

# Dane panelowe

## Część 2. Estymatory

- Do szacowania parametrów modeli panelowego wykorzystuje się jeden z dwóch estymatorów

- Do szacowania parametrów modeli panelowego wykorzystuje się jeden z dwóch estymatorów
  - estymator efektów losowych *random effects*

- Do szacowania parametrów modeli panelowego wykorzystuje się jeden z dwóch estymatorów
  - estymator efektów losowych *random effects*
  - estymator efektów stałych *fixed effects*

- Do szacowania parametrów modeli panelowego wykorzystuje się jeden z dwóch estymatorów
  - estymator efektów losowych *random effects*
  - estymator efektów stałych *fixed effects*
- Pierwotnie zakładano różną naturę efektów indywidualnych

- Do szacowania parametrów modeli panelowego wykorzystuje się jeden z dwóch estymatorów
  - estymator efektów losowych *random effects*
  - estymator efektów stałych *fixed effects*
- Pierwotnie zakładano różną naturę efektów indywidualnych
- Obecnie efekt indywidualny uważa się za zmienną losową, to który estymator należy wykorzystać zależy od ewentualnej korelacji między efektem indywidualnym a charakterystykami obiektu

- Estymator efektów losowych jest zgodny, gdy

- Estymator efektów losowych jest zgodny, gdy
  - ① Dla każdego obiektu  $i$  zmienne objaśniające są nieskorelowane z równoczesnym błędem czysto losowym  $\varepsilon_{it}$

$$\forall i, t \text{ cov}(x_{it}, \varepsilon_i) = 0$$

- Estymator efektów losowych jest zgodny, gdy
  - 1 Dla każdego obiektu  $i$  zmienne objaśniające są nieskorelowane z równoczesnym błędem czysto losowym  $\varepsilon_{it}$

$$\forall i, t \text{ cov}(x_{it}, \varepsilon_i) = 0$$

- 2 Dla każdego obiektu  $i$  zmienne objaśniające są nieskorelowane z efektem indywidualnym  $u_{it}$

$$\forall i, t \text{ cov}(x_{it}, u_i) = 0$$

- Estymator efektów losowych jest efektywny, gdy dodatkowo

- Estymator efektów losowych jest efektywny, gdy dodatkowo
  - 3 błąd czysto losowy i efekt indywidualny są nieskorelowane

$$\forall i, t \text{ cov}(u_i, \varepsilon_{it}) = 0$$

- Estymator efektów losowych jest efektywny, gdy dodatkowo
  - 3 błąd czysto losowy i efekt indywidualny są nieskorelowane

$$\forall i, t \text{ cov}(u_i, \varepsilon_{it}) = 0$$

- 4 Efekty indywidualne i błędy czysto losowe są homoscedastyczne i nieskorelowane

$$\forall i \text{ var}(u_i) = \sigma_u^2 \mathbb{I}$$

$$\forall i, t \text{ var}(\varepsilon_{it}) = \sigma_\varepsilon^2$$

- Model zapisujemy w formie

$$y_{it} = X_{it}\beta + v_{it} \quad v_{it} = u_i + \varepsilon_{it}$$

- Model zapisujemy w formie

$$y_{it} = X_{it}\beta + v_{it} \quad v_{it} = u_i + \varepsilon_{it}$$

- Przypomnijmy, że

$$\text{var}(v_{it}) = \text{var}(u_i + \varepsilon_{it}) = \text{var}(u_i) + \text{var}(\varepsilon_{it}) = \sigma_u^2 + \sigma_\varepsilon^2$$

- Model zapisujemy w formie

$$y_{it} = X_{it}\beta + v_{it} \quad v_{it} = u_i + \varepsilon_{it}$$

- Przypomnijmy, że

$$\text{var}(v_{it}) = \text{var}(u_i + \varepsilon_{it}) = \text{var}(u_i) + \text{var}(\varepsilon_{it}) = \sigma_u^2 + \sigma_\varepsilon^2$$

- oraz

$$\text{cov}(v_{it}, v_{js}) = \text{cov}(u_i, u_j) + \text{cov}(u_i, \varepsilon_{js}) + \text{cov}(u_j, \varepsilon_{it}) + \text{cov}(\varepsilon_{it}, \varepsilon_{js})$$

- Model zapisujemy w formie

$$y_{it} = X_{it}\beta + v_{it} \quad v_{it} = u_i + \varepsilon_{it}$$

- Przypomnijmy, że

$$\text{var}(v_{it}) = \text{var}(u_i + \varepsilon_{it}) = \text{var}(u_i) + \text{var}(\varepsilon_{it}) = \sigma_u^2 + \sigma_\varepsilon^2$$

- oraz

$$\text{cov}(v_{it}, v_{js}) = \text{cov}(u_i, u_j) + \text{cov}(u_i, \varepsilon_{js}) + \text{cov}(u_j, \varepsilon_{it}) + \text{cov}(\varepsilon_{it}, \varepsilon_{js})$$

- więc

$$\text{cov}(v_{it}, v_{js}) = \sigma_u^2$$

- W rezultacie macierz wariancji kowariancji łącznego błędu losowego ma postać

$$\Omega_i = E(v_i, v_i') = \begin{bmatrix} \sigma_u^2 + \sigma_\varepsilon^2 & \sigma_u^2 & \dots & \sigma_u^2 \\ \sigma_u^2 & \sigma_u^2 + \sigma_\varepsilon^2 & \dots & \sigma_u^2 \\ \vdots & & \ddots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \dots & \sigma_u^2 + \sigma_\varepsilon^2 \end{bmatrix}$$

- Model można zapisać z skondensowanej postaci

$$y = X\beta + \nu$$

- Model można zapisać z skondensowanej postaci

$$y = X\beta + \nu$$

- Na mocy założenia 3 efekt indywidualny i błąd czysto losowy są nieskorelowane

- Model można zapisać z skondensowanej postaci

$$y = X\beta + \nu$$

- Na mocy założenia 3 efekt indywidualny i błąd czysto losowy są nieskorelowane
- W rezultacie macierz  $\Omega$  ma postać blokowo-diagonalną

$$\Omega = E(\nu_i, \nu_i') = \begin{bmatrix} \Omega & & 0 \\ & \ddots & \\ 0 & & \Omega \end{bmatrix}$$

- Zatem estymatorem efektywnym jest estymator UMNK

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

- Zatem estymatorem efektywnym jest estymator UMNK

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

- Ponieważ  $\Omega$  jest macierzą blokowo diagonalną to

$$X'\Omega^{-1}X = \sum_{i=1}^N X_i'\Omega_i^{-1}X_i$$

$$X'\Omega^{-1}y = \sum_{i=1}^N X_i'\Omega_i^{-1}y_i$$

- Zatem estymatorem efektywnym jest estymator UMNK

$$\hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

- Ponieważ  $\Omega$  jest macierzą blokowo diagonalną to

$$X'\Omega^{-1}X = \sum_{i=1}^N X_i'\Omega_i^{-1}X_i$$

$$X'\Omega^{-1}y = \sum_{i=1}^N X_i'\Omega_i^{-1}y_i$$

- Więc estymatorem efektów losowych jest

$$\hat{\beta}_{RE} = \left( \sum_{i=1}^N X_i'\Omega_i^{-1}X_i \right)^{-1} \left( \sum_{i=1}^N X_i'\Omega_i^{-1}y_i \right)$$

- Bezpośrednie wykorzystanie wzoru nie jest możliwe, gdyż nie jest znana postać macierzy kowariancji  $\Omega$

- Bezpośrednie wykorzystanie wzoru nie jest możliwe, gdyż nie jest znana postać macierzy kowariancji  $\Omega$
- W przypadku estymatora efektów losowych jej oszacowanie jest stosunkowo proste

- Bezpośrednie wykorzystanie wzoru nie jest możliwe, gdyż nie jest znana postać macierzy kowariancji  $\Omega$
- W przypadku estymatora efektów losowych jej oszacowanie jest stosunkowo proste
- Idea bazuje na wykorzystaniu dekompozycji wariancji

- W pierwszym kroku na podstawie całej próby szacowane są parametry modelu. Oszacowanie wariancji błędu losowego wynosi

$$\hat{\sigma}_v^2 = \frac{1}{NT - k} \sum_{i=1}^N \sum_{t=1}^T \hat{v}_{it}^2$$

- W pierwszym kroku na podstawie całej próby szacowane są parametry modelu. Oszacowanie wariancji błędu losowego wynosi

$$\hat{\sigma}_v^2 = \frac{1}{NT - k} \sum_{i=1}^N \sum_{t=1}^T \hat{v}_{it}^2$$

- gdzie  $v_{it}$  to reszty z regresji MNK

- W pierwszym kroku na podstawie całej próby szacowane są parametry modelu. Oszacowanie wariancji błędu losowego wynosi

$$\hat{\sigma}_v^2 = \frac{1}{NT - k} \sum_{i=1}^N \sum_{t=1}^T \hat{v}_{it}^2$$

- gdzie  $v_{it}$  to reszty z regresji MNK
- Oszacowaniem  $\sigma_u^2$  jest

$$\hat{\sigma}_u^2 = \frac{1}{NT \frac{T-1}{2} - k} \sum_{i=1}^N \sum_{t=1}^{t-1} \sum_{s=1}^T \hat{v}_{it} \hat{v}_{is}$$

- W pierwszym kroku na podstawie całej próby szacowane są parametry modelu. Oszacowanie wariancji błędu losowego wynosi

$$\hat{\sigma}_v^2 = \frac{1}{NT - k} \sum_{i=1}^N \sum_{t=1}^T \hat{v}_{it}^2$$

- gdzie  $v_{it}$  to reszty z regresji MNK
- Oszacowaniem  $\sigma_u^2$  jest

$$\hat{\sigma}_u^2 = \frac{1}{NT \frac{T-1}{2} - k} \sum_{i=1}^N \sum_{t=1}^{t-1} \sum_{s=1}^T \hat{v}_{it} \hat{v}_{is}$$

- Ponieważ  $\sigma_v^2 = \sigma_u^2 + \sigma_\varepsilon^2$  to

$$\hat{\sigma}_\varepsilon^2 = \hat{\sigma}_v^2 - \hat{\sigma}_u^2$$

- W pierwszym kroku na podstawie całej próby szacowane są parametry modelu. Oszacowanie wariancji błędu losowego wynosi

$$\hat{\sigma}_v^2 = \frac{1}{NT - k} \sum_{i=1}^N \sum_{t=1}^T \hat{v}_{it}^2$$

- gdzie  $v_{it}$  to reszty z regresji MNK
- Oszacowaniem  $\sigma_u^2$  jest

$$\hat{\sigma}_u^2 = \frac{1}{NT \frac{T-1}{2} - k} \sum_{i=1}^N \sum_{t=1}^{t-1} \sum_{s=1}^T \hat{v}_{it} \hat{v}_{is}$$

- Ponieważ  $\sigma_v^2 = \sigma_u^2 + \sigma_\varepsilon^2$  to

$$\hat{\sigma}_\varepsilon^2 = \hat{\sigma}_v^2 - \hat{\sigma}_u^2$$

- Niestety, nie ma gwarancji że ostatni estymator będzie dodatni

- Estymator efektów losowych (RE) jest bardziej efektywny niż estymator MNK, jeżeli efekty indywidualne są istotne w analizowanej populacji

- Estymator efektów losowych (RE) jest bardziej efektywny niż estymator MNK, jeżeli efekty indywidualne są istotne w analizowanej populacji
- Ich występowanie można weryfikować wykorzystując test LM Breuscha - Pagana na istnienie efektów indywidualnych

- Estymator efektów losowych (RE) jest bardziej efektywny niż estymator MNK, jeżeli efekty indywidualne są istotne w analizowanej populacji
- Ich występowanie można weryfikować wykorzystując test LM Breuscha - Pagana na istnienie efektów indywidualnych
- Hipoteza zerowa - brak efektów indywidualnych, co jest równoważne zerowej wariancji efektu indywidualnego  $\sigma_u^2 = 0$

- Estymator efektów losowych (RE) jest bardziej efektywny niż estymator MNK, jeżeli efekty indywidualne są istotne w analizowanej populacji
- Ich występowanie można weryfikować wykorzystując test LM Breuscha - Pagana na istnienie efektów indywidualnych
- Hipoteza zerowa - brak efektów indywidualnych, co jest równoważne zerowej wariancji efektu indywidualnego  $\sigma_u^2 = 0$
- Hipoteza alternatywna  $\sigma_u^2 > 0$

- Estymator efektów losowych (RE) jest bardziej efektywny niż estymator MNK, jeżeli efekty indywidualne są istotne w analizowanej populacji
- Ich występowanie można weryfikować wykorzystując test LM Breuscha - Pagana na istnienie efektów indywidualnych
- Hipoteza zerowa - brak efektów indywidualnych, co jest równoważne zerowej wariancji efektu indywidualnego  $\sigma_u^2 = 0$
- Hipoteza alternatywna  $\sigma_u^2 > 0$
- Odrzucenie  $H_0$  wskazuje na istotność efektu indywidualnego

# Przykład

Breusch and Pagan Lagrangian multiplier test for random effects

$$\ln\text{placa}[\text{ID}, t] = Xb + u[\text{ID}] + e[\text{ID}, t]$$

Estimated results:

|             | Var      | sd = sqrt(Var) |
|-------------|----------|----------------|
| -----+----- |          |                |
| lnplaca     | .2516836 | .5016808       |
| e           | .0113868 | .1067089       |
| u           | .1737    | .4167734       |

Test: Var(u) = 0

chi2(1) = 8858.38  
 Prob > chi2 = 0.0000

# Przykład

Random-effects GLS regression  
Group variable: ID

Number of obs = 9902  
Number of groups = 3309

R-sq: within = 0.0226  
between = 0.3270  
overall = 0.3232

Obs per group: min = 1  
avg = 3.0  
max = 4

Random effects u\_i ~ Gaussian  
corr(u\_i, X) = 0 (assumed)

Wald chi2(14) = 1668.42  
Prob > chi2 = 0.0000

| lnplaca | Coef.     | Std. Err. | z      | P> z  | [95% Conf. Interval] |           |
|---------|-----------|-----------|--------|-------|----------------------|-----------|
| wiek    | .5441276  | .0438852  | 12.40  | 0.000 | .4581142             | .630141   |
| wiek2   | -.0178531 | .0016521  | -10.81 | 0.000 | -.0210912            | -.014615  |
| wiek3   | .0002553  | .0000264  | 9.66   | 0.000 | .0002035             | .0003071  |
| wiek4   | -1.34e-06 | 1.52e-07  | -8.84  | 0.000 | -1.64e-06            | -1.05e-06 |
| plec    | -.2245747 | .0153559  | -14.62 | 0.000 | -.2546717            | -.1944777 |

# Przykład

| lnplaca  | Coef.     | Std. Err.                         | z      | P> z  | [95% Conf. Interval] |           |
|----------|-----------|-----------------------------------|--------|-------|----------------------|-----------|
| wykszt_2 | -.2194375 | .0186776                          | -11.75 | 0.000 | -.2560449            | -.1828301 |
| wykszt_3 | -.3824129 | .0215858                          | -17.72 | 0.000 | -.4247203            | -.3401055 |
| wykszt_4 | -.62673   | .0294189                          | -21.30 | 0.000 | -.6843899            | -.56907   |
| klm_2    | -.0844443 | .0312316                          | -2.70  | 0.007 | -.1456571            | -.0232315 |
| klm_3    | -.1280112 | .0259817                          | -4.93  | 0.000 | -.1789344            | -.077088  |
| klm_4    | -.180314  | .027437                           | -6.57  | 0.000 | -.2340894            | -.1265385 |
| rok_2    | .0012748  | .003125                           | 0.41   | 0.683 | -.0048502            | .0073997  |
| rok_3    | .0143942  | .0033968                          | 4.24   | 0.000 | .0077366             | .0210518  |
| rok_4    | .0223004  | .0034923                          | 6.39   | 0.000 | .0154557             | .0291452  |
| _cons    | 1.285498  | .4190395                          | 3.07   | 0.002 | .4641958             | 2.1068    |
| sigma_u  | .41677338 |                                   |        |       |                      |           |
| sigma_e  | .10670893 |                                   |        |       |                      |           |
| rho      | .93847863 | (fraction of variance due to u_i) |        |       |                      |           |

- W przypadku danych panelowych mówimy o dwóch rodzajach wariancji

- W przypadku danych panelowych mówimy o dwóch rodzajach wariancji
  - zmienność wewnątrzgrupowa

- W przypadku danych panelowych mówimy o dwóch rodzajach wariancji
  - zmienność wewnątrzgrupowa
  - zmienność międzygrupowa

- W przypadku danych panelowych mówimy o dwóch rodzajach wariancji
  - zmienność wewnątrzgrupowa
  - zmienność międzygrupowa
- Możliwe jest zdefiniowanie współczynników dopasowania  $R^2$  obrazujących oba typy zmienności

- Zdefiniujemy następujące wartości dopasowane

- Zdefiniujemy następujące wartości dopasowane
  - Wartość dopasowana z modelu

$$\hat{y}_{it} = x_{it}\beta$$

- Zdefiniujemy następujące wartości dopasowane
  - Wartość dopasowana z modelu

$$\hat{y}_{it} = x_{it}\beta$$

- Średnia z wartości dopasowanych dla obiektu

$$\hat{\bar{y}}_i = \bar{x}_i\beta$$

- Zdefiniujemy następujące wartości dopasowane
  - Wartość dopasowana z modelu

$$\hat{y}_{it} = x_{it}\beta$$

- Średnia z wartości dopasowanych dla obiektu

$$\hat{\bar{y}}_i = \bar{x}_i\beta$$

- Odchylenie wartości dopasowanej od średniej

$$\hat{\hat{y}}_{it} = (\hat{y}_{it} - \hat{\bar{y}}_i) = (x_{it} - \bar{x}_i)\beta$$

- Współczynnik  $R^2$  to kwadrat współczynnika korelacji między wartościami zmiennej zależnej a wartościami dopasowanymi

- Współczynnik  $R^2$  to kwadrat współczynnika korelacji między wartościami zmiennej zależnej a wartościami dopasowanymi
- Całkowity  $R^2$  to kwadrat współczynnika korelacji między  $y_t$  a  $\hat{y}_{it}$

- Współczynnik  $R^2$  to kwadrat współczynnika korelacji między wartościami zmiennej zależnej a wartościami dopasowanymi
- Całkowity  $R^2$  to kwadrat współczynnika korelacji między  $y_t$  a  $\hat{y}_{it}$
- Międzygrupowy  $R^2$  to kwadrat współczynnika korelacji między  $y_t$  a  $\hat{y}_i$

- Współczynnik  $R^2$  to kwadrat współczynnika korelacji między wartościami zmiennej zależnej a wartościami dopasowanymi
- Całkowity  $R^2$  to kwadrat współczynnika korelacji między  $y_t$  a  $\hat{y}_{it}$
- Międzygrupowy  $R^2$  to kwadrat współczynnika korelacji między  $y_t$  a  $\hat{y}_i$
- Wewnątrzgrupowy  $R^2$  to kwadrat współczynnika korelacji między  $y_t$  a  $\hat{y}_{it}$

- Relacja między wartościami  $R^2$  pokazuje czy lepiej wyjaśniane jest kształtowanie zjawiska w czasie, czy zróżnicowanie zjawiska między obiektami

- Relacja między wartościami  $R^2$  pokazuje czy lepiej wyjaśniane jest kształtowanie zjawiska w czasie, czy zróżnicowanie zjawiska między obiektami
- Analiza panelowa umożliwia dekompozycję wariancji łącznego błędu na wariancję powodowaną przez efekt indywidualny i błąd czysto losowy

$$\rho = \frac{\sigma_u^2}{\sigma_v^2} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_\varepsilon^2} = \frac{\text{cov}(u_{is}, u_{it})}{\text{var}(u_{it})}$$

## Przykład - Dekompozycja wariancji

```
R-sq:  within  = 0.0226  
       between = 0.3270  
       overall = 0.3232
```

## Przykład - Dekompozycja wariancji

```
-----+-----  
sigma_u | .41677338  
sigma_e | .10670893  
rho      | .93847863   (fraction of variance due to u_i)  
-----
```

- Estymator efektów stałych jest zgodny, nawet w przypadku niezerowej korelacji między efektem indywidualnym a charakterystykami obiektów

- Estymator efektów stałych jest zgodny, nawet w przypadku niezerowej korelacji między efektem indywidualnym a charakterystykami obiektów
- Niezbędne jednak są dwa założenia

- Estymator efektów stałych jest zgodny, nawet w przypadku niezerowej korelacji między efektem indywidualnym a charakterystykami obiektów
- Niezbędne jednak są dwa założenia
  - 1 Dla każdego obiektu  $i$  efekt indywidualny jest nieskorelowany z błędem czysto losowym

$$\forall i, t \text{ cov}(u_i, \varepsilon_{it}) = 0$$

- Estymator efektów stałych jest zgodny, nawet w przypadku niezerowej korelacji między efektem indywidualnym a charakterystykami obiektów
- Niezbędne jednak są dwa założenia
  - 1 Dla każdego obiektu  $i$  efekt indywidualny jest nieskorelowany z błędem czysto losowym

$$\forall i, t \text{ cov}(u_i, \varepsilon_{it}) = 0$$

- 2 Dla każdego obiektu  $i$  błąd czysto losowy jest nieskorelowany z charakterystykami obiektu

$$\forall i, t \text{ cov}(\varepsilon_{it}, x_{it}) = 0$$

- Estymator efektów stałych jest zgodny, nawet w przypadku niezerowej korelacji między efektem indywidualnym a charakterystykami obiektów
- Niezbędne jednak są dwa założenia
  - 1 Dla każdego obiektu  $i$  efekt indywidualny jest nieskorelowany z błędem czysto losowym

$$\forall i, t \text{ cov}(u_i, \varepsilon_{it}) = 0$$

- 2 Dla każdego obiektu  $i$  błąd czysto losowy jest nieskorelowany z charakterystykami obiektu

$$\forall i, t \text{ cov}(\varepsilon_{it}, x_{it}) = 0$$

- Aby estymator był efektywny konieczna jest dodatkowo

- Estymator efektów stałych jest zgodny, nawet w przypadku niezerowej korelacji między efektem indywidualnym a charakterystykami obiektów
- Niezbędne jednak są dwa założenia
  - 1 Dla każdego obiektu  $i$  efekt indywidualny jest nieskorelowany z błędem czysto losowym

$$\forall i, t \text{ cov}(u_i, \varepsilon_{it}) = 0$$

- 2 Dla każdego obiektu  $i$  błąd czysto losowy jest nieskorelowany z charakterystykami obiektu

$$\forall i, t \text{ cov}(\varepsilon_{it}, x_{it}) = 0$$

- Aby estymator był efektywny konieczna jest dodatkowo
  - 3 homoscedastyczność składnika losowego

$$\text{var}(\varepsilon_{it}) = \sigma_{\varepsilon}^2 \mathbb{I}$$

- Zapiszmy model efektów nieobserwowanych

$$y_{it} = X_{it}\beta + u_i + \varepsilon_{it}$$

- Zapiszmy model efektów nieobserwowanych

$$y_{it} = X_{it}\beta + u_i + \varepsilon_{it}$$

- Uśredniając po czasie uzyskujemy

$$\bar{y}_{it} = \bar{X}_{it}\beta + u_i + \bar{\varepsilon}_i$$

- Zapiszmy model efektów nieobserwowanych

$$y_{it} = X_{it}\beta + u_i + \varepsilon_{it}$$

- Uśredniając po czasie uzyskujemy

$$\bar{y}_{it} = \bar{X}_{it}\beta + u_i + \bar{\varepsilon}_i$$

- Odejmując równania stronami uzyskujemy

$$y_{it} - \bar{y}_{it} = (X_{it} - \bar{X}_{it})\beta + u_i + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

- Zapiszmy model efektów nieobserwowanych

$$y_{it} = X_{it}\beta + u_i + \varepsilon_{it}$$

- Uśredniając po czasie uzyskujemy

$$\bar{y}_{it} = \bar{X}_{it}\beta + u_i + \bar{\varepsilon}_i$$

- Odejmując równania stronami uzyskujemy

$$y_{it} - \bar{y}_{it} = (X_{it} - \bar{X}_{it})\beta + u_i + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

- W literaturze też spotykane jest oznaczenie

$$\ddot{y}_{it} = \ddot{X}_{it}\beta + u_i + \ddot{\varepsilon}_{it}$$

- Zapiszmy model efektów nieobserwowanych

$$y_{it} = X_{it}\beta + u_i + \varepsilon_{it}$$

- Uśredniając po czasie uzyskujemy

$$\bar{y}_{it} = \bar{X}_{it}\beta + u_i + \bar{\varepsilon}_i$$

- Odejmując równania stronami uzyskujemy

$$y_{it} - \bar{y}_{it} = (X_{it} - \bar{X}_{it})\beta + u_i + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

- W literaturze też spotykane jest oznaczenie

$$\ddot{y}_{it} = \ddot{X}_{it}\beta + u_i + \ddot{\varepsilon}_{it}$$

- Przekształcenie zmiennych doprowadziło do eliminacji efektu indywidualnego z modelu. Nazywamy je przekształceniem efektów stałych.

- Aby uzyskać zgodne oszacowania wystarczy zastosować MNK do przekształconych zmiennych

- Aby uzyskać zgodne oszacowania wystarczy zastosować MNK do przekształconych zmiennych
- Dzieje się tak, ponieważ błąd losowy jest nieskorelowany z przekształconymi zmiennymi niezależnymi

$$\text{cov}(\ddot{\varepsilon}_{it}, \ddot{X}_{it}) = \text{cov}\left(\varepsilon_{it} - \frac{\sum_t \varepsilon_{it}}{T}, X_{it} - \frac{\sum_t X_{it}}{T}\right) =$$

- Aby uzyskać zgodne oszacowania wystarczy zastosować MNK do przekształconych zmiennych
- Dzieje się tak, ponieważ błąd losowy jest nieskorelowany z przekształconymi zmiennymi niezależnymi

$$\begin{aligned} cov(\ddot{\varepsilon}_{it}, \ddot{X}_{it}) &= cov\left(\varepsilon_{it} - \frac{\sum_t \varepsilon_{it}}{T}, X_{it} - \frac{\sum_t X_{it}}{T}\right) = \\ &= \underbrace{cov(\varepsilon_{it}, X_{it})}_0 - \frac{1}{T} \sum_s \underbrace{cov(\varepsilon_{is}, X_{it})}_0 \end{aligned}$$

- Aby uzyskać zgodne oszacowania wystarczy zastosować MNK do przekształconych zmiennych
- Dzieje się tak, ponieważ błąd losowy jest nieskorelowany z przekształconymi zmiennymi niezależnymi

$$\begin{aligned} cov(\ddot{\varepsilon}_{it}, \ddot{X}_{it}) &= cov\left(\varepsilon_{it} - \frac{\sum_t \varepsilon_{it}}{T}, X_{it} - \frac{\sum_t X_{it}}{T}\right) = \\ &= \underbrace{cov(\varepsilon_{it}, X_{it})}_0 - \frac{1}{T} \sum_s \underbrace{cov(\varepsilon_{is}, X_{it})}_0 \\ &\quad - \frac{1}{T} \sum_s \underbrace{cov(\varepsilon_{it}, X_{is})}_0 + \frac{1}{T^2} \sum_s \sum_t \underbrace{cov(\varepsilon_{it}, X_{is})}_0 \end{aligned}$$

- Aby uzyskać zgodne oszacowania wystarczy zastosować MNK do przekształconych zmiennych
- Dzieje się tak, ponieważ błąd losowy jest nieskorelowany z przekształconymi zmiennymi niezależnymi

$$\begin{aligned} cov(\ddot{\varepsilon}_{it}, \ddot{X}_{it}) &= cov\left(\varepsilon_{it} - \frac{\sum_t \varepsilon_{it}}{T}, X_{it} - \frac{\sum_t X_{it}}{T}\right) = \\ &= \underbrace{cov(\varepsilon_{it}, X_{it})}_0 - \frac{1}{T} \sum_s \underbrace{cov(\varepsilon_{is}, X_{it})}_0 \\ &\quad - \frac{1}{T} \sum_s \underbrace{cov(\varepsilon_{it}, X_{is})}_0 + \frac{1}{T^2} \sum_s \sum_t \underbrace{cov(\varepsilon_{it}, X_{is})}_0 \\ cov(\ddot{\varepsilon}_{it}, \ddot{X}_{it}) &= 0 \end{aligned}$$

- Zatem estymatorem efektu stałego (FE) jest

$$b_{FE} = (\ddot{X}'\ddot{X})^{-1}\ddot{X}'y$$

- Zatem estymatorem efektu stałego (FE) jest

$$b_{FE} = (\ddot{X}'\ddot{X})^{-1}\ddot{X}'y$$

- Oszacowania uzyskuje się przeprowadzając regresję  $y_{it}$  na  $X_{it}$  oraz zmienne zero-jedynkowe reprezentujące stałe dla każdego obiektu

- Zatem estymatorem efektu stałego (FE) jest

$$b_{FE} = (\ddot{X}'\ddot{X})^{-1}\ddot{X}'y$$

- Oszacowania uzyskuje się przeprowadzając regresję  $y_{it}$  na  $X_{it}$  oraz zmienne zero-jedynkowe reprezentujące stałe dla każdego obiektu
- Ten estymator jest nazywany estymatorem LSDV *Least Squares Dummy Variables*

- Zatem estymatorem efektu stałego (FE) jest

$$b_{FE} = (\ddot{X}'\ddot{X})^{-1}\ddot{X}'y$$

- Oszacowania uzyskuje się przeprowadzając regresję  $y_{it}$  na  $X_{it}$  oraz zmienne zero-jedynkowe reprezentujące stałe dla każdego obiektu
- Ten estymator jest nazywany estymatorem LSDV *Least Squares Dummy Variables*
- Estymator efektów stałych jest również nazywany estymatorem wewnątrzgrupowym, ponieważ wykorzystuje wyłącznie zróżnicowanie wewnątrzgrupowe

- Zatem estymatorem efektu stałego (FE) jest

$$b_{FE} = (\ddot{X}'\ddot{X})^{-1}\ddot{X}'y$$

- Oszacowania uzyskuje się przeprowadzając regresję  $y_{it}$  na  $X_{it}$  oraz zmienne zero-jedynkowe reprezentujące stałe dla każdego obiektu
- Ten estymator jest nazywany estymatorem LSDV *Least Squares Dummy Variables*
- Estymator efektów stałych jest również nazywany estymatorem wewnątrzgrupowym, ponieważ wykorzystuje wyłącznie zróżnicowanie wewnątrzgrupowe
- W rezultacie w modelu nie można oszacować parametrów dla stałych w czasie charakterystyk

# Przykład

- Szacujemy parametry modelu

$$\ln(\text{placa})_{it} = \beta_0 + \beta_1 \text{plec} + \beta_2 \text{wiek} + u_t + \varepsilon_{it}$$

## Przykład

- Szacujemy parametry modelu

$$\ln(\text{placa})_{it} = \beta_0 + \beta_1 \text{plec} + \beta_2 \text{wiek} + u_t + \varepsilon_{it}$$

- Po przekształceniu efektów stałych model ma postać

$$\ln(\text{placa})_{it} = \beta_2(\text{wiek} - \bar{\text{wiek}}) + \tilde{\varepsilon}_{it}$$

## Przykład

- Szacujemy parametry modelu

$$\ln(\text{placa})_{it} = \beta_0 + \beta_1 \text{plec} + \beta_2 \text{wiek} + u_t + \varepsilon_{it}$$

- Po przekształceniu efektów stałych model ma postać

$$\ln(\text{placa})_{it} = \beta_2 (\text{wiek} - \bar{\text{wiek}}) + \tilde{\varepsilon}_{it}$$

- Zatem po przekształceniu w modelu nie ma stałej i zmiennej płeć!

# Przykład

Fixed-effects (within) regression  
Group variable: ID

R-sq: within = 0.0327  
between = 0.0421  
overall = 0.0381

corr(u\_i, Xb) = 0.0870

Number of obs = 9902  
Number of groups = 3309

Obs per group: min = 1  
avg = 3.0  
max = 4

F(11,6582) = 20.23  
Prob > F = 0.0000

| lnplaca | Coef.     | Std. Err. | t     | P> t  | [95% Conf. Interval] |           |
|---------|-----------|-----------|-------|-------|----------------------|-----------|
| wiek    | .3773752  | .0704308  | 5.36  | 0.000 | .2393081             | .5154424  |
| wiek2   | -.013465  | .0027318  | -4.93 | 0.000 | -.0188202            | -.0081097 |
| wiek3   | .0002031  | .0000449  | 4.52  | 0.000 | .000115              | .0002911  |
| wiek4   | -1.10e-06 | 2.66e-07  | -4.13 | 0.000 | -1.62e-06            | -5.78e-07 |
| plec    | (dropped) |           |       |       |                      |           |

## Przykład

| lnplaca  | Coef.     | Std. Err.                         | t     | P> t  | [95% Conf. Interval] |          |
|----------|-----------|-----------------------------------|-------|-------|----------------------|----------|
| wykszt_2 | -.0132047 | .032923                           | -0.40 | 0.688 | -.0777444            | .051335  |
| wykszt_3 | .0216162  | .0547571                          | 0.39  | 0.693 | -.0857256            | .1289579 |
| wykszt_4 | .0470884  | .0878841                          | 0.54  | 0.592 | -.1251929            | .2193697 |
| klm_2    | (dropped) |                                   |       |       |                      |          |
| klm_3    | -.0225189 | .047031                           | -0.48 | 0.632 | -.1147148            | .0696771 |
| klm_4    | (dropped) |                                   |       |       |                      |          |
| rok_2    | .0040024  | .0033602                          | 1.19  | 0.234 | -.0025847            | .0105895 |
| rok_3    | .0246789  | .0060117                          | 4.11  | 0.000 | .0128941             | .0364637 |
| rok_4    | .0369567  | .007174                           | 5.15  | 0.000 | .0228933             | .05102   |
| _cons    | 3.079557  | .6677399                          | 4.61  | 0.000 | 1.77057              | 4.388543 |
| sigma_u  | .50805056 |                                   |       |       |                      |          |
| sigma_e  | .10670893 |                                   |       |       |                      |          |
| rho      | .95774878 | (fraction of variance due to u_i) |       |       |                      |          |

F test that all u\_i=0: F(3308, 6582) = 42.42 Prob > F = 0.0000

- Analogicznie jak dla estymatora efektów losowych możemy dokonać dekompozycji wariancji

- Analogicznie jak dla estymatora efektów losowych możemy dokonać dekompozycji wariancji
  - wariancja całkowita

$$S_y^2 = \frac{1}{NT} \sum_N \sum_T (y_{it} - \bar{y})^2$$

- Analogicznie jak dla estymatora efektów losowych możemy dokonać dekompozycji wariancji
  - wariancja całkowita

$$S_y^2 = \frac{1}{NT} \sum_N \sum_T (y_{it} - \bar{y})^2$$

- wariancja wewnątrzgrupowa

$$\bar{S}_y^2 = \frac{1}{N} \sum_N (\bar{y}_i - \bar{y})^2$$

- Analogicznie jak dla estymatora efektów losowych możemy dokonać dekompozycji wariancji
  - wariancja całkowita

$$S_y^2 = \frac{1}{NT} \sum_N \sum_T (y_{it} - \bar{y})^2$$

- wariancja wewnątrzgrupowa

$$\bar{S}_y^2 = \frac{1}{N} \sum_N (\bar{y}_i - \bar{y})^2$$

- wariancja międzygrupowa

$$\ddot{S}_y^2 = \sum_N \sum_T (y_{it} - \bar{y}_i)^2$$

- Analogicznie jak dla estymatora efektów losowych możemy dokonać dekompozycji wariancji

- wariancja całkowita

$$S_y^2 = \frac{1}{NT} \sum_N \sum_T (y_{it} - \bar{y})^2$$

- wariancja wewnątrzgrupowa

$$\bar{S}_y^2 = \frac{1}{N} \sum_N (\bar{y}_i - \bar{y})^2$$

- wariancja międzygrupowa

$$\ddot{S}_y^2 = \sum_N \sum_T (y_{it} - \bar{y}_i)^2$$

- Oczywiście  $S_y^2 = \bar{S}_y^2 + \ddot{S}_y^2$

- Nieobciążonym estymatorem wariancji błędu składnika losowego w przypadku estymatora efektów stałych (FE) jest

$$\hat{\sigma}_{\varepsilon}^2 = \frac{\ddot{e}'\ddot{e}}{N(T-1) - K} = \frac{RSS}{N(T-1) - K}$$

# Test efektów indywidualnych

- Testem weryfikującym występowanie efektów indywidualnych jest porównanie estymatora POLS i estymatora efektów stałych

# Test efektów indywidualnych

- Testem weryfikującym występowanie efektów indywidualnych jest porównanie estymatora POLS i estymatora efektów stałych
- Formalnie statystyka testowa ma postać

$$\frac{(RSS_{FE} - RSS_{POLS})/(N - 1)}{RSS_{POLS}/(NT - K - 1)} \sim F(N - 1, NT - K - 1)$$

## Test efektów indywidualnych

- Testem weryfikującym występowanie efektów indywidualnych jest porównanie estymatora POLS i estymatora efektów stałych
- Formalnie statystyka testowa ma postać

$$\frac{(RSS_{FE} - RSS_{POLS})/(N - 1)}{RSS_{POLS}/(NT - K - 1)} \sim F(N - 1, NT - K - 1)$$

- Przykład

F test that all  $u_i=0$ :  $F(3308, 6582) = 42.42$

- Wszystkie estymatory dla danych panelowych przy spełnionych założeniach są zgodne dla  $N \rightarrow \infty$

- Wszystkie estymatory dla danych panelowych przy spełnionych założeniach są zgodne dla  $N \rightarrow \infty$
- W obu przypadkach możliwe jest oszacowanie wielkości efektów indywidualnych

- Wszystkie estymatory dla danych panelowych przy spełnionych założeniach są zgodne dla  $N \rightarrow \infty$
- W obu przypadkach możliwe jest oszacowanie wielkości efektów indywidualnych
- W przypadku estymatora efektów losowych

$$\hat{u}_t = \frac{1}{T} \sum_t \hat{v}_i t = \bar{v}_i = \bar{y}_i - \bar{x}_i \beta_{FE}$$

- Ten estymator nie jest zgodny ponieważ

$$\text{plim}_{N \rightarrow \infty} \hat{u}_i = \text{plim} \bar{\tilde{v}}_i = \bar{v}_i$$

- Ten estymator nie jest zgodny ponieważ

$$\text{plim}_{N \rightarrow \infty} \hat{u}_i = \text{plim} \bar{\tilde{v}}_i = \bar{v}_i$$

- ale  $v_{it} = u_i + \varepsilon_{it}$ , oraz  $E(\varepsilon_{it} | x_{it}) = 0$ , zatem

$$E(\bar{v}_i | u_i) = u_i$$

- Ten estymator nie jest zgodny ponieważ

$$\text{plim}_{N \rightarrow \infty} \hat{u}_i = \text{plim} \bar{\tilde{v}}_i = \bar{v}_i$$

- ale  $v_{it} = u_i + \varepsilon_{it}$ , oraz  $E(\varepsilon_{it}|x_{it}) = 0$ , zatem

$$E(\bar{v}_i|u_i) = u_i$$

- ale

$$\text{var}(\bar{v}_i|u_i) = \text{var}\left(\frac{1}{T} \sum_t \varepsilon_{it}\right) = \frac{1}{T} \sigma_T^2 > 0 \text{ dla skończonego } T$$

- Ten estymator nie jest zgodny ponieważ

$$\text{plim}_{N \rightarrow \infty} \hat{u}_i = \text{plim} \bar{\tilde{v}}_i = \bar{v}_i$$

- ale  $v_{it} = u_i + \varepsilon_{it}$ , oraz  $E(\varepsilon_{it}|x_{it}) = 0$ , zatem

$$E(\bar{v}_i|u_i) = u_i$$

- ale

$$\text{var}(\bar{v}_i|u_i) = \text{var}\left(\frac{1}{T} \sum_t \varepsilon_{it}\right) = \frac{1}{T} \sigma_T^2 > 0 \text{ dla skończonego } T$$

- Estymator będzie zgodny jedynie dla  $T \rightarrow \infty$

- W przypadku estymatora efektów stałych oszacowaniem efektu indywidualnego jest

$$\hat{u}_t = \bar{y}_i - \bar{x}_i \beta_{FE}$$

- W przypadku estymatora efektów stałych oszacowaniem efektu indywidualnego jest

$$\hat{u}_t = \bar{y}_i - \bar{x}_i \beta_{FE}$$

- Efekty te można oszacować wartości współczynników przy zmiennych zero-jedynkowych dla estymatora LSDV

- Jest to ogólny test wykorzystywany do porównań konkurencyjnych estymatorów

- Jest to ogólny test wykorzystywany do porównań konkurencyjnych estymatorów
- zakładamy, że jeden z estymatorów jest zgodny i efektywny przy prawdziwej  $H_0$ , a drugi jest zawsze zgodny, ale nie jest efektywny

|       | A                   | B                      |
|-------|---------------------|------------------------|
| $H_0$ | zgodny<br>efektywny | zgodny<br>nieefektywny |
| $H_A$ | -                   | zgodny                 |

- Intuicja: jeżeli  $H_0$  jest prawdziwa to oba estymatory dadzą zbliżony wynik, w przeciwnym przypadku pojawi się różnica wynikająca z braku zgodności estymatora A.

- Intuicja: jeżeli  $H_0$  jest prawdziwa to oba estymatory dadzą zbliżony wynik, w przeciwnym przypadku pojawi się różnica wynikająca z braku zgodności estymatora A.
- Test przeprowadzamy porównując oszacowania uzyskane z wykorzystaniem obu estymatorów

- Intuicja: jeżeli  $H_0$  jest prawdziwa to oba estymatory dadzą zbliżony wynik, w przeciwnym przypadku pojawi się różnica wynikająca z braku zgodności estymatora A.
- Test przeprowadzamy porównując oszacowania uzyskane z wykorzystaniem obu estymatorów
- Statystyka testowa ma postać formy kwadratowej

$$(\beta_A - \beta_B)' \Sigma_{\beta_A - \beta_B}^{-1} (\beta_A - \beta_B) \xrightarrow{D} \chi^2(r(\Sigma))$$

- Intuicja: jeżeli  $H_0$  jest prawdziwa to oba estymatory dadzą zbliżony wynik, w przeciwnym przypadku pojawi się różnica wynikająca z braku zgodności estymatora A.
- Test przeprowadzamy porównując oszacowania uzyskane z wykorzystaniem obu estymatorów
- Statystyka testowa ma postać formy kwadratowej

$$(\beta_A - \beta_B)' \Sigma_{\beta_A - \beta_B}^{-1} (\beta_A - \beta_B) \xrightarrow{D} \chi^2(r(\Sigma))$$

- w praktyce przyjmuje się, że  $\Sigma_{\beta_A - \beta_B}^{-1} = \Sigma_{\beta_A}^{-1} - \Sigma_{\beta_B}^{-1}$

- Test Hausmana wykorzystany do badania założenia o braku korelacji między efektem indywidualnym i zmiennymi objaśniającymi

|       | RE                  | FE                     |
|-------|---------------------|------------------------|
| $H_0$ | zgodny<br>efektywny | zgodny<br>nieefektywny |
| $H_A$ | -                   | zgodny                 |

|             | ---- Coefficients ---- |           |            |                     |
|-------------|------------------------|-----------|------------|---------------------|
|             | (b)                    | (B)       | (b-B)      | sqrt(diag(V_b-V_B)) |
|             | fe_res                 | re_res    | Difference | S.E.                |
| AGE         | .3773752               | .5441276  | -.1667524  | .0550871            |
| AGE2        | -.013465               | -.0178531 | .0043881   | .0021756            |
| AGE3        | .0002031               | .0002553  | -.0000523  | .0000363            |
| AGE4        | -1.10e-06              | -1.34e-06 | 2.43e-07   | 2.19e-07            |
| _IEDUCAT__2 | -.0132047              | -.2194375 | .2062328   | .0271122            |
| _IEDUCAT__5 | .0216162               | -.3824129 | .4040291   | .0503229            |
| _IEDUCAT__7 | .0470884               | -.62673   | .6738184   | .0828139            |
| _ITOWN__2   | -.0225189              | -.1280112 | .1054923   | .0392029            |
| _IOBSNUM_2  | .0040024               | .0012748  | .0027276   | .001235             |
| _IOBSNUM_3  | .0246789               | .0143942  | .0102847   | .0049601            |
| _IOBSNUM_4  | .0369567               | .0223004  | .0146563   | .0062666            |

b = consistent under  $H_0$  and  $H_a$ ; obtained from xtreg  
B = inconsistent under  $H_a$ , efficient under  $H_0$ ; obtained from xtreg

Test:  $H_0$ : difference in coefficients not systematic

$$\begin{aligned}\chi^2(9) &= (b-B)' [(V_b-V_B)^{-1}] (b-B) \\ &= 142.27\end{aligned}$$

- Konstrukcję zbioru panelowego można wykorzystać do analizy zmian stanów

- Konstrukcję zbioru panelowego można wykorzystać do analizy zmian stanów
- Patrzymy gdzie obecnie znajdują się jednostki które były w okresie  $t - 1$  w stanie  $k$

- Konstrukcję zbioru panelowego można wykorzystać do analizy zmian stanów
- Patrzymy gdzie obecnie znajdują się jednostki które były w okresie  $t - 1$  w stanie  $k$
- Od strony matematycznej budujemy macierz prawdopodobieństw przejścia pomiędzy stanami

| $t / t-1$  | Pra   | Bez  | Nie   |
|------------|-------|------|-------|
| Pracujący  | 97,95 | 0,87 | 1,18  |
| Bezrobotny | 16,02 | 0,70 | 13,84 |
| Nieaktywny | 1,72  | 1,09 | 97,19 |