівнюють:

$$t\hat{a}_0 = 3{,}904; \qquad t\hat{a}_1 = 2{,}650;$$

$$t\hat{a}_2 = 2{,}868; \qquad t\hat{a}_3 = 1{,}065;$$

$$t\hat{a}_4 = 0{,}100.$$

Табличне значення t – критерію для рівня значущості α = 0,05 та ступені свободи n – m =12 дорівнює 1,782. Оскільки $t\hat{a}_0$, $t\hat{a}_1$ і $t\hat{a}_2$ перевищують ці значення. Звідси $t\hat{a}_3$ і $t\hat{a}_4$ - недостовірні. Таким чином, достовірність моделі в цілому досягається за рахунок трьох перших параметрів.

4. Виконаємо порівняльний аналіз характеристик взаємозв'язку, здобутих методом Ейткена.

Метод 1МНК

$$\hat{A} = \begin{pmatrix} 60{,}146 \\ 0{,}316 \\ -1{,}936 \\ -2{,}316 \\ -0{,}182 \end{pmatrix};$$

$R^2 = 0{,}902;$
$R = 0{,}949;$
$F = 27{,}5;$

$$t\hat{a}_j = \begin{pmatrix} 5{,}59 \\ 2{,}42 \\ -3{,}88 \\ -1{,}00 \\ -0{,}35 \end{pmatrix};$$

$$y_{пр} = \begin{pmatrix} 66{,}27 \\ 64{,}40 \\ 64{,}15 \\ 62{,}15 \end{pmatrix};$$

Метод Ейткена

$$\hat{A} = \begin{pmatrix} 50{,}869 \\ 0{,}404 \\ -1{,}672 \\ -2{,}479 \\ -0{,}04822 \end{pmatrix};$$

$R^2 = 0{,}858;$
$R = 0{,}926;$
$F = 18{,}16.$

$$t\hat{a}_j = \begin{pmatrix} 3{,}90 \\ 2{,}65 \\ -2{,}86 \\ -1{,}05 \\ -0{,}10 \end{pmatrix};$$

$$y_{пр} = \begin{pmatrix} 67{,}26 \\ 65{,}74 \\ 65{,}49 \\ 64{,}22 \end{pmatrix}.$$

Порівнюючи оцінки параметрів моделі, здобутих за допомогою 1МНК і методом Ейткена, робимо висновок про те, що загальна тенденція впливу чинників на продуктивність праці не змінилася, але рівень оцінок став дещо іншим. Ці зміни є незначними. Деякі параметри збільшилися, а деякі зменшилися. Такі зміни зумовлені сутністю методу Ейткена, який уточнює параметри моделі за наявності навіть незначних взаємозв'язків залишків.

Як бачимо з наведених даних, усі характеристики дисперсійного аналізу за методом Ейткена дещо нижчі, ніж за методом 1МНК, оскільки метод Ейткена уточнює дисперсію залишків та дисперсії оцінок параметрів моделі, які в разі застосування 1МНК завжди будуть нижчими від істинних, коли в моделі залишки автокорельовані.

І нарешті, очікуваний рівень залежної змінної на основі економетричної моделі, побудованої за методом Ейткена, дещо вищий, ніж за методом 1МНК. Це зумовлюється тим, що метод Ейткена вимагає включення системної складової залишків.

Теорія Моделі часових рядів. [19], с. 45 – 49.

Питання для самоперевірки

1 Що таке автокореляція?

2 Яке припущення МНК порушується при автокореляції?

3 Які наслідки автокореляції?

4 У чому полягає сутність автокореляційного процесу збурень першого порядку?

5 Який вигляд має коваріаційна матриця збурень для авторегресійного процесу першого порядку?

6 Як можна діагностувати авторегресійний процес першого порядку?

7 Яка особливість критерію Дарбіна-Уотсона?

8 За якою схемою проводяться діагностики автокореляції за критерієм Дарбіна-Уотсона?

9 Як можна оцінити параметри моделі з автокорельованими збуреннями при відомій матриці K_0?

10 Як можна оцінити параметри моделі при невідомій матриці K_0?

11 За якою схемою проводиться оцінка регресійного параметра ρ за методом Кохрейна-Оркатта?

ПРАКТИЧНЕ ЗАНЯТТЯ 8

Моделі часових рядів

Приклад 1 За даними 18 місяців побудовано рівняння регресії залежності прибутку підприємства y (млн. грн.) від цін на сировину x_1 тис. грн. за 1 т) і продуктивності праці x_2 (од. продукції на 1 працівника) :

При аналізі залишкових величин були використані значенні приведені в таблиці 8.1.

Таблиця 8.1

Вихідні дані

№	y	x_1	x_2
1	210	800	300
2	720	1000	500
3	300	1500	600
…	…	…	…

$\sum \varepsilon_t^2 = 10500$, $\sum (\varepsilon_t - \varepsilon_{t-1})^2 = 40000$

Треба:

1 По трьом позиціям розрахувати $\hat{y}_t$, ε_t, ε_{t-1}, ε_t^2, $(\varepsilon_t - \varepsilon_{t-1})^2$.

2 Розрахувати критерій Дарбина - Уотсона.

3 Оцінити отриманий результат при 5% - ному рівні значущості.

4 Вказати, чи придатне рівняння для прогнозу.

Розв'язання.

1 $\hat{y}_t$ визначається шляхом підстановки фактичних значень x_1 і x_2 у рівняння регресії:

$$\hat{y}_1 = 200 - 1,5 \cdot 800 + 4,0 \cdot 300 = 200;$$
$$\hat{y}_2 = 200 - 1,5 \cdot 1000 + 4,0 \cdot 500 = 700;$$
$$\hat{y}_3 = 200 - 1,5 \cdot 1500 + 4,0 \cdot 600 = 350.$$

Залишки ε_t розраховуються за формулою $\varepsilon_t = y_t - \hat{y}_t$.

Отже,

$\varepsilon_1 = 210 - 200 = 10$, $\varepsilon_2 = 720 - 700 = 20$, $\varepsilon_3 = 300 - 350 = -50$, $\varepsilon_1^2 = 100$, $\varepsilon_2^2 = 400$, $\varepsilon_3 = 2500$;

ε_{t-1} - ті ж значення, що і ε_t, але зі зміщенням на один місяць.

Результати обчислень оформимо у вигляді таблиці 6.2.

Таблиця 8.2

Результати обчислень

№	$\hat{y}_t$	ε_t	ε_{t-1}	$(\varepsilon_t - \varepsilon_{t-1})$	$(\varepsilon_t - \varepsilon_{t-1})^2$	ε_t^2
1	200	10	-	-	-	100
2	700	20	10	10	100	400
3	350	-50	20	-70	4900	2500
…	…	…	…	…	…	…
Σ					40 000	10 500

2 Критерій Дарбина - Уотсона розраховується за формулою:

3

$$d = \frac{\sum(\varepsilon_t - \varepsilon_{t-1})^2}{\sum \varepsilon_t^2} = \frac{40000}{10500} = 3,81.$$

4 Фактичне значення d порівнюємо з табличними значеннями при 5% рівні значущості. При n = 18 місяців і m = 2 (число чинників) нижнє значення d' дорівнює 1,05, а верхнє - 1,53. Оскільки фактичне значення d близько до 4, можна вважати, що автокореляція в залишках характеризується від'ємною величиною. Щоб перевірити значущість від'ємного коефіцієнта автокореляції, знайдемо величину:

$$4 - d = 4 - 3,81 = 0,19$$

що значно менше, ніж d'. Це означає наявність в залишках автокореляції.

5 Рівняння регресії не може бути використане для прогнозу, оскільки в ньому не усунена автокореляція в залишках, яка може мати різні причини. Автокореляція в залишках може означати, що в рівняння не включений який-небудь істотний чинник. Можливо також, що форма зв'язку неточна, а може бути, в рядах динаміки є загальна тенденція.

Приклад 2

Є наступні дані про величину доходу на одного члена сім'ї і витрати на товар *А* (таблиця 8.3).

Таблиця 8.3

Вихідні дані

Показник	1985 р.	1986 р.	1987 р.	1988 р.	1989 р.	1990 р.
Витрати на товар А, крб.	30	35	39	44	50	53
Дохід на одного члена сім'ї, % до 1985 р.	100	103	105	109	115	118

Треба:

1 Визначити щорічні абсолютні прирости прибутків і витрат і зробити висновки про тенденцію розвитку кожного ряду.

2 Перерахувати основні шляхи усунення тенденції для побудови моделі попиту на товар *A* залежно від доходу.

3 Побудувати лінійну модель попиту, використовуючи перші різниці рівнів початкових динамічних рядів.

4 Пояснити економічний сенс коефіцієнта регресії.

5 Побудувати лінійну модель попиту на товар *A*, включивши в неї чинник часу. Інтерпретувати отримані параметри.

Розв'язання. Позначимо витрати на товар *A* через *y*, а прибутки одного члена сім'ї - через *x*. Щорічні абсолютні прирости визначаються за формулами: $\Delta y_t = y_t - y_{t-1}$, $\Delta x_t = x_t - x_{t-1}$.

Розрахунки можна оформити у вигляді таблиці (таблиця 8.4).

Таблиця 8.4

Числові розрахунки

y_t	Δy_t	x_t	Δx_t
30	-	100	-
35	5	103	3
39	4	105	2
44	5	109	4
50	6	115	6
53	3	118	3

Значення Δy не мають чітко вираженої тенденції, вони варіюють навколо середнього рівня, що означає наявність у ряді динаміки лінійного тренду (лінійній тенденції). Аналогічний висновок можна зробити і по ряду *x*: абсолютні прирости не мають систематичної спрямованості, вони приблизно стабільні, а отже, ряд характеризується лінійною тенденцією.

2 Оскільки ряди динаміки мають загальну тенденцію до росту, то для побудови регресійної моделі попиту на товар *A* залежно від доходу необхідно усунути тенденцію. З цією метою модель може будуватися по перших різницях, тобто $\Delta y = f(\Delta x)$, якщо ряди динаміки характеризуються лінійною тенденцією.

Інший можливий шлях обліку тенденції при побудові моделей - знайти по кожному ряду рівняння тренду: $\hat{y}_t = f(t)$ і $\hat{x}_t = f(t)$ та відхилення від нього: $dy = y_t - \hat{y}_t$; $dx = x_t - \hat{x}_t$. Далі модель будується по відхиленнях від тренду: $dy = f(dx)$.

При побудові економетричних моделей частіше використовується інший шлях обліку тенденції – включення в модель чинника часу. Іншими словами, модель будується за початковими даними, але в неї в якості самостійного чинника включається час, тобто $\hat{y}_t = f(x,t)$.

3 Модель має вигляд $\Delta\hat{y} = a + b\Delta x$.

Для визначення параметрів *a* і *b* застосовується МНК. Система нормальних рівнянь наступна:

$$\begin{cases} \sum \Delta y = na + b\sum \Delta x \\ \sum \Delta y \Delta x = a\sum \Delta x + b\sum \Delta^2 x \end{cases}.$$

Стосовно наших даних маємо $\begin{cases} 23 = 5a + 18b \\ 88 = 18a + 74b \end{cases}$.

Розв'язуючи цю систему, отримаємо: *a* = 2,565 і *b*= 0,565, звідки модель має вигляд

$$\Delta\hat{y} = 2,565 + 0,565\Delta x.$$

4 Коефіцієнт регресії *b* - 0,565 грн. Він означає, що з ростом приросту душового доходу на 1% витрати на товар *A* збільшуються з середнім прискоренням, рівним 0,565 крб.

5 Модель має вигляд $\Delta\hat{y} = a + bx + ct$.

Застосовуючи МНК, отримаємо систему нормальних рівнянь:

$$\begin{cases} \sum y = na + b\sum x + c\sum t, \\ \sum yx = a\sum x + b\sum x^2 + c\sum xt, \\ \sum yt = a\sum t + b\sum xt + c\sum t^2. \end{cases}$$

Розрахунки оформимо у вигляді таблиці 8.5.

Таблиця 8.5

Числові розрахунки

t	y	x	Yx	y_t	x_t	x^2	t^2
1	30	100	3000	зо	100	10000	1
2	35	103	3605	70	206	10609	4
3	39	105	4095	117	315	11025	9
4	44	109	4796	176	436	11881	16
5	50	115	5750	250	575	13225	25
6	53	118	6254	318	708	13924	36
21	**251**	**650**	**27500**	**961**	**2340**	**70664**	**91**

Система рівнянь приймає вид

$$\begin{cases} 251 = 6a + 650b + 21c, \\ 27500 = 650a + 70664b + 2340c, \\ 961 = 21a + 2340b + 91c. \end{cases}$$

Розв'язуючи її отримаємо $a = -5,42$, $b = 0,322$, $c = 3,516$.

Рівняння регресії має вигляд $y = -5,42 + 0,322x + 3,516t$.

Параметр b = 0,322 фіксує силу зв'язку y і x. Його величина означає, що з ростом доходу на одного члена родини на 1% пункт за умови незмінної тенденції витрати на товар A зростають в середньому на 0,322 грн. Параметр c=3,516 характеризує середньорічний абсолютний приріст витрат на товар A під впливом інших чинників за умови незмінного доходу.

Приклад 3

За даними 30 місяців деякого часового ряду x_t були отримані значення коефіцієнтів автокореляції рівнів:

r_1=0,63; r_2=0,38; r_3=0,72; r_4=0,97; r_5=0,55; r_6=0,40; r_7=0,65;

r_i - коефіцієнти автокореляції i – го порядку.

Треба:

1 Охарактеризувати структуру цього ряду, використовуючи графічне зображення.

2 Для прогнозування значень x_t у майбутні періоди передбачається побудувати рівняння авторегресії. Вибрати найкраще рівняння, обґрунтувати вибір. Вказати загальний вигляд цього рівняння.

Розв'язання.

1 Оскільки значення всіх коефіцієнтів автокореляції достатньо високі, ряд містить тенденцію. Оскільки найбільше абсолютне значення має коефіцієнт автокореляції 4-го порядку r_4, ряд містить періодичні коливання, цикл цих коливань дорівнює 4.

Графік цього ряду можна представити на рисунку 8.1.

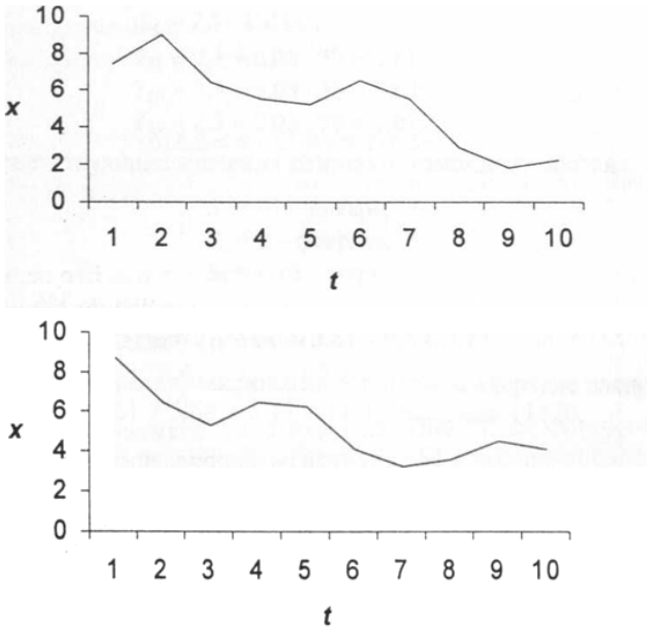

Рис. 8.1

На рисунку 8.1 зображено графіки, що характеризують спадаючу тенденцію при різних можливих періодичних коливаннях.

2 Найбільш доцільна побудова рівняння авторегресії:

$$y_t = a + b \cdot y_{t-4} + u_t,$$

оскільки значення r_4 = 0,97 свідчить про наявність дуже тісного зв'язку між рівнями ряду з лагом в 4 місяці.

Крім того, можлива побудова і множинного рівняння авторегресії y_t від y_{t-3} і y_{t-4}, оскільки r_3 = 0,72:

$$y_t = a + b_1 \cdot y_{t-3} + b_2 \cdot y_{t-4} + u_t,$$

Порівняти отримані рівняння і вибрати найкращий розв’язок можна за допомогою скоректованого коефіцієнта детерміації.

Приклад 4

Побудова адитивної моделі часового ряду.

Нехай маємо данні про об'єм споживання електроенергії мешканцями району за останні 16 кварталів (чотири роки), які представлено в таблиці 8.6.

Таблиця 8.6

Данні про об'єм споживання електроенергії

t	y_t	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}	t	y_t	y_{t-1}	y_{t-2}	y_{t-3}	y_{t-4}
1	**6**	-	-	-	-	**9**	**8**	10	6	4,8	7,2
2	**4,4**	6	-	-	-	**10**	**5,6**	8	10	6	4,8
3	**5**	4,4	6	-	-	**11**	**6,4**	5,6	8	10	6
4	**9**	5	4,4	6	-	**12**	**11**	6,4	5,6	8	10
5	**7,2**	9	5	4,4	6	**13**	**9**	11	6,4	5,6	8
6	**4,8**	7,2	9	5	4,4	**14**	**6,6**	9	11	6,4	5,6
7	**6**	4,8	7,2	9	5	**15**	**7**	6,6	9	11	6,4
8	**10**	6	4,8	7,2	9	**16**	**10,8**	7	6,6	9	11

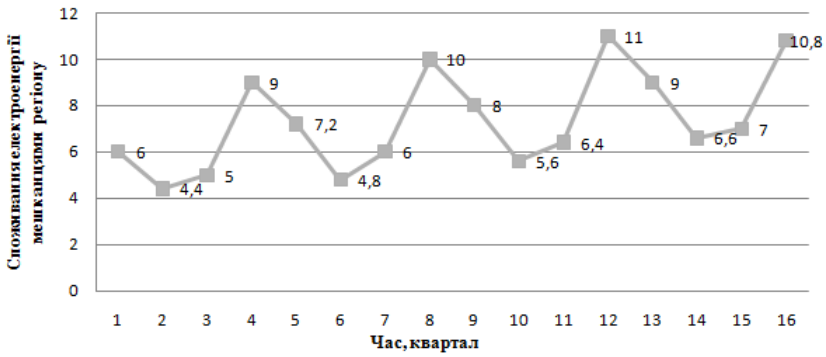

Рис. 8.2

Знайдемо коефіцієнт автокореляції першого порядку цього ряду за формулою:

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)\cdot(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^{n}(y_{t-1} - \bar{y}_2)^2}},$$

де $\bar{y}_1 = \frac{\sum_{t=2}^{n} y_t}{n-1}$; $\bar{y}_2 = \frac{\sum_{t=2}^{n} y_{t-1}}{n-1}$

t	y_t	y_{t-1}	$(y_t - \bar{y}_1)$	$(y_{t-1} - \bar{y}_2)$	$(y_t - \bar{y}_1)\cdot(y_{t-1} - \bar{y}_2$	$(y_t - \bar{y}_1)^2$	$(y_{t-1} - \bar{y}_2)^2$
1	6						
2	4,4	6	-2,98666667	-1,06667	3,185778	8,920178	1,137778
3	5	4,4	-2,38666667	-2,66667	6,364444	5,696178	7,111111
4	9	5	1,613333333	-2,06667	-3,33422	2,602844	4,271111
5	7,2	9	-0,18666667	1,933333	-0,36089	0,034844	3,737778
6	4,8	7,2	-2,58666667	0,133333	-0,34489	6,690844	0,017778
7	6	4,8	-1,38666667	-2,26667	3,143111	1,922844	5,137778
8	10	6	2,613333333	-1,06667	-2,78756	6,829511	1,137778
9	8	10	0,613333333	2,933333	1,799111	0,376178	8,604444
10	5,6	8	-1,78666667	0,933333	-1,66756	3,192178	0,871111
11	6,4	5,6	-0,98666667	-1,46667	1,447111	0,973511	2,151111
12	11	6,4	3,613333333	-0,66667	-2,40889	13,05618	0,444444
13	9	11	1,613333333	3,933333	6,345778	2,602844	15,47111
14	6,6	9	-0,78666667	1,933333	-1,52089	0,618844	3,737778
15	7	6,6	-0,38666667	-0,46667	0,180444	0,149511	0,217778
16	10,8	7	3,413333333	-0,06667	-0,22756	11,65084	0,004444
Середнє	7,3867	7,0667					
Сума			0	0	9,813333	65,31733	54,05333

$$r_1 = \frac{9,813333}{\sqrt{65,31733 \cdot 54,05333}} = 0,165154$$

Аналогічно знайдемо коефіцієнт автокореляції другого порядку за формулою:

$$r_2 = \frac{\sum_{t=3}^{n}(y_t - \bar{y}_3)\cdot(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \bar{y}_3)^2 \cdot \sum_{t=3}^{n}(y_{t-2} - \bar{y}_4)^2}}$$ де $\bar{y}_3 = \frac{\sum_{t=3}^{n} y_t}{n-2}$; $\bar{y}_4 = \frac{\sum_{t=3}^{n} y_{t-2}}{n-2}$

t	y_t	y_{t-2}	$(y_t - \overline{y}_3)$	$(y_{t-2} - \overline{y}_4)$	$(y_t - \overline{y}_3)\cdot(y_{t-2} - \overline{y}_4)$	$(y_t - \overline{y}_3)^2$	$(y_{t-2} - \overline{y}_4)^2$
1	6						
2	4,4						
3	5	6	-2,6	-1,07143	2,785714	6,76	1,147959
4	9	4,4	1,4	-2,67143	-3,74	1,96	7,136531
5	7,2	5	-0,4	-2,07143	0,828571	0,16	4,290816
6	4,8	9	-2,8	1,928571	-5,4	7,84	3,719388
7	6	7,2	-1,6	0,128571	-0,20571	2,56	0,016531
8	10	4,8	2,4	-2,27143	-5,45143	5,76	5,159388
9	8	6	0,4	-1,07143	-0,42857	0,16	1,147959
10	5,6	10	-2	2,928571	-5,85714	4	8,576531
11	6,4	8	-1,2	0,928571	-1,11429	1,44	0,862245
12	11	5,6	3,4	-1,47143	-5,00286	11,56	2,165102
13	9	6,4	1,4	-0,67143	-0,94	1,96	0,450816
14	6,6	11	-1	3,928571	-3,92857	1	15,43367
15	7	9	-0,6	1,928571	-1,15714	0,36	3,719388
16	10,8	6,6	3,2	-0,47143	-1,50857	10,24	0,222245
Середнє	7,6	7,0714286					
Сума					-31,12	55,76	54,04857

$$r_2 = \frac{-31,12}{\sqrt{55,76 \cdot 54,04857}} = -0,566873$$

За знаком коефіцієнта автокореляції не треба робити висновки про зростаючу або спадаючу тенденцію в рівняннях ряду.

Аналогічно знаходяться і решта коефіцієнтів автокореляції r_3, r_4,… . Продовживши обчислювати коефіцієнти автокореляції таким чином, побудуємо автокореляційну функцію, яку надамо у вигляді таблиці

Лаг	Коефіцієнти автокореляції
1	0,1651548
2	-0,566873
3	0,113558
4	0,983025
5	0,1187113
6	-0,722046
7	-0,00337
8	0,973848
…	…

Аналізуючи значення автокореляційної функції робимо висновок, що у часовому ряді, який вивчається, по-перше, має місце лінійна тенденція, по-друге, мають місце сезонні коливання з періодичністю у 4 квартали. Наш висновок підтверджує і рисунок 8.2.

Знайдемо компоненти нашої моделі.

Крок 1 Проведемо вирівнювання початкових рівнів ряду методом ковзаючою середньою. Для цього:

а) підсумуємо рівні ряду послідовно за кожні чотири квартали із зрушенням на один момент часу і визначимо умовні річні об'єми споживання електроенергії (колонка 3 таблиці 8.7);

б) розділивши отримані суми на 4, знайдемо ковзаючі середні (кол. 4 таблиці 8.7). Відмітимо, що отримані таким чином вирівняні значення вже не містять сезонної компоненти;

в) приведемо ці значення у відповідність з фактичними моментами часу, для чого знайдемо середні значення з двох послідовних ковзаючих середніх - центровані ковзаючі середні (кол. 5 таблиця 8.7).

Таблиця 8.7

Вирівнювання початкових рівнів

№ кварталу, t	Споживання електроенергії, y_t	Разом за 4 квартали	Ковзаюча середня за 4 квартали	Центрована ковзаюча середня	Оцінка сезонної компоненти
1	2	3	4	5	6
1	6	-	-	-	-
2	4,4	24,4	6,1	-	-
3	5	25,6	6,4	6,25	-1,25
4	9	26	6,5	6,45	2,55
5	7,2	27	6,75	6,625	0,575
6	4,8	28	7	6,875	-2,075
7	6	28,8	7,2	7,1	-1,1
8	10	29,6	7,4	7,3	2,7
9	8	30	7,5	7,45	0,55
10	5,6	31	7,75	7,625	-2,025
11	6,4	32	8	7,875	-1,475
12	11	33	8,25	8,125	2,875
13	9	33,6	8,4	8,325	0,675
14	6,6	33,4	8,35	8,375	-1,775
15	7	-	-	-	-
16	10,8	-	-	-	-

Крок 2 Знайдемо оцінки сезонної компоненти як різницю між фактичними рівнями ряду і центрованими ковзаючими середніми (кол. 6 таблиці 8.7). Використовуємо ці оцінки для розрахунку значень сезонної компоненти S (таблиця 8.8). Для цього знайдемо середні за кожен квартал (по усіх роках) оцінки сезонної компоненти. В моделях з сезонною компонентою зазвичай передбачається, що сезонний вплив за період взаємознищується. У аддитивній моделі це виражається в тому, що сума значень сезонної компоненти по усіх кварталах має дорівнювати нулю.

Таблиця 8.8

Розрахунок значень сезонної компоненти *S*

Показники	Рік	№ кварталу, і			
		I	II	III	IV
	1	-	-	-1,250	2,550
	2	0,575	-2,075	1,100	2,700
	3	0,550	2,025	1,475	2,875
	4	0,675	1,775	-	-
Разом за і-й квартал (за усі роки)		1,800	-5,875	-3,825	8,125
Середня оцінка сезонної компоненти для і-го кварталу, $\overline{S}_i$		0,600	-1,958	-1,275	2,708
Скоректована сезонна компонента, S_i		0,581	-1,977	-1,294	2,690

Для даної моделі маємо: 0,6-1,958-1,275+2,708=0,075.

Визначимо коректуючий коефіцієнт: *k*=0,075/4=0,01875.

Розрахуємо скоректовані значення сезонної компоненти як різницю між її середньою оцінкою і коректуючим коефіцієнтом *k*: $S_i = \overline{S}_i - k$, де $i = \overline{1,4}$.

Перевіримо умову рівності нулю суми значень сезонної компоненти: 0,581-1,977-1,294+2,69=0.

Таким чином, отримані наступні значення сезонної компоненти :

I квартал: S_1 = 0,581; II квартал: S_2= - 1,979;

III квартал: S_3 = - 1,294; IV квартал: S_4 = 2,690.

Занесемо отримані значення в таблиці 8.8 для відповідних кварталів кожного року (рядок 3).

Крок 3 Віднімаючи, значення сезонної компоненти з кожного рівня початкового часового ряду. Отримаємо величини *T+E=Y-S* (кол. 4 таблиці 8.9). Ці значення розраховуються за кожен момент часу і містять тільки тенденцію і випадкову компоненту.

Таблиця 8.9

t	y_t	S_i	$T+E=$ $=y_t-S_i$	T	$T+S$	$E=y_t-$ $-(T+S)$	$(E)^2$
1	2	3	4	5	6	7	8
1	6,0	0,581	5,419	5,902	6,483	-0,483	0,2333
2	4,4	-1,977	6,337	6,088	4,111	0,289	0,0835
3	5,0	-1,294	6,294	6,275	4,981	0,019	0,0004
4	9,0	2,690	6,310	6,461	9,151	-0,151	0,0228
5	7,2	0,581	6,619	6,648	7,229	-0,029	0,0008
6	4,8	-1,977	6,777	6,834	4,857	-0,057	0,0032

7	6,0	-1,294	7,294	7,020	5,727	0,273	0,0745
8	10,0	2,690	7,310	7;207	9,896	0,104	0,0108
9	8,0	0,581	7,419	7,393	7,974	0,026	0,0007
10	5,6	-1,977	7,577	7,580	5,603	-0,030	0,0009
11	6,4	-1,294	7,694	7,766	6,472	-0,072	0,0052
12	11	2,690	8,310	7,952	10,642	0,358	0,1282
13	9,0	0,581	8,419	8,139	8,720	0,280	0,0784
14	6,6	-1,977	8,577	8,325	6,348	0,252	0,0635
15	7,0	-1,294	8,294	8,519	7,218	-0,218	0,0475
16	10,8	2,690	8,110	8,698	11,388	-0,588	0,3457

Крок 4 Визначимо компоненту *T* цієї моделі. Для цього проведемо аналітичне вирівнювання ряду (*T*+*E*) за допомогою лінійного тренду. Результати аналітичного вирівнювання наступні:

Знайдемо коефіцієнти *a* та *b* для рівняння лінійного тренду методом найменших квадратів. Маємо систему: $\begin{cases} na + b\sum t = \sum (T+E) \\ a\sum t + b\sum t^2 = \sum ((T+E)t) \end{cases}$ з якої знаходимо *a*=5,715416, *b*=0,186421

Обчислимо також:

стандартну помилку коефіцієнта регресії 0,015188,

R^2 0,914971

Число спостережень 16

Число ступенів свободи 14

Таким чином, маємо наступний лінійний тренд: $T = 5{,}715+0{,}186t$. Підставляючи в це рівняння значення t = 1,…,16, знайдемо рівні *T* для кожного моменту часу (кол. 5 таблиці 8.9). Графік рівняння тренду приведений на рисунку 8.3.

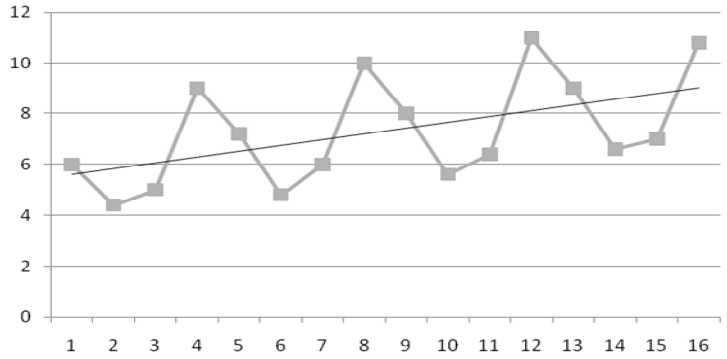

Рис. 8.3

Крок 5 Знайдемо значення рівнів ряду, отримані по адитивній моделі. Для цього додамо до рівнів *T* значення сезонної компоненти для відповідних кварталів. Графічно значення (*T*+S) представлені на малюнку.

Крок 6 Відповідно до методики побудови адитивної моделі розрахунок помилки робиться за формулою $E = Y - (T + S)$.

Це абсолютна помилка. Чисельні значення абсолютних помилок приведені в колонці 7 таблиці 8.9.

По аналогії з моделлю регресії для оцінки якості побудови моделі або для вибору найкращої моделі можна застосувати суму квадратів отриманих абсолютних помилок. Для даної адитивної моделі сума квадратів абсолютних помилок дорівнює 1,1. По відношенню до загальної суми квадратів відхилень рівнів ряду від його середнього рівня, рівною 71,59, ця величина складає трохи більше 1,5%:

$$\left(1 - \frac{1,10}{71,59}\right) \cdot 100 = 1,536$$

.

Отже, можна сказати, що адитивна модель пояснює 98,5% загальної варіації рівнів часового ряду споживання електроенергії за останні 16 кварталів.

Приклад 5

Побудова мультиплікативної моделі часового ряду.

Нехай є поквартальні дані про прибуток компанії за останні чотири роки (таблиця 8.10).

Таблиця 8.10

Поквартальні дані про прибуток компанії

Квартал / Рік	I	II	III	IV
1	72	100	90	64
2	70	92	80	58
3	62	80	68	48
4	52	60	50	30

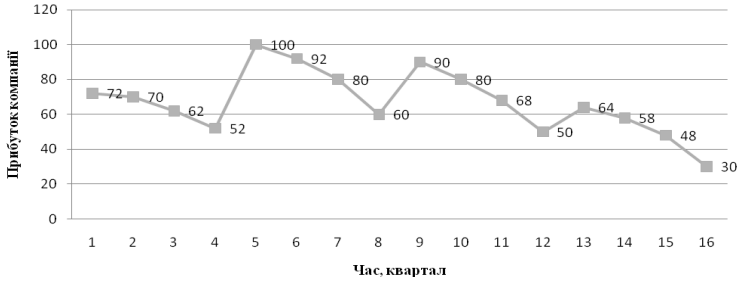

Рис. 8.4

Графік цього тимчасового ряду (рис.8.4) свідчить про наявність сезонних коливань (період коливань дорівнює 4) і загальної спадаючої тенденції рівнів ряду. Прибуток компанії у весняно-літній період вищий, ніж в осінньо-зимовий період. Оскільки амплітуда сезонних коливань зменшується, можна припустити існування мультиплікативної моделі. Визначимо її компоненти.

Крок 1 Проведемо вирівнювання початкових рівнів ряду методом ковзаючою середньою. Методика, вживана на цьому кроці, повністю співпадає з методикою адитивної моделі. Результати розрахунків оцінок сезонної компоненти представлені в таблиці 8.11.

Таблиця 8.11

Результати розрахунків оцінок сезонної компоненти

№ кварталу, t	Споживання електро-енергії, y_t	Разом за 4 квартали	Ковзаюча середня за 4 квартали	Центрована ковзаюча середня	Оцінка сезонної компоненти
1	**2**	**3**	**4**	**5**	**6**
1	72	-	-	-	-
2	70	326	81,5	-	-
3	62	324	81	81,25	1,108
4	52	316	79	80	0,8
5	100	306	76,5	77,75	0,9
6	92	300	75	75,75	1,215
7	80	292	73	74	1,081
8	60	280	70	71,5	0,811
9	90	268	67	68,5	0,905
10	80	258	64,5	65,75	1,217
11	68	248	62	63,25	1,075
12	50	228	57	59,5	0,807
13	64	210	52,5	54,75	0,95
14	58	192	48	50,25	1,194
15	48	-	-	-	-
16	30	-	-	-	-

Крок 2 Знайдемо оцінки сезонної компоненти як частку від ділення фактичних рівнів ряду на центровані ковзаючі середні (кол. 6 таблиці 8.11). Використовуємо ці оцінки для розрахунку значень сезонної компоненти S (таблиця 8.12). Для цього знайдемо середні за кожен квартал оцінки сезонної компоненти S_i. Взаємознищування сезонних впливів в мультиплікативній моделі виражається у тому, що сума значень сезонної компоненти по усіх кварталах має дорівнювати числу періодів у циклі. У нашому випадку число періодів одного циклу (рік) дорівнює 4 (чотири квартали).

Таблиця 8.12

Розрахунок значень сезонної компоненти S

Показники	Рік	№ кварталу, i			
		I	II	III	IV
	1	-	-	1,108	0,800
	2	0,900	1,215	1,081	0,817
	3	0,905	1,217	1,075	0,807
	4	0,950	1,194	-	-
Разом за і-й квартал (за усі роки)		2,755	3,626	3,264	2,424
Середня оцінка сезонної		0,918	1,209	1,088	0,808

компоненти для і-го кварталу, $\overline{S}_i$					
Скоректована сезонна компонента, S_i		0,913	1,202	1,082	0,803

Маємо: 0,918+1,209+1,088+0,808=4,023

Визначимо коректуючий коефіцієнт: k=4/4,023=0,9943.

Визначимо скоректовані значення сезонної компоненти, помноживши її середні оцінки на коректуючий коефіцієнт k.

$S_i = \overline{S}_i \cdot k$, де $i = \overline{1,4}$.

Перевіримо умову рівності чотирьом суми значень сезонної компоненти:

$$0{,}913 + 1{,}202 + 1{,}082 + 0{,}803 = 4.$$

Отримаємо наступні значення сезонної компоненти :

I квартал: S_1 = 0,913; II квартал: S_2= 1,202;

III квартал: S_3 = 1,082; IV квартал: S_4 = 0,803.

Занесемо отримані значення до таблиці 8.13 для відповідних кварталів кожного року (кол. 3).

Таблиця 8.13

t	y_t	S_i	$TE=y_t/S_i$	T	TS	$E=y_t:(T+S)$	$E=y_t-(T+S)$	$(E)^2$
1	2	3	4	5	6	7	8	9
1	72	0,913	78,86	87,80	80,16	0,898	-8,16	66,66
2	100	1,202	83,19	85,03	102,20	0,978	-2,20	4,86
3	90	1,082	83,18	82,25	89,00	1,011	1,00	1,00
4	64	0,803	79,70	79,48	63,82	1,003	0,18	0,03
5	70	0,913	76,67	76,70	70,03	1,000	-0,03	0,00
6	92	1,202	76,54	73,93	88,86	1,035	3,14	9,85
7	80	1,082	73,94	71,15	76,99	1,039	3,01	9,08
8	58	0,803	72,23	68,38	54,91	1,056	3,09	9,57
9	62	0,913	67,91	65,60	59,90	1,035	2,10	4,43
10	80	1,202	66,56	62,83	75,52	1,059	4,48	20,08
11	68	1,082	62,85	60,05	64,98	1,047	3,02	9.14
12	48	0,803	59,78	57,28	45,99	1,044	2,01	4,03
13	52	0,913	56,96	54,50	49,76	1,045	2,24	5,02
14	60	1,202	49,92	51,73	62,18	0,965	-2,18	4,73
15	50	1,082	46,21	48,95	52,97	0,944	-2,97	8,79
16	30	0,803	37,36	46,18	37,08	0,809	-7.08	50,12

Крок 4 Визначимо компоненту T в мультиплікативній моделі. Для цього розрахуємо параметри лінійного тренду, використовуючи рівні (T-E). Результати аналітичного вирівнювання цього ряду представлені нижче:

Константа a 90,585150

Коефіцієнт регресії b - 2,773250

Стандартна помилка коефіцієнта регресії 0,225556

R^2 0,915239

Число спостережень 16

Число ступенів свободи 14

Рівняння тренду має наступний вигляд: $T = 90{,}59 - 2{,}773t$.

Підставляючи в це рівняння значення t = 1,…,16, знайдемо рівні T для кожного моменту часу (кол. 5 таблиці 8.13). Графік рівняння тренду приведений на рисунку 8.5.

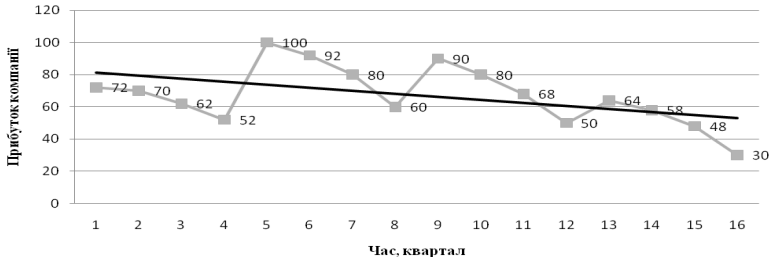

Рис. 8.5

Крок 5 Знайдемо рівні ряду по мультиплікативній моделі, помноживши рівні T на значення сезонної компоненти для відповідних кварталів. Графічно значення $(T \cdot S)$ представлені на рис. 8.5.

Крок 6 Розрахунок помилки в мультиплікативній моделі виконується за формулою $E=Y:(T \cdot S)$

Чисельні значення помилки приведені в колонці 7 таблиці 8.13. Якщо часовий ряд помилок не містить автокореляції, його можна використовувати замість початкового ряду для вивчення його взаємозв'язку з іншими часовими рядами. Для того, щоб порівняти мультиплікативну модель і інші моделі часового ряду, можна по аналогії з адитивною моделлю використовувати суму квадратів абсолютних помилок. Абсолютні помилки в мультиплікативній моделі визначаються як $E=y_t - (T \cdot S)$.

У цій моделі сума квадратів абсолютних помилок складає 207,40. Загальна сума квадратів відхилень фактичних рівнів цього ряду від середнього значення дорівнює 5023. Таким чином, доля поясненої дисперсії рівнів ряду дорівнює: (1 - 207,40/5023) = 0,959, або 95,9%.

Виявлення і усунення сезонного ефекту використовуються в двох напрямах. По-перше, дію сезонних коливань слід усувати на етапі попередньої обробки початкових даних при вивченні взаємозв'язку декількох часових рядів. Тому в російських і міжнародних статистичних збірках часто публікуються дані, в яких усунений вплив сезонної компоненти (якщо це щомісячна або поквартальна статистика), наприклад показники обсягів виробництва в окремих галузях промисловості, рівня безробіття і т.д.. По-друге, це аналіз структури одномірних часових рядів з метою прогнозування рівнів ряду у майбутні моменти часу.

Вправи

Задача 1 Адміністрація банку вивчає динаміку депозитів фізичних осіб за кілька років (млн дол. в порівнянних цінах). Початкові дані представлені нижче:

								Сума
Час, років	1	2	3	4	5	6	7	28
Депозити фізичних осіб, х	2	6	7	3	10	12	13	53

Відоме також наступне $\sum x^2 = 511$.

Побудуйте рівняння лінійного тренду і дайте інтерпретацію його параметрів. Визначите коефіцієнт детермінації для лінійного тренду.

Адміністрація банку припускає, що середньорічний абсолютний приріст депозитів фізичних осіб складає не менше 2,5 млн дол. Чи підтверджується це припущення результатами, які ви отримали?

Задача 2 Вивчається динаміка споживання м'яса в регіоні. Дня цього були зібрані дані про об'єми средньодушевого споживання м'яса y_t (кг) за 7 місяців. Попередня обробка даних шляхом логарифмування привела до отримання наступних результатів :

Місяць	1	2	3	4	5	6	7
ln y_t	2,10	2,11	2,13	2,17	2,22	2,28	2,31

Побудуйте рівняння експоненціального тренду. Дайте інтерпретацію його параметрів.

Задача 3 Маємо дані про врожайність зернових в господарствах області :

Рік	1	2	3	4	5	6	7	8
Врожайність зернових, ц/га	10,2	10,7	11,7	13,1	14,9	17,2	20,0	23,2

Обґрунтуйте вибір типу рівняння тренду. Розрахуйте параметри рівняння тренду. Дайте прогноз врожайності зернових наступного року.

Задача 4 Маємо наступні дані про рівень безробіття у, (%) за 8 місяців:

Місяць	1	2	3	4	5	6	7	8
y_t	8,8	8,6	8,4	8,1	7,9	7,6	7,4	7,0

Визначите коефіцієнти автокореляції рівнів цього ряду першого і другого порядку. Обґрунтуйте вибір рівняння тренду і визначите його параметри. Інтерпретуйте отримані результати.

Задача 5 Нехай маємо наступний часовий ряд:

Місяць	1	2	3	4	5	6	7	8
x_t	20	…	…	…	…	…	…	10

Відомо також, що $\sum x_t = 150$, $\sum x_t^2 = 8100$, $\sum x_t x_{t-1} = 7350$.

Визначите коефіцієнт автокореляції рівнів цього ряду першого порядку. Встановіть, чи включає досліджуваний часовий ряд тенденцію.

Задача 6 Експорт, імпорт, зовнішньоторговий обіг Австрії і Бельгії за 1961 - 1995 рр. характеризуються такими даними.

	Австралія, млн шилінгів			Бельгія, млн франків		
Рік	Експорт	Імпорт	Зовнішньоторговий обіг	Експорт	Імпорт	Зовнішньоторговий обіг
1961	44	43	87	202	209	411
1962	47	46	93	219	221	440
1963	51	51	102	239	248	487
1964	56	56	112	278	283	561
1965	62	63	125	306	305	611
1966	67	71	138	328	337	665
1967	72	74	146	352	351	703
1968	79	80	159	402	400	802
1969	95	91	186	483	474	957
1970	117	131	248	562	533	1095
1971	129	126	255	609	581	1190
1972	146	144	290	683	633	1316
1973	166	164	330	846	811	1657
1974	204	206	410	1116	1109	2225
1975	209	205	414	1065	1061	2126
1976	236	247	483	1266	1261	2527
1977	257	278	535	1474	1499	2973
1978	281	280	561	1540	1570	3110
1979	328	332	660	1798	1866	3664
1980	366	386	752	2026	2125	4151
1981	405	419	824	2286	2357	4643
1982	431	412	843	2640	2694	5334
1983	450	434	884	2924	2864	5788
1984	498	496	994	3337	3277	6614
1985	549	547	1096	3479	3379	6858
1986	523	510	1033	3367	3187	6554
1987	527	520	1047	3477	3334	6811
1988	590	584	1174	3900	3719	7619
1989	669	661	1330	4498	4320	8818
1990	737	720	1457	4660	4506	9166
1991	775	758	1533	4846	4658	9504
1992	792	772	1564	4980	4713	9693
1993	787	773	1560	5012	4674	9686
1994	835	842	1677	5491	5108	10599
1995	887	911	1798	5764	5377	11141

По кожному ряду побудуйте графік динаміки. Проведіть розрахунок параметрів трендів різної форми. Оцініть якість кожного тренду через середню помилку апроксимації, лінійний коефіцієнт автокореляції відхилень. Оцініть статистичну значущість трендів через F-критерій, значущість параметрів тренду – через t-критерій. Виберіть кращу форму тренду і виконаєте по ній точковий прогноз на 1998 г. Оцініть помилку прогнозу і побудуйте довірчий інтервал прогнозу для рівня значущості 0,05.

Питання для самоперевірки

1 Перелічити основні елементи часового ряду.

2 Що таке автокореляція рівнів часового ряду?

3 Дати означення автокореляційної функції часового ряду.

4 Перелічити основні види трендів.

5 Як інтерпретувати параметри лінійного і експоненційного трендів?

6 Записати загальний вигляд мультиплікативної і адитивної моделей часового ряду.

7 Перелічити етапи побудови мультиплікативної і адитивної моделей часового ряду.

Бібліографічний список

1 Аршава О.О., Поклонський Є.В., Стасенко О.М., Харченко А.П., Щелкунова Л.І.. Економетрія в прикладах і задачах. Навчально-методичний посібник. – Харків, ХНУБА, 2015.

2 Боровиков В.П. Statistica: искусство анализа данных на компьютере. – Питер, 2003.

3 Бородич С.А. Эконометрика – Минск: Новое знание, 2001.

4 Грубер Й. Эконометрия: Учебное пособие для студентов экономических специальностей. Т. 1: Введение в эконометрию. – К., 1996.

5 Дорохина Е.Ю. и др. Сборник задач по эконометрике. – М.:Экзамен, 2003.

6 Доугерти К. Введение в эконометрию. – М.: ИНФРА, 2009.

7 Елисеева И.И. и др. Практикум по эконометрике.– М.: Финансы и статистика, 2001.

8 Клебанова Т., Дубровина Н., Раевнева Е. Эконометрия - Х., 2005.

9 Леоненко М.М., Мішура Ю.С., Пархоменко В.М., Ядренко М.Й. Теоретико-ймовірнісні та статистичні методи в економетриці та фінансовій математиці. – К.: Інформтехніка, 1995.

10 Лондар С.Л., Юринець Р.В. Економетрія засобами MS Excel – К, 2004.

11 Лук’яненко І., Краснікова Л. Економетрия: Підручник. – К.:Знання, 1998.

12 Лук’яненко І., Краснікова Л. Економетрия: Практикум з використанням комп’ютера. – К: Знання, 1998.

13 Лук'яненко І.Г., Городніченко Ю.О. Сучасні економетричні методи в фінансах. – К.:Літера, 2003.

14 Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Навчальный курс -М.: Дело, 2004.

15 Назаренко О.М. Основи економетрики. – Київ, 2004.

16 Наконечний С.І. і ін. Економетрія – Київ, 2004.

17 Наконечний С.І., Терещенко Т.О., Водзянова Н.К., Роскач О.С. Практикум з економетрії: Навч. посібник. – К.: КНЕУ, 1998.

18 Черняк О.І., Ставицький А.В. Динамічна економетрія: Навчальний посібник. – К.: КНЕУ, 2000.

19 Аршава О.О. Тексти лекцій з дисципліни «Економіко-математичні методи та моделі (економетрика)» для студентів спеціальності 051 «Економіка». – ХНУБА, 2018.

ЗМІСТ

Навчальне видання

Методичні вказівки до практичних занять з дисципліни «Економіко-математичні методи та моделі (економетрика)» для студентів спеціальності 051 «Економіка»

Укладач: Аршава Олена Олександрівна

Відповідальний за випуск А.П. Харченко

Редактор Пуцик В.І.

План 2019 р., поз. 87
Підп. до друку 11.09.18
Надруковано на різографі.
Тираж 50 прим.

Формат 60х84 1/16.
Обл. -вид. арк. 2,5
Ум. друк. арк. 2,3
Зам. № 5355

Папір друк. №2.
Безкоштовно.

ХНУБА, 61002, Харків, вул. Сумська, 40

Підготовлено та надруковано РВВ Харківського національного університету будівництва та архітектури