Тема 1.3.
ЛІНІЙНІ ЕКОНОМЕТРИЧНІ МОДЕЛІ
План

1. Загальні поняття і визначення.
2. Оцінювання параметрів моделі.
3. Верифікація моделі.
4. Прогнозування і економіко-математичний аналіз.
5. Методи побудови загальної лінійної економетричної моделі.

Теоретична (“канонічна”) загальна лінійна економетрична модель

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_m x_m + \varepsilon \qquad (1)$$

де y – залежна (пояснювана) змінна моделі,

x_1, x_2, … , x_m – незалежні (пояснюючі) змінні моделі або фактори,

β_0, β_1, …. , β_m – параметри моделі (коефіцієнти регресії),

ε – стохастична складова моделі,

m – кількість пояснюючих змінних моделі.

Вибіркова (емпірична) загальна лінійна економетрична модель

$$y = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_m x_m + e \qquad (2)$$

де y – залежна (пояснювана) змінна моделі,

x_1, x_2, … , x_m – незалежні (пояснюючі) змінні моделі (фактори),

b_0, b_1, b_m – параметри вибіркової моделі,

e – залишки моделі.

Вибіркова (емпірична) функція регресії

для загальної лінійної економетричної моделі

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \cdots + b_m x_m \qquad (3)$$

де $\hat{y}$ – оцінка математичного сподівання залежної (пояснюваної) змінної моделі,

$x_1, x_2, \ldots, x_m$ – незалежні (пояснюючі) змінні моделі (фактори),

b_0, b_1, b_m – параметри вибіркової регресії.

ВИБІРКОВА МОДЕЛЬ У МАТРИЧНОМУ ВИГЛЯДІ

$$Y = XB + e$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix}, \dim Y = n \times 1 \quad X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{m1} \\ 1 & x_{12} & x_{22} & \cdots & x_{m2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{mn} \end{pmatrix}, \dim X = n \times k \quad B = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \cdots \\ b_m \end{pmatrix}, \dim B = k \times 1 \quad e = \begin{pmatrix} e_1 \\ e_2 \\ \cdots \\ e_n \end{pmatrix}, \dim e = n \times 1$$

де Y – вектор спостережень за залежною змінною моделі ;

X – матриця спостережень за пояснюючими змінними моделі;

B – вектор оцінок параметрів моделі (вектор параметрів вибіркової моделі) ;

e – вектор залишків моделі.

n – розмір статистичної вибірки (кількість спостережень в статистичній вибірці);

m – число незалежних (пояснюючих) змінних моделі;

k = **m** + **1** – число параметрів моделі.

Теоретична ("канонічна") модель парної лінійної регресії

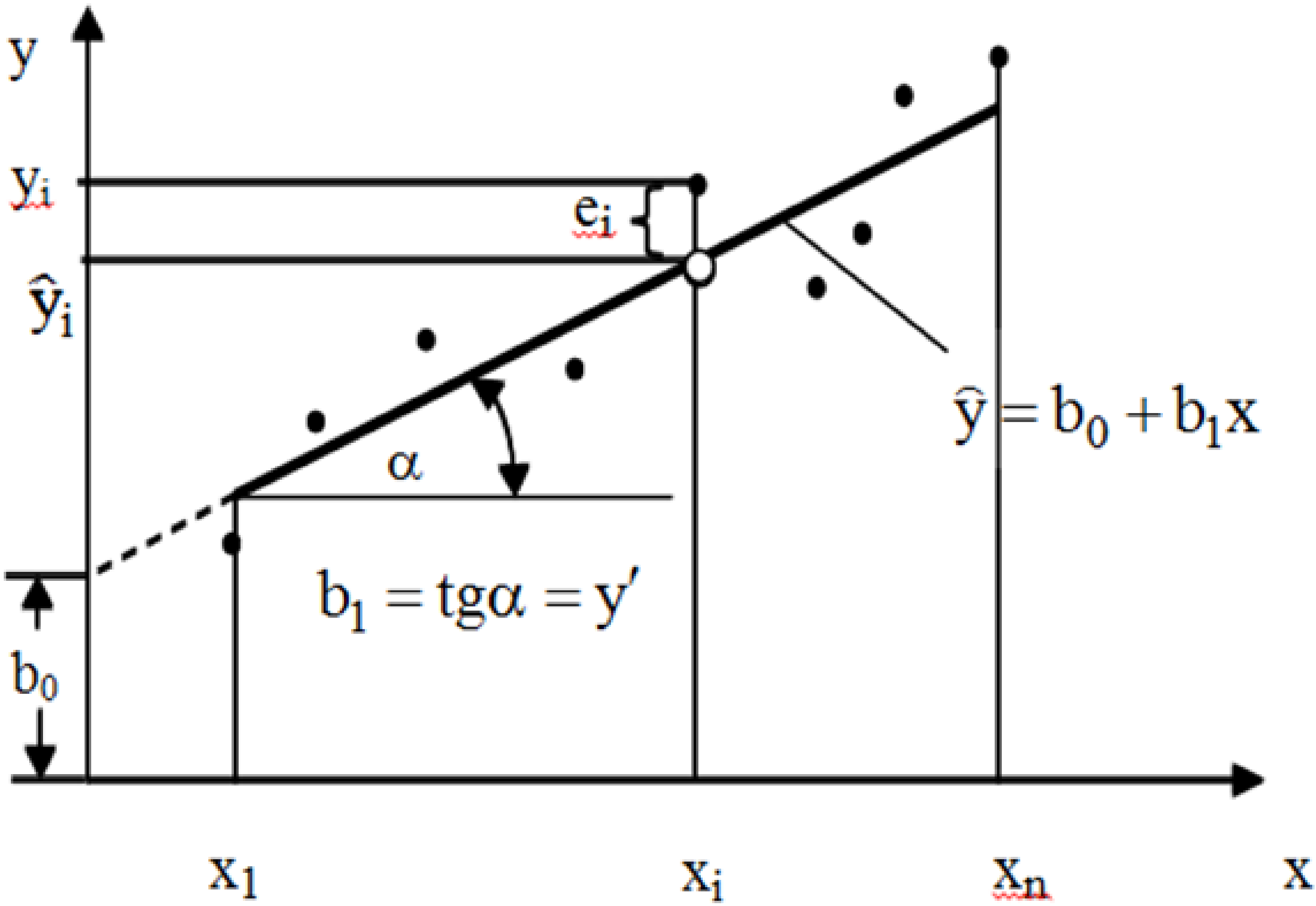
$$y = \beta_0 + \beta_1 x + \varepsilon \tag{4}$$

Вибіркова (емпірична) модель парної лінійної регресії

$$y = b_0 + b_1 x + e \tag{5}$$

Вибіркова функція парної лінійної регресії

$$\widehat{y} = b_0 + b_1 x \tag{6}$$



x	y
x_1	y_1
x_2	y_2
…	…
x_i	y_i
…	…
…	…
x_n	y_n

Рис. 1. Парна лінійна регресія

❖**<u>Припущення 1.</u>** ***Математичне сподівання стохастичної складової моделі дорівнює нулю для всіх спостережень***

$$M(\varepsilon_i) = 0\,,\ i = 1,2,\ \ldots\ ,n$$

❖**<u>Припущення 2.</u>** ***Гомоскедастичність***

$$\operatorname{var}(\varepsilon_i) = \sigma_\varepsilon^2 = \text{const}\,,\ i = 1,2,\ \ldots\ ,n$$

❖**<u>Припущення 3.</u>** ***Відсутність автокореляції залишків***

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0 \ \text{ для } \ i \neq j$$

❖**<u>Припущення 4.</u>** ***Незалежні (пояснюючи) змінні не пов'язані із стохастичною складовою моделі***

$$\operatorname{cov}(\varepsilon_i, x_{ji}) = 0 \ \text{ для } \ i = 1,2,\ \ldots\ ,n;\ j = 1,2,\ldots,m$$

❖**<u>Припущення 5.</u>** ***Матриця спостережень X не є стохастичною***

❖**<u>Припущення 6.</u>** ***Відсутність мультиколіарності***

❖**<u>Припущення 7.</u>** ***Випадкова величина*** $\varepsilon_i\,, i = 1,2,\ldots,n$ ***має нормальний закон розподілу з математичним сподіванням О і сталою дисперсією***

$$\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$$

КРИТЕРІЙ МЕТОДУ НАЙМЕНШИХ КВАДРАТІВ (1МНК)

$$\sum_{i=1}^{n} e_i^2 \to \min \qquad (7)$$

$$e_i = y_i - \widehat{y}_i, \quad \left(i = \overline{1, n}\right) \qquad (8)$$

Подамо рівняння $Y = XB + e$ у вигляді

$$e = Y - XB$$

Тоді суму квадратів залишків **e** можна записати наступним чином

$$\sum_{i=1}^{n} e_i^2 = e'e = (Y - XB)'(Y - XB) = Y'Y - 2B'X'Y + B'X'XB$$

$$\frac{\partial \sum_{i=1}^{n} e_i^2}{\partial B} = \frac{\partial (e'e)}{\partial B} = -2X'Y + 2X'XB = 0$$

$$X'XB = X'Y$$

Помножимо обидві частини виразу на обернену матрицю $(X'X)^{-1}$

$$(X'X)^{-1}X'XB = (X'X)^{-1}X'Y$$

Оскільки

$$(X'X)^{-1}X'X = E$$

де E – одинична матриця

Остаточно отримуємо

$$B = (X'X)^{-1}X'Y \tag{9}$$

$$(X'X) = \begin{pmatrix} n & \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{2i} & \cdots & \sum_{i=1}^{n} x_{mi} \\ \sum_{i=1}^{n} x_{1i} & \sum_{i=1}^{n} x_{1i}^2 & \sum_{i=1}^{n} x_{1i}x_{2i} & \cdots & \sum_{i=1}^{n} x_{1i}x_{mi} \\ \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{2i}x_{1i} & \sum_{i=1}^{n} x_{2i}^2 & \cdots & \sum_{i=1}^{n} x_{2i}x_{mi} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum_{i=1}^{n} x_{mi} & \sum_{i=1}^{n} x_{mi}x_{1i} & \sum_{i=1}^{n} x_{mi}x_{2i} & \cdots & \sum_{i=1}^{n} x_{mi}^2 \end{pmatrix}$$

$\dim(X'X) = k \times k$.

$$X'Y = \begin{pmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_{1i} \\ \sum_{i=1}^{n} y_i x_{2i} \\ \cdots \\ \sum_{i=1}^{n} y_i x_{mi} \end{pmatrix}, \dim X'Y = k \times 1$$

У випадку парної лінійної регресії

$$b_1 = \frac{\operatorname{cov}(x, y)}{\operatorname{var}(x)}$$

$$b_0 = \bar{y} - b_1\bar{x}$$

де $\operatorname{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ – вибірковий коефіцієнт коваріації,

$\operatorname{var}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ – вибіркова дисперсія пояснюючої змінної моделі,

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ – середнє значення пояснюючої змінної **x**,

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ – середнє значення залежної змінної **y**.

Властивості рівняння регресії,

параметри якого оцінені методом найменших квадратів

1. Пряма регресії проходить через середню точку з координатами $\bar{y},\ \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_m$.
2. Середнє значення оцінки залежної змінної дорівнює фактичному середньому значенню

$$\bar{\hat{y}} = \bar{y}$$

3. Сума залишків моделі дорівнює нулю

$$\sum_{i=1}^{n} e_i = 0$$

ВЛАСТИВОСТІ ОЦІНОК ПАРАМЕТРІВ МОДЕЛІ, ОТРИМАНИХ 1МНК

BLUE–оцінки

(best linear unbiased estimators)

найкращі незміщені лінійні оцінки

1.Вони є **незміщеними оцінками**

$$M(b_j)=\beta_j, \ \left(j=\overline{0,m}\right)$$

2.Вони є **ефективними**

3.Вони є **обгрунтованими** оцінками

$$\lim P\left\{\left|b_j-\beta_j\right|<\delta\right\}=1$$

при довільному $\delta > 0$

ВЕРИФІКАЦІЯ МОДЕЛІ

<u>Показники якості моделі</u>

А) **<u>Стандартна похибка рівняння регресії</u>**

$$\hat{\sigma}_{\varepsilon} = \sqrt{\hat{\sigma}_{\varepsilon}^2} \qquad (10)$$

де $\hat{\sigma}_{\varepsilon}^2$ - незміщена оцінка дисперсії залишків

$$\hat{\sigma}_{\varepsilon}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-k} \qquad (11)$$

$$e_i = y_i - \hat{y}_i\,, \quad \left(i = \overline{1,n}\right)$$

де y_i – фактичні значення залежної змінної моделі,

$\hat{y}_i$ – розрахункові значення залежної змінної моделі

$$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + \cdots + b_m x_{mi}\,, \quad \left(i = \overline{1,n}\right)$$

Б) **<u>Стандартні похибки параметрів моделі</u>**

$$\hat{\sigma}_{b_j} = \sqrt{\hat{\sigma}^2_{b_j}}\,, \quad \left(j = \overline{1, m}\right) \tag{12}$$

де $\hat{\sigma}^2_{b_j}$ – оцінка дисперсії j–го параметра моделі

Дисперсійно–коваріаційна матриця оцінок параметрів моделі

$$\operatorname{var}-\operatorname{cov}(b) = \begin{pmatrix} \hat{\sigma}^2_{b_0} & \operatorname{cov}(b_0, b_1) & \cdots & \operatorname{cov}(b_0, b_m) \\ \operatorname{cov}(b_1, b_0) & \hat{\sigma}^2_{b_1} & \cdots & \operatorname{cov}(b_1, b_m) \\ \cdots & \cdots & \cdots & \cdots \\ \operatorname{cov}(b_m, b_0) & \operatorname{cov}(b_m, b_1) & \cdots & \hat{\sigma}^2_{b_m} \end{pmatrix} \tag{13}$$

$$\operatorname{var}-\operatorname{cov}(b) = \hat{\sigma}^2_{\varepsilon} \left(X' X\right)^{-1} \tag{14}$$

У випадку парної лінійної регресії

$$\hat{\sigma}_{b_0}^2 = \hat{\sigma}_{\varepsilon}^2 \frac{\sum_{i=1}^{n} x_i^2}{n \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\sigma}_{b_1}^2 = \hat{\sigma}_{\varepsilon}^2 \frac{1}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

В) <u>Коефіцієнт множинної кореляції</u>

$$R = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^{n} \left(\hat{y}_i - \bar{\hat{y}}\right)^2}} \qquad (15)$$

де y_i – фактичні (статистичні) значення залежної змінної,

$\hat{y}_i$ – розрахункові значення.

$$-1 \le R \le 1 \qquad (16)$$

Коефіцієнта парної кореляції

$$r_{yx} = \frac{\operatorname{cov}(x, y)}{\sqrt{\operatorname{var}(x)}\,\sqrt{\operatorname{var}(y)}} \qquad (17)$$

де $\operatorname{cov}(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$ – вибірковий коефіцієнт коваріації,

$\operatorname{var}(x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$ – вибіркова дисперсія пояснюючої змінної моделі,

$\operatorname{var}(y) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$ – вибіркова дисперсія залежної змінної моделі,

$$-1 \le r_{yx} \le 1 \qquad (18)$$

Г) <u>Коефіцієнт детермінації</u>

$$R^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \qquad (19)$$

для множинної регресії

$$R^2 = (R)^2 \qquad (20)$$

для парної регресії

$$R^2 = (r_{yx})^2 \qquad (21)$$

$$0 \le R^2 \le 1 \qquad (22)$$

Перевірка статистичної значущості і інтервали довіри

А) **Перевірка адекватності (статистичної значущості) моделі у цілому**

нульова гіпотеза

$$H_0 : \quad R^2 = 0$$

(23)

альтернативна гіпотеза

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

$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_m = 0,$$

H_1 : значення хоча б одного параметра відмінне від нуля.

F–статистика Фішера

$$F^* = \frac{R^2}{1-R^2}\cdot\frac{n-k}{m} \qquad (24)$$

Ступені вільності

$$\nu_1 = m \qquad\qquad \nu_2 = n-k$$

Рівень значущості

α (α = 0,05 або α = 0,01)

Критичне значення критерію Фішера

$$F_{кр}$$

Умова перевірки адекватності (статистичної значущості) моделі у цілому

$$F^* > F_{кр}$$

Б) **<u>Перевірка статистичної значущості параметрів моделі.</u>**

<u>Інтервали довіри для параметрів моделі</u>

нульова гіпотеза

$$H_0 : \beta_j = 0, (j = \overline{0, m}),$$

(25)

альтернативна гіпотеза

$$H_1 : \beta_j \neq 0, (j = \overline{0, m}).$$

t–статистика Ст'юдента

$$t^{*}_{b_j} = \frac{b_j}{\hat{\sigma}_{b_j}}, \left(j = \overline{0, m}\right) \qquad (26)$$

де b_j – оцінка параметра β_j теоретичної регресії

$\hat{\sigma}_{b_j}$ – стандартна похибка j-го параметра моделі

<u>Ступень вільності</u>

$$\nu = n - k$$

Критичне значення критерію Ст'юдента

$$t_{\text{кр}} = t_{\alpha/2}$$

Умова перевірки статистичної значущості параметрів моделі

$$\left| t^{*}_{b_j} \right| > t_{\kappa p}$$

Рівень довіри

$$p = 1 - \alpha$$

Довірчі інтервали для параметрів моделі

$$b_j - t_{\alpha/2} \cdot \hat{\sigma}_{b_j} < \beta_j < b_j + t_{\alpha/2} \cdot \hat{\sigma}_{b_j}, \left(j = \overline{0, m}\right), \qquad (27)$$

або

$$\beta_j = b_j \pm t_{\alpha/2} \cdot \hat{\sigma}_{b_j}, \quad \left(j = \overline{0, m}\right)$$

де b_j – оцінка параметра моделі (вибірковий параметр моделі)

$\hat{\sigma}_{b_j}$ – стандартна похибка цього параметра,

$t_{\alpha/2}$ – критичне значення критерію Ст'юдента для рівня значущості α і ступеня вільності $\nu = n - k$.

В) <u>Перевірка статистичної значущості коефіцієнта кореляції</u>.

нульова гіпотеза

$$H_0 : \quad \rho = 0,$$

(28)

альтернативна гіпотеза

$$H_1 : \quad \rho \neq 0,$$

де ρ – коефіцієнт кореляції всієї генеральної сукупності (точніше його гіпотетичне значення).

$$t_R^* = \frac{R\sqrt{n-k}}{\sqrt{1-R^2}} , \qquad (29)$$

де R – вибірковий коефіцієнт кореляції,

R^2 - вибірковий коефіцієнт детермінації.

$$\left| t_R^* \right| > t_{кр} ,$$

вектор прогнозних значень пояснюючих змінних

$$X_{pr} = \begin{pmatrix} 1 \\ x_{1,pr} \\ x_{2,pr} \\ \dots \\ x_{m,pr} \end{pmatrix}, \qquad (30)$$

де $x_{1,pr}, x_{2,pr}, \dots, x_{m,pr}$ – прогнозні значення пояснюючих змінних.

Точковий прогноз

$$\hat{y}_{pr} = X'_{pr} \cdot B, \qquad (31)$$

де X'_{pr} – транспонований вектор прогнозних значень пояснюючих змінних моделі,

B – вектор оцінок параметрів моделі.

<u>Інтервальний прогноз</u>

Інтервальний прогноз для математичного сподівання

$$M(y_{pr}) = \hat{y}_{pr} \pm \hat{\sigma}_{\varepsilon} \cdot t_{\alpha/2} \sqrt{X'_{pr}(X'X)^{-1} X_{pr}} \qquad (32)$$

Інтервальний прогноз для індивідуального значення

$$y_{pr} = \hat{y}_{pr} \pm \hat{\sigma}_{\varepsilon} \cdot t_{\alpha/2} \sqrt{1 + X'_{pr}(X'X)^{-1} X_{pr}} \qquad (33)$$

де $\hat{y}_{pr}$ – точкове прогнозне значення,

$\hat{\sigma}_{\varepsilon}$ – стандартна похибка рівняння регресії ,

$t_{\alpha/2}$ – критичне (табличне) значення критерію Ст'юдента для рівня значимості α і ступеня вільності $\nu = n - k$.

У випадку парної лінійної регресії

$$\widehat{y}_{pr} = b_0 + b_1 \cdot x_{pr}$$

$$M(y_{pr}) = \widehat{y}_{pr} \pm \widehat{\sigma}_{\varepsilon} \cdot t_{\alpha/2} \sqrt{\frac{1}{n} + \frac{\left(x_{pr} - \bar{x}\right)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

$$y_{pr} = \widehat{y}_{pr} \pm \widehat{\sigma}_{\varepsilon} \cdot t_{\alpha/2} \sqrt{1 + \frac{1}{n} + \frac{\left(x_{pr} - \bar{x}\right)}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

Економіко-математичний аналіз

А. ***Показники середньої ефективності впливу пояснюючих змінних на залежну змінну***

$$A_j = \frac{\overline{\hat{\bar{y}}}}{\bar{x}_j}, \left(j = \overline{1, m}\right) \qquad (34)$$

де $\bar{\hat{\bar{y}}}$ – середнє значення залежної змінної, обчислене для масиву розрахункових значень залежної змінної,

$\bar{x}_j$ – середнє значення j – ї пояснюючої змінної.

Б. ***Показники граничної ефективності впливу пояснюючих змінних на залежну змінну***

$$M_j = \frac{\partial \widehat{y}}{\partial x_j} = b_j, \left(j = \overline{1, m}\right) \qquad (35)$$

де b_j – параметр вибіркового рівняння регресії при пояснюючій змінній $\bar{x}_j$.

B. ***Показники відносного впливу пояснюючих змінних на залежну змінну***

$$\overline{E}_{y/x_j} = \frac{M_j}{A_j} = b_j \frac{\overline{x}_j}{\overline{\hat{y}}}, \left(j = \overline{1, m} \right) \qquad (36)$$

$$p = \sum_{j=1}^{m} \overline{E}_{y/x_j} \qquad (37)$$

Г. ***Показники сили впливу пояснюючих змінних на залежну змінну***

$$b_j^* = b_j \frac{\hat{\sigma}_{x_j}}{\hat{\sigma}_y}, \left(j = \overline{1, m}\right) \quad (38)$$

де b_j – коефіцієнт регресії при пояснюючій змінній x_j,

$\hat{\sigma}_{x_j}$ – стандартна похибка пояснюючої змінної x_j,

$\hat{\sigma}_y$ – стандартна похибка залежної змінної моделі.

- Метод усіх можливих регресій
- Метод покрокової регресії
- Метод виключень

- коефіцієнт детермінації R^2
- стандартна похибка $\hat{\sigma}_{\varepsilon}$
- критерій Меллоуза $\mathrm{C_p}$

Частковий F- критерій

$$F_{p,n-k} = \frac{SSE_1 - SSE_2}{SSE_1} \cdot \frac{n-k}{p}$$

Оцінений коефіцієнт детермінації

Оцінений коефіцієнт детермінації, скоригований за Тейлом

$$\overline{R}_T^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \cdot \frac{n-1}{n-k}$$

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

Оцінений коефіцієнт детермінації, скоригований за Амемією

$$\overline{R}_A^2 = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \cdot \frac{n+m}{n-k}$$

$$\overline{R}_A^2 = 1 - (1 - R^2) \cdot \frac{n+m}{n-k}$$

Кореляційна матриця

$$r = \begin{pmatrix} r_{yy} & r_{yx_1} & r_{yx_2} & \cdots & r_{yx_m} \\ r_{x_1y} & r_{x_1x_1} & r_{x_1x_2} & \cdots & r_{x_1x_m} \\ r_{x_2y} & r_{x_2x_1} & r_{x_2x_2} & \cdots & r_{x_2x_m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{x_my} & r_{x_mx_1} & r_{x_mx_2} & \cdots & r_{x_mx_m} \end{pmatrix} \qquad \dim r = (m+1)\times(m+1)$$

Частковий коефіцієнт кореляції

$$r_{jk,12\ldots m+1} = \frac{c_{jk}}{\sqrt{c_{jj}\cdot c_{kk}}},\ (j=\overline{1,m+1}),\ (k=\overline{1,m+1})$$

Міністерство освіти і науки України
Національний університет водного господарства та природокористування

102-40

Курс лекцій

з дисципліни „Економетрія”
для студентів напряму підготовки
„Економіка і підприємництво”

Рекомендовано методичною радою
факультету економіки і підприємництва.
Протокол №10 від 15 травня 2006 р.

Рівне – 2006

Тема 1.
ПРЕДМЕТ, МЕТОДИ І ЗАВДАННЯ ДИСЦИПЛІНИ
Стор. 4–11

Тема 2.
ОСНОВНІ ПОНЯТТЯ І ЗАГАЛЬНІ ЗАСАДИ ЕКОНОМЕТРИЧНОГО МОДЕЛЮВАННЯ
Стор. 11–28

Тема 3.
ЗАГАЛЬНА ЛІНІЙНА ЕКОНОМЕТРИЧНА МОДЕЛЬ
Стор. 29–66