HETEROSKEDASTYCZNOŚĆ

EKONOMETRIA WNE 2022/23

Sebastian Zalas

15 grudnia 2022

MODEL - PRZYPOMNIENIE

- Macierz $\boldsymbol{X}$ $(n \times k)$ obserwacji k zmiennych niezależnych
- wektor $\boldsymbol{y}$ $(n \times 1)$ obserwacji zm. zależnej
- wektor $\boldsymbol{\varepsilon}$ $(n \times 1)$ składników losowych.
- wektor $\boldsymbol{\beta}$ $(k + 1 \times 1)$ wektor $k + 1$ parametrów do oszacowania
- Zapis macierzowy:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} 1 & x_{1,1} & \dots & x_{1,k} \\ 1 & x_{2,1} & \dots & x_{2,k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & \dots & x_{n,k} \end{bmatrix}_{n\times k+1} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}_{k+1\times 1} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}_{n\times 1}$$

$$\boldsymbol{y} = \boldsymbol{X} \boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

WARIANCJA ESTYMATORA MNK I

- Definicja wariancji. Niech $\boldsymbol{Z}$ będzie wektorem losowym:

$$\mathrm{Var}[\boldsymbol{Z}] = \mathbb{E}[(\boldsymbol{Z} - \mathbb{E}[\boldsymbol{Z}])(\boldsymbol{Z} - \mathbb{E}[\boldsymbol{Z}])^{\mathsf{T}}]$$

- Wariancja warunkowa:

$$\mathrm{Var}[\boldsymbol{X} \mid \boldsymbol{Z}] = \mathbb{E}[(\boldsymbol{Z} - \mathbb{E}[\boldsymbol{Z}])(\boldsymbol{Z} - \mathbb{E}[\boldsymbol{Z}])^{\mathsf{T}} \mid \boldsymbol{X}]$$

- Wariancja składnika losowego:

$$\mathrm{Var}[\varepsilon \mid \boldsymbol{X}] = \mathbb{E}[\varepsilon\varepsilon^{\mathsf{T}} \mid \boldsymbol{X}]$$

WARIANCJA ESTYMATORA MNK II

Przywołajmy dwa fakty:

1. $\mathbb{E}[\hat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$ - nieobciążoność
2. $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\varepsilon}$

- Wyprowadzimy formułę wariancji estymatora MNK, czyli $\mathrm{Var}[\hat{\boldsymbol{\beta}}]$:

$$
\begin{aligned}
\mathrm{Var}[\hat{\boldsymbol{\beta}} \mid \boldsymbol{X}] &= \mathbb{E}[(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mid \boldsymbol{X}] \\
&= \mathbb{E}[((\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\varepsilon})((\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\varepsilon})^\top \mid \boldsymbol{X}] \\
&= \mathbb{E}[(\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^\top \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \mid \boldsymbol{X}] \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \, \mathbb{E}[\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^\top \mid \boldsymbol{X}] \, \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1}
\end{aligned}
$$

- wariancja estymatora $\hat{\boldsymbol{\beta}}$ zależy od wariancji składnika losowego $\mathrm{Var}[\boldsymbol{\varepsilon} \mid \boldsymbol{X}]$

WARIANCJA ESTYMATORA MNK - HOMOSKEDASTYCZNOŚĆ

- Na mocy założenia tw. Gaussa-Markova obowiązuje **sferyczność składnika losowego**:

$$\mathrm{Var}[\varepsilon \mid \boldsymbol{X}] = \mathbb{E}[\varepsilon\varepsilon^\top \mid \boldsymbol{X}] = \boldsymbol{\Omega} = \begin{bmatrix} \sigma^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma^2 \end{bmatrix} = \sigma^2 \boldsymbol{I}$$

  - homoskedastyczność $\Rightarrow$ stałość wariancji $\Rightarrow$ takie same elementy na diagonali macierzy $\boldsymbol{\Omega}$
  - brak autokorelacji $\Rightarrow$ zera poza diagonalą macierzy $\boldsymbol{\Omega}$

- W takim przypadku, wariancja **estymatora MNK** przyjmuje postać:

$$\mathrm{Var}[\hat{\beta} \mid \boldsymbol{X}] = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top \sigma^2 \boldsymbol{I}\, \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$$

- Estymator wariancji estymatora MNK:

$$\widehat{\mathrm{Var}[\hat{\beta} \mid \boldsymbol{X}]} = s^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1} = \frac{\hat{\varepsilon}^\top\hat{\varepsilon}}{n-k}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$$

WARIANCJA ESTYMATORA MNK - HETEROSKEDASTYCZNOŚĆ

- Analizowane dane mogą jednak nie spełniać założenia o sferyczności składnika losowego. Wtedy jego wariancja ma postać:

$$\mathrm{Var}[\varepsilon \mid \boldsymbol{X}] = \mathbb{E}[\varepsilon\varepsilon^\top \mid \boldsymbol{X}] = \boldsymbol{\Omega} = \begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_n^2 \end{bmatrix}$$

  - heteroskedastyczność $\Rightarrow$ brak stałości wariancji $\Rightarrow$ różne elementy na diagonali macierzy $\boldsymbol{\Omega}$
  - utrzymujemy założenia o braku autokorelacji $\Rightarrow$ zera poza diagonalą macierzy $\boldsymbol{\Omega}$

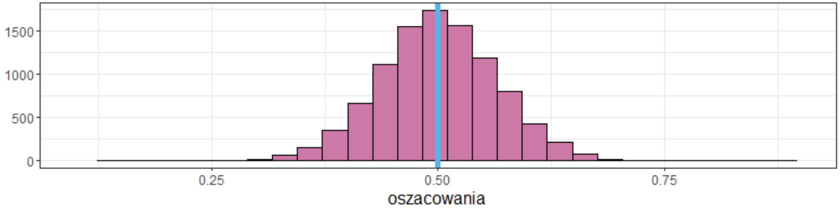
- W takim przypadku, wariancja **estymatora MNK** przyjmuje postać:

$$\mathrm{Var}[\hat{\beta} \mid \boldsymbol{X}] = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Omega}\,\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$$

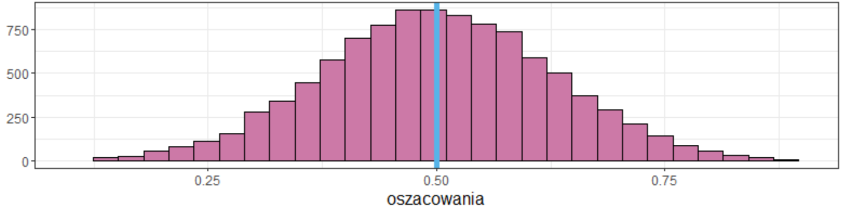
KONSEKWENCJE HETEROSKEDASTYCZNOŚCI

- Nieobciążoność $\hat{\beta}$ pozostaje nienaruszona.
- Jednak estymator $\hat{\beta}$ będzie nieefektywny tzn. możemy znaleźć estymator o niższej wariancji.
- Oznacza to, że precyzja estymatora się zmniejsza, co wypływa także na jakość wnioskowania statystycznego z oszacowanego modelu.

SYMULACJA - HOMOSKEDASTYCZNOŚĆ VS HETEROSKEDASTYCZNOŚĆ



(a) Rozkład $\hat{\beta}$ w modelu z **homoskedastycznym** składnikiem losowym.



(b) Rozkład $\hat{\beta}$ w modelu z **heteroskedastycznym** składnikiem losowym.

TEST BREUSCH'A - PAGAN'A

- Hipoteza zerowa

$$H_0 : \mathbb{E}[\varepsilon^2 \mid x_1, x_2, \ldots, x_k] = \mathbb{E}[\varepsilon^2] = \sigma^2.$$

- Oszacuj model, uzyskaj kwadraty reszt, $\widehat{u}^2$:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

- Oszacuj model, policz $R^2_{\widehat{u}^2}$ z tej regresji:

$$\widehat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + +\delta_k x_k + e$$

- Oblicz statystykę, uzyskaj *p-value*:

$$F = \frac{R^2_{\widehat{u}^2} \frac{1}{k}}{(1 - R^2_{\widehat{u}^2}) \frac{1}{n-k-1}} \sim F_{k,n-k-1}$$

  lub skorzystaj ze statystyki:

$$LM = n \times R^2_{\widehat{u}^2} \sim \chi^2_k$$

TEST WHITE'A

- Hipoteza zerowa

$$H_0 : \mathbb{E}[\varepsilon^2 \mid x_1, x_2, \ldots, x_k] = \mathbb{E}[\varepsilon^2] = \sigma^2.$$

- Oszacuj model, uzyskaj kwadraty reszt, $\widehat{u}^2$:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

- Oszacuj model (z kwadratami i interakcjami), policz $R^2_{\widehat{u}^2}$ z tej regresji:

$$\widehat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \ldots + \delta_k x_k + \sum_{i=1}^{k} \sum_{j=1}^{k} x_i x_j + e$$

- Oblicz statystykę i uzyskaj *p-value* tak jak w przypadku testu BP.

ESTYMATOR WARIANCJI ODPORNY NA HETEROSKEDASTYCZNOŚĆ (I)

- Gdy występuje problem heteroskedastyczności, zwykły estymator wariancji może być obciążony. W takiej sytuacji, musimy skonstruować taki estymator wariancji, który nie wymaga założenia o homoskedastyczności.

- Idealnie byłoby, gdybyśmy mogli mieć:

$$\widehat{\mathrm{Var}[\hat{\beta} \mid \boldsymbol{X}]}_{ideal} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \, \mathbb{E}[\varepsilon \varepsilon^\top \mid \boldsymbol{X}] \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \begin{bmatrix} \varepsilon_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \varepsilon_n^2 \end{bmatrix} \boldsymbol{X} (\boldsymbol{X}^\top \boldsymbol{X})^{-1}$$

- Otrzymanie takiego estymatora nie jest możliwe ponieważ mamy za mało stopni swobody.

ESTYMATOR WARIANCJI ODPORNY NA HETEROSKEDASTYCZNOŚĆ (II)

- White (1980) pokazał, że poniższy estymator wariancji estymatora MNK:

$$\widehat{\mathrm{Var}[\hat{\beta} \mid \boldsymbol{X}]} = (\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\Big(\sum_{i=1}^{n} X_i X_i^{\top} \widehat{\varepsilon}_i^2\Big)(\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}$$

jest zgodnym estymatorem $\widehat{\mathrm{Var}[\hat{\beta} \mid \boldsymbol{X}]}_{ideal}$, przez co jest odporny na heteroskedastyczność.

Pytania? Wątpliwości?

Dziękuję!

e: s.zalas@uw.edu.pl