Metoda najmniejszych kwadratów

Przykład wstępny. W ekonomicznej teorii produkcji rozważa się funkcję produkcji Cobba–Douglasa:

$$z = AL^{\alpha}K^{\beta}$$

gdzie z oznacza wielkość produkcji, L jest nakładem pracy, K nakładem kapitału. Liczby A, α, β są pewnymi stałymi. Ekonomiści interesują się wielkościami A, α i β. W celu oceny nieznanych wielkości prowadzone są obserwacje wielkości produkcji Z_i przy różnych nakładach L_i oraz K_i, przy czym zakłada się, że obserwacje te obarczone są błędami losowymi ε_i, tzn.:

$$Z_i = AL_i^{\alpha}K_i^{\beta} + \varepsilon_i, \quad i = 1, \ldots, n$$

Jako oszacowania nieznanych parametrów przyjmuje się takie wartości, przy których błędy losowe są małe

$$\sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (Z_i - AL_i^{\alpha}K_i^{\beta})^2 = \min!$$

. . . .

Obserwujemy zmienne losowe $Y_1, \ldots, Y_n$ takie, że

$$EY_i = g_i(\theta), \quad i = 1, \ldots, n$$

$$\theta \subset \mathbb{R}^k; \quad g_i : \Theta \to \mathbb{R}^1$$

$$S(\theta) = \sum_{i=1}^{n} (Y_i - g_i(\theta))^2$$

Definicja 8.1 Estymatorem najmniejszych kwadratów parametru θ ($EMNK[\theta]$) nazywamy wielkość minimalizującą $S(\theta)$. $\square$

Resztowa suma kwadratów ($EMNK[\theta] \overset{\text{ozn}}{=\!=} \hat{\theta}$)

$$S(\hat{\theta}) = \sum_{i=1}^{n} (Y_i - g_i(\hat{\theta}))^2$$

Definicja 8.2 Modelem liniowym nazywamy model statystyczny, w którym obserwacje $Y_1, \ldots, Y_n$ mają postać

$$Y_i = \beta_1 x_{1i} + \cdots + \beta_p x_{pi} + \varepsilon_i, \quad i = 1, \ldots, n$$

gdzie x_{ji} są ustalonymi liczbami, β_j są nieznanymi parametrami modelu, ε_i są niezależnymi „błędami losowymi" takimi, że $E\varepsilon_i = 0$ oraz $D^2\varepsilon_i = \sigma^2$. $\square$

Zapis macierzowy

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

$$\boldsymbol{Y}' = (Y_1, \ldots, Y_n)$$

$$\boldsymbol{\beta}' = (\beta_0, \beta_1, \ldots \beta_p)$$

$$\boldsymbol{\varepsilon}' = (\varepsilon_1, \ldots, \varepsilon_n)$$

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & \cdots & x_{p1} \\ x_{12} & \cdots & x_{p2} \\ \vdots & \cdots & \vdots \\ x_{1n} & \cdots & x_{pn} \end{bmatrix}$$

Założenie: macierz $\boldsymbol{X}$ jest pełnego rzędu

Resztowa suma kwadratów

$$\begin{aligned} S(\boldsymbol{\beta}) &= \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{p} \beta_j x_{ji} \right)^2 \\ &= (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) \\ &= \|\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 \end{aligned}$$

$\boldsymbol{X}\boldsymbol{\beta}$ jest rzutem $\boldsymbol{Y}$ na $\{\boldsymbol{X}\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathbb{R}^p\}$

$$(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{X} = \mathbf{0}$$

$$EMNK[\boldsymbol{\beta}] = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} \overset{\text{ozn}}{=\!=\!=} \hat{\boldsymbol{\beta}}$$

Jeżeli $\boldsymbol{c} \in \mathbb{R}^p$, to

$$EMNK[\boldsymbol{c}'\boldsymbol{\beta}] = \boldsymbol{c}'\hat{\boldsymbol{\beta}}$$

Przykład . Na podstawie obserwacji $Y_1, \ldots, Y_n$ szacujemy wartość oczekiwaną μ. Model

$$Y_i = \mu + \varepsilon_i, \quad i = 1, \ldots, n$$

W zapisie macierzowym

$$\beta = \mu, \quad \boldsymbol{X} = \mathbf{1}_n$$

$$EMNK[\mu] = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = \bar{Y}$$

. . . .

Przykład . Rozważmy model

$$Y_i = x_{i1}\mu_1 + x_{i2}\mu_2 + \varepsilon_i, \quad i = 1, \ldots, n$$

przy czym
$x_{i1} = 1$, $x_{i2} = 0$ dla $i = 1, \ldots, n_1$
$x_{i1} = 0$, $x_{i2} = 1$ dla $i = n_1 + 1, \ldots, n$

Szacujemy różnicę $\mu_1 - \mu_2$.

Zapis macierzowy

$$\boldsymbol{\beta}' = (\mu_1, \mu_2), \quad \boldsymbol{X} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{0}_{n-n_1} & \mathbf{1}_{n-n_1} \end{bmatrix}$$

$$EMNK[\boldsymbol{\beta}] = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

$$= \begin{bmatrix} n_1 & 0 \\ 0 & n_2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^{n_1} Y_i \\ \sum_{i=n_1+1}^{n} Y_i \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i \\ \frac{1}{n-n_1} \sum_{i=n_1+1}^{n} Y_i \end{bmatrix}$$

Jeżeli $\boldsymbol{c}' = (1, -1)$, to $\boldsymbol{c}'\boldsymbol{\beta} = \mu_1 - \mu_2$

$$EMNK[\mu_1 - \mu_2] = \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i - \frac{1}{n - n_1} \sum_{i=n_1+1}^{n} Y_i$$

Jeżeli $\boldsymbol{c}' = (1, 0)$, to $\boldsymbol{c}'\boldsymbol{\beta} = \mu_1$

$$EMNK[\mu_1] = \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i$$

. . . .

Twierdzenie 8.1 *$EMNK[\boldsymbol{\beta}]$ jest estymatorem nieobciążonym o wariancji $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$* $\blacksquare$

Definicja 8.3 Funkcję parametryczną $\boldsymbol{c}'\boldsymbol{\beta}$ nazywamy estymowalną, jeżeli istnieje jej estymator nieobciążony postaci $\boldsymbol{b}'\boldsymbol{Y}$. $\square$

Twierdzenie 8.2 *Funkcja parametryczna $\boldsymbol{c}'\boldsymbol{\beta}$ jest estymowalna wtedy i tylko wtedy, gdy $\boldsymbol{c} \in \mathrm{Im}\boldsymbol{X}'$* $\blacksquare$

Twierdzenie 8.3 *(Gaussa–Markowa) Jeżeli błędy losowe $\varepsilon_1, \ldots, \varepsilon_n$ są nieskorelowanymi zmiennymi losowymi o zerowej wartości oczekiwanej i takiej samej wariancji, to dla każdej estymowalnej funkcji parametrycznej $\boldsymbol{c}'\boldsymbol{\beta}$ i dla każdego nieobciążonego estymatora liniowego $\boldsymbol{b}'\boldsymbol{Y}$ tej funkcji zachodzi*

$$D^2_{\boldsymbol{\beta}}(\boldsymbol{c}'\hat{\boldsymbol{\beta}}) \leq D^2_{\boldsymbol{\beta}}(\boldsymbol{b}'\boldsymbol{Y}), \quad \forall \boldsymbol{\beta} \in \mathbb{R}^p$$

$\blacksquare$

Estymacja wariancji σ^2

Resztowa suma kwadratów

$$\begin{aligned}
&\|\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2 = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \\
&= (\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y})'(\boldsymbol{Y} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}) \\
&= \boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{X}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X})(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{Y} \\
&= \boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{Y}
\end{aligned}$$

$$\begin{aligned}
&E_{\boldsymbol{\beta}}\|\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}\|^2 \\
&= E_{\boldsymbol{\beta}}\boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{Y} \\
&= \mathrm{tr}E_{\boldsymbol{\beta}}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{Y}\boldsymbol{Y}' \\
&= \mathrm{tr}\left\{(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')E_{\boldsymbol{\beta}}(\boldsymbol{Y}\boldsymbol{Y}')\right\} \\
&= \sigma^2\mathrm{tr}(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}') \\
&= (n - \mathrm{rz}\boldsymbol{X})\sigma^2
\end{aligned}$$

$$\hat{\sigma}^2 = \frac{1}{(n - \mathrm{rz}\boldsymbol{X})}\boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{Y}$$

Przykład . Rozważamy model

$$Y_i = \mu + \varepsilon_i, \quad i = 1, \ldots, n$$

W zapisie macierzowym

$$\beta = \mu, \quad \boldsymbol{X} = \mathbf{1}_n$$

W modelu rz$\boldsymbol{X} = 1$

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-1}\boldsymbol{Y}'(\boldsymbol{I} - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n')\boldsymbol{Y} \\
&= \frac{1}{n-1}\left\{\boldsymbol{Y}'\boldsymbol{Y} - \frac{1}{n}\left(\boldsymbol{Y}'\mathbf{1}_n\right)\left(\mathbf{1}_n'\boldsymbol{Y}\right)\right\} \\
&= \frac{1}{n-1}\left\{\sum_{i=1}^{n} Y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} Y_i\right)^2\right\} \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2
\end{aligned}$$

. . . .

Przykład . Rozważmy model

$$Y_i = x_{i1}\mu_1 + x_{i2}\mu_2 + \varepsilon_i, \quad i = 1, \ldots, n$$

przy czym
$x_{i1} = 1$, $x_{i2} = 0$ dla $i = 1, \ldots, n_1$
$x_{i1} = 0$, $x_{i2} = 1$ dla $i = n_1 + 1, \ldots, n$

Zapis macierzowy ($n_2 = n - n_1$)

$$\boldsymbol{\beta}' = (\mu_1, \mu_2), \quad \boldsymbol{X} = \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} \end{bmatrix}$$

W modelu rz$\boldsymbol{X} = 2$

$$\begin{aligned}
&(n-2)\hat{\sigma}^2 \\
&= \boldsymbol{Y}'\left(\boldsymbol{I} - \begin{bmatrix} \mathbf{1}_{n_1} & \mathbf{0}_{n_1} \\ \mathbf{0}_{n_2} & \mathbf{1}_{n_2} \end{bmatrix} \begin{bmatrix} n_1 & 0 \\ 0 & n_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1}'_{n_1} & \mathbf{0}'_{n_2} \\ \mathbf{0}'_{n_1} & \mathbf{1}'_{n_2} \end{bmatrix}\right)\boldsymbol{Y} \\
&= \sum_{i=1}^{n_1}\left(Y_i - \frac{1}{n_1}\left(\sum_{i=1}^{n_1} Y_i\right)\right)^2 + \\
&\qquad + \sum_{i=n_1+1}^{n}\left(Y_i - \frac{1}{n_2}\left(\sum_{i=n_1+1}^{n} Y_i\right)\right)^2
\end{aligned}$$

. . . .

Przykład (praktyczny). Niech Y_1, Y_2, Y_3, Y_4 będą lotniczymi pomiarami kątów $\theta_1, \theta_2, \theta_3, \theta_4$ pewnego czworokąta na powierzchni ziemi. Zakładając, że obserwacje obciążone są niezależnymi błędami o zerowej wartości oczekiwanej i takiej samej wariancji σ^2, wyznaczyć $EMNK$ wielkości θ. Wyznaczyć nieobciążony estymator wariancji σ^2.

Model

$$\begin{aligned} Y_1 &= \theta_1 + \varepsilon_1 \\ Y_2 &= \theta_2 + \varepsilon_2 \\ Y_3 &= \theta_3 + \varepsilon_3 \\ Y_4 &= \theta_4 + \varepsilon_4 \\ 2\pi &= \theta_1 + \theta_2 + \theta_3 + \theta_4 \end{aligned}$$

Zapis macierzowy

$$\boldsymbol{Y}' = (Y_1, Y_2, Y_3, Y_4, 2\pi); \quad \boldsymbol{\beta}' = (\theta_1, \theta_2, \theta_3, \theta_4)$$

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, 0)$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} 0.8 & -0.2 & -0.2 & -0.2 \\ -0.2 & 0.8 & -0.2 & -0.2 \\ -0.2 & -0.2 & 0.8 & -0.2 \\ -0.2 & -0.2 & -0.2 & 0.8 \end{bmatrix}$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' = \begin{bmatrix} 0.8 & -0.2 & -0.2 & -0.2 & 0.2 \\ -0.2 & 0.8 & -0.2 & -0.2 & 0.2 \\ -0.2 & -0.2 & 0.8 & -0.2 & 0.2 \\ -0.2 & -0.2 & -0.2 & 0.8 & 0.2 \end{bmatrix}$$

Estymatory kątów:

$$\hat{\theta}_i \stackrel{?}{=} Y_i + \frac{1}{5}\left(2\pi - \sum_{i=1}^{4} Y_i\right), \quad i = 1, 2, 3, 4$$

Estymator wariancji:

$$\hat{\sigma}^2 \stackrel{?}{=} \frac{1}{5}\left(2\pi - \sum_{i=1}^{4} Y_i\right)^2$$

Wiadomo, iż dany czworokąt jest równoległobokiem takim, że $\theta_1 = \theta_3$ oraz $\theta_2 = \theta_4$. Wyznaczyć EMNK kątów i estymator wariancji σ^2.

Model

$$\begin{aligned}
Y_1 &= \theta_1 + \varepsilon_1 \\
Y_2 &= \theta_2 + \varepsilon_2 \\
Y_3 &= \theta_3 + \varepsilon_3 \\
Y_4 &= \theta_4 + \varepsilon_4 \\
2\pi &= \theta_1 + \theta_2 + \theta_3 + \theta_4 \\
0 &= \theta_1 - \theta_3 \\
0 &= \theta_2 - \theta_4
\end{aligned}$$

Zapis macierzowy

$$\boldsymbol{Y}' = (Y_1, Y_2, Y_3, Y_4, 2\pi, 0, 0); \quad \boldsymbol{\beta}' = (\theta_1, \theta_2, \theta_3, \theta_4)$$

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, 0, 0, 0)$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$$

. . . .

Przykład (Regresja liniowa). Model

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i,\ i = 1, \ldots, n,$$

ε_i są niezależnymi zmiennymi losowymi

$$E\varepsilon_i = 0, \quad D^2\varepsilon_i = \sigma^2, \qquad i = 1, \ldots, n$$

Znaleźć takie β_0 i β_1 by

$$\sum_{i=1}^{n} (Y_i - (\beta_0 + \beta_1 x_i))^2 = \min$$

Zapis macierzowy

$$\boldsymbol{Y}' = (Y_1, \ldots, Y_n);\ \boldsymbol{\beta}' = (\beta_0, \beta_1);\ \boldsymbol{\varepsilon}' = (\varepsilon_1, \ldots, \varepsilon_n)$$

$$\boldsymbol{X}' = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix}$$

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

$$(\boldsymbol{X}'\boldsymbol{X})^{-1} = \frac{1}{n\sum x_i^2 - \left(\sum x_i\right)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}$$

$$\boldsymbol{X}'\boldsymbol{Y} = \begin{bmatrix} \sum Y_i \\ \sum x_i Y_i \end{bmatrix}$$

$$\begin{bmatrix} \hat\beta_0 \\ \hat\beta_1 \end{bmatrix} = EMNK \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} =$$

$$\begin{bmatrix} \frac{1}{n}\sum Y_i - \frac{1}{n}\left(\sum x_i\right) \frac{\sum x_i Y_i - \frac{1}{n}\left(\sum x_i\right)\left(\sum Y_i\right)}{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2} \\ \frac{\sum x_i Y_i - \frac{1}{n}\left(\sum x_i\right)\left(\sum Y_i\right)}{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2} \end{bmatrix}$$

$$\begin{cases} \hat\beta_1 = \dfrac{\sum x_i y_i - n\bar{x}\bar{Y}}{\sum x_i^2 - n\bar{x}^2} \\ \hat\beta_0 = \bar{Y} - \hat\beta_1 \bar{x} \end{cases}$$

. . . .

Rozkłady prawdopodobieństwa estymatorów

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}.$$

$$\boldsymbol{\beta}' = (\beta_0, \beta_1, \dots \beta_p)$$

$$\boldsymbol{\varepsilon}' = (\varepsilon_1, \dots, \varepsilon_n) \sim N_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$$

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

$$s^2 = \hat{\sigma}^2 = \frac{1}{(n - \mathrm{rz}\boldsymbol{X})}\boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{Y}$$

Twierdzenie 8.4 *Jeżeli* $\boldsymbol{Y} \sim N_n(\boldsymbol{X}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n)$ *oraz* $\mathrm{rz}\boldsymbol{X} = p$, *to*

1. $\hat{\boldsymbol{\beta}} \sim N_p(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$

2. $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \sim \sigma^2\chi^2_p$

3. $\hat{\boldsymbol{\beta}}$ *jest niezależne od* s^2

4. $(n-p)s^2 \sim \sigma^2\chi^2_{n-p}$

■