

Analiza danych panelowych cz. III

Stanisław Cichocki

Natalia Nehrebecka

Plan zajęć

1. Wprowadzenie do danych panelowych

- a) Charakterystyka danych panelowych
- b) Zalety i ograniczenia

2. Modele ekonometryczne danych panelowych

- a) Model efektów nieobserwowalnych
- b) Model efektów losowych
- c) Model efektów stałych
- d) Modele z dwukierunkowym komponentem błędu

3. Testy

- a) Test istotności efektów stałych i losowych
- b) Test Hausmana

Ekonometryczne modele danych panelowych – Fixed Effects model

- ▶ Zupełnie inne podejście wykorzystywane jest w przypadku, gdy **zmienne objaśniane są skorelowane z efektem indywidualnym jednostki**.
 - Estymatory otrzymane z modelu z efektami losowymi lub z metody *pooled regression* nie są w tej sytuacji estymatorami zgodnymi.
- ▶ Jednym z rozwiązań jest użycie modelu z efektami stałymi - *FE* (ang. *fixed effects model*).
- ▶ Postać funkcyjna modelu przedstawiona jest w równaniu.

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \underbrace{u_i + \varepsilon_{it}}_{v_{it}}$$

Ekonometryczne modele danych panelowych – Fixed Effects model

- ▶ Założenia modelu efektów stałych są następujące:
- ▶ Aby estymator modelu był **zgodny** konieczne jest założenie braku korelacji pomiędzy błędem losowym i efektem indywidualnym oraz zmiennymi objaśniającymi pochodzącymi z różnych okresów czasu:

$$Cov(\varepsilon_{it}, u_i) = 0$$

$$Cov(\varepsilon_{it}, \mathbf{x}_{is}) = 0$$

- ▶ Natomiast warunkiem wystarczającym dla **efektywności** estymatora jest założenie o braku autokorelacji i heteroskedastyczności składnika losowego:

$$Var(\boldsymbol{\varepsilon}) = \sigma_{\varepsilon}^2 \mathbf{I}$$

Ekonometryczne modele danych panelowych – Fixed Effects model

- ▶ Model efektów stałych ma postać:

$$y_{it} = x_{it}\beta + \underbrace{u_i + \varepsilon_{it}}_{v_{it}}$$

- ▶ Co macierzowo można zapisać:

$$y = X\beta + v = X\beta + Z_u u + \varepsilon$$

- ▶ Gdzie:
 - Z_u – macierz zero-jedynkowych zmiennych „dobierających” do odpowiednich równań modelu efekty indywidualne dla poszczególnych jednostek.

Ekonometryczne modele danych panelowych – Fixed Effects model

$$y = X\beta + v = X\beta + Z_u u + \varepsilon$$

- ▶ Wiedząc jak wyglądają poszczególne macierze, warto zauważyć, iż model można oszacować *metodą najmniejszych kwadratów*.
- ▶ Estymator **MNK** byłby dany wzorem:

$$\hat{y} = (C'C)^{-1}C'y$$

- $C = [X \quad Z_u]$
- Wymiar C jest $NT \times (K + N)$.
- Wymiar $C'C$ jest $(K + N) \times (K + N)$
- ▶ **Dla dużej liczby jednostek (N) zaczyna być problematyczne!**
 - Z uwagi na duży wymiar macierzy $C'C$, w której główną część stanowią zmienne zerojedynekowe związane z efektami indywidualnymi, estymator \hat{y} nazwiemy estymatorem **Least Squares Dummy Variable (LSDV)**.

Ekonometryczne modele danych panelowych – Fixed Effects model

- *Dane:*
- *unit* - Firm Number
- *year*
- *c* - Total cost, in thousands
- *q* - Output in revenue passenger miles (index)
- *pf* - Fuel price
- *lf* - Load Factor (utilization index) (%)

Ekonometryczne modele danych panelowych – Fixed Effects model

xtset unit year

```
panel variable:  unit (strongly balanced)
time variable:  year, 1970 to 1984
delta: 1 unit
```

xtdes

```
unit: 1, 2, ..., 6                      n = 6
year: 1970, 1971, ..., 1984             T = 15
Delta(year) = 1 unit
Span(year) = 15 periods
(unit*year uniquely identifies each observation)
```

Distribution of T _i :	min	5%	25%	50%	75%	95%	max
	15	15	15	15	15	15	15

Freq.	Percent	Cum.	Pattern
6	100.00	100.00	111111111111111
6	100.00		XXXXXXXXXXXXXXXXX

Ekonometryczne modele danych panelowych – Fixed Effects model

► *Least squares dummy variable model (LSDV)*

Source	SS	df	MS	Number of obs = 90		
Model	113.74827	8	14.2185338	F(8, 81) = 3935.79		
Residual	.292622868	81	.003612628	Prob > F = 0.0000		
Total	114.040893	89	1.28135835	R-squared = 0.9974		
				Adj R-squared = 0.9972		
				Root MSE = .06011		

l_c	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
l_q	.9192846	.0298901	30.76	0.000	.8598126	.9787565
l_pf	.4174918	.0151991	27.47	0.000	.3872503	.4477333
l_f	-1.070396	.20169	-5.31	0.000	-1.471696	-.6690963
_Iunit_2	-.0412359	.0251839	-1.64	0.105	-.0913441	.0088722
_Iunit_3	-.2089211	.0427986	-4.88	0.000	-.2940769	-.1237653
_Iunit_4	.1845557	.0607527	3.04	0.003	.0636769	.3054345
_Iunit_5	.0240547	.0799041	0.30	0.764	-.1349293	.1830387
_Iunit_6	.0870617	.0841995	1.03	0.304	-.080469	.2545924
_cons	9.705942	.193124	50.26	0.000	9.321686	10.0902

Ekonometryczne modele danych panelowych – Fixed Effects model

► test _Iunit_2 _Iunit_3 _Iunit_4 _Iunit_5 _Iunit_6

```
( 1)  _Iunit_2 = 0
( 2)  _Iunit_3 = 0
( 3)  _Iunit_4 = 0
( 4)  _Iunit_5 = 0
( 5)  _Iunit_6 = 0
```

```
F( 5,      81) =    57.73
Prob > F =      0.0000
```

Ekonometryczne modele danych panelowych – Fixed Effects model

Snapshots

	unit	year	_Iunit_2	_Iunit_3	_Iunit_4	_Iunit_5	_Iunit_6
1	1	1970	0	0	0	0	0
2	1	1971	0	0	0	0	0
3	1	1972	0	0	0	0	0
4	1	1973	0	0	0	0	0
5	1	1974	0	0	0	0	0
6	1	1975	0	0	0	0	0
7	1	1976	0	0	0	0	0
8	1	1977	0	0	0	0	0
9	1	1978	0	0	0	0	0
10	1	1979	0	0	0	0	0
11	1	1980	0	0	0	0	0
12	1	1981	0	0	0	0	0
13	1	1982	0	0	0	0	0
14	1	1983	0	0	0	0	0
15	1	1984	0	0	0	0	0
16	2	1970	1	0	0	0	0
17	2	1971	1	0	0	0	0
18	2	1972	1	0	0	0	0
19	2	1973	1	0	0	0	0
20	2	1974	1	0	0	0	0
21	2	1975	1	0	0	0	0
22	2	1976	1	0	0	0	0
23	2	1977	1	0	0	0	0
24	2	1978	1	0	0	0	0
25	2	1979	1	0	0	0	0
26	2	1980	1	0	0	0	0
27	2	1981	1	0	0	0	0
28	2	1982	1	0	0	0	0
29	2	1983	1	0	0	0	0
30	2	1984	1	0	0	0	0
31	3	1970	0	1	0	0	0
32	3	1971	0	1	0	0	0
33	3	1972	0	1	0	0	0
34	3	1973	0	1	0	0	0
35	3	1974	0	1	0	0	0

Ekonometryczne modele danych panelowych – Fixed Effects model

- ▶ O ile dla małych N , estymator **LSDV** można wyznaczyć w standardowy sposób, o tyle dla dużych N (setki, tysiące, ... , setki tysięcy), należy poszukać efektywniejszego sposobu estymacji.
- ▶ W celu oszacowania wartości estymatora modelu należy przeprowadzić odpowiednie przekształcenia wzoru.

$$y_{it} = \alpha + x_{it}\beta + \underbrace{u_i + \varepsilon_{it}}_{v_{it}}$$

- ▶ Uśredniając po czasie wartości modelu $\bar{y}_{i.} = \alpha + \bar{x}_{i.}\beta + u_i + \bar{\varepsilon}_{i.}$
 - otrzymujemy **transformację between**
- a następnie odejmując je od równania $y_{it} = \alpha + x_{it}\beta + u_i + \varepsilon_{it}$, otrzymuje się model pozbawiony efektu indywidualnego. Ostateczna postać równania:

$$(y_{it} - \bar{y}_{i.}) = (x_{it} - \bar{x}_{i.})\beta + (\varepsilon_{it} - \bar{\varepsilon}_{i.})$$

Jest to tzw. **transformacja within**, która eliminuje z równania efekty indywidualne.

Ekonometryczne modele danych panelowych – Fixed Effects model

- ▶ Możemy teraz estymator wektora parametrów $\boldsymbol{\beta}$ wyznaczyć z równania:

$$(y_{it} - \bar{y}_{i.}) = (x_{it} - \bar{x}_{i.})\boldsymbol{\beta} + (\varepsilon_{it} - \bar{\varepsilon}_{i.})$$

$$\ddot{y}_{it} = \ddot{x}_{it}\boldsymbol{\beta} + \ddot{\varepsilon}_{it}$$

- gdzie:

$$\ddot{y}_{it} = y_{it} - \bar{y}_{i.}$$

$$\ddot{x}_{it} = x_{it} - \bar{x}_{i.}$$

$$\ddot{\varepsilon}_{it} = \varepsilon_{it} - \bar{\varepsilon}_{i.}$$

Ekonometryczne modele danych panelowych – Fixed Effects model

- ▶ Powyższa transformacja jest nazywana przekształceniem **efektów stałych**.

- We wzorze

$$\ddot{y}_{it} = \ddot{x}_{it}\boldsymbol{\beta} + \ddot{\varepsilon}_{it}$$

- nie występuje korelacja pomiędzy składnikiem czysto losowym a zmienną objaśniającą, co oznacza, że można dokonać oszacowania zgodnego estymatora używając *Metody Najmniejszych Kwadratów*.

- ▶ **Brak zależności** przedstawiony jest jako:

$$\text{Cov}(\ddot{x}_{it}, \ddot{\varepsilon}_{it}) = 0$$

- ▶ Ogólnie postać estymatora przekształconego modelu ma formę estymatora używanego dla *Metody Najmniejszych Kwadratów*:

$$\hat{\boldsymbol{\beta}}_{FE} = (\ddot{X}'\ddot{X})^{-1}\ddot{X}'\ddot{y}$$

Ekonometryczne modele danych panelowych – Fixed Effects model

- ▶ Efekty indywidualne uzyskujemy, przyjmując również standardowe założenie:

$$\sum_{i=1}^N u_i = 0$$

- ▶ Dodatkowo uśredniając równanie $y_{it} = \alpha + x_{it}\beta + u_i + \varepsilon_{it}$ po wszystkich obserwacjach, mamy:

$$\bar{y}_{..} = \alpha + \bar{x}_{..}\beta + \bar{\varepsilon}_{..}$$

Ekonometryczne modele danych panelowych – Fixed Effects model

- Po wyznaczeniu estymatora FE , z równania $\bar{y}_{..} = \alpha + \bar{x}_{..}\beta + \bar{\varepsilon}_{..}$ możemy z kolei wyznaczyć estymator stałej, α , za pomocą wzoru:

$$\tilde{\alpha} = \bar{y}_{..} - \bar{x}_{..}\hat{\beta}_{FE}$$

- Efekty indywidualne możemy wyznaczyć z równania

$$\bar{y}_{i.} = \alpha + \bar{x}_{i.}\beta + u_i + \bar{\varepsilon}_{i.}$$

$$\tilde{u}_i = \bar{y}_{i.} - \tilde{\alpha} - \bar{x}_{i.}\hat{\beta}_{FE}$$

Ekonometryczne modele danych panelowych – Fixed Effects model

Fixed-effects (within) regression
Group variable: unit

Number of obs = 90
Number of groups = 6

R-sq: within = 0.9926
between = 0.9856
overall = 0.9873

Obs per group: min = 15
avg = 15.0
max = 15

$\text{corr}(u_i, Xb) = -0.3475$

$F(3, 81) = 3604.80$
Prob > F = 0.0000

l_c	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
l_q	.9192846	.0298901	30.76	0.000	.8598126	.9787565
l_pf	.4174918	.0151991	27.47	0.000	.3872503	.4477333
lf	-1.070396	.20169	-5.31	0.000	-1.471696	-.6690963
_cons	9.713528	.229641	42.30	0.000	9.256614	10.17044
sigma_u	.13207751					
sigma_e	.06010514					
rho	.82843656	(fraction of variance due to u_i)				

F test that all $u_i=0$: $F(5, 81) = 57.73$ Prob > F = 0.0000

Ekonometryczne modele danych panelowych – Fixed Effects model

Fixed-effects (within) regression
Group variable: idcode

Number of obs = 28089
Number of groups = 4695

R-sq: within = 0.1606
between = 0.3292
overall = 0.2384

Obs per group: min = 1
avg = 6.0
max = 15

F(5,23389) = 894.83
Prob > F = 0.0000

corr(u_i, Xb) = 0.1881

ln_wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	.0463734	.0027572	16.82	0.000	.0409691	.0517778
age_2	-.0008695	.0000464	-18.75	0.000	-.0009604	-.0007786
black	0	(omitted)				
not_smsa	-.093142	.0095725	-9.73	0.000	-.1119046	-.0743793
grade	0	(omitted)				
ttl_exp	.0340055	.0014642	23.23	0.000	.0311357	.0368753
tenure	.0106481	.0009177	11.60	0.000	.0088493	.0124469
_cons	.8837677	.0408061	21.66	0.000	.8037851	.9637503
sigma_u	.3608354					
sigma_e	.29278824					
rho	.6029913	(fraction of variance due to u_i)				

F test that all u_i=0: F(4694, 23389) = 6.97 Prob > F = 0.0000

Ekonometryczne modele danych panelowych – Fixed Effects model

Variable		Mean	Std. Dev.	Min	Max	Observations
black						
	overall	.2821744	.4500657	0	1	N = 28532
	between		.4497183	0	1	n = 4709
	within		0	.2821744	.2821744	T-bar = 6.05904
grade						
	overall	12.53249	2.323954	0	18	N = 28530
	between		2.566951	0	18	n = 4707
	within		0	12.53249	12.53249	T-bar = 6.06119

Ekonometryczne modele danych panelowych – Fixed Effects model

- ▶ Estymator efektów stałych:
- ▶ **zalety:**
 - zgodny nawet wtedy, gdy efekt indywidualny skorelowany ze zmiennymi objaśniającymi
- ▶ **wady:**
 - nieefektywny, jeśli efekt indywidualny nieskorelowany ze zmiennymi objaśniającymi
 - nie można za jego pomocą oszacować wpływu na zmienną objaśnianą zmiennych objaśniających, które dla jednostek nie zmieniają się w czasie

Modele z dwukierunkowym komponentem błędu - fixed effects model

```
Fixed-effects (within) regression
Group variable: idcode
R-sq:  within = 0.1658
        between = 0.3336
        overall = 0.2423
corr(u_i, Xb) = 0.1784

Number of obs   = 28089
Number of groups = 4695
Obs per group: min = 1
                avg  = 6.0
                max  = 15
F(19,23375)    = 244.47
Prob > F       = 0.0000
```

ln_wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	.0771794	.0105333	7.33	0.000	.0565334	.0978255
age_2	-.0011094	.0000597	-18.58	0.000	-.0012264	-.0009924
black	0	(omitted)				
grade	0	(omitted)				
not_smsa	-.0913919	.0095496	-9.57	0.000	-.1101098	-.072674
t1l_exp	.0334607	.0014958	22.37	0.000	.0305288	.0363925
tenure	.0106458	.0009201	11.57	0.000	.0088424	.0124492
_Iyear_69	.0505931	.0156074	3.24	0.001	.0200015	.0811847
_Iyear_70	-.0174251	.0230286	-0.76	0.449	-.0625627	.0277124
_Iyear_71	-.0091309	.0319901	-0.29	0.775	-.0718336	.0535717
_Iyear_72	-.0343158	.0414354	-0.83	0.408	-.1155318	.0469003
_Iyear_73	-.0692245	.0510547	-1.36	0.175	-.1692951	.0308461
_Iyear_75	-.1221112	.0701862	-1.74	0.082	-.2596806	.0154583
_Iyear_77	-.1296354	.0898219	-1.44	0.149	-.3056922	.0464214
_Iyear_78	-.137558	.1001423	-1.37	0.170	-.3338434	.0587275
_Iyear_80	-.1961825	.1195946	-1.64	0.101	-.4305958	.0382308
_Iyear_82	-.2463163	.1393786	-1.77	0.077	-.5195074	.0268748
_Iyear_83	-.2573793	.1491913	-1.73	0.085	-.549804	.0350455
_Iyear_85	-.2609034	.1690675	-1.54	0.123	-.5922868	.07048
_Iyear_87	-.2907455	.1890538	-1.54	0.124	-.6613034	.0798124
_Iyear_88	-.2640814	.2025622	-1.30	0.192	-.6611165	.1329537
_cons	.3506171	.1954761	1.79	0.073	-.0325287	.733763
sigma_u	.35896829					
sigma_e	.29196749					
rho	.60185091	(fraction of variance due to u_i)				

```
F test that all u_i=0: F(4694, 23375) = 6.93 Prob > F = 0.0000
```

Plan zajęć

1. Wprowadzenie do danych panelowych

- a) Charakterystyka danych panelowych
- b) Zalety i ograniczenia

2. Modele ekonometryczne danych panelowych

- a) Model efektów nieobserwowalnych
- b) Model efektów losowych
- c) Model efektów stałych
- d) Modele z dwukierunkowym komponentem błędu

3. Testy

- a) Test istotności efektów stałych i losowych
- b) Test Hausmana

Testowanie istotności stałych efektów indywidualnych

- ▶ W modelu *FE* można sprawdzić istnienie efektów stałych.
- ▶ **Hipoteza zerowa** sprowadza się do założenia, że **wszystkie efekty indywidualne w modelu wynoszą zero**.
- ▶ Wzór przedstawia statystykę testową, w której używa się sumy kwadratów reszt z modelu *Pooled* oraz sumy kwadratów reszt z modelu efektów stałych:

$$F = \frac{(RSS_{\text{Pooled}} - RSS_{\text{FE}})/(N - 1)}{RSS_{\text{Pooled}}/(TN - (N - 1) - K)} \sim F(N - 1, TN - (N - 1) - K)$$

Ekonometryczne modele danych panelowych – Fixed Effects model

Fixed-effects (within) regression
Group variable: idcode

Number of obs = 28089
Number of groups = 4695

R-sq: within = 0.1606
between = 0.3292
overall = 0.2384

Obs per group: min = 1
avg = 6.0
max = 15

F(5,23389) = 894.83
Prob > F = 0.0000

corr(u_i, Xb) = 0.1881

ln_wage	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	.0463734	.0027572	16.82	0.000	.0409691	.0517778
age_2	-.0008695	.0000464	-18.75	0.000	-.0009604	-.0007786
black	0	(omitted)				
not_smsa	-.093142	.0095725	-9.73	0.000	-.1119046	-.0743793
grade	0	(omitted)				
ttl_exp	.0340055	.0014642	23.23	0.000	.0311357	.0368753
tenure	.0106481	.0009177	11.60	0.000	.0088493	.0124469
_cons	.8837677	.0408061	21.66	0.000	.8037851	.9637503
sigma_u	.3608354					
sigma_e	.29278824					
rho	.6029913	(fraction of variance due to u_i)				

F test that all u_i=0: F(4694, 23389) = 6.97 Prob > F = 0.0000

Testowanie istotności stałych efektów indywidualnych

- ▶ Jeżeli efekty indywidualne są istotne w badanej próbie to estymator efektów losowych jest efektywniejszy niż estymator dla *pooled regression*.
- ▶ W celu zidentyfikowania występowania efektów indywidualnych stosuje się **test Breuscha – Pagana** wykorzystujący mnożniki Lagrange'a.
 - Hipoteza zerowa testu zakłada, że efekty indywidualne nie występują

$$H_0: \sigma_u^2 = 0$$

- w związku z tym odpowiedni jest model *pooled regression*.
- W przypadku gdy hipoteza zerowa zostanie odrzucona wnioskuje się, że efekty indywidualne istnieją, więc efektywniejszy będzie estymator efektów losowych.
- ▶ Statystyka testowa ma postać:
 - v_{it} -reszty z estymacji *pooled regression*

$$LM = \frac{NT}{2(T-1)} \left[\frac{\sum_{i=1}^N \left(\sum_{t=1}^T v_{it} \right)^2}{\sum_{i=1}^N \sum_{t=1}^T (v_{it}^2)} - 1 \right]^2 \sim \chi^2(1)$$

Ekonometryczne modele danych panelowych – random effects model

Random-effects GLS regression

Group variable: idcode

R-sq: within = 0.1595
 between = 0.4676
 overall = 0.3567

Number of obs = 28089
 Number of groups = 4695

Obs per group: min = 1
 avg = 6.0
 max = 15

Wald chi2(7) = 8629.06
 Prob > chi2 = 0.0000

corr(u_i, X) = 0 (assumed)

ln_wage	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.0452213	.0026291	17.20	0.000	.0400685	.0503742
age_2	-.0008463	.0000437	-19.36	0.000	-.0009319	-.0007606
black	-.0783429	.009869	-7.94	0.000	-.0976857	-.059
grade	.0664253	.0017922	37.06	0.000	.0629126	.069938
not_smsa	-.1430699	.0071603	-19.98	0.000	-.1571039	-.1290359
ttl_exp	.0329041	.0011253	29.24	0.000	.0306986	.0351096
tenure	.0131911	.0008403	15.70	0.000	.011544	.0148381
_cons	.0923273	.0440066	2.10	0.036	.006076	.1785786
sigma_u	.26027187					
sigma_e	.29278824					
rho	.44140898					(fraction of variance due to u_i)

Ekonometryczne modele danych panelowych – random effects model

xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

```
ln_wage[idcode,t] = Xb + u[idcode] + e[idcode,t]
```

Estimated results:

	Var	sd = sqrt(Var)
ln_wage	.2281486	.477649
e	.085725	.2927882
u	.0677414	.2602719

Test: Var(u) = 0

chibar2(01) = 15215.93
Prob > chibar2 = 0.0000

Pytania teoretyczne

1. Wyjaśnić, w jaki sposób w estymatorze efektów stałych pozbywamy się wpływu efektu indywidualnego.
2. Wyjaśnić, dlaczego za pomocą estymatora efektów stałych nie można oszacować parametrów przy zmiennych nie zmieniających się w czasie.

Dziękuję za uwagę