Cechy prób panelowych

- Dane panelowe są to dane, w które mają charakter zarówno próby przekrojowej jak i szeregu czasowego.

$$\begin{array}{c|cccc}
\text{Czas} & & \text{Obiekty} & & \\
\hline
1 & \boxed{1} & \boxed{2} & \cdots & \boxed{N} \\
 & \downarrow & \downarrow & & \downarrow \\
2 & \boxed{1} & \boxed{2} & \cdots & \boxed{N} \\
\vdots & \downarrow & \downarrow & & \downarrow \\
T & \boxed{1} & \boxed{2} & \cdots & \boxed{N}
\end{array}$$

przy czym $\boxed{i}$ oznacza obiekt badany o numerze i ten sam dla wszystkich okresów badania.

- Dane panelowe należy odróżnić od prób przekrojowo-czasowych, dla których dysponujemy szeregiem *różnych* prób przekrojowych dla kolejnych momentów czasu.

Przykład Próba Badania Ekonomicznej Aktywności Ludności jest w istocie rotacyjnym panelem. Każde gospodarstwo domowe wylosowane do próby jest badane czterokrotnie: w 1 i 2 kwartale badania, oraz w 5 i 6 kwartale badania. Przy każdym nowym badaniu jedna czwarta próby jest na nowo losowana. W rezultacie na podstawie $BAEL$ można skonstruować panele dla gospodarstw z ilością przekrojów równą 4. Podobna sytuacja istnieje w przypadku badań budżetów gospodarstw.

1995				1996				1997			
1	2	3	4	1	2	3	4	1	2	3	4
x	-	-	x	x							
x	x	-	-	x	x						
	x	x	-	-	x	x					
		x	x	-	-	x	x				
			x	x	-	-	x	x			
				x	x	-	-	x	x		
					x	x	-	-	x	x	
						x	x	-	-	x	x
							x	x	-	-	x

Table 1: The smaple structure of Polish LFS

Liniowy model efektów nieobserwowalnych

- Model efektów nieobserwowalnych (indeks i dla jednostki, t dla czasu):

$$y_{it} = \boldsymbol{x}_{it}\,\boldsymbol{\beta} + u_i + \varepsilon_{it}$$

- u_i jest nieobserwowalny i stały w czasie

- u_i jest nazywany *charakterystyką (efektem) nieobserwowalną (indywidualną)* bądź *nieobserwowalnym (indywidualnym) zróżnicowaniem*

- ε_{it} jest nazywany *błędem czysto losowym*

- **Panele zbilansowane i niezbilansowane**

- Wszystkie (po czasie) obserwacje dostępne dla wszystkich jednostek

- Forma estymatorów dla paneli niezbilansowanych jest nieco bardziej skoplikowana niż dla paneli zbilansowanych niemniej np. w STAT'cie poprawki konieczne przy braku zbilansowania stosowane są automatyczne.

Przykład Struktura panelu BAEL, fala 13, 1995 kw.1, 1995 kw.2, 1996 kw.1, 1996 kw. 2.

```
      ID:  1300007, 1300015, ..., 1372365                  n =       5137
    DATE:  140, 141, ..., 145                              T =          4
           Delta(DATE) = 1; (145-140)+1 = 6
           (ID*DATE uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       1       2         4         4       4       4

     Freq.  Percent    Cum. |  Pattern
 ---------------------------+---------
     3082     60.00   60.00 |  11..11
      404      7.86   67.86 |  ....11
      371      7.22   75.08 |  11....
      281      5.47   80.55 |  .1..11
      224      4.36   84.91 |  .....1
      181      3.52   88.44 |  1.....
      146      2.84   91.28 |  11..1.
      126      2.45   93.73 |  .1....
      102      1.99   95.72 |  ....1.
      220      4.28  100.00 | (other patterns)
 ---------------------------+---------
     5137    100.00         |  XX..XX
```

- **Analiza asymptotyczna w panelach**

- Zwykle liczba obserwacji w przekrojach N jest znacznie większa niż liczba przekrojów T

- Z tego powodu jest bardzo ważne udowodnienie, że estymator jest zgodny dla $N \to \infty$ nawet jeśli T jest skończone (małe)

- **Korzyści związane z używaniem paneli**

- Efektywność: panelowa struktura danych może być użyta do efektywniejszego szacowania parametrów dzięki uwzględnieniu korelacji między nieobserwowalnymi efektami indywidualnymi (estymacja z efektami losowymi)

- Zgodność: jeśli efekty indywidualne są skorelowane ze zmiennymi objaśniającymi, to struktura danych panalowych może posłużyć do uzyskania zgodnego estymatora dzięki uwzględnieniu faktu, że nieobserwowalny efekt indywidualny jest stały w czasie (estymacja z efektami stałymi)

Estymacja w próbach panelowych z pomocą estymatora MNK

- **Regresja liniowa na pełnej próbie (P**ooled **O**rdinary **L**east **S**quares - $POLS$)

- Policz MNK na wszystkich obserwacjach (wszystkich przekrojach)

- u_i jest nieobserwowalny, łączny błąd losowy równy $v_{it} = u_i + \boldsymbol{\varepsilon}_{it}$

- Estymator MNK będzie zgodny jeśli:

$$\mathrm{E}\left(u_i \middle| \boldsymbol{x}_{it}\right) = 0$$

- Estymator MNK będzie więc zgodny tylko wtedy, gdy efekty indywidualne nie będą skorelowane ze zmiennymi objaśniającymi ($\mathrm{Cov}\left(u_i, \boldsymbol{x}_{it}\right) = 0$) ponieważ tylko wtedy łączny błąd losowy będzie nieskorelowany z $\boldsymbol{x}_{it}$ ($\mathrm{Cov}\left(v_{it}, \boldsymbol{x}_{it}\right) = 0$)

  – **Intuicja**: indywidualne nieobserwowalne chrakterystyki reprezentują zazwyczaj pominięte (nieobserwowalne) zmienne. Estymator MNK jest zgodny tylko wtedy, gdy pominięte zmienne są nieskorelowane z błędem losowym.

- Jeśli spełniony jest ten warunek, to estymator MNK daje zgodne choć nieefektywne oszacowanie parametrów.

- Problem: v_{it} będzie wykazywała autokorelację z racji na występowanie u_i w każdym z okresów

  ($\mathrm{Cov}\left(v_{it}, v_{it-1}\right) = \boldsymbol{\sigma}_u^2$)

- Z tego powodu powinniśmy zastosować w tym przypadku odporną macierz wariancji. W tym przypadku korelacja występuje tylko między obserwacjami.

- Dla tego przypadku macierz warancji kowariancji można uzyskać definiując jednostki jako tak zwane warstwy (*clusters*) - grupy obserwacji o powiązanych niezerowych kowariancjach między błędami losowymi.

Przykład Determinanty płacy. Szacunek na podstawie panelu dla BAEL rozpoczętej w 1995.1. Estymator MNK.

```
Regression with robust standard errors                 Number of obs =   16178
                                                       F( 18,  5053) =  325.53
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.3465
Number of clusters (ID) = 5054                         Root MSE      =  .33666

------------------------------------------------------------------------------
             |               Robust
     LNETPAY |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         AGE |    .023474   .0029087     8.07   0.000     .0177716    .0291764
        AGE2 |  -.0002125   .0000386    -5.51   0.000    -.0002881   -.0001369
     _ISEX_2 |  -.2944018   .0094636   -31.11   0.000    -.3129546   -.2758491
  _IEDUCAT_2 |  -.2471495   .0231851   -10.66   0.000    -.2926024   -.2016966
  _IEDUCAT_3 |  -.2508185    .017512   -14.32   0.000    -.2851495   -.2164874
  _IEDUCAT_4 |  -.2017467   .0220539    -9.15   0.000    -.2449818   -.1585115
  _IEDUCAT_5 |  -.3873151   .0171888   -22.53   0.000    -.4210127   -.3536176
  _IEDUCAT_7 |  -.4826221   .0191834   -25.16   0.000    -.5202299   -.4450144
  _IEDUCAT_8 |  -.4696502   .1126117    -4.17   0.000     -.690418   -.2488824
   _ITOWN2_1 |   -.014375   .0201918    -0.71   0.477    -.0539596    .0252096
```

```
   _ITOWN2_2 |   .0238628   .0223628     1.07   0.286     -.019978    .0677035
   _ITOWN2_3 |  -.0180127   .0208025    -0.87   0.387    -.0587947    .0227693
   _ITOWN2_4 |  -.0663333    .019641    -3.38   0.001    -.1048381   -.0278285
   _ITOWN2_5 |  -.0432959    .022122    -1.96   0.050    -.0866646    .0000727
   _ITOWN2_6 |  -.1104966   .0276596    -3.99   0.000    -.1647214   -.0562718
   _ITOWN2_7 |  -.1238275   .0286291    -4.33   0.000    -.1799529    -.067702
   _ITOWN2_9 |   -.139221    .017152    -8.12   0.000    -.1728464   -.1055955
      OBSNUM |   .1080488   .0017707    61.02   0.000     .1045776    .1115201
       _cons |    5.70988   .0565267   101.01   0.000     5.599064    5.820697
------------------------------------------------------------------------------
```

Efekty stałe i efekty losowe

- W zastosowaniach ekonomicznych prawie zawsze efekty indywidualne u_i są losowe

- Różnica między efektami stałymi i losowymi wiąże się z założeniami dotyczącymi korelacji między efektami indywidualnymi a zmiennymi objaśniającymi:

  – jeśli u_i jest nieskorelowany z $\boldsymbol{x}_i$ używamy *estymatora efektów losowych (random effects)*
  – jeśli u_i jest skorelowany z $\boldsymbol{x}_i$ używamy *estymatora efektów stałych (fixed effects)*

Estymator efektów losowych

- Warunki, przy których estymator efektów losowych (random effects) jest zgodny:

  1. Dokładna ezgogeniczność: $\mathrm{E}\left(\varepsilon_{it}\middle|\, \boldsymbol{x}_{it}, u_i\right) = 0$ dla $t = 1, \ldots, T$
     – brak korelacji między $\boldsymbol{x}_{it}$, u_i oraz obecnymi, przyszłymi i zaszłymi ε_{it}
  2. Brak korelacji między u_i i $\boldsymbol{x}_i$, $\mathrm{E}\left(u_i\middle|\, \boldsymbol{x}_i\right) = 0$, gdzie $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots, \boldsymbol{x}_{iT})$

- Model

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + v_{it}$$

- $v_{it} = u_i + \varepsilon_{it}$, gdzie v_{it} jest łącznym błędem losowym

- W zapisie macierzowym

$$\boldsymbol{y}_i = \boldsymbol{X}_i\,\boldsymbol{\beta} + \boldsymbol{v}_i$$

- O błędzie czysto losowym ε_{it} zakładamy, że jest homoskedastyczny i nieskorelowany:
  - $\mathrm{E}\left(\varepsilon_{it}^2\right) = \boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2$
  - $\mathrm{E}\left(\varepsilon_{it}\varepsilon_{js}\right) = 0,\ \ (i,t) \neq (j,s)$ (próba losowa)

- O u_i także zakładamy, że jest homoskedastyczny i nieskorelowany:
  - $\mathrm{E}\left(u_i^2\right) = \boldsymbol{\sigma}_u^2$
  - $\mathrm{E}\left(u_i u_j\right) = 0,\ \ i \neq j$ (próba losowa)

- Przy tych założeniach wariancja błędu łącznego wynosi:

$$\mathrm{E}\left(v_{it}^2\right) = \mathrm{E}\left(u_i^2\right) + 2\,\mathrm{E}\left(u_i\varepsilon_{it}\right) + \mathrm{E}\left(\varepsilon_{it}^2\right) = \boldsymbol{\sigma}_u^2 + \boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2$$

  gdzie $\boldsymbol{\sigma}_u^2$ jest wariancją u_i

- Dla $t = s$, kowariancja między v_{it} i v_{is} jest równa

$$\mathrm{E}\left(v_{it}v_{is}\right) = \mathrm{E}\left[\left(u_i + \boldsymbol{\varepsilon}_{it}\right)\left(u_i + \boldsymbol{\varepsilon}_{is}\right)\right] = \mathrm{E}\left(u_i^2\right) = \boldsymbol{\sigma}_u^2$$

- W rezultacie

$$\boldsymbol{\Omega} = \mathrm{E}\left(\boldsymbol{v}_i\boldsymbol{v}_i'\right) = \begin{bmatrix} \boldsymbol{\sigma}_u^2 + \boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2 & \boldsymbol{\sigma}_u^2 & \cdots & \boldsymbol{\sigma}_u^2 \\ \boldsymbol{\sigma}_u^2 & \boldsymbol{\sigma}_u^2 + \boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2 & & \vdots \\ \vdots & & \ddots & \boldsymbol{\sigma}_u^2 \\ \boldsymbol{\sigma}_u^2 & & & \boldsymbol{\sigma}_u^2 + \boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2 \end{bmatrix}$$

- Jeśli powyższe założenia są prawdziwe to estymator efektów losowych jest estymatorem efektywny

Estymator UMNK dla efektów losowych

- Uogólniona Metoda Najmniejszych Kwadratów $(UMNK)$ - przypadek ogólny

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\overline{\boldsymbol{\Omega}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\overline{\boldsymbol{\Omega}}^{-1}\boldsymbol{y}$$

gdzie $\overline{\boldsymbol{\Omega}} = Var\left(\boldsymbol{v}\right) = \mathrm{E}\left(\boldsymbol{v}\boldsymbol{v}'\right)$, i $\boldsymbol{v} = \left(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N\right)$

- Dla paneli skorelowane są jedynie v_{it} and v_{is} dla $t \neq s$ ale v_{it} i v_{js} dla $i \neq j$ są nieskorelowane

- W związku z tym macierz $\overline{\boldsymbol{\Omega}}$ będzie miała postać blokowo diagonalną

$$\overline{\boldsymbol{\Omega}} = \begin{bmatrix} \boldsymbol{\Omega} & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{\Omega} \end{bmatrix}$$

  a estymator $UMNK$ będzie miał specjalną postać:

$$\widehat{\boldsymbol{\beta}}_{RE} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{\Omega}^{-1} \boldsymbol{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \boldsymbol{\Omega}^{-1} \boldsymbol{y}_i \right)$$

- $\boldsymbol{\Omega}$ jest nieznana ale jako macierz $T \times T$ może być wyestymowana (T jest zazwyczaj małe)

- Załóżmy, że dysponujemy zgodnymi estymatorami $\boldsymbol{\sigma}_u^2$ and $\boldsymbol{\sigma}_v^2$

- Stosowalny estymator Uogólnionej Metody Najmniejszych Kwadratów ($SUMNK$) ma więc postać

$$\widehat{\boldsymbol{\beta}}_{RE} = \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \widehat{\boldsymbol{\Omega}}^{-1} \boldsymbol{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \boldsymbol{X}_i' \widehat{\boldsymbol{\Omega}}^{-1} \boldsymbol{y}_i\right)$$

- Aby użyć tego wzoru musimy znaleźć zgodne estymatory dla $\boldsymbol{\sigma}_u^2$ and $\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2$

- Ponieważ MNK jest zgodnym estymatorem w przypadku prawdziwości założeń koniecznych do zgodności estymatora efektów losowych więc estymator MNK dla błędu losowego regresji (s^2) będzie estymatorem zgodnym $\widehat{\boldsymbol{\sigma}}_v^2$, gdzie $\boldsymbol{\sigma}_v^2 = \boldsymbol{\sigma}_u^2 + \boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2$

- Z drugiej strony kowariancja między v_{it}, v_{is} jest równa $\boldsymbol{\sigma}_u^2$. Kowariancję tę można oszacować na podstawie kowariancji reszt uzyskanych w MNK:

$$\widehat{\boldsymbol{\sigma}}_u^2 = \frac{1}{[NT\,(T-1)/\,2-K]} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \widehat{\widehat{v}}_{it} \widehat{\widehat{v}}_{is}$$

- Ponieważ $\boldsymbol{\sigma}_v^2 = \boldsymbol{\sigma}_u^2 + \boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2$, więc estymator $\widehat{\boldsymbol{\sigma}}_{\boldsymbol{\varepsilon}}^2$ można policzyć jako $\widehat{\boldsymbol{\sigma}}_{\boldsymbol{\varepsilon}}^2 = \widehat{\boldsymbol{\sigma}}_v^2 - \widehat{\boldsymbol{\sigma}}_u^2$

Uwaga Niekiedy oszacowana w ten sposób wariancja $\widehat{\boldsymbol{\sigma}}_{\boldsymbol{\varepsilon}}^2$ może wyjść ujemna. Jednym z proponowanych rozwiązań jest przyjęcie, że $\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2 = 0$

- Można także znaleźć oszacowanie dla metody efektów zmiennych za pomocą MNW. W tym przypadku nie ma problemu z ujemnymi oszacowaniami $\boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2$.

- Przy założeniu, że $\mathrm{E}\left(\boldsymbol{v}_i\boldsymbol{v}_i'\middle|\, \boldsymbol{x}_i, \boldsymbol{u}_i\right) = \boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2\boldsymbol{I}_T$ i $\mathrm{E}\left(u_i^2\middle|\, \boldsymbol{x}_i\right) = \boldsymbol{\sigma}_u^2$, to estymator efektów losowych jest najbardziej efektywny

- Najważniejszą hipotezą jaką testuje się w kontekście metody efektów losowych jest hipoteza o istnieniu efektów indywidualnych.

- W kontekście tej metody testowanie istnienia efektów indywidualnych sprowadza się do testowania zerowści wariancji u_i.
  - jeśli $\mathrm{Var}\left(u_i\right) = \boldsymbol{\sigma}_u^2 \neq 0$, efekty indywidualne istnieją, estymator efektów losowych efektywniejszy od estymatora MNK
  - jeśli $\mathrm{Var}\left(u_i\right) = \boldsymbol{\sigma}_u^2 = 0$, efekty indywidualne nie są istotne, estymator MNK równie efektywny jak estymator efektów losowych

Przykład Determinanty płacy. Estymator $SUMNK$.

```
Random-effects GLS regression                   Number of obs      =     16178
Group variable (i): ID                          Number of groups   =      5054

R-sq:  within  = 0.4063                         Obs per group: min =         1
       between = 0.3262                                        avg =       3.2
       overall = 0.3457                                        max =         4

Random effects u_i ~Gaussian                   Wald chi2(18)      =  10088.51
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
     LNETPAY |      Coef.   Std. Err.       z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         AGE |   .0249031   .0028113     8.86   0.000     .0193929    .0304132
        AGE2 |  -.0002209   .0000372    -5.94   0.000    -.0002938   -.0001479
     _ISEX_2 |  -.2843309   .0096685   -29.41   0.000    -.3032807   -.2653811
  _IEDUCAT_2 |  -.2404923   .0256165    -9.39   0.000    -.2906996   -.1902849
  _IEDUCAT_3 |  -.2506126   .0157783   -15.88   0.000    -.2815374   -.2196877
  _IEDUCAT_4 |  -.2144521   .0202829   -10.57   0.000    -.2542059   -.1746983
  _IEDUCAT_5 |  -.3898099   .0156276   -24.94   0.000    -.4204395   -.3591803
  _IEDUCAT_7 |  -.4959456   .0181669   -27.30   0.000    -.5315522   -.4603391
```

```
 _IEDUCAT_8 |  -.4941375   .1234323    -4.00   0.000    -.7360604   -.2522147
  _ITOWN2_1 |  -.0182726   .0191882    -0.95   0.341    -.0558807    .0193355
  _ITOWN2_2 |   .0095936   .0207216     0.46   0.643     -.03102     .0502071
  _ITOWN2_3 |  -.0315726   .0200406    -1.58   0.115    -.0708515    .0077063
  _ITOWN2_4 |    -.07844   .0190514    -4.12   0.000    -.1157801   -.0410999
  _ITOWN2_5 |  -.0593752   .0218263    -2.72   0.007     -.102154   -.0165965
  _ITOWN2_6 |  -.1322534   .0278677    -4.75   0.000    -.1868731   -.0776336
  _ITOWN2_7 |  -.1465557   .0323892    -4.52   0.000    -.2100374   -.0830741
  _ITOWN2_9 |  -.1460398     .01687    -8.66   0.000    -.1791044   -.1129752
     OBSNUM |   .1082641   .0012704    85.22   0.000     .1057742     .110754
      _cons |   5.666961   .0542192   104.52   0.000     5.560694    5.773229
-------------+----------------------------------------------------------------
    sigma_u |  .30293482
    sigma_e |   .1662991
        rho |  .76842858   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

- R^2_{within} jest procentem zróżnicowania wewnątrzgrupowego (a więc zróżnicowania dotyczącego poszczególnych jenostek), które udało nam się wytłumaczyć zmianami w czasie charakterystyk tych jednostek

- $R^2_{between}$ jest procentem zróżnicowania międzygrupowego (a więc zróżnicowania między jednostkami), które udało nam się wytłumaczyć różnicami w średnich wielkościach charakterystyk dla tych jednostek

- $R^2_{overall}$ jest procentem całkowitego zróżnicowania zmiennej zależnej, które udało nam się wytłumaczyć

- $\boldsymbol{\rho} = \frac{\mathrm{Cov}(v_{it}, v_{is})}{\mathrm{Var}(v_{it})} = \frac{\boldsymbol{\sigma}_u^2}{\boldsymbol{\sigma}_v^2} = \frac{\boldsymbol{\sigma}_u^2}{\boldsymbol{\sigma}_u^2 + \boldsymbol{\sigma}_{\boldsymbol{\varepsilon}}^2} = \frac{.30303724^2}{.30303724^2 + .16615063^2} = .\,768\,87$ jest częścią łącznego błędu losowego, który jest związana z efektami indywidaulnymi.

- W tym analizowanym przypadku, 77% błędu losowego można wytłumaczyć, niezmiennymi w czasie charakterystykami indywidualnymi.

- Wielkości współczynników intepretujemy identycznie jak w zwykłej regresji.

- W analizowanym przez nas przypadku różnice w wielkościach oszacowanych współczynników i błędach standardowych wyszły bardzo niewielkie.

Przykład Determinanty płacy. Wynik testu na istnienie efektów losowych

```
Breusch and Pagan Lagrangian multiplier test for random effects:

        LNETPAY[id,t] = Xb + u[id] + e[id,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                 LNETPAY |   .1732811       .4162705
                       e |    .027606       .1661506
                       u |   .0918316       .3030372

        Test:   Var(u) = 0
                             chi2(1) = 10552.09
                         Prob > chi2 =     0.0000
```

- Hipotezę zerową o braku efektów losowych odrzucamy!

Estymator efektów stałych

- Model
$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + u_i + \boldsymbol{\varepsilon}_{it}$$

- Zakładamy, że błąd losowy nie zależy od przeszłych, obecnych i przyszłych wartości $\boldsymbol{x}_i$
$$\mathrm{E}\left(\boldsymbol{\varepsilon}_{it} \middle| \boldsymbol{x}_i, u_i\right) = 0 \text{ for } t = 1, \ldots, T$$

- Efekt indywidualny u_i może być skorelowany z $\boldsymbol{x}_{it}$

- Można pokazać, że estymator efektów, jest zgodny nawet jeśli $\mathrm{E}\left(u_i \middle| \boldsymbol{x}_i\right) \neq 0$, to jest nawet w sytuacji kiedy estymatory efektów zmiennych i MNK nie są zgodne

- W tym sensie estymator efektów stałych jest bardziej odporny

- **Przekształcenie efektów stałych** (przekształcenie wewnątrz obiektowe):

- Uśredniamy po czasie równanie

$$y_{it} = \boldsymbol{x}_{it}\boldsymbol{\beta} + u_i + \boldsymbol{\varepsilon}_{it}$$

  i otrzymujemy

$$\overline{y}_i = \overline{\boldsymbol{x}}_i\boldsymbol{\beta} + u_i + \overline{\boldsymbol{\varepsilon}}_i$$

  gdzie $\overline{y}_i = \frac{\sum_{t=1}^{T} y_{it}}{T}$, $\overline{\boldsymbol{x}}_i = \frac{\sum_{t=1}^{T} \boldsymbol{x}_{it}}{T}$, $\overline{\boldsymbol{\varepsilon}}_i = \frac{\sum_{t=1}^{T} \boldsymbol{\varepsilon}_{it}}{T}$

- Odejmując od siebie stronami model i uśredniając po czasie otrzymujemy:

$$y_{it} - \overline{y}_i = \left(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i\right)\boldsymbol{\beta} + \left(\boldsymbol{\varepsilon}_{it} - \overline{\boldsymbol{\varepsilon}}_i\right)$$

lub

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

- Zauważmy, że stosując te przekształcenie eliminujemy efekt indywidualny!

- Do przekształconego równania możemy zastosować zwykłe MNK, ponieważ

$$\mathrm{E}\left(\boldsymbol{\varepsilon}_{it}\middle|\,\boldsymbol{x}_i, u_i\right) = 0 \Longrightarrow \mathrm{E}\left(\overline{\varepsilon}_{it}\middle|\,\overline{\boldsymbol{x}}_i, u_i\right) = 0$$

  a więc

$$\mathrm{E}\left(\boldsymbol{\varepsilon}_{it}\middle|\,\boldsymbol{x}_i, u_i\right) = 0 \Longrightarrow \mathrm{E}\left(\ddot{\varepsilon}_{it}\middle|\,\ddot{\boldsymbol{x}}_{it}, u_i\right) = 0$$

  co wystarcza, aby MNK zastosowane do przekształconego modelu dało zgodny estymator.

- Zauważmy, że w przypadku estymatora efektów stałych nie musimy robić żadnych założeń na temat relacji między tym efektami indywidualnymi a czystym błędem losowym.

- Z drugiej strony, tak samo jak w modelu efektów losowych musimy założyć, że ε_{it} jest nieskorelowane z przeszłymi, obecnymi i przyszłymi wartościami $\boldsymbol{x}_i$.

- **Podstawowy problem**: jeśli wśród zmiennych pojawia się zmienna, która nie zmienia się w czasie, to jej wpływu nie da się oszacować, ponieważ eliminuje ją przekształcenie efektów stałych.

Przykład Załóżmy, że chcemy oszacować wpływ płci i wieku na płacę.

$$wage_{it} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 sex_i + \boldsymbol{\beta}_2 age_{it} + u_i + \boldsymbol{\varepsilon}_{it}$$

Przy zmiennej sex_i jest tylko jeden indeks ponieważ nie zmienia się ona w czasie. Zastosujmy przekształcenie efektów stałych. Otrzymamy

$$wage_{it} - \overline{wage_i} = \boldsymbol{\beta}_2 \left(age_{it} - \overline{age}_i \right) + \ddot{\boldsymbol{\varepsilon}}_{it}$$

Stała i zmienna sex_i całkowicie znikły z modelu po jego przekształceniu!

- Estymator efektów stałych (FE - **F**ixed **E**ffect) liczymy za pomocą MNK przeprowadzając regresję $\ddot{y}_{it}$ on $\ddot{\boldsymbol{x}}_{it}$

- Jedyną korektą jaką należy zastosować jest korekta liczby stopni swobody we wzorze na $\widehat{\boldsymbol{\sigma}}_u^2$.

- Aby uzyskać niobciążony estymator $\boldsymbol{\sigma}_u^2$ należy zstosować wzór $\widehat{\boldsymbol{\sigma}}_u^2 = \frac{RSS}{[N(T-1)-K]}$

- Estymator efektów stałych nazywany jest także estymatorem wewnątrzobiektowym (*within estimator*), ponieważ wykorzystuje on do estymacji parametrów jedynie zróżnicowanie wewnątrzobiektowe (odchylenia obserwacji dla jednostki od średniej dla tej jednostki)

- Przy założeniu, że $\mathrm{E}\left(\boldsymbol{u}_i\boldsymbol{u}_i'\middle|\,\boldsymbol{x}_i\right) = \boldsymbol{\sigma}_u^2\boldsymbol{I}_T$ estymator efektów stałych jest estymatorem efektywnym.

- Testowanie występowania efektów indywidualnych w przypadku stosowania estymatora efektów stałych polega na testowaniu parametrycznej hipotezy, że wszystkie efekty indywidualne równe są zeru.

- Test taki można przeprowadzić porównując sumy kwadratów reszt w modelu wyestymowanym za pomocą MNK zastosowanego do próby wymieszanej i sumy kwadratów reszt w modelu wyestymowanym za pomocą estymatora efektów stałych.

- Uzyskana statystyka będzie miała postać

$$\frac{\left(RSS_{MNK} - RSS_{FE}\right)/\left(N-1\right)}{\left(RSS_{MNK}\right)/\left(N-K-1\right)} \sim F\left(N-1, N-K-1\right)$$

Przykład Płaca netto a charakterystyki respondenta - kontynuacja.

```
Fixed-effects (within) regression               Number of obs      =     16178
Group variable (i): ID                          Number of groups   =      5054

R-sq:  within  = 0.4254                         Obs per group: min =         1
       between = 0.0502                                        avg =       3.2
       overall = 0.0495                                        max =         4

                                                F(8,11116)         =   1028.50
corr(u_i, Xb)  = -0.9457                        Prob > F           =    0.0000

------------------------------------------------------------------------------
     LNETPAY |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         AGE |   .1408278    .012854    10.96   0.000     .1156316     .166024
        AGE2 |  -.0002616   .0001496    -1.75   0.080    -.0005548    .0000316
     _ISEX_2 |  (dropped)
  _IEDUCAT_2 |  -.1363578   .1326115    -1.03   0.304    -.3962998    .1235843
  _IEDUCAT_3 |  -.1170238   .0845124    -1.38   0.166     -.282683    .0486355
  _IEDUCAT_4 |  -.1765529   .0737762    -2.39   0.017    -.3211673   -.0319385
  _IEDUCAT_5 |  -.1683508   .0990707    -1.70   0.089     -.362547    .0258455
  _IEDUCAT_7 |  -.1236027   .1132831    -1.09   0.275    -.3456577    .0984522
```

```
  _IEDUCAT_8 |  (dropped)
   _ITOWN2_1 |  (dropped)
   _ITOWN2_2 |  (dropped)
   _ITOWN2_3 |  (dropped)
   _ITOWN2_4 |  (dropped)
   _ITOWN2_5 |  (dropped)
   _ITOWN2_6 |  (dropped)
   _ITOWN2_7 |  (dropped)
   _ITOWN2_9 |  (dropped)
      OBSNUM |   .0638621   .0026914    23.73   0.000     .0585866    .0691377
       _cons |   1.133222   .3066956     3.69   0.000     .5320439      1.7344
-------------+----------------------------------------------------------------
     sigma_u |  1.2139867
     sigma_e |   .1662991
         rho |   .9815805   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(5053, 11116) =    10.91         Prob > F = 0.0000
```

- Zauważmy, że wszystkie zmienne, SEX (płeć) i TOWN2 (wielkość miasta w którym zamieszkuje respondent) wypadły z regresji. Płeć nie zmienia się w czasie zaś wielkość miasta, w którym mieszka respondent nie może się zmienić z racji na konstrukcję badania - jeśli respondent zmieni miejsce zamieszkania to wypada z próby

- Oszacowania przy zmiennej EDUCAT (wykształcenie) są bardzo nieprecyzyjne - prawdopodobnym powodem jest mały udział zaobserwowanych w próbie osób, które w trakcie trwania badania ($1,5$ roku) podniosły swój poziom wykształcenia.

- Wynik testu na istnienie efektów indywidualnych `F(5053, 11116) = 10.91`, sugeruje udrzucenie hipotezy o nieistotności efektów losowych

- Estymator efektów stałych można policzyć także przeprowadzając regresję y_{it} na $\boldsymbol{x}_{it}$ oraz na zmiennych $d_{1it}, d_{2it}, \ldots, d_{Nit}$, gdzie zmienne d_{kit} jest zmienną zerojedynkową przyjmującą wartość 1 jeśli $i = k$ i 0 w przeciwnym przypadku.

- Ta forma estymatora efektów stałych określana jest w literaturze $LSDV$ (**L**east **S**quares **D**ummy **V**ariables)

- W praktyce stosowanie tego estymatora ma sens tylko wtedy, gdy liczba jednostek jest mała, w przeciwnym razie powstają problemy z odwracaniem macierzy $\boldsymbol{X}$ - ponieważ zawiera ona bardzo dużo ($(K+N) \times (K+N)$) elementów.

- Jedyną zaletą estymatora $LSDV$ jest to, że uzyskujemy za jego pomocą bezpośrednio oszacowania efektów indywidualnych dla poszczególnych jednostek jako oszacowania parametrów stojących przed zmiennymi zero jedynkowymi odnoszącymi się do tych jednostek.

- Oszacowania te, choć nieobciążone, nie są jednak zgodne, chyba że $T \to \infty$ (ale jako powiedziano zazwyczaj T jest małe)

- Aby oszacować efekty indywidualne dla poszczególnych jednostek w przypadku, kiedy stosujemy standardową formę estymatora efektów stałych możemy zastosować następujący wzór:

$$\widehat{u}_i = \overline{y}_i - \overline{\boldsymbol{x}}_i \widehat{\boldsymbol{\beta}}_{FE}$$

Porównanie estymatorów i test na nieobciążoność estymatora efektów losowych

- Estymator MNK ($POLS$) i estymator efektów losowych wykorzystują zróżnicowanie wewnątrzobiektowe

$$y_{it} - \overline{y}_i = \left(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_i\right)\boldsymbol{\beta} + \left(\boldsymbol{\varepsilon}_{it} - \overline{\boldsymbol{\varepsilon}}_i\right)$$

gdzie $\overline{y}_i = \frac{\sum_{t=1}^{T} y_{it}}{T}$, $\overline{\boldsymbol{x}}_i = \frac{\sum_{t=1}^{T} \boldsymbol{x}_{it}}{T}$, $\overline{\boldsymbol{\varepsilon}}_i = \frac{\sum_{t=1}^{T} \boldsymbol{\varepsilon}_{it}}{T}$

oraz zróżnicowanie międzyobiektowe

$$y_{it} - \overline{y}_t = \left(\boldsymbol{x}_{it} - \overline{\boldsymbol{x}}_t\right)\boldsymbol{\beta} + \left(\boldsymbol{\varepsilon}_{it} - \overline{\boldsymbol{\varepsilon}}_t\right)$$

gdzie $\overline{y}_t = \frac{\sum_{i=1}^{N} y_{it}}{N}$, $\overline{\boldsymbol{x}}_t = \frac{\sum_{i=1}^{N} \boldsymbol{x}_{it}}{N}$, $\overline{\boldsymbol{\varepsilon}}_t = \frac{\sum_{i=1}^{N} \boldsymbol{\varepsilon}_{it}}{N}$

- Estymator efektów stałych wykorzystuje jedynie zróżnicowanie wewnątrzobiektowe
- Estymator efektów losowych jest efektywny i zgodny jeśli u_i i $\boldsymbol{x}_i$ są nieskorelowane
- Estymator efektów stałych jest nieefektywny ale zgodny jeśli u_i i $\boldsymbol{x}_i$ są nieskorelowane
- Estymator efektów losowych jest nie jest zgodny jeśli u_i i $\boldsymbol{x}_i$ są skorelowane
- Estymator efektów stałych jest efektywny i zgodny jeśli u_i i $\boldsymbol{x}_i$ są skorelowane
- Za pomocą estymatora efektów stałych nie da się oszacować wpływu zmiennych stałych w czasie na zmienną zależną

Test Hausmana

- Testowanie poprawności założeń modelu efektów losowych
  - H_0: założenie o braku skorelowania efektów indywidualnych ze zmiennymi objaśniającymi jest prawdziwe
    - estymatory efektów losowych i stałych są zgodne - różnica między szacowaniami nie powinna być statystycznie istotna
  - H_1: założenie o braku korelacji między efektami indywidualnymi i zmienymi objaśniającymi jest fałszywe
    - estymator efektów losowych nie jest zgodny, estymator efektów stałych zgodny - różnica między oszacowaniami może być statystycznie istotna
- Wniosek: jeśli różnica między oszacowaniami wektora parametrów $\boldsymbol{\beta}$

uzyskanymi za pomocą estymatora efektów losowych i zmiennych jest statystycznie istotna powinniśmy odrzucić H_0

- Istotne uproszczenie postaci statystki można uzyskać dzięki temu, że przy H_0 estymator RE jest efektywny, ponieważ w takiej sytuacji $\mathrm{Cov}\left(\widetilde{\boldsymbol{\beta}}_{FE}, \widetilde{\boldsymbol{\beta}}_{RE}\right) \xrightarrow{p} \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}_{RE}}$

- Test (nazywany testem Hausmana) przyjmuje w tej sytuacji typową postać formy kwadratowej z udziałem odwrotności macierzy wariancji

$$\left(\widetilde{\boldsymbol{\beta}}_{FE} - \widetilde{\boldsymbol{\beta}}_{RE}\right)\left(\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}_{FE}} - \widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\beta}_{RE}}\right)^{-1}\left(\widetilde{\boldsymbol{\beta}}_{FE} - \widetilde{\boldsymbol{\beta}}_{RE}\right) \xrightarrow{D} \boldsymbol{\chi}^2_{K^*+1},$$

- Uwagi: w przypadku, kiedy w modelu występują zmienne stałe w czasie, zarówno estymatory jak i macierz wariancji kowariancji powinny dotyczyć jedynie oszacowań przy parametrach zmieniających się w czasie.

Przykład Wynik testu Hausmana na poprawność modelu efektów losowych

```
                ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       fe           re         Difference          S.E.
-------------+----------------------------------------------------------------
         AGE |    .1408278     .0249031         .1159247         .0125428
        AGE2 |   -.0002616    -.0002209        -.0000407         .0001449
  _IEDUCAT_2 |   -.1363578    -.2404923         .1041345         .1301138
  _IEDUCAT_3 |   -.1170238    -.2506126         .1335888         .0830264
  _IEDUCAT_4 |   -.1765529    -.2144521         .0378992         .0709333
  _IEDUCAT_5 |   -.1683508    -.3898099         .2214591         .0978304
  _IEDUCAT_7 |   -.1236027    -.4959456         .3723429         .1118169
      OBSNUM |    .0638621     .1082641        -.044402          .0023727
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(7) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =      369.23
```

```
Prob>chi2 =      0.0000
```

- Wielkość statystyki testu Hausmana `chi2(7) = 369.23` bardzo silnie sugeruje odrzucenie H_0 o prawidłowości założeń modelu efektów losowych

- Powinniśmy więc stosować model efektów stałych

- Z racji na wspomniane wady modelu efektów stałych opracowano inne estymatory, nie będziemy ich jednak tutaj bliżej opisywać.

Inne zastosowania paneli

- Jednym z bardzo ważnych zastosowań paneli jest możliwość obserwacji zmiany stanów jednostek
- Powiedzmy, że jesteśmy zainteresowani problemem bezrobocia
- Bardzo często w literaturze pojawia się opinia, że szczegolnie trudny społecznie problemem jest bezrobocie długotrwałe
- Jeśli chcemy zbadać problem występowania bezrobocia długotrwałego możemy albo oprzeć się na:
  - ankietach retrospektywnych: pytania dotyczące przeszłości zawodowej

– panelu: obserwujemy wtedy także zmiany w czasie charakterystyk respondenta

- Na podstawie próby przekrojowej bez pytań retrospektywnych nie da się prześledzić dynamiki zmian

- Na podstawie panelu można między innymi oszacować prawdopodobieństwa zmiany stanów np. z bezrobotnego na emeryta, lub pracującego na bezrobotnego.

Przykład Główne źródło utrzymania. Podział w ramach całej populacji badanej.

```
1 praca najemna
2 gosp.rolne
3 praca na r-k wlasny
4 renta, emeryt
5 zasilek
6 inne niezarobkowe
```

zrodlo utrzymania	Freq.	Percent	Cum.
praca najemna	37,964	52.76	52.76
gosp.rolne	8,373	11.64	64.39
praca na r-k wlasny	4,654	6.47	70.86
renta, emeryt	17,919	24.90	95.76
zasilek	1,950	2.71	98.47
inne niezarobkowe	1,099	1.53	100.00
Total	71,959	100.00	

- Na podstawie tego rodzaju tabeli możemy ustalić udziały poszczególnych grup w ogóle populacji nie możemy jednak powiedzieć nic na temat przepływów między grupami.

- Pytanie - na ile prawdopodobna jest zmiana głównego źródła utrzymania i w jakim kierunku zachodzą te zmian?

- Oszacowanie prawdopodobieństw zmian stanów między okresem t i $t+1$

zrodlo utrzymania	zrodlo utrzymania 1	2	3	4	5	6	Total
1	96.11	0.23	0.85	1.15	1.42	0.24	100.00
2	0.86	97.87	0.47	0.74	0.06	0.00	100.00
3	3.46	0.27	93.25	1.30	1.14	0.57	100.00
4	1.47	0.16	0.30	97.74	0.23	0.10	100.00
5	24.36	0.49	2.75	1.62	66.38	4.38	100.00
6	22.30	1.30	2.97	3.32	4.27	65.84	100.00
Total	52.63	11.69	6.57	25.03	2.74	1.35	100.00

- Najbardziej stabilnymi stanami jest osiąganie głównego źródła dochodów z rolnictwa lub z emerytury lub renty - blisko 98% osób zaobserwowanych w tym stanie znajdowało się w tym samym stanie w następnym badaniu

- Mniej stabilnymi stanami jest utrzymywanie się z zasiłku i źródeł niezarobkowych.

- 66% osób, dla których głównym źródłem był zasiłek, w następnym okresie było także na zasiłku, 24% udało się zdobyć pracę

- Szerszej analizy przejść między stanami dokonuje się w ramach analizy łańcuchów Markowa.