

Regresja liniowa. Estymator metody najmniejszych kwadratów. Właściwości asymptotyczne. Twierdzenie Gaussa-Markowa

Jacek Suda na podstawie prezentacji dr Jakuba Mućka
Szkola Główna Handlowa w Warszawie

O kursie

■ Jacek Suda

- ▶ Katedra Ekonomii Ilościowej, Kolegium Analiz Ekonomicznych
- ▶ *E-mail*: jsuda@sgh.waw.pl
- ▶ <https://jaceksuda.com>

■ Paweł Galiński

- ▶ Katedra Ekonomii Ilościowej, Kolegium Analiz Ekonomicznych
- ▶ *E-mail*: pgalin@sgh.waw.pl

1. Regresja liniowa. Estymator metody najmniejszych kwadratów. Właściwości asymptotyczne. Twierdzenie Gaussa-Markowa
2. Testowanie hipotez ekonomicznych. Testowanie wielu hipotez. Hipotezy liniowe i nieliniowe. Przedziały ufności. Metoda delta.
3. Weryfikacja kluczowych założeń: normalność, współliniowość i postać funkcyjna. Dopasowanie modelu.
4. Heteroskedastyczność i autokorelacja. Estymator uogólnionej metody najmniejszych kwadratów. Ważona metoda najmniejszych kwadratów. Błędy standardowe odporne i klastrowe.
5. Endogeniczność. Estymacja metodą zmiennych instrumentalnych. Właściwości zmiennych instrumentalnych.
6. Modele równań jednoczesnych (Simultaneous equations model). Problem identyfikacji parametrów. Metody estymacji SEM.
7. Szeregi czasowe. Stacjonarność, regresja pozorna i kointegracja.
8. Modele autoregresyjne z rozłożonymi opóźnieniami (autoregressive distributed lags). Modele wektorowej autoregresji (VAR). Strukturalne modele VAR (SVAR).

9. Dane panelowe. Zmienność między-grupowa i wewnątrz-grupowa. Modele z efektami losowymi i stałymi. Regresja międzygrupowa (estymator *between*). Estymator Hausmana-Taylora.
10. Ograniczona zmienna zależna (limited dependent variable). Modele dla binarnej i wielomianowej zmiennej objaśnianej. Estymator największej wiarygodności. Dane panelowe i ograniczona zmienna zależna.
11. Modele danych zliczeniowych. Regresja tobitowa.
12. Uogólniona metoda momentów (UMM/GMM). Wybrane zastosowania
13. Dynamiczne modele panelowe. Obciążenie Nickella. Estymator Andersona-Hsiao. Estymator Arellano-Bonda. Systemowy estymator GMM.
14. Estymacja efektu traktowania (treatment effect). Difference-in-difference.
15. Metoda regresji nieciągłej (Regression discontinuity design)

Podręczniki ekonometrii:

1. Wooldridge J. M., *Econometric Analysis of Cross Section and Panel Data*.
2. Pesaran M. H., *Time Series and Panel Data Econometrics*.
3. Greene W. H., *Econometric Analysis*.
4. Hall R. C., Griffiths W. E., Lim G. C., *Principles of Econometrics*.
5. Wooldridge J. M., *Introductory Econometrics: A Modern Approach*.

Oprogramowanie:

- Stata.
- R.

- Egzamin
- Prace domowe ($\times 2$) i aktywność na zajęciach.

Wprowadzenie do ekonometrii

Ekonometria

to zastosowanie technik statystycznych w ekonomii w badaniu problemów, analizie danych oraz opracowywaniu i testowaniu teorii i modeli.

Ekonometria

to zastosowanie technik statystycznych w ekonomii w badaniu problemów, analizie danych oraz opracowywaniu i testowaniu teorii i modeli.

Używamy ekonometrii

- do szacowania parametrów ekonomicznych (np. elastyczności),
- do prognozowania zmiennych/procesów ekonomicznych,
- do weryfikacji hipotez ekonomicznych.

- **Model ekonomiczny** reprezentuje ilościowe relacje między zbiorem zmiennych ekonomicznych. Na przykład, krańcowa skłonność do konsumpcji:

$$C = \beta_0 + \beta_1 Y \quad (1)$$

gdzie C to wydatki konsumpcyjne a Y to dochód do dyspozycji.

- **Model ekonomiczny** reprezentuje ilościowe relacje między zbiorem zmiennych ekonomicznych. Na przykład, krańcowa skłonność do konsumpcji:

$$C = \beta_0 + \beta_1 Y \quad (1)$$

gdzie C to wydatki konsumpcyjne a Y to dochód do dyspozycji.

- **Model ekonometryczny** jest dodatkowo rozszerzony o komponent stochastyczny:

$$C = \beta_0 + \beta_1 Y + \varepsilon, \quad (2)$$

gdzie ε to (zaburzenie/szok) składnik losowy.

Ogólny jednorównaniowy liniowy model ekonometryczny:

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad i = 1, 2, \dots, N \quad (3)$$

gdzie

- y_i to zmienna zależna (wynikowa),
- $x_{1i}, \dots, x_{ki}$ to zbiór k zmiennych objaśniających (niezależnych),
- $\beta_0, \beta_1, \dots, \beta_k$ to parametry modelu,
- ε_i to składnik losowy,
- i to indeks obserwacji.

- **Dane przekrojowe** są zbierane dla jednostek z próby (osób) w danym okresie czasu.

$$y_i, \text{ gdzie} \\ i \in \{1, \dots, N\}.$$

- **Szeregi czasowe:** są zbierane w dyskretnych odstępach czasu:

$$y_t, \text{ gdzie} \\ t \in \{1, \dots, T\}.$$

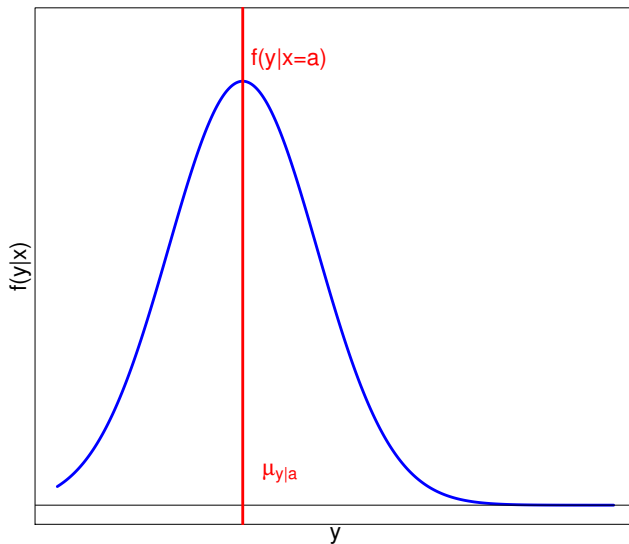
- **Dane panelowe lub przekrojowo-czasowe** są zbierane dla jednostek w czasie:

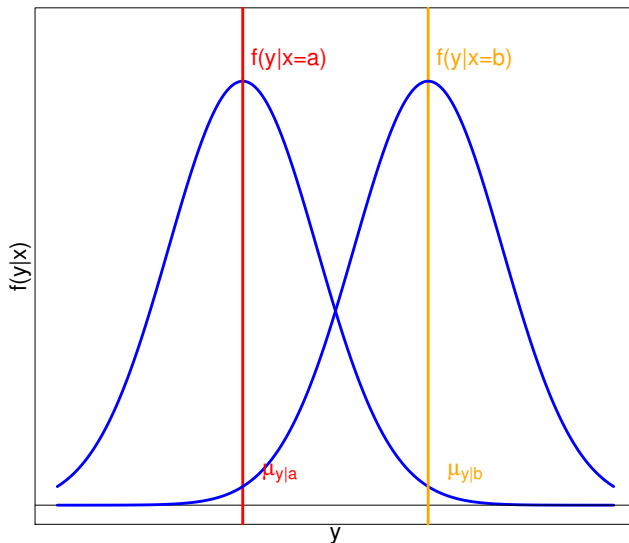
$$y_{it}, \text{ gdzie} \\ i \in \{1, \dots, N\} \\ t \in \{1, \dots, T\}.$$

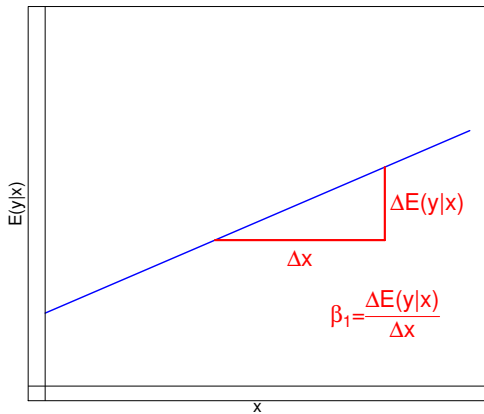
- Dane eksperymentalne
- Dane mikroekonomiczne
- Dane makroekonomiczne
- Dane finansowe

1. **Sformułowanie hipotez badawczych.** Na podstawie problemu badawczego i przeglądu literatury.
2. **Wybór modelu ekonomicznego**, który prowadzi do **modelu ekonometrycznego**. Obejmuje to wybór formy funkcyjnej oraz zbioru zmiennych objaśniających.
3. **Zbieranie danych.** Uzyskanie próby i wybór metody, która umożliwia zastosowanie wnioskowania statystycznego.
4. **Szacowanie/estymacja parametrów.**
5. **Diagnostyka modelu.** Sprawdzenie poprawności założeń.

Prosty model regresji liniowej







Regresja prosta:

$$\mathbb{E}(y|x) = \beta_0 + \beta_1 x$$

■ Prosty model regresji liniowej :

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (4)$$

gdzie

- ▶ y to zmienna zależna (wynikowa);
 - ▶ x to zmienna niezależna;
 - ▶ ε to składnik losowy.
- Zmienna zależna jest wyjaśniana komponentami, które zmieniają się wraz ze zmianami **zmiennej zależnej** i **składnikiem losowym**.
 - β_0 to wyraz wolny (stała).
 - β_1 to współczynnik (nachylenie) przy x .

■ Prosty model regresji liniowej :

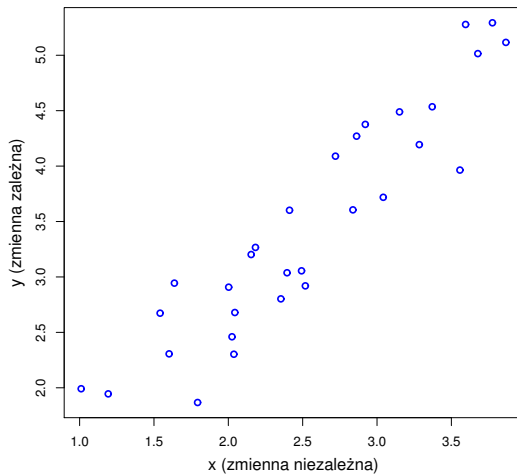
$$y = \beta_0 + \beta_1 x + \varepsilon \quad (4)$$

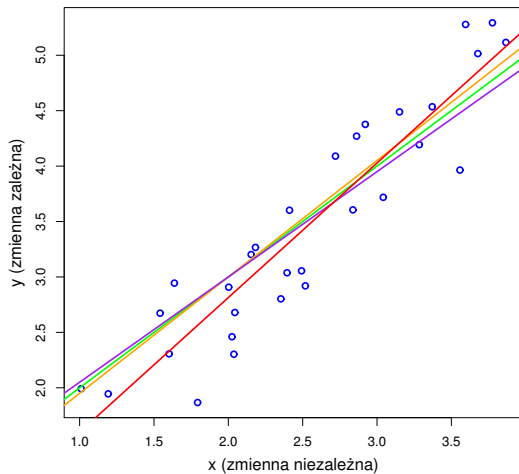
gdzie

- ▶ y to zmienna zależna (wynikowa);
 - ▶ x to zmienna niezależna;
 - ▶ ε to składnik losowy.
- Zmienna zależna jest wyjaśniana komponentami, które zmieniają się wraz ze zmianami **zmiennej zależnej** i **składnikiem losowym**.
 - β_0 to wyraz wolny (stała).
 - β_1 to współczynnik (nachylenie) przy x .

β_1 mierzy wpływ zmiany x na wartość oczekiwaną y (*ceteris paribus*).

Estymator metody najmniejszych kwadratów (MNK)





- **Założenie #1:** prawdziwy proces generowania danych (DGP):

$$y = \beta_0 + \beta_1 x + \varepsilon. \quad (5)$$

- **Założenie #2:** wartość oczekiwana składnika losowego wynosi zero:

$$\mathbb{E}(\varepsilon) = 0, \quad (6)$$

co implikuje że $\mathbb{E}(y) = \beta_0 + \beta_1 x$.

- **Założenie #3:** stała wariancja składnika losowego i zerowa kowariancja między obserwacjami. W szczególności:

- ▶ wariancja składnika losowego równa się σ^2 :

$$\text{var}(\varepsilon) = \sigma^2 = \text{var}(y). \quad (7)$$

- ▶ kowariancja między dowolną parą ε_i i ε_j wynosi zero:

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0. \quad (8)$$

- **Założenie #4: Egzogeniczność.** Zmienna niezależna jest **nielosowa** i dlatego nie jest skorelowana ze składnikiem losowym.
- **Założenie #5:** zmienna niezależna przyjmuje co najmniej dwie wartości.
- **Założenie #6 (opcjonalne):** składnik losowy o rozkładzie normalnym:

$$\varepsilon \sim \mathcal{N}(0, \sigma^2). \quad (9)$$

■ **Wartości dopasowane zmiennej zależnej ($\hat{y}_i$):**

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (10)$$

gdzie $\hat{\beta}_0$ i $\hat{\beta}_1$ to estymatory wyrazu wolnego i nachylenia.

■ **Reszty ($\hat{e}_i$):**

$$\hat{e}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i, \quad (11)$$

to różnice między obserwowanymi (empirycznymi) a dopasowanymi wartościami zmiennej zależnej.

■ Suma kwadratów reszt (*SSE*):

$$SSE = \sum_i^N \hat{e}_i^2 = \sum_i^N (y_i - \hat{y}_i)^2. \quad (12)$$

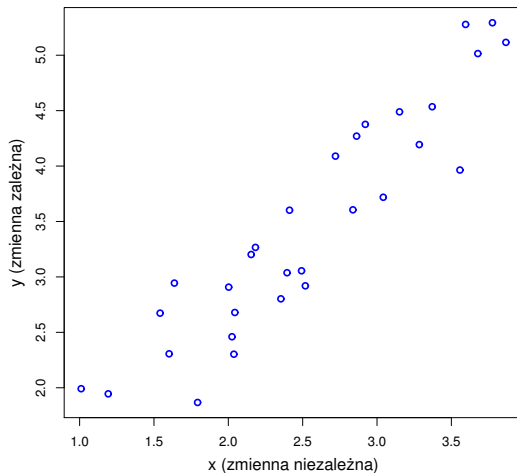
■ *SSE* można wyrazić jako funkcję parametrów β_0 i β_1 :

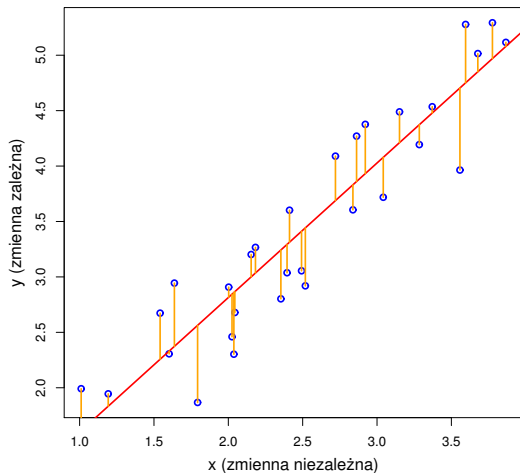
$$SSE(\beta_0, \beta_1) = \sum_i^N \hat{e}_i^2 = \sum_i^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2. \quad (13)$$

■ **Zasada najmniejszych kwadratów** to metoda wyboru parametrów, która zapewnia najniższą *SSE*:

$$\min_{\beta_0, \beta_1} \sum_i^N (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2. \quad (14)$$

Innymi słowy, metoda najmniejszych kwadratów minimalizuje *SSE*.





Estymator MNK minimalizuje sumę kwadratów *reszt* (*SSE*).

- Estymator metody najmniejszych kwadratów dla modelu regresji prostej:

$$\hat{\beta}_0^{MNK} = \bar{y} - \hat{\beta}_1^{MNK} \bar{x}, \quad (15)$$

$$\hat{\beta}_1^{MNK} = \frac{\sum_i^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^N (x_i - \bar{x})^2}. \quad (16)$$

gdzie $\bar{y}$ i $\bar{x}$ to średnie arytmetyczne zmiennej zależnej i niezależnej.

Twierdzenie Gaussa-Markowa

Przy założeniach #1–#5 modelu regresji liniowej, estymatory metody najmniejszych kwadratów $\hat{\beta}_0^{MНК}$ i $\hat{\beta}_1^{MНК}$ mają najmniejszą wariancję ze wszystkich liniowych i nieobciążonych estymatorów β_0 i β_1 .

$\hat{\beta}_0^{MНК}$ i $\hat{\beta}_1^{MНК}$ to Najlepsze Liniowe Nieobciążone Estymatory (NLNE) / Best Linear Unbiased Estimators (BLUE) parametrów β_0 i β_1 .

1. Estymatory $\hat{\beta}_0^{MNK}$ i $\hat{\beta}_1^{MNK}$ są najlepsze gdy **porównamy je z liniowymi i nieobciążonymi estymatorami**.

Na podstawie twierdzenia Gaussa-Markowa nie możemy twierdzić, że estymatory $\hat{\beta}_0^{MNK}$ i $\hat{\beta}_1^{MNK}$ to **najlepsze ze wszystkich możliwych estymatorów**.

2. Dlaczego estymatory $\hat{\beta}_0^{MNK}$ i $\hat{\beta}_1^{MNK}$ są *najlepsze*?
Ponieważ mają minimalną wariancję.
3. Twierdzenie Gaussa-Markowa obowiązuje jeśli spełnione są założenia #1–#5.
Jeśli nie, to $\hat{\beta}_0^{MNK}$ i $\hat{\beta}_1^{MNK}$ to **nie są NLNE/BUE**.
4. Twierdzenie Gaussa-Markowa nie wymaga założenia normalności (#6)
5. Poza tym estymator metody najmniejszych kwadratów jest zgodny jeśli spełnione są założenia #1–#5.

- Estymator metody najmniejszych kwadratów dla β_1 :

$$\hat{\beta}_1^{MNK} = \frac{\sum_i^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_i^N (x_i - \bar{x})^2} \quad (17)$$

można przepisać jako:

$$\hat{\beta}_1^{MNK} = \sum_{i=1}^N w_i y_i, \quad (18)$$

gdzie $w_i = (x_i - \bar{x}) / \sum (x_i - \bar{x})^2$.

- Po transformacji otrzymujemy:

$$\hat{\beta}_1^{MNK} = \beta_1 + \sum_{i=1}^N w_i \varepsilon_i. \quad (19)$$

Ponieważ w_i są znane, to jest liniowa funkcja zmiennej losowej (ε).

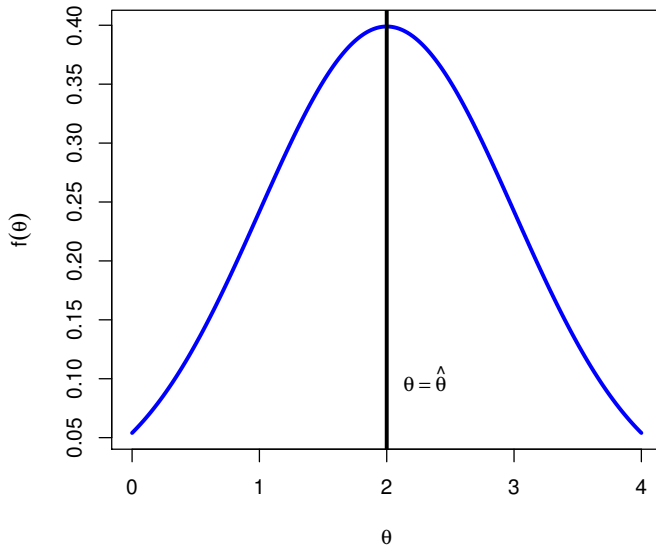
- Estymator jest nieobciążony jeśli jego wartość oczekiwana równa się prawdziwej wartości, tj.

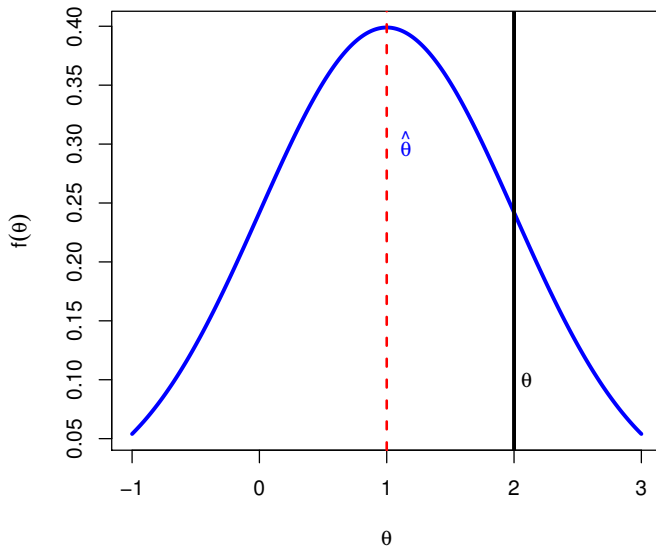
$$\mathbb{E}(\hat{\beta}) = \beta. \quad (20)$$

- Dla estymatora metody najmniejszych kwadratów:

$$\begin{aligned} \mathbb{E}(\hat{\beta}_1^{MNK}) &= \mathbb{E}\left(\beta_1 + \sum_{i=1}^N w_i \varepsilon_i\right) = \mathbb{E}(\beta_1) + \mathbb{E}\left(\sum_{i=1}^N w_i \varepsilon_i\right) \\ &= \beta_1 + \sum_{i=1}^N w_i \mathbb{E}(\varepsilon_i) = \beta_1. \end{aligned}$$

- W powyższej transformacji korzystamy z dwóch założeń: (i) $\mathbb{E}(\varepsilon_i) = 0$, oraz (ii) $\mathbb{E}(w_i \varepsilon_i) = w_i \mathbb{E}(\varepsilon_i)$. To ostatnie założenie jest równoważne egzogeniczności zmiennej niezależnej.
- Nieobciążoność dotyczy głównie średniej naszych estymatorów z wielu próbek (pobranych z tej samej populacji).





- Ogólnie, wariancja mierzy efektywność.
- Jeśli założenia #1-#5 są spełnione, to:

$$\begin{aligned} \text{var}(\hat{\beta}_0^{MNK}) &= \sigma^2 \left[\frac{\sum_{i=1}^N x_i^2}{N \sum_{i=1}^N (x_i - \bar{x})^2} \right] \\ \text{var}(\hat{\beta}_1^{MNK}) &= \frac{\sigma^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \\ \text{cov}(\hat{\beta}_0^{MNK}, \hat{\beta}_1^{MNK}) &= \sigma^2 \left[\frac{-\bar{x}}{\sum_{i=1}^N (x_i - \bar{x})^2} \right] \end{aligned}$$

- Im większa wariancja składnika losowego (σ^2), tj. większa rola składnika losowego, tym większa wariancja i kowariancja estymatorów.
- Im większa zmienność zmiennej niezależnej $\sum_{i=1}^N (x_i - \bar{x})^2$, tym mniejsza wariancja estymatorów MNK.
- Im większa liczba obserwacji (N) tym mniejsza wariancja estymatorów MNK.
- Im większy $\sum_{i=1}^N x_i^2$ tym większa wariancja estymatora wyrazu wolnego.
- Kowariancja estymatora ma znak przeciwny do znaku $\bar{x}$ i im większy $\bar{x}$ tym większa wartość kowariancji.

- Jeśli założenie normalności (#6) jest spełnione, to:

$$\hat{\beta}_0^{MNK} \sim \mathcal{N}(\hat{\beta}_0^{MNK}, \text{var}(\hat{\beta}_0^{MNK})) \quad (21)$$

$$\hat{\beta}_1^{MNK} \sim \mathcal{N}(\hat{\beta}_1^{MNK}, \text{var}(\hat{\beta}_1^{MNK})) \quad (22)$$

- Co jeśli założenie normalności nie jest spełnione?

Jeśli założenia #1–#5 są spełnione i jeśli próba (N) jest wystarczająco duża, estymatory metody najmniejszych kwadratów, tj. β_0^{MNK} i β_1^{MNK} , mają rozkład, który przybliża rozkłady normalne opisane powyżej.

- Wariancja składnika losowego:

$$\text{var}(\varepsilon_i) = \sigma^2 = \mathbb{E}[\varepsilon_i - \mathbb{E}(\varepsilon_i)]^2 = \mathbb{E}(\varepsilon_i^2) \quad (23)$$

ponieważ założyliśmy że $\mathbb{E}(\varepsilon_i) = 0$.

- Estymacja wariancji składnika losowego na podstawie reszt:

$$\hat{\sigma}^2 = \frac{1}{N-2} \sum_{i=1}^N \hat{e}_i^2. \quad (24)$$

gdzie $\hat{e}_i = y_i - \hat{y}_i$.

► dowód nieobciążoności

- $\hat{\sigma}^2$ można bezpośrednio użyć do oszacowania wariancji/kowariancji estymatora MNK.

- Do uzyskania estymatorów $var(\hat{\beta}_0^{MNK})$ i $var(\hat{\beta}_1^{MNK})$ używamy estymatora wariancji składnika losowego ($\hat{\sigma}^2$):

$$\begin{aligned} \hat{var}(\hat{\beta}_0^{MNK}) &= \hat{\sigma}^2 \left[\frac{\sum_{i=1}^N x_i^2}{N \sum_{i=1}^N (x_i - \bar{x})^2} \right] \\ \hat{var}(\hat{\beta}_1^{MNK}) &= \frac{\hat{\sigma}^2}{\sum_{i=1}^N (x_i - \bar{x})^2} \\ \hat{cov}(\hat{\beta}_0^{MNK}, \hat{\beta}_1^{MNK}) &= \hat{\sigma}^2 \left[\frac{-\bar{x}}{\sum_{i=1}^N (x_i - \bar{x})^2} \right] \end{aligned}$$

- Na podstawie wariancji możemy obliczyć błędy standardowe, które są odchyleniami standardowymi estymatorów:

$$\hat{se}(\hat{\beta}_0^{MNK}) = \sqrt{\hat{var}(\hat{\beta}_0^{MNK})} \quad \text{oraz} \quad \hat{se}(\hat{\beta}_1^{MNK}) = \sqrt{\hat{var}(\hat{\beta}_1^{MNK})}. \quad (25)$$

Regresja wielowymiarowa

■ Regresja wielowymiarowa :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K + \varepsilon \quad (26)$$

gdzie

- ▶ y to zmienna zależna (wynikowa);
 - ▶ $x_1, x_2, \dots, x_K$ to zbiór zmiennych niezależnych;
 - ▶ ε to składnik losowy.
- Zmienna zależna jest wyjaśniana komponentami, które zmieniają się wraz ze zmianami **zmiennych zależnych** i **składnikiem losowym**.
 - β_0 to wyraz wolny.
 - $\beta_1, \beta_2, \dots, \beta_K$ to współczynniki regresji dla zmiennych $x_1, x_2, \dots, x_K$.

■ Regresja wielowymiarowa :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_K x_K + \varepsilon \quad (26)$$

gdzie

- ▶ y to zmienna zależna (wynikowa);
 - ▶ $x_1, x_2, \dots, x_K$ to zbiór zmiennych niezależnych;
 - ▶ ε to składnik losowy.
- Zmienna zależna jest wyjaśniana komponentami, które zmieniają się wraz ze zmianami **zmiennych zależnych** i **składnikiem losowym**.
 - β_0 to wyraz wolny.
 - $\beta_1, \beta_2, \dots, \beta_K$ to współczynniki regresji dla zmiennych $x_1, x_2, \dots, x_K$.

$\beta_1, \beta_2, \dots, \beta_K$ mierzą wpływ zmiany $x_1, x_2, \dots, x_K$ na wartość oczekiwaną y (*ceteris paribus*).

Forma ogólna:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon. \quad (27)$$

Forma macierzowa:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon \quad (28)$$

gdzie

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}_{N \times 1}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,K} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,K} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{N,1} & x_{N,2} & \dots & x_{N,K} \end{bmatrix}_{N \times (K+1)},$$

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}_{(K+1) \times 1}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}_{N \times 1},$$

K – liczba zmiennych objaśniających; N – liczba obserwacji.

- **Założenie #1:** prawdziwy proces generowania danych (DGP):

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon. \quad (29)$$

- **Założenie #2:** wartość oczekiwana składnika losowego wynosi zero:

$$\mathbb{E}(\varepsilon) = 0, \quad (30)$$

co implikuje że $\mathbb{E}(y) = \mathbf{X}\beta$.

- **Założenie #3:** Sferyczna macierz wariancji-kowariancji składnika losowego.

$$\text{var}(\varepsilon) = \mathbb{E}(\varepsilon\varepsilon') = I\sigma^2. \quad (31)$$

W szczególności:

- ▶ wariancja składnika losowego równa się σ^2 :

$$\text{var}(\varepsilon) = \sigma^2 = \text{var}(y). \quad (32)$$

- ▶ kowariancja między dowolną parą ε_i i ε_j wynosi zero:

$$\text{cov}(\varepsilon_i, \varepsilon_j) = 0. \quad (33)$$

- **Założenie #4: Egzogeniczność.** Zmienne niezależne są **nielosowe** i dlatego nie są skorelowane ze składnikiem losowym.

$$\mathbb{E}(\mathbf{X}\varepsilon) = 0. \quad (34)$$

- **Założenie #5:** pełny rząd macierzy zmiennych objaśniających (brak tzw. współliniowości):

$$\text{rank}(\mathbf{X}) = K + 1 \leq N. \quad (35)$$

- **Założenie #6 (opcjonalne):** składnik losowy o rozkładzie normalnym:

$$\varepsilon \sim \mathcal{N}(0, \sigma^2). \quad (36)$$

- Punkt wyjścia to DGP:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \quad (37)$$

- Podobnie jak wcześniej, estymator MNK jest uzyskiwany przez minimizowanie sumy kwadratów reszt:

$$\hat{\beta}^{MNK} = \arg \min_{\beta} \mathbf{e}'\mathbf{e}, \quad (38)$$

gdzie $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\beta}$.

- SSE można wyrazić jako funkcję nieznanymi parametrów:

$$SSE(\beta) = \mathbf{e}'\mathbf{e} = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) = (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta), \quad (39)$$

- Po manipulacji otrzymujemy:

$$SSE(\beta) = \mathbf{y}\mathbf{y}' - 2\mathbf{y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta, \quad (40)$$

- Warunek pierwszego rzędu (FOC):

$$\frac{\partial SSE(\beta)}{\partial \beta} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\beta = 0, \quad (41)$$

- który możemy zapisać jako

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\beta, \quad (42)$$

Ostatecznie, wykorzystując założenie o pełnym rzędzie macierzy $\mathbf{X}$ otrzymujemy:

$$\hat{\beta}^{MNK} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}. \quad (43)$$

- Ogólna wariancja estymatora MNK ($\hat{\beta}^{MNK}$):

$$Var(\hat{\beta}^{MNK}) = \mathbb{E} \left[(\hat{\beta}^{MNK} - \beta) (\hat{\beta}^{MNK} - \beta)' \right]. \quad (44)$$

- Estymator MNK możemy zapisać jako:

$$\hat{\beta}^{MNK} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'(\mathbf{X}\beta + \varepsilon) = \beta + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\varepsilon. \quad (45)$$

- Wtedy wariancja estymatora to

$$\begin{aligned} Var(\hat{\beta}^{MNK}) &= \mathbb{E} \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\varepsilon \left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\varepsilon \right)' \right] \\ &= \mathbb{E} \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\varepsilon\varepsilon'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \right] \\ &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbb{E} [\varepsilon\varepsilon'] \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

- Przy założeniu #3 dotyczącym **sferycznej macierzy wariancji-kowariancji błędów**, tj. $\mathbb{E}(\varepsilon\varepsilon') = \sigma^2 I$, powyższe wyrażenie można uprościć i napisać:

$$Var(\hat{\beta}^{MNK}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}. \quad (46)$$

- Wariancja estymatora MNK można obliczyć, używając estymatorów wariancji składnika losowego ($\mathbb{S}_\varepsilon^2$):

$$\text{Var}(\hat{\beta}^{MNK}) = \mathbb{S}_\varepsilon^2 (\mathbf{X}'\mathbf{X})^{-1}, \quad (47)$$

gdzie

$$\mathbb{S}_\varepsilon^2 = \frac{\mathbf{e}'\mathbf{e}}{N - (K + 1)} = \frac{SSE(\hat{\beta}^{MNK})}{df} \quad (48)$$

gdzie $SSE(\hat{\beta}^{MNK})$ to suma kwadratów reszt, a df oznacza **stopnie swobody**.

- Elementy diagonalne macierzy wariancji-kowariancji (oznaczane jako $\hat{d}_{ii}$) mierzą wariancję odpowiednich parametrów. Wtedy **błąd standardowy**:

$$\mathbb{S}(\hat{\beta}_i) = \sqrt{d_{ii}}. \quad (49)$$

- Relatywne błędy standardowe**

$$\left| \frac{\mathbb{S}(\hat{\beta}_i)}{\hat{\beta}_i} \right|. \quad (50)$$

Twierdzenie Gaussa-Markowa (regresja wielowymiarowa)

Przy założeniach #1–#5 modelu regresji liniowej wielowymiarowej, estymator MNK, $\hat{\beta}^{MNK}$, ma najmniejszą wariancję ze wszystkich liniowych i nieobciążonych estymatorów β .

$\hat{\beta}^{MNK}$ to Najlepszy Liniowy Nieobciążony Estymator (NLNE) / Best Linear Unbiased Estimator (BLUE) parametrów β .

Właściwości asymptotyczne.

W dodatku.

- GREENE, W. H. (2003): *Econometric Analysis*. Prentice Hall, Upper Saddle River, NJ, 5 edn.
- HILL, R. C., W. E. GRIFFITHS, AND G. C. LIM (2012): *Principles of Econometrics*. John Wiley & Sons, Hoboken, NJ, 4 edn.
- PESARAN, M. H. (2015): *Time Series and Panel Data Econometrics*. Oxford University Press, Oxford.
- WOOLDRIDGE, J. M. (2008): *Introductory Econometrics: A Modern Approach*. South-Western Cengage Learning, Mason, OH, 4 edn.
- (2010): *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA, 2 edn.

Dodatek. Podstawy teoretyczne

Zmienna losowa

Zmienna losowa to funkcja o wartościach rzeczywistych określona na przestrzeni zdarzeń elementarnych.

- **Zmienna losowa** to funkcja, która przypisuje wartości liczbowe wynikami doświadczenia losowego.
- **Dyskretna zmienna losowa** przyjmuje skończoną lub przeliczalnie nieskończoną liczbę wartości.
- **Wskaźnikowa zmienna losowa** przyjmuje tylko dwie wartości:
 - 1 – jeśli dane zdarzenie zajdzie,
 - 0 – jeśli dane zdarzenie nie zajdzie.
- **Ciągła zmienna losowa** może przyjąć dowolną wartość z przedziału).

■ Funkcja gęstości prawdopodobieństwa (pdf):

$$f(x) \geq 0 \quad \text{dla wszystkich } x \quad (51)$$

- Dla ciągłych zmiennych losowych $P(X = x) = 0$ i , funkcja gęstości prawdopodobieństwa (pdf) dla ciągłej zmiennej losowej może być określona jedynie dla przedziału wartości

$$P(a \leq X \leq b) = \int_a^b f(x) dx \quad (52)$$

- Całka funkcji gęstości prawdopodobieństwa na całej dziedzinie wynosi 1:

$$\int_{-\infty}^{\infty} f(x) dx = 1. \quad (53)$$

- Dla dyskretnej zmiennej losowej X , funkcja masy prawdopodobieństwa:

$$p(x) = P(X = x) \geq 0 \quad \text{i} \quad \sum_x p(x) = 1. \quad (54)$$

- **Dystrybuanta** (cumulative distribution function - cdf):

$$F(x) = \int_{-\infty}^x f(t)dt = P(X \leq x). \quad (55)$$

- Własności dystrybuanty:

- ▶ $0 \leq F(x) \leq 1$ dla wszystkich x
- ▶ $F(x)$ jest funkcją niemalejącą
- ▶ $\lim_{x \rightarrow -\infty} F(x) = 0$ i $\lim_{x \rightarrow \infty} F(x) = 1$
- ▶ Dla zmiennych dyskretnych: $F(x) = \sum_{t \leq x} p(t)$

- Dystrybuanta jest przydatna do obliczania prawdopodobieństw. Na przykład:

$$P(X > x) = 1 - P(X \leq x) = 1 - F(x), \quad (56)$$

$$P(a < X \leq b) = F(b) - F(a). \quad (57)$$

- Związek między pdf a cdf:

$$f(x) = \frac{dF(x)}{dx}. \quad (58)$$

- W przypadku (co najmniej dwóch) zmiennych losowych użyteczna jest analiza **rozkładu łącznego**.
- **Łączna funkcja gęstości prawdopodobieństwa** podsumowuje informacje o możliwych wynikach analizowanych zmiennych losowych oraz odpowiadających im prawdopodobieństwach.
- Dla dyskretnych zmiennych losowych:

$$f(x, y) = P(X = x, Y = y)$$

- Dla ciągłych zmiennych losowych:

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y) dy dx \quad (59)$$

- **Rozkład brzegowy** pozwala na wyznaczenie rozkładu pojedynczej zmiennej losowej. Dla zmiennej dyskretnej otrzymujemy go sumując po wszystkich wartościach drugiej zmiennej:

$$f_X(x) = P(X = x) = \sum_y f(x, y)$$

- **Rozkład warunkowy** to rozkład prawdopodobieństwa zmiennej Y pod warunkiem, że wartość zmiennej X jest już znana:

$$f(y|x) = P(Y = y|X = x) = \frac{f(x, y)}{f_X(x)} \quad (60)$$

- Zmienne losowe są **niezależne** wtedy i tylko wtedy, gdy ich łączna funkcja gęstości prawdopodobieństwa (gęstość łączna) jest iloczynem gęstości brzegowych.
- Dla dwóch zmiennych losowych warunek ten zapisujemy jako:

$$f(x, y) = f_X(x)f_Y(y) \quad (61)$$

- **Wartość oczekiwana** zmiennej losowej X to średnia ważona (przez gęstość prawdopodobieństwa) wszystkich możliwych wartości przyjmowanych przez zmienną X .
- **Wartość oczekiwana** (średnia) zmiennej losowej ciągłej:

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x f(x) dx = \mu. \quad (62)$$

- Dla zmiennej losowej dyskretnej:

$$\mathbb{E}(X) = \sum_{i=1}^k x_i p(x_i), \quad (63)$$

gdzie $p(x_i)$ to funkcja masy prawdopodobieństwa.

- Wartość oczekiwana bywa również nazywana **średnią populacyjną** (co istotne: **≠ średnia z próby**).

- Dla dowolnej stałej c :

$$E(c) = c$$

- Dla dowolnych stałych a i b :

$$E(aX + b) = aE(X) + b$$

- Dla dowolnych stałych $a_1, \dots, a_k$ oraz zmiennych losowych $X_1, \dots, X_k$:

$$E\left(\sum_{i=1}^k a_i X_i\right) = \sum_{i=1}^k a_i E(X_i)$$

, a gdy $a_i = 1$ dla wszystkich i , wartość oczekiwana sumy jest po prostu równa sumie wartości oczekiwanych.

- Dla funkcji $g(\cdot)$ tworzącej nową zmienną losową (w przypadku dyskretnym):

$$E(g(X)) = \sum_{i=1}^k g(x_i) f_X(x_i)$$

- oraz dla ciągłej zmiennej losowej:

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)f_X(x) dx$$

- W kontekście rozkładu łącznego, **warunkowa wartość oczekiwana** to wartość oczekiwana zmiennej X , gdy znana jest wartość zmiennej Y . Dla zmiennych dyskretnych:

$$E(X|Y = y) = \sum_{i=1}^k x_i f(x_i|y)$$

a dla zmiennych ciągłych

$$\mathbb{E}(X|Y = y) = \int_{-\infty}^{\infty} x f(x|y) dx.$$

- **Wariancja** to jedna z miar zmienności:

$$\text{var}(X) = \mathbb{E}[(X - \mu)^2], \quad (64)$$

gdzie $\mathbb{E}(X) = \mu$.

- Wariancja jest zwykle oznaczana σ^2 (lub σ_X^2 dla X).
- Alternatywnie, wariancję można wyrazić jako:

$$\begin{aligned} \text{var}(X) &= \mathbb{E}(X^2 - 2X\mu + \mu^2) = \mathbb{E}(X^2) - 2\mu\mathbb{E}(X) + \mu^2 \\ &= \mathbb{E}(X^2) - \mu^2. \end{aligned}$$

- Dla dowolnych stałych a i b :

$$\text{var}(aX + b) = a^2 \text{var}(X). \quad (65)$$

- **Odchylenie standardowe** (*standard deviation, sd*) to pierwiastek kwadratowy z wariancji:

$$\text{sd}(X) = \sqrt{\text{var}(X)} \quad (66)$$

- Dla dowolnych stałych a i b :

$$\text{var}(aX + bY) = a^2 \text{var}(X) + b^2 \text{var}(Y) + 2ab \cdot \text{cov}(X, Y). \quad (67)$$

- **Kowariancja** to miara związku między zmiennymi losowymi.
Dla zmiennych losowych X i Y :

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)], \quad (68)$$

gdzie $\mathbb{E}(Y) = \mu_Y$ i $\mathbb{E}(X) = \mu_X$.

- Kowariancja między X i Y jest czasami oznaczana jako σ_{XY} .
- Kowariancja można także wyrazić jako:

$$\begin{aligned} \text{cov}(X, Y) &= \mathbb{E}[(X - \mu_X)(Y - \mu_Y)] = \mathbb{E}[X(Y - \mu_Y)] \\ &= \mathbb{E}[(X - \mu_X)Y] = \mathbb{E}(XY) - \mu_X\mu_Y \end{aligned}$$

- Ważne: gdy X i Y są niezależne, to $\text{cov}(X, Y) = 0$ i $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.
- Bazując na kowariancji, trudno ocenić wielkość związku między dwiema zmiennymi losowymi. Jednak **korelacja** (ρ) uwzględnia różnice w wariancjach:

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}, \quad (69)$$

gdzie $\rho \in [-1, 1]$.

- Jeśli X jest zmienną losową o rozkładzie normalnym ze średnią μ i wariancją σ^2 , tj. $X \sim \mathcal{N}(\mu, \sigma^2)$, to *funkcja gęstości prawdopodobieństwa* (pdf) X jest:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right], \quad -\infty < x < \infty. \quad (70)$$

- **Standardowy rozkład normalny** występuje, gdy $\mu = 0$ i $\sigma^2 = 1$, ($\mathcal{N}(0, 1)$).
- **Standaryzacja**. Gdy $X \sim \mathcal{N}(\mu, \sigma^2)$, to

$$Z = \frac{X - \mu}{\sigma} \sim \mathcal{N}(0, 1). \quad (71)$$

- Przykład:

$$P(a \leq X \leq b) = F(b) - F(a) = \Phi \left(\frac{b - \mu}{\sigma} \right) - \Phi \left(\frac{a - \mu}{\sigma} \right), \quad (72)$$

gdzie $\Phi(\cdot)$ to dystrybuanta (cdf) standardowego rozkładu normalnego.

- Niech $Z_1, \dots, Z_k$ będą niezależnymi zmiennymi losowymi o standardowym rozkładzie normalnym. Wtedy suma zmiennych Z_i podniesionych do kwadratu ma rozkład chi-kwadrat:

$$V = Z_1^2 + \dots + Z_k^2 \sim \chi^2(k), \quad (73)$$

gdzie k oznacza liczbę stopni swobody. Ważne:

$$\begin{aligned} \mathbb{E}(V) &= k \\ \text{var}(V) &= 2k. \end{aligned}$$

- Jeśli Z jest zmienną losową o standardowym rozkładzie normalnym i $V \sim \chi^2(k)$, to

$$t = \frac{Z}{\sqrt{V/k}} \sim t_k. \quad (74)$$

Dodatek. Właściwości asymptotyczne estymatora MNK

- Nieobciążoność estymatora:

$$\mathbb{E}(\hat{\beta}^{MNK}) = \beta. \quad (75)$$

- Używając prawdziwego DGP, tj. $\mathbf{y} = \mathbf{X}\beta + \varepsilon$, możemy zapisać estymator metody najmniejszych kwadratów:

$$\hat{\beta}^{MNK} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \varepsilon). \quad (76)$$

- Używając $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{I}$:

$$\hat{\beta}^{MNK} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon. \quad (77)$$

- Wartość oczekiwana:

$$\mathbb{E}(\hat{\beta}^{MNK}) = \mathbb{E}(\beta) + \mathbb{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon], \quad (78)$$

- Używając założenia, że $\mathbb{E}(\beta) = \beta$ i że zmienne objaśniające nie są losowe:

$$\mathbb{E}(\hat{\beta}^{MNK}) = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}(\varepsilon), \quad (79)$$

- Przy założeniu egzogeniczności $\mathbb{E}(\varepsilon|\mathbf{X}) = 0$ otrzymujemy:

$$\mathbb{E}(\hat{\beta}^{MNK}) = \beta. \quad (80)$$

- Z poprzednich wyprowadzeń:

$$\hat{\beta}^{MNK} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon. \quad (81)$$

- Niech $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Wtedy:

$$\hat{\beta}^{MNK} = \beta + \mathbf{A}\varepsilon (= \mathbf{A}y), \quad (82)$$

- Tak więc $\hat{\beta}^{MNK}$ jest liniową funkcją składnika losowego, który jest zmienną losową. W rezultacie estymator $\hat{\beta}^{MNK}$ jest liniowy.

- Rozważmy pewien liniowy, nieobciążony estymator β , np. $\hat{\beta} = \mathbf{C}\mathbf{y}$. Wtedy:

$$\mathbb{E}(\hat{\beta}) = \mathbb{E}(\mathbf{C}\mathbf{X}\beta + \mathbf{C}\varepsilon) = \beta \quad (83)$$

- Można zauważyć, że $\mathbf{C}\mathbf{X} = \mathbb{I}$ (ponieważ założyliśmy, że $\hat{\beta}$ jest nieobciążony).
- Wariancja estymatora $\hat{\beta}$:

$$\text{Var}(\hat{\beta}) = \sigma^2 \mathbf{C}\mathbf{C}'. \quad (84)$$

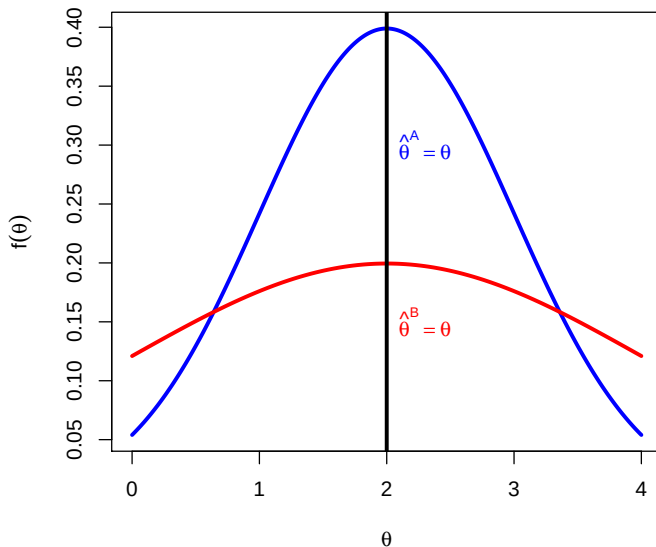
- Zdefiniujmy $\mathbf{D} = \mathbf{C} - (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$.
- Wariancja estymatora $\hat{\beta}$:

$$\text{Var}(\hat{\beta}) = \sigma^2 \left[\left(\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right) \left(\mathbf{D} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \right)' \right], \quad (85)$$

- $\mathbf{D}\mathbf{X} = 0$, ponieważ $\mathbf{C}\mathbf{X} = \mathbb{I} = \mathbf{D}\mathbf{X} + (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X})$.
- Wariancje można zapisać jako:

$$\text{Var}(\hat{\beta}) = \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} + \sigma^2 \mathbf{D}\mathbf{D}' = \text{Var}(\hat{\beta}^{MNK}) + \sigma^2 \mathbf{D}\mathbf{D}'. \quad (86)$$

- Jeśli $\mathbf{D}$ jest macierzą zerową, to wariancja $\hat{\beta}$ jest najmniejsza i rozważanym estymatorem jest estymator metody najmniejszych kwadratów.



- **Estymator zgodny** zbiega według prawdopodobieństwa do prawdziwej wartości nieznanego parametru.
- Zgodność estymatora MNK:

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}^{MNK} = \beta \quad (87)$$

- Z poprzednich wyprowadzeń ($\mathbb{E}(\hat{\beta}^{MNK}) = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}(\varepsilon)$):

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}^{MNK} = \beta + \text{plim}_{N \rightarrow \infty} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbb{E}(\varepsilon). \quad (88)$$

- Mnożymy przez jeden, tj. $1 = 1/N \times N$:

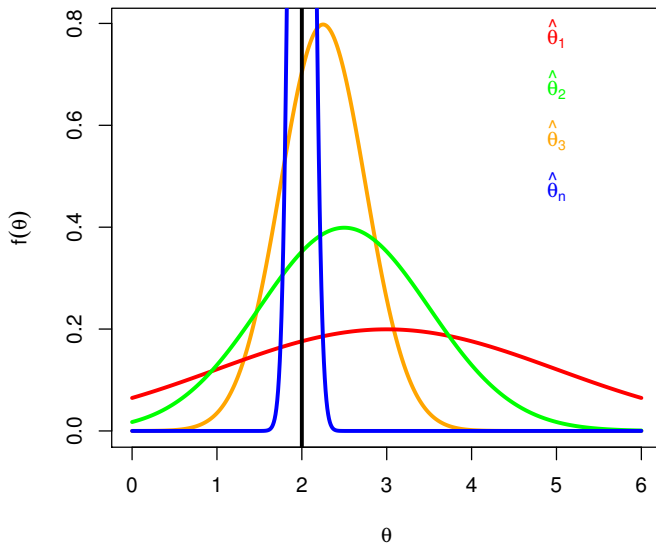
$$\text{plim}_{N \rightarrow \infty} \hat{\beta}^{MNK} = \beta + \text{plim}_{N \rightarrow \infty} \left(\frac{1}{N} \mathbf{X}'\mathbf{X} \right)^{-1} \frac{1}{N} \mathbf{X}'\mathbb{E}(\varepsilon). \quad (89)$$

- Używając założenia egzogeniczności:

$$\text{plim}_{N \rightarrow \infty} \frac{1}{N} \mathbf{X}'\mathbb{E}(\varepsilon) = 0, \quad (90)$$

- Trudno ograniczyć wyrażenie $\text{plim}_{N \rightarrow \infty} (\mathbf{X}'\mathbf{X})$, ale wyrażenie $\text{plim}_{N \rightarrow \infty} (1/N \mathbf{X}'\mathbf{X})$ można ograniczyć przez niektóre $\mathcal{C}$. Wtedy:

$$\text{plim}_{N \rightarrow \infty} \hat{\beta}^{MNK} = \beta + \mathcal{C} \times 0 = \beta. \quad (91)$$



- Rozważamy model prostej regresji liniowej: $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$.
- Estymatorem MNK wyrazu wolnego jest $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.
- Resztę $\hat{\varepsilon}_i = y_i - \hat{y}_i$ możemy przedstawić jako:

$$\hat{\varepsilon}_i = y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i = (y_i - \bar{y}) - \hat{\beta}_1 (x_i - \bar{x})$$

- Prawdziwy model w ujęciu odchyień od średniej to:

$$y_i - \bar{y} = \beta_1 (x_i - \bar{x}) + (\varepsilon_i - \bar{\varepsilon})$$

- Podstawiając go do wzoru na resztę i grupując wyrazy, uniezależniamy reszty od y :

$$\hat{\varepsilon}_i = (\varepsilon_i - \bar{\varepsilon}) - (\hat{\beta}_1 - \beta_1)(x_i - \bar{x})$$

- Aby wyznaczyć sumę kwadratów reszt, podnosimy równanie do kwadratu i sumujemy po wszystkich n obserwacjach:

$$\sum_{i=1}^n \hat{e}_i^2 = \sum_{i=1}^n [(\varepsilon_i - \bar{\varepsilon}) - (\hat{\beta}_1 - \beta_1)(x_i - \bar{x})]^2$$

- Rozwijając zgodnie ze wzorem skróconego mnożenia:

$$\sum_{i=1}^n \hat{e}_i^2 = \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 - 2(\hat{\beta}_1 - \beta_1) \sum_{i=1}^n (x_i - \bar{x})(\varepsilon_i - \bar{\varepsilon}) + (\hat{\beta}_1 - \beta_1)^2 \sum_{i=1}^n (x_i - \bar{x})^2$$

- Z definicji estymatora $\hat{\beta}_1$ zachodzi tożsamość:

$$\sum (x_i - \bar{x})\varepsilon_i = (\hat{\beta}_1 - \beta_1) \sum (x_i - \bar{x})^2.$$

- Suma kwadratów reszt

$$\sum_{i=1}^n \hat{e}_i^2 = \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 - (\hat{\beta}_1 - \beta_1)^2 \sum_{i=1}^n (x_i - \bar{x})^2$$

- Wartość oczekiwana uproszczonego równania sumy kwadratów reszt:

$$E \left(\sum_{i=1}^n \hat{e}_i^2 \right) = E \left[\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 \right] - E \left[(\hat{\beta}_1 - \beta_1)^2 \sum_{i=1}^n (x_i - \bar{x})^2 \right]$$

- Pierwszy człon to wartość oczekiwana wariancji błędu losowego dla próby:

$$E \left[\sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2 \right] = (n-1)\sigma^2$$

- W drugim członie znajduje się definicja wariancji estymatora $Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}$:

$$E \left[(\hat{\beta}_1 - \beta_1)^2 \right] \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = \sigma^2$$

- Łącząc wyniki otrzymujemy:

$$E \left(\sum_{i=1}^n \hat{e}_i^2 \right) = (n-1)\sigma^2 - \sigma^2 = (n-2)\sigma^2 \quad \Rightarrow \quad E \left(\frac{\sum_{i=1}^n \hat{e}_i^2}{n-2} \right) = \sigma^2$$