Metoda Najmniejszych Kwadratów

Stanisław Cichocki

Natalia Nehrebecka

Wykład 2

Plan wykładu

- 1. Model liniowy
  - Postać modelu liniowego
- 2. Estymacja modelu
  - Wstęp
  - Wartość teoretyczna (dopasowana)
  - Reszty
  - Metoda Najmniejszych Kwadratów
- 3. MNK dla modelu z jedną zmienną

Plan wykładu

- 1. Model liniowy
  - Postać modelu liniowego
- 2. Estymacja modelu
  - Wstęp
  - Wartość teoretyczna (dopasowana)
  - Reszty
  - Metoda Najmniejszych Kwadratów
- 3. MNK dla modelu z jedną zmienną

Postać modelu liniowego

teoria ekonomiczna ↔ dane empiryczne

↓

zależności ilościowe między zmiennymi

↓

badanie ekonometryczne

Postać modelu liniowego

$$y_i = \beta_1 + x_{2i}\beta_2 + \ldots + x_{Ki}\beta_K + \varepsilon_i$$

- y_i – zmienna objaśniana (zależna, endogeniczna),
- $x_2, \ldots, x_K$ – zmienne objaśniające (niezależne, egzogeniczne),
- ε_i – błąd losowy,
- $\beta_1, \ldots, \beta_K$ – nieznane parametry,
- $i=1, \ldots, N$
- i – indeks obserwacji,
- N – liczba obserwacji

Postać modelu liniowego

y	x
Zmienna objaśniana	Zmienna objaśniająca
Zmienna zależna	Zmienna niezależna
Zmienna endogeniczna	Zmienna egzogeniczna
Regresant	Regresor
	Zmienna kontrolna
	Predyktor

Postać modelu liniowego

- Zatem:
  - zależność między analizowanymi zmiennymi jest **liniowa** (równanie regresji liniowej wyznacza hiperpłaszczyznę regresji)
  - istnieje zależność przyczynowo-skutkowa między zmiennymi (≠korelacja)
    - zmienne objaśniające są przyczyną zmienności zmiennej objaśnianej
    - zależność zwykle wynika z teorii (powinna)
  - pewna część zmienności zmiennej objaśnianej pozostaje niewyjaśniona, bo:
    - nieuwzględnienie pewnych zmiennych objaśniających
    - losowy charakter czynników wpływających na zmienną objaśnianą

Postać modelu liniowego

- Który z modeli jest poprawny i dlaczego?
- Co jest zmienną objaśnianą a co objaśniającą?

$$wydatki_i = \beta_1 + \beta_2 dochód_i + \varepsilon_i$$

$$dochód_i = \beta_1 + \beta_2 wydatki_i + \varepsilon_i$$

Postać modelu liniowego

- Związek przyczynowo-skutkowy ≠ korelacja

Przykład

Stwierdzono dodatnią korelację między wielkością spożycia lodów w danym dniu i liczbą utonięć w tym dniu. Czy po zjedzeniu lodów nie powinno się wchodzić do wody?

Odpowiedź

Więcej utonięć zdarza się w ciepłe dni (kąpie się wtedy więcej osób). W takie dni jest też większe spożycie lodów. Jednak spożycie lodów nie powoduje utonięcia

Czyli: występuje korelacja między zdarzeniami ale nie ma między nimi związku przyczynowo-skutkowego

Plan wykładu

- 1. Model liniowy
  - Postać modelu liniowego
- 2. Estymacja modelu
  - Wstęp
  - Wartość teoretyczna (dopasowana)
  - Reszty
  - Metoda Najmniejszych Kwadratów
- 3. MNK dla modelu z jedną zmienną

Estymacja - wstęp

- Teoria zwykle nie dostarcza informacji nt. wielkości parametrów modelu $(\beta_1, \ldots, \beta_K)$.

$$\Downarrow$$

- Wielkość nieznanych parametrów należy oszacować (estymować) na podstawie danych empirycznych (próby).

$$\Downarrow$$

- Oszacowane wielkości parametrów (estymatory) $(b_1, \ldots, b_K)$ są niedokładne (losowe), zależą od próby.

Wartość teoretyczna (dopasowana)

- **Wartości dopasowane:** wartości zmiennej objaśnianej (y_i) przewidywane na podstawie oszacowanego modelu - regresji liniowej y_i na $x_{2i}, \ldots, x_{Ki}$:

$$\hat{y}_i = b_1 + x_{2i} b_2 + \ldots + x_{Ki} b_K$$

- Różnią się od wartości rzeczywistych, bo:
  - zamiast nieznanych prawdziwych wielkości parametrów ($\beta_1, \ldots, \beta_K$) używamy ich estymatorów ($b_1, \ldots, b_K$)
  - pomijamy błąd losowy (ε_i)

Reszty

- **Reszty:** różnica między wartością rzeczywistą a dopasowaną zmiennej objaśnianej, są to oszacowania (ε_i) :

$$e_i = y_i - \hat{y}_i = y_i - b_1 - x_{2i}b_2 - \ldots - x_{Ki}b_K$$

- Zależność między resztami, obserwacjami i oszacowaniami parametrów:

$$y_i = \hat{y}_i + e_i = b_1 + x_{2i}b_2 + \ldots + x_{Ki}b_K + e_i$$

Estymatory i reszty

- **Estymatory** ($b_1, \ldots, b_K$) są to oszacowania ($\beta_1, \ldots, \beta_K$) ale nie są im równe

- **Reszty** (e_i) są to oszacowania (ε_i) ale nie są im równe

Metoda Najmniejszych Kwadratów

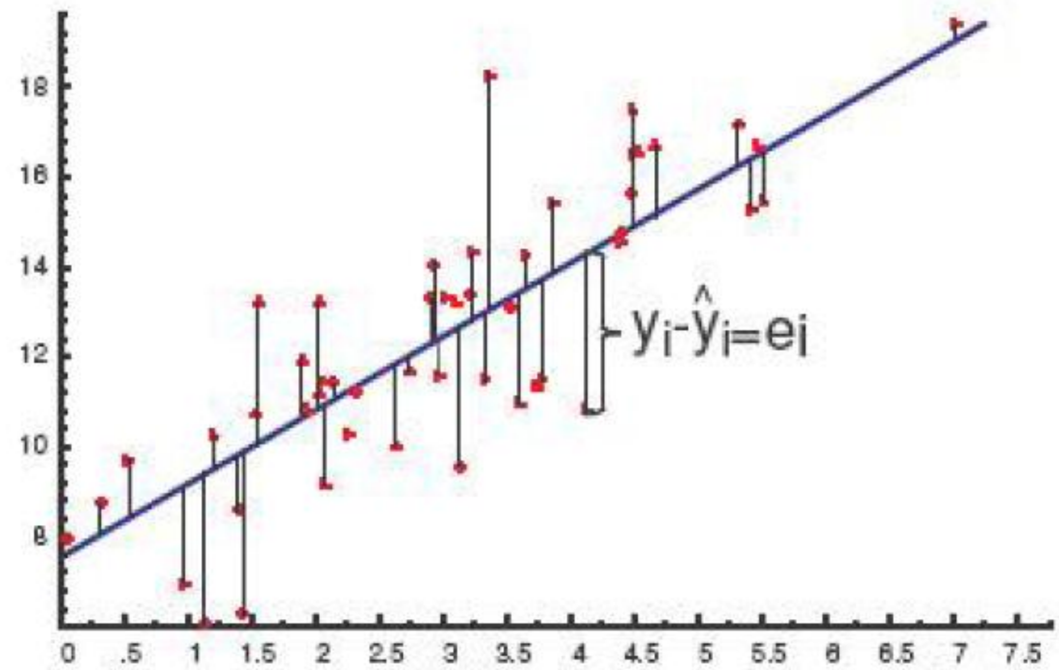
- Im mniejsza jest odległość wartości rzeczywistych od teoretycznych tym lepszy model $\Longrightarrow$

  estymatory parametrów modelu minimalizują sumę odległości y_i od $\hat{y}_i$:

$$\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{N} e_i^2$$

Metoda Najmniejszych Kwadratów

Rysunek: Reszty i prosta regresji



Źródło: J. Mycielski, *skrypt do ekonometrii...*

Plan wykładu

- 1. Model liniowy
  - Postać modelu liniowego
- 2. Estymacja modelu
  - Wstęp
  - Wartość teoretyczna (dopasowana)
  - Reszty
  - Metoda Najmniejszych Kwadratów
- 3. MNK dla modelu z jedną zmienną

MNK dla modelu z jedną zmienną

- **Kroki**:

1. Zapisujemy wzór na sumę kwadratów reszt i podstawiamy za reszty
2. Liczymy warunki pierwszego rzędu (pochodne)
3. Przekształcamy tak aby uzyskać oszacowania parametrów

MNK dla modelu z jedną zmienną

- Oszacowania $\mathbf{b_1}$ i $\mathbf{b_2}$ powinny być dobrane tak, by suma kwadratów reszt była jak najmniejsza.

$$S(b_1, b_2) = \sum_{i=1}^{N} e_i = \sum_{i=1}^{N} \left(y_i - b_1 - x_i b_2\right)^2 =$$

$$= \sum_{i=1}^{N} \left(y_i^2 - 2y_i b_1 - 2y_i b_2 x_i + 2b_1 b_2 x_i + b_1^2 + b_2^2 x_i^2\right)$$

MNK dla modelu z jedną zmienną

- Policz pochodne cząstkowe względem parametrów $\mathbf{b_1}$ i $\mathbf{b_2}$ powyższego równania i przyrównaj je do zera.

$$\begin{cases} \dfrac{\partial S(b_1, b_2)}{\partial b_1} = 0 \\ \dfrac{\partial S(b_1, b_2)}{\partial b_2} = 0 \end{cases}$$

Warunki pierwszego rzędu

MNK dla modelu z jedną zmienną

- Licząc pochodne dla poszczególnych równań uzyskujemy układ równań zwany **układem równań normalnych**.

$$
\begin{cases}
\sum_{i=1}^{N} \left[-2y_i + 2b_1 + 2b_2 x_i\right] = 0 \\
\sum_{i=1}^{N} \left[-2y_i x_i + 2b_1 x_i + 2b_2 x_i^2\right] = 0
\end{cases}
$$

MNK dla modelu z jedną zmienną

$$b_1 = \overline{y} - b_2 \overline{x}$$

$$b_2 = \frac{\dfrac{\sum_{i=1}^{N} y_i x_i}{N} - \overline{yx}}{\dfrac{\sum_{i=1}^{N} x_i^2}{N} - \overline{x}^2}$$

MNK dla modelu z jedną zmienną

- Przypomnij wzór na wariancję (s^2_x) i kowariancję (s_{xy}) empiryczną.

$$b_1 = \overline{y} - b_2 \overline{x}$$

$$b_2 = \frac{S_{yx}}{S_x^2}$$

Dziękuję za uwagę