

Metody badań w zarządzaniu finansami – WZ – UW – 2019/2020

## **ZAJĘCIA 3**

# Ostatnio

- Po co nam ekonometria?
  - Co to jest model?
  - Jak sprawdzić czy model jest poprawny?
  - Jak porównywać modele między sobą
- Jak zbudować dobry model ekonometryczny?
- Jakie dane wybrać? Czym się różnią?

# Global Financial Development Database

	A	B	C	D	E	F	G	DC	DD	DE	DF	DG	DH	DI	DJ	DK	DL	DM	DN	DO	DP	DQ	DR	DT	
1	iso3	iso2	imfn	country	region	income	year	gfdDOI	gfdDI	gfdDE	gfdDF	gfdDG	gfdDH	gfdDI	gfdDJ	gfdDK	gfdDL	gfdDM	gfdDN	ny_gdp	ny_gdp	ny_gdp	sp_pop	ptl	
386	ARG	AR	213	Argentina	Latin Ame	High inco	1996	63,6225	0,02388	18,6408	24			0,087	0	4,17064	32,6953	38,8227	38,8227	2,70E+11	7994,24	2,70E+11	3,50E+07		
387	ARG	AR	213	Argentina	Latin Ame	High inco	1997	69,5216	0,02243	26,1354	29			0,351	0	3,81407	33,9425	38,95	38,95	2,90E+11	8543,03	2,90E+11	3,60E+07		
388	ARG	AR	213	Argentina	Latin Ame	High inco	1998	60,2187	0,02291	27,0699	32			0,403	0	3,63249	-25,969	39,2088	39,2088	3,00E+11	8772,06	2,90E+11	3,60E+07		
389	ARG	AR	213	Argentina	Latin Ame	High inco	1999	55,2539	0,02268	31,8426	37		0,344	0,524	0	3,42774	-12,258	38,4963	38,4963	2,80E+11	8381,25	2,80E+11	3,60E+07		
390	ARG	AR	213	Argentina	Latin Ame	High inco	2000	56,0247	0,03038	32,1934	37		0,246	0,561	0	3,30885	3,52911	38,2148	38,2148	2,80E+11	8224,11	2,80E+11	3,70E+07		
391	ARG	AR	213	Argentina	Latin Ame	High inco	2001	35,5733	0,07056	27,2582	37		0,156	0,34	1	3,11195	-27,197	37,625	37,625	2,70E+11	7776,14	2,60E+11	3,70E+07		
392	ARG	AR	213	Argentina	Latin Ame	High inco	2002	124,971	0,21144	39,7988	34			0,073	1	2,91919	8,81705	53,0335	53,0335	9,80E+10	6854,29	9,30E+10	3,80E+07		
393	ARG	AR	213	Argentina	Latin Ame	High inco	2003	98,7048	0,2143	28,0789	35			0,068	1	2,78304	83,2605	54,9726	54,9726	1,30E+11	7380,47	1,20E+11	3,80E+07		
394	ARG	AR	213	Argentina	Latin Ame	High inco	2004	76,0238	0,18935	18,7698	33	29	0,36	0,074	0	2,67588	51,2389	58,3265	58,3265	1,60E+11	7962,41	1,50E+11	3,80E+07		
395	ARG	AR	213	Argentina	Latin Ame	High inco	2005	58,9286	0,21742	13,8278	32	27	0,19	0,146	0	2,57116	34,7541	65,5161	65,5161	2,00E+11	8577,87	1,80E+11	3,90E+07		
396	ARG	AR	213	Argentina	Latin Ame	High inco	2006	55,5577	0,23247	14,9099	32	26	0,27	0,166	0	2,57064	16,8388	71,9617	71,9617	2,30E+11	9174,5	2,20E+11	3,90E+07		
397	ARG	AR	213	Argentina	Latin Ame	High inco	2007	53,76	0,21093	13,1085	32	27	2,92	0,17232	0	2,67108	23,589	78,0612	78,0612	2,90E+11	9901,51	2,80E+11	4,00E+07		
398	ARG	AR	213	Argentina	Latin Ame	High inco	2008	46,7523	0,19503	9,81198	33	28	0,43	0,14383	0	2,66965	-17,713	83,7114	83,7114	3,60E+11	10201,5	3,50E+11	4,00E+07		
399	ARG	AR	213	Argentina	Latin Ame	High inco	2009	48,5127	0,18877	10,4104	33	28	0,37	0,13976	0	2,49489	-8,3268	90,1606	90,1606	3,30E+11	9502,24	3,20E+11	4,00E+07		
400	ARG	AR	213	Argentina	Latin Ame	High inco	2010	38,2314	0,15209	9,35257	33	24	0,25	0,10942	0	2,47619	56,416	100	100	4,20E+11	10386	4,10E+11	4,10E+07		
401	ARG	AR	213	Argentina	Latin Ame	High inco	2011	29,0865	0,13154	9,14851	33	29	0,32	0,12455	0	2,39933	22,1574	109,533	109,533	5,30E+11	10883,3	5,20E+11	4,10E+07		
402	ARG	AR	213	Argentina	Latin Ame	High inco	2012	26,8998	0,10578	8,76722	32	27	0,17	0,14449	0	2,42013	-19,837	121,398	121,398	5,50E+11	10649,8	5,30E+11	4,20E+07		
403	ARG	AR	213	Argentina	Latin Ame	High inco	2013	25,4607	0,09692	7,83024	32	25	0,8	0,20579	0	2,29842	60,007	134,668	134,668	5,50E+11	10784,6	5,40E+11	4,20E+07		
404	ARG	AR	213	Argentina	Latin Ame	High inco	2014	27,431	0,09602	7,96745			0,06	0,32745	0	2,22641	103,15	113,38	105,507	5,30E+11	10398,7	5,20E+11	4,30E+07		
405	ARG	AR	213	Argentina	Latin Ame	High inco	2015	23,125	0,08313	7,23731			0,18	0,28923	0	2,15617	36,64		120,604	5,90E+11	10568,2	5,80E+11	4,30E+07		
406	ARG	AR	213	Argentina	Latin Ame	High inco	2016	25,649	0,07023	7,81101			0,11	0,37698	0	2,1335	31,1342			5,60E+11	10239,5	5,50E+11	4,40E+07		
407	ARG	AR	213	Argentina	Latin Ame	High inco	2017	18,5366	0,06994	9,01434			0,14	0,40333	0	2,1796	56,3341			6,40E+11	10404,3	6,30E+11	4,40E+07		
3808	FIN	FI	172	Finland	Europe &	High inco	1996	31,1002	0,11011		14			18,793	0	22,0506	5,92921	80,4963	80,4836	1,30E+11	33059,1	1,30E+11	5,10E+06		
3809	FIN	FI	172	Finland	Europe &	High inco	1997	35,5805	0,18548	0,02838	14			25,541	0	24,1253	57,7962	82,016	81,4434	1,30E+11	35021,6	1,20E+11	5,10E+06		
3810	FIN	FI	172	Finland	Europe &	High inco	1998	34,4019	0,28029	0,00971	14			27,366	0	25,0315	41,2068	82,6879	82,5832	1,30E+11	36824,8	1,30E+11	5,20E+06		
3811	FIN	FI	172	Finland	Europe &	High inco	1999	32,3611	0,32805	29,7724	14		0,659	4,174	0	28,4582	72,2627	84,3675	83,543	1,40E+11	38372,2	1,30E+11	5,20E+06		
3812	FIN	FI	172	Finland	Europe &	High inco	2000	43,1971	0,37656	43,7104	14			0,779	5,292	0	30,5243	90,8097	86,9145	86,0845	1,30E+11	40450,4	1,20E+11	5,20E+06	
3816	FIN	FI	172	Finland	Europe &	High inco	2004	68,008	0,33869	67,4972	13		1,05	6,308	0	25,6304	7,50736	90,8574	90,6479	2,00E+11	44277,8	2,00E+11	5,20E+06		
3817	FIN	FI	172	Finland	Europe &	High inco	2005	60,4821	0,3388	61,5009	13	72	0,87	6,058	0	25,5428	17,4228	91,3958	91,2134	2,00E+11	45353,3	2,10E+11	5,20E+06		
3818	FIN	FI	172	Finland	Europe &	High inco	2006	66,375	0,32995	75,9869	22	85	1	6,75	0	25,445	24,6206	93,4025	92,6424	2,20E+11	47011,7	2,20E+11	5,30E+06		
3819	FIN	FI	172	Finland	Europe &	High inco	2007	61,3723	0,31564	83,1869	22	85	1,24	6,77944	0	24,5806	24,4946	95,8653	94,9684	2,60E+11	49239,2	2,60E+11	5,30E+06		
3820	FIN	FI	172	Finland	Europe &	High inco	2008	54,3939	0,33804	73,6359	22	84	1,04	6,53059	0	23,7136	-23,834	99,149	98,8297	2,80E+11	49363,7	2,80E+11	5,30E+06		
3821	FIN	FI	172	Finland	Europe &	High inco	2009	63,0305	0,34396	86,3896	22	82	0,72	5,93902	0	23,4132	-32,936	98,6017	98,8297	2,50E+11	45065,8	2,50E+11	5,30E+06		

# Global Financial Development Database

- ❑ 214 krajów
- ❑ 1960-2017
- ❑ 115 zmiennych (charakterystyk finansowych) dotyczących 6 „bloków tematycznych”: **Access, Depth, Efficiency, Stability, Other, Other Economic**
  
- ❑ <https://datacatalog.worldbank.org/dataset/global-financial-development> - więcej informacji nt bazy danych

# Co chcemy zbadać?

- **Co ma wpływ na tempo wzrostu gospodarczego?**
  - **Badamy tylko korelację**
  - Wykorzystamy wskaźniki finansowe publikowane przez Global Finance Development Database
- **Jak mierzyć tempo wzrostu gospodarczego?**
  - przyrost realnego PKB na mieszkańca (rate of real per capita GDP growth)
  - wzrost kapitału na osobę (growth in capital stock per person),
  - wzrost całkowitej wydajności (TFP, Total productivity growth)

# Przykład

- Global Financial Development Database, m.in.:

**Depth** → Liquid liabilities to GDP (%) (GFDD.DI.05)

**Bank** → Deposit money bank assets to deposit money bank assets and central bank assets (%) (GFDD.DI.04)

**Privy** → Private credit by deposit money banks to GDP (%) (GFDD.DI.01)

**Assets** → Assets held by deposit money banks as a share of GDP (GFDD.DI.02)

***GDP growth***

$$= \beta_0 + \beta_1 \text{Depth} + \beta_2 \text{Bank} + \beta_3 \text{Privy} + \beta_4 \text{Assets} + \varepsilon$$

# Zadanie R

- Używając jako zmiennej objaśnianej GDP\_growth oraz zmiennych objaśniających wskaźników finansowych gfdddi01, gfdddi02, gfdddi04, gfdddi05 stwórz model *pooled ols*
- **R:** `plm(GDP_growth~gfdddi01 + gfdddi02 + gfdddi04 + gfdddi05, data=pdata, model="pooling")`
- **R:** `lm(GDP_growth~gfdddi01 + gfdddi02 + gfdddi04 + gfdddi05, data=...)`
- **STATA:** `reg $ylist $xlist`

# Zadanie R

```
> ##POOLED OLS estimator
> pooling<-plm(GDP_growth~gfdddi01 + gfdddi02 + gfdddi04 + gfdddi05,data=pdata,model="pooling")
> summary(pooling)
Pooling Model

Call:
plm(formula = GDP_growth ~ gfdddi01 + gfdddi02 + gfdddi04 + gfdddi05,
     data = pdata, model = "pooling")

Unbalanced Panel: n = 36, T = 14-38, N = 1174

Residuals:
    Min.   1st Qu.   Median     3rd Qu.    Max.
-17.07975  -1.35826   0.11148   1.48064   22.29291

Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
(Intercept)  1.67562106  0.81969083   2.0442 0.0411564 *
gfdddi01     0.00612910  0.00758401   0.8082 0.4191626
gfdddi02    -0.02488922  0.00697015  -3.5708 0.0003703 ***
gfdddi04     0.03035677  0.00919419   3.3017 0.0009899 ***
gfdddi05    -0.00075014  0.00133684  -0.5611 0.5748153
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

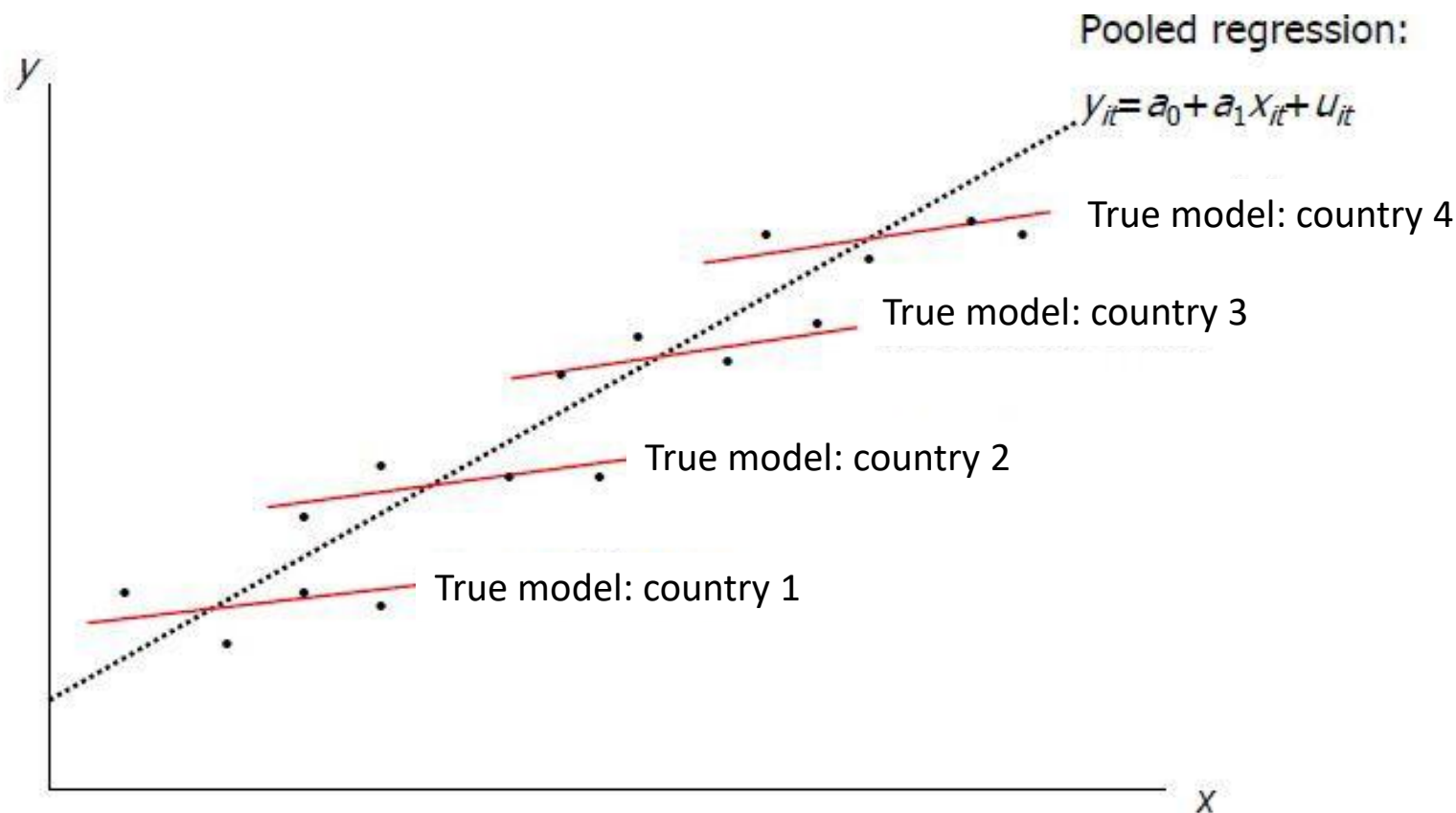
Total sum of Squares:    11434
Residual sum of Squares: 10617
R-Squared:               0.071513
Adj. R-Squared:         0.068336
F-statistic: 22.5095 on 4 and 1169 DF, p-value: < 2.22e-16
> |
```



**CZY TO BYŁO „DOBRCZE”?**

# OLS gdy dane są panelowe ...

investment depends on profit



# OLS – co zakładamy, gdy dane panelowe?

- $GDP\ growth_{it} = \beta_0 + \beta_1 Depth_{it} + \beta_2 Bank_{it} + \beta_3 Privy_{it} + \beta_4 Assets_{it} + \varepsilon_{it}$
- W praktyce nakładamy na model silne restrykcje, często nieprawdziwe:
  - Wyraz wolny i nachylenie (dla każdej zmiennej objaśniającej) jest stałe w czasie ...
  - ...i takie same dla wszystkich krajów ...
  - ... a wszystkie różnice w czasie i między krajami „wepchnięte” są w błąd  $\varepsilon$

# Co się dzieje, gdy OLS a dane panelowe?

- Dokonujemy regresji na wszystkich dostępnych obserwacjach, tak jakby były danymi przekrojowymi (stosujemy zwykłe OLS)

$$y_{it} = \beta_0 + X_{it}\beta' + \varepsilon_{it}$$

- Kiedy wolno to stosować? TYLKO wtedy, gdy wiemy, że
  - ➔ **brak efektów indywidualnych**
  - ➔ **brak zmian analizowanego zjawiska w czasie**

(powiązania między x i y są takie same dla **wszystkich** i w **czasie**)

**Wszyscy** = firmy, kraje, osoby, działy w firmie, typy produktów, itp

# Do przemyślenia

- Podaj przykład danych panelowych, gdzie zależności są stałe w czasie, ale różne dla wszystkich.
  - Czy wtedy da się oszacować model? Jak to zrobić? Czy to jest sensowne?
- Podaj przykład danych panelowych, gdzie zależności są takie same dla wszystkich, ale różne w czasie.
  - Czy wtedy da się oszacować model? Jak to zrobić? Czy to jest sensowne?
- Podaj przykład danych panelowych, gdzie zależności są stałe w czasie i takie same dla wszystkich.
  - Czy wtedy da się oszacować model? Jak to zrobić? Czy to jest sensowne?

Podpowiedź: zastanów się co jest źródłem zmienności (wariancji  $x$  i wariancji  $y$ ) w każdym z tych trzech przypadków.

# Dane panelowe

- opisują pewną **zbiorowość** w więcej niż jednym okresie **czasu**
- połączenie danych **przekrojowych** (opisujących zbiorowość w pojedynczym momencie) i szeregów **czasowych** (opisujących jednostkę w różnych okresach).

1	A	B	C	D	E	F	G	DC	DD	DE	DF	DG	DH	DI	DJ	DK	DL	DM	DN	DO	DP	DQ	DR
386	ARG	AR	213	Argentina	Latin Ame High inco	1996	63,6225	0,02388	18,6408	24				0,087	0	4,17064	32,6953	38,8227	38,8227	2,70E+11	7994,24	2,70E+11	3,50E+07
387	ARG	AR	213	Argentina	Latin Ame High inco	1997	69,5216	0,02243	26,1354	29				0,351	0	3,81407	33,9425	38,95	38,95	2,90E+11	8543,03	2,90E+11	3,60E+07
388	ARG	AR	213	Argentina	Latin Ame High inco	1998	60,2187	0,02291	27,0699	32				0,403	0	3,63249	-25,969	39,2088	39,2088	3,00E+11	8772,06	2,90E+11	3,60E+07
389	ARG	AR	213	Argentina	Latin Ame High inco	1999	55,2539	0,02268	31,8426	37			0,344	0,524	0	3,42774	-12,258	38,4963	38,4963	2,80E+11	8381,25	2,80E+11	3,60E+07
390	ARG	AR	213	Argentina	Latin Ame High inco	2000	56,0247	0,03038	32,1934	37			0,246	0,561	0	3,30885	3,52911	38,2148	38,2148	2,80E+11	8224,11	2,80E+11	3,70E+07
391	ARG	AR	213	Argentina	Latin Ame High inco	2001	35,5733	0,07056	27,2582	37			0,156	0,34	1	3,11195	-27,197	37,625	37,625	2,70E+11	7776,14	2,60E+11	3,70E+07
392	ARG	AR	213	Argentina	Latin Ame High inco	2002	124,9711	0,21144	39,7988	34				0,073	1	2,91919	8,81705	53,0335	53,0335	9,80E+10	6854,29	9,30E+10	3,80E+07
393	ARG	AR	213	Argentina	Latin Ame High inco	2003	98,7048	0,2143	28,0789	35				0,068	1	2,78304	83,2605	54,9726	54,9726	1,30E+11	7380,47	1,20E+11	3,80E+07
394	ARG	AR	213	Argentina	Latin Ame High inco	2004	76,0238	0,18935	18,7698	33	29	0,36	0,074	0	2,67588	51,2389	58,3265	58,3265	1,60E+11	7962,41	1,50E+11	3,80E+07	
395	ARG	AR	213	Argentina	Latin Ame High inco	2005	58,9286	0,21742	13,8278	32	27	0,19	0,146	0	2,57116	34,7541	65,5161	65,5161	2,00E+11	8577,87	1,80E+11	3,90E+07	
396	ARG	AR	213	Argentina	Latin Ame High inco	2006	55,5577	0,23247	14,9099	32	26	0,27	0,166	0	2,57064	16,8388	71,9617	71,9617	2,30E+11	9174,5	2,20E+11	3,90E+07	
397	ARG	AR	213	Argentina	Latin Ame High inco	2007	53,76	0,21093	13,1085	32	27	2,92	0,17232	0	2,67108	23,589	78,0612	78,0612	2,90E+11	9901,51	2,80E+11	4,00E+07	
398	ARG	AR	213	Argentina	Latin Ame High inco	2008	46,7523	0,19503	9,81198	33	28	0,43	0,14383	0	2,66965	-17,713	83,7114	83,7114	3,60E+11	10201,5	3,50E+11	4,00E+07	
399	ARG	AR	213	Argentina	Latin Ame High inco	2009	48,5127	0,18877	10,4104	33	28	0,37	0,13976	0	2,49489	-8,3268	90,1606	90,1606	3,30E+11	9502,24	3,20E+11	4,00E+07	
400	ARG	AR	213	Argentina	Latin Ame High inco	2010	38,2314	0,15209	9,35257	33	24	0,25	0,10942	0	2,47619	56,416	100	100	4,20E+11	10386	4,10E+11	4,10E+07	
401	ARG	AR	213	Argentina	Latin Ame High inco	2011	29,0865	0,13154	9,14851	33	29	0,32	0,12455	0	2,39933	22,1574	109,533	109,533	5,30E+11	10883,3	5,20E+11	4,10E+07	
402	ARG	AR	213	Argentina	Latin Ame High inco	2012	26,8998	0,10578	8,76722	32	27	0,17	0,14449	0	2,42013	-19,837	121,398	121,398	5,50E+11	10649,8	5,30E+11	4,20E+07	
403	ARG	AR	213	Argentina	Latin Ame High inco	2013	25,4607	0,09692	7,83024	32	25	0,8	0,20579	0	2,29842	60,007	134,668	134,668	5,50E+11	10784,6	5,40E+11	4,20E+07	
404	ARG	AR	213	Argentina	Latin Ame High inco	2014	27,431	0,09602	7,96745				0,06	0,32745	0	2,22641	103,15	113,38	105,507	5,30E+11	10398,7	5,20E+11	4,30E+07
405	ARG	AR	213	Argentina	Latin Ame High inco	2015	23,125	0,08313	7,23731				0,18	0,28923	0	2,15617	36,64	120,604	5,90E+11	10568,2	5,80E+11	4,30E+07	
406	ARG	AR	213	Argentina	Latin Ame High inco	2016	25,649	0,07023	7,81101				0,11	0,37698	0	2,1335	31,1342	5,60E+11	10239,5	5,50E+11	4,40E+07		
407	ARG	AR	213	Argentina	Latin Ame High inco	2017	18,5366	0,06994	9,01434				0,14	0,40333	0	2,1796	56,3341	6,40E+11	10404,3	6,30E+11	4,40E+07		
3808	FIN	FI	172	Finland	Europe & High inco	1996	31,1002	0,11011			14			18,793	0	22,0506	5,92921	80,4963	80,4836	1,30E+11	33059,1	1,30E+11	5,10E+06
3809	FIN	FI	172	Finland	Europe & High inco	1997	35,5805	0,18548	0,02838		14			25,541	0	24,1253	57,7962	82,016	81,4434	1,30E+11	35021,6	1,20E+11	5,10E+06
3810	FIN	FI	172	Finland	Europe & High inco	1998	34,4019	0,28029	0,00971		14			27,366	0	25,0315	41,2068	82,6879	82,5832	1,30E+11	36824,8	1,30E+11	5,20E+06
3811	FIN	FI	172	Finland	Europe & High inco	1999	32,3611	0,32805	29,7724		14		0,659	4,174	0	28,4582	72,2627	84,3675	83,543	1,40E+11	38372,2	1,30E+11	5,20E+06
3812	FIN	FI	172	Finland	Europe & High inco	2000	43,1971	0,37656	43,7104		14		0,779	5,292	0	30,5243	90,8097	86,9145	86,0845	1,30E+11	40450,4	1,20E+11	5,20E+06
3816	FIN	FI	172	Finland	Europe & High inco	2004	68,008	0,33869	67,4972		13		1,05	6,308	0	25,6304	7,50736	90,8574	90,6479	2,00E+11	44277,8	2,00E+11	5,20E+06
3817	FIN	FI	172	Finland	Europe & High inco	2005	60,4821	0,3388	61,5009		13	72	0,87	6,058	0	25,5428	17,4228	91,3958	91,2134	2,00E+11	45353,3	2,10E+11	5,20E+06
3818	FIN	FI	172	Finland	Europe & High inco	2006	66,375	0,32995	75,9869		22	85	1	6,75	0	25,445	24,6206	93,4025	92,6424	2,20E+11	47011,7	2,20E+11	5,30E+06
3819	FIN	FI	172	Finland	Europe & High inco	2007	61,3723	0,31564	83,1869		22	85	1,24	6,77944	0	24,5806	24,4946	95,8653	94,9684	2,60E+11	49239,2	2,60E+11	5,30E+06
3820	FIN	FI	172	Finland	Europe & High inco	2008	54,3939	0,33804	73,6359		22	84	1,04	6,53059	0	23,7136	-23,834	99,149	98,8297	2,80E+11	49363,7	2,80E+11	5,30E+06
3821	FIN	FI	172	Finland	Europe & High inco	2009	63,0305	0,34396	86,3896		22	82	0,72	5,93902	0	23,4132	-32,936	98,6017	98,8297	2,50E+11	45065,8	2,50E+11	5,30E+06

# Rodzaje danych panelowych

- **Mikropanel** - dane zebrane z obserwacji pojedynczych jednostek (osób, gospodarstw domowych, przedsiębiorstw, etc.)
- **Makropanel** - dane, w których jednostką obserwacji jest pewna zbiorowość (państwo, sektor gospodarki, jednostka samorządowa, etc.)
- **Przykłady**
  - Badanie Aktywności Ekonomicznej Ludności (BAEL)
  - Panel Polskich Przedsiębiorców
  - Informacje o wzroście PKB różnych krajów w kolejnych latach
  - Codzienne informacje o liczbie zgonów z powodu koronawirusa w różnych krajach
  - Bazy danych firmy ubezpieczeniowych o swoich klientach

# Dane Panelowe - ZALETY

- Obserwowanie różnorodności poszczególnych jednostek oraz zmian na przestrzeni czasu (badania społeczne, epidemiologia/ medycyna, analizy zachowań klientów – ubezpieczenia, banki)
- Identyfikacja i pomiar efektów, których po prostu nie można wykryć używając danych przekrojowych lub szeregów czasowych
- Dostarczają więcej informacji z danych dzięki czemu pozwalają na bardziej efektywne (dokładniejsze) oszacowania parametrów (np. redukcja obciążeń oszacowań wynikająca z agregacji danych indywidualnych)
- Umożliwiają zwiększenie mocy statystycznej mimo małego zbioru (w krajach UE dysponujemy jedynie 27 obserwacjami, ale kilka lat dla każdego z krajów wielokrotnia wielkość próby)
- Używając danych panelowych możemy (czasem) rozwiązać problem zmiennej pominiętej



# Dane Panelowe - PROBLEMY

- Nie każde dane zawierające informacje o pewnej zbiorowości w kolejnych okresach są danymi panelowymi. **(Obserwowane muszą być te same jednostki)**
- Zbieranie danych mikropanelowych jest zwykle bardzo kosztowne i kłopotliwe (odmowy kontynuacji uczestnictwa w badaniu, trudność z dotarciem do jednostek badania).
- **Dane niezbilansowane – nie posiadamy wszystkich wartości zmiennych w czasie**

**JAK SZACOWAĆ MODEL, GDY DANE  
PANELOWE?**

# Model z efektami stałymi (Fixed effect)

- Najprostszym założeniem jest występowanie stałych, nieznanymi (nieobserwowalnych), ale **stałych w czasie** różnic pomiędzy jednostkami:

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

$\alpha_i$  - jest stałym w czasie efektem indywidualnym dla obserwacji  $i$ .

- Kiedy stosować?

→ **jednostki się różnią**

(nieznane (nieobserwowalne) różnice są stałe w czasie)

→ Aby zbadać czy efekty indywidualne są istotne można wykorzystać **test Breuscha-Pagana**

# Panel fixed effects

$$\begin{aligned} &GDP\ growth_{it} \\ &= \beta_0 i + \beta_1 Depth_{it} + \beta_2 Bank_{it} + \beta_3 Privy_{it} + \beta_4 Assets_{it} \\ &+ \varepsilon_{it} \end{aligned}$$

- **Wyraz wolny (stała) różni się pomiędzy krajami**
- ... nachylenie jest stałe w czasie i takie same dla wszystkich krajów ...
- ... w błąd wpychamy wszystkie **pozostałe** różnice w czasie i między krajami

# Co w praktyce zakładamy?

$$\begin{aligned} &GDP\ growth_{it} \\ &= \beta_0 + v_i + \beta_1 Depth_{it} + \beta_2 Bank_{it} + \beta_3 Privy_{it} \\ &+ \beta_4 Assets_{it} + \varepsilon_{it} \end{aligned}$$

- **Nowy komponent  $v_i$  to element, który jest różny dla krajów.**
- .. do tego jest stary dobry (wspólny dla wszystkich) wyraz wolny
- ... i wspólne dla wszystkich nachylenie oraz zmienność w czasie
- ... a wszystko co nie pasuje, upychane jest w błąd

# Fixed effect – zadanie R

- Używając jako zmiennej objaśnianej GDP\_growth oraz zmiennych objaśniających wskaźników finansowych gfdddi01, gfdddi02, gfdddi04, gfdddi05 stwórz model *fixed effect (within estimator)*
- **R:** `plm(GDP_growth~gfdddi01 + gfdddi02 + gfdddi04 + gfdddi05, data=pdata, model="within")`
- **STATA:** `xtreg $ylist $xlist, fe`

# Fixed effect – zadanie R

```
> #FIXED EFFECT OR WITHIN ESTIMATOR
> fixed <-plm(GDP_growth~gfdddi01 + gfdddi02 + gfdddi04 + gfdddi05,data=pdata,model="within")
> summary(fixed)
Oneway (individual) effect within Model

Call:
plm(formula = GDP_growth ~ gfdddi01 + gfdddi02 + gfdddi04 + gfdddi05,
     data = pdata, model = "within")

Unbalanced Panel: n = 36, T = 14-38, N = 1174

Residuals:
    Min.    1st Qu.    Median     3rd Qu.     Max.
-18.07221  -1.17821    0.28778    1.58118    20.46692

Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
gfdddi01 -0.0099786   0.0128449  -0.7769   0.4374
gfdddi02 -0.0072959   0.0127133  -0.5739   0.5662
gfdddi04  0.0142661   0.0111940   1.2744   0.2028
gfdddi05 -0.0091153   0.0021799  -4.1816 3.118e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total sum of squares:    9607.8
Residual sum of squares: 9060
R-Squared:              0.057011
Adj. R-Squared:        0.02458
F-statistic: 17.1396 on 4 and 1134 DF, p-value: 1.1699e-13
< |
```

# Model z efektami losowymi (Random effects)

- każdej jednostce przypisujemy pewną zmienną losową, której realizacja odpowiada za efekt indywidualny w danym okresie
  - efekty indywidualne nie są jednakowe w kolejnych okresach
    - zakładamy brak korelacji między efektami indywidualnymi  $\alpha_i$  a zmiennymi objaśniającymi  $X_{it}$
  - nie traktujemy efektów indywidualnych jak parametrów i nie szacujemy ich wartości

$$y_{it} = \beta_0 + X_{it}\beta + \vartheta_{i,t}$$

$\vartheta_{i,t} = \alpha_i + \varepsilon_{it}$  - zmienna losowa stanowiąca sumę indywidualnego składnika losowego  $\alpha_i$  i białego szumu  $\varepsilon_{it}$



# Random effect– zadanie R

- Używając jako zmiennej objaśnianej GDP\_growth oraz zmiennych objaśniających wskaźników finansowych gfdddi01, gfdddi02, gfdddi04, gfdddi05 stwórz model *random effect*
- **R:** `plm(GDP_growth~gfdddi01 + gfdddi02 + gfdddi04 + gfdddi05, data=pdata, model=" random ")`
- **STATA:** `xtreg $ylist $xlist, re`

# Random effect – zadanie R

```
> #RANDOM EFFECTS ESTIMATOR
> random<- plm(GDP_growth~gfdddi01 + gfdddi02 + gfdddi04 + gfdddi05,data=pdata,model="random")
> summary(random)
Oneway (individual) effect Random Effect Model
(Swamy-Arora's transformation)

Call:
plm(formula = GDP_growth ~ gfdddi01 + gfdddi02 + gfdddi04 + gfdddi05,
     data = pdata, model = "random")

Unbalanced Panel: n = 36, T = 14-38, N = 1174

Effects:
              var std.dev share
idiosyncratic 7.989   2.827 0.883
individual     1.062   1.031 0.117
theta:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.4088 0.5843  0.5935  0.5721 0.5935  0.5935

Residuals:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-17.5766 -1.1861  0.1671 -0.0088  1.4781  21.1826

Coefficients:
              Estimate Std. Error z-value Pr(>|z|)
(Intercept)  2.8095635  0.9627313  2.9183 0.003519 **
gfdddi01     -0.0026179  0.0107189 -0.2442 0.807055
gfdddi02     -0.0158181  0.0103779 -1.5242 0.127457
gfdddi04      0.0204707  0.0105527  1.9399 0.052398 .
gfdddi05     -0.0049180  0.0017812 -2.7611 0.005761 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total sum of squares:    10014
Residual sum of squares: 9401.1
R-Squared:                0.061253
Adj. R-Squared:          0.058041
Chisq: 76.2299 on 4 DF, p-value: 1.0945e-15
> |
```

# Model z efektami losowymi (Random effect)

- Założenia dla wszystkich **jednostek (krajów)  $i$** :
  - $\alpha_i$  - *iid* (niezależne o tym samym rozkładzie)
  - Brak zależności między  $X_{it}$  i  $\alpha_i$  (efekty indywidualne)
  - Brak zależności między  $X_{it}$  i  $\varepsilon_{it}$  (składnik losowy)
  - Brak zależności między  $\varepsilon_{it}$  (składnik losowy) i  $\alpha_i$  (efekty indywidualne)
  
- Problem do przemyślenia:
  - W jakiej sytuacji estymator RE będzie taki sam jak estymator FE?
  - Który estymator i w jakich warunkach jest bardziej efektywny?

# Modele random effect vs. fixed effect

## Zalety

efektywny, jeśli efekt indywidualny nieskorelowany ze zmiennymi objaśniającymi

zgodny nawet wtedy, gdy efekt indywidualny skorelowany ze zmiennymi objaśniającymi

## Wady

nie jest zgodny, jeśli efekt indywidualny skorelowany ze zmiennymi objaśniającymi

nieefektywny, jeśli efekt indywidualny nieskorelowany ze zmiennymi objaśniającymi

nie można modelować cech niezmiennych w czasie  
(np. geografia a handel, albo wykształcenie a zarobki)

# Fixed effect vs. Random effect

## Ogólne zasady

1. Przy dużych  $t$  i małych  $i \rightarrow$  FE
2. Przy dużych  $i$  i małych  $t$ 
  - Jeżeli grupy przekrojowe są losową próbą populacji (tzn. indywidualne efekty są losowe)  $\rightarrow$  RE
  - w pozostałych przypadkach FE
3. Jeżeli komponent  $v_i$  jest skorelowany z  $x \rightarrow$  FE  
(oszacowania RE są obciążone)

# Fixed effect vs. Random effect

- ❑ Który z modeli lepszy? **Test Hausmana**
- ❑ Test weryfikujący czy oszacowania uzyskane za pomocą **FE** są statystycznie (istotnie) różne niż oszacowania **RE**
- ❑ **Koncepcja testu** polega na założeniu, że jeżeli w modelu **nie występuje korelacja pomiędzy efektami indywidualnymi a zmiennymi objaśniającymi**, to estymatory obu metod są zgodne, lecz estymator modelu z efektami stałymi jest nieefektywny.

# Test Hausmana

<i>Hipoteza</i>	<i>Założenie</i>	<i>Estymator efektów stałych</i>	<i>Estymator efektów losowych</i>
$H_0:$	$Cov(u_i, X_i) = 0$	<b>zgodny, nieefektywny</b>	<b>zgodny, efektywny</b>
$H_1:$	$Cov(u_i, X_i) \neq 0$	<b>zgodny</b>	<b>niezgodny</b>

- $H_0$ : indywidualne efekty są losowe → oszacowania **RE** są efektywne i zgodne (oszacowania **FE** są nieefektywne)
- Jeżeli parametry oszacowane za pomocą RE są statystycznie różne niż te oszacowane za pomocą FE → odrzucam  $H_0$

# Test Hausmana – zadanie w R

- Wykorzystując test Hausmana sprawdź, który model (FE vs RE) jest lepszy?

R: `phtest(fixed,random)`

STATA: `hausman fixed random`

```
> phtest(fixed,random)

      Hausman Test

data:  GDP_growth ~ gfdddi01 + gfdddi02 + gfdddi04 + gfdddi05
chisq = 13.603, df = 4, p-value = 0.008677
alternative hypothesis: one model is inconsistent
```



# Pierwsze różnice (First Difference)

$$y_{it} = \alpha_i + X_{it}\beta + \varepsilon_{it}$$
$$y_{it-1} = \alpha_i + X_{it-1}\beta + \varepsilon_{it-1}$$

Odejmujemy równania od siebie:

$$\Delta y_{it} = y_{it} - y_{it-1} = \Delta X_{it}\beta + \Delta \varepsilon_{it}$$

- Tracimy  $\alpha_i \rightarrow$  pomijamy *efekty* charakterystyczne dla poszczególnych jednostek (krajów)
- Metoda używana przy problemie zmiennych pominiętych
- UWAGA: Jeśli  $E[\varepsilon_{it} - \varepsilon_{it-1} | x_{it} - x_{it-1}] = \mathbf{0}$  wtedy estymatory są nieobciążone i zgodne.
- Jeżeli  $T=2 \rightarrow$  **FD** = **FE**

# First Difference – zadanie w R

- Używając jako zmiennej objaśnianej GDP\_growth oraz zmiennych objaśniających wskaźników finansowych gfdddi01, gfdddi02, gfdddi04, gfdddi05 stwórz model *first difference*
- **R:** `plm(GDP_growth~gfdddi01 + gfdddi02 + gfdddi04 + gfdddi05, data=pdata, model=" fd")`
- **STATA:** `reg D.($ylist $xlist)`

# First Difference – zadanie w R

```
> firstdiff<-plm(GDP_growth~gfdddi01 + gfdddi02 + gfdddi04 + gfdddi05,data=pdata,model="fd")
> summary(firstdiff)
Oneway (individual) effect First-Difference Model

Call:
plm(formula = GDP_growth ~ gfdddi01 + gfdddi02 + gfdddi04 + gfdddi05,
     data = pdata, model = "fd")

Unbalanced Panel: n = 36, T = 14-38, N = 1174
Observations used in estimation: 1138

Residuals:
    Min.    1st Qu.    Median    3rd Qu.    Max.
-22.147649  -1.427742   0.050671   1.449954   16.943494

Coefficients:
              Estimate Std. Error t-value Pr(>|t|)
(Intercept)  0.1592297  0.0956817  1.6642  0.09636 .
gfdddi01    -0.0170390  0.0405172  -0.4205  0.67417
gfdddi02    -0.0652540  0.0386016  -1.6904  0.09122 .
gfdddi04     0.0053693  0.0472826  0.1136  0.90961
gfdddi05    -0.0194720  0.0049600  -3.9258 9.164e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    11978
Residual Sum of Squares: 11321
R-Squared:                0.054823
Adj. R-Squared:          0.051486
F-statistic: 16.4294 on 4 and 1133 DF, p-value: 4.3055e-13
> |
```

# First Difference vs. Fixed Effect

- Jeżeli  $T=2 \rightarrow \text{FD} = \text{FE}$
- Jeżeli  $T>3$  ???
- Jeżeli  $\text{cov}(\varepsilon_{it}, \varepsilon_{it-1}) = 0$ ,  $\varepsilon_{it}$  - iid  $\rightarrow \text{FE}$

Ogólna zasada:

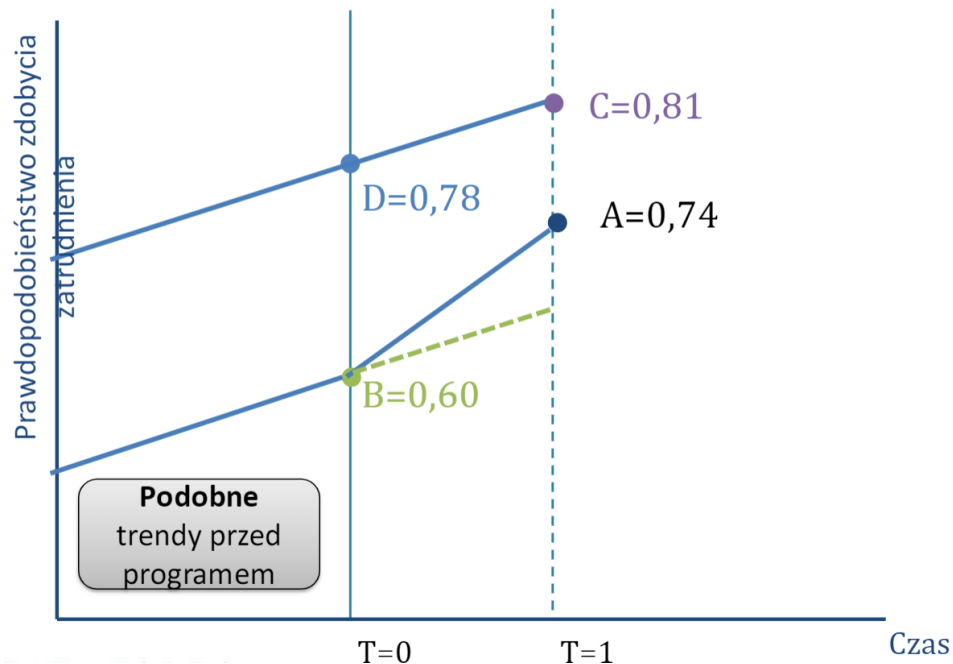
**Fixed effects**  $\rightarrow$  jeśli mamy *niezależne idiosynkratyczne błędy*  $\varepsilon_{it}$  o *tym samym rozkładzie*

**First differences**  $\rightarrow$  jeśli idiosynkratyczne błędy  $\varepsilon_{it}$  mają własności procesu błędzenia losowego

# Metoda różnicy w różnicach (Difference in Difference)

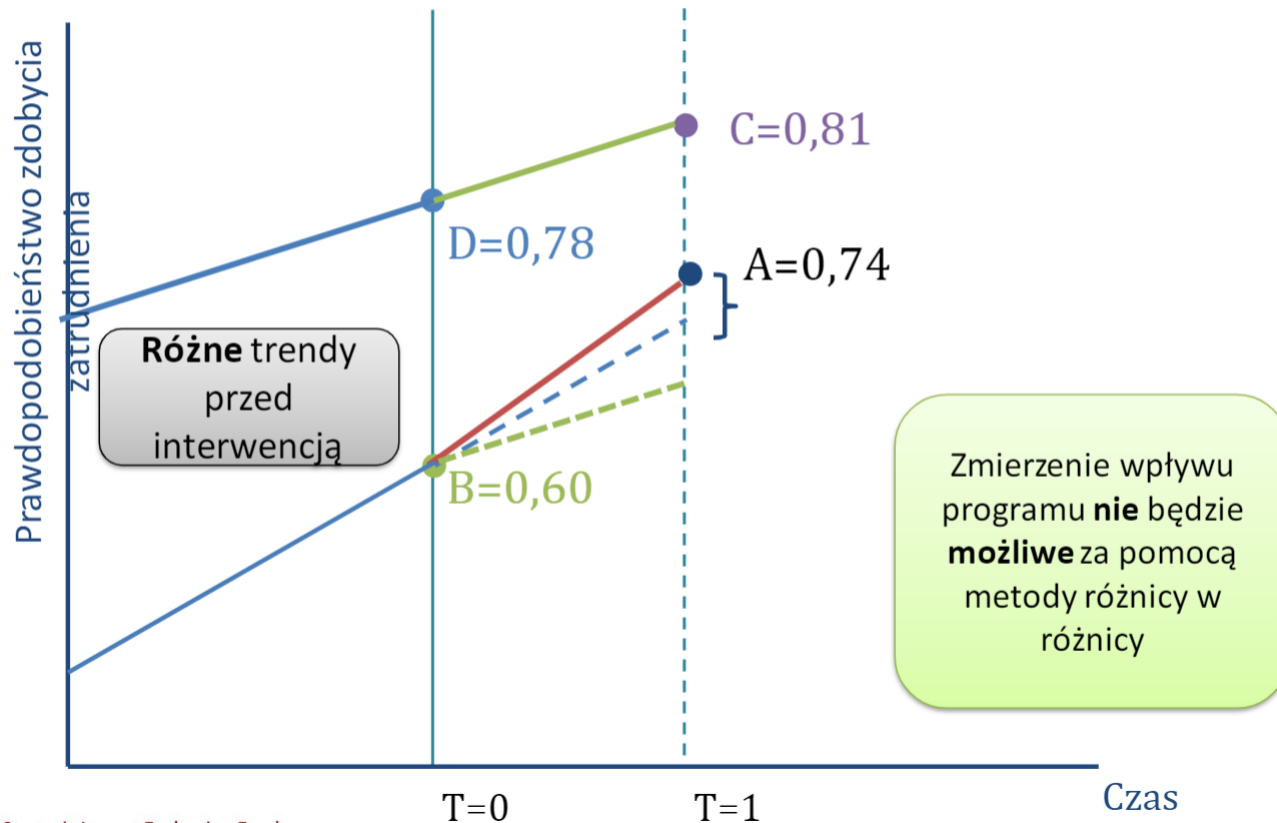
- Takie same trendy przed programem
- Zmiany w zaobserwowanych wynikach wśród osób nieprzyjętych byłyby takie same wśród osób przyjętych.

Równe trendy



# Metoda różnicy w różnicach (Difference in Difference)

## Różne trendy



# Metoda różnicy w różnicach (Difference in Difference)

- Dwie zmienne:  $s$  (np. kraj zamieszkania),  $t$  (rok)

$$y_{ist} = \beta_0 + \beta_1 s + \beta_2 t + \beta_3 s * t + \varepsilon_{ist}$$

- $y_{ist}$  - poziom zmiennej zależnej dla osoby  $i$  warunkowo względem  $s$  i  $t$
- $\beta_1$  i  $\beta_2$  oszacowują więc różnice związane tylko z upływem czasu, lub innym miejscem zamieszkania
- Współczynnik  $\beta_3$  oszacowuje **różnice w różnicach** – wielkość związaną jednocześnie z położeniem i upływem czasu
  
- badanie Carda i Krugera (1994)

# Jak wybrać wskaźniki?

- ❑ Global Financial Development Database podaje wartości 115 wskaźników finansowych - **Które wybrać?**
  - W oparciu o wcześniejsze badania
  - Wiedza ekspercka/ dziedzinowa
  - Metoda Hellwiga (dla chętnych)



# Jakie wskaźniki można wykorzystać?

□ King and Levine (1993)

$$\text{Depth} = \frac{\text{liquid liabilities of the financial system}}{\text{GDP}}$$

$$\text{Bank} = \frac{\text{bank credit}}{(\text{bank credit} + \text{central bank domestic assets})}$$

$$\text{Privy} = \frac{\text{credit to private enterprises}}{\text{GDP}}$$

$$\text{Private} = \frac{\text{claims on the nonfinancial private sector}}{\text{total domestic credit (excluding credit to money banks)}}$$

# Jakie wskaźniki można wykorzystać?

## □ Levine and Zervos (1998)

$$\text{Size – Capitalization} = \frac{\text{value of listed domestic shares on domestic exchanges}}{\text{GDP}}$$

$$\text{Turnover} = \frac{\text{value of the trades of domestic shares on domestic exchanges}}{\text{value of listed domestic shares}}$$

$$\text{Value Traded} = \frac{\text{value of the trades of domestic shares on domestic exchanges}}{\text{GDP}}$$

$$\text{Banking Develop.} = \frac{\text{value of loans (commercial banks and other) to private sec.}}{\text{GDP}}$$

$$\text{Financial Depth} = \frac{\text{stock of broad money (M2)}}{\text{GDP}}$$

# Dla chętnych – metoda Hellwiga

# Jak wybrać wskaźniki? – Metoda Hellwiga

- ❑ Zdzisław Hellwig (1925-2013) - *Metoda optymalnego wyboru predyktant (metoda wskaźników pojemności informacji)*
- ❑ Idea:
  - ❑ Zmienne które wybieramy powinny być **silnie skorelowane ze zmienną objaśnianą, a słabo skorelowane między sobą**
  - ❑ Mając  $m$  potencjalnych zmiennych objaśniających, to liczba wszystkich kombinacji jest równa
$$L = 2^m - 1$$

# Jak wybrać wskaźniki? – Metoda Hellwiga

- Indywidualna pojemność nośników informacji -  $h_{kj}$  - mierzy **wielkość informacji** jaką wnosi zmienna  $X_j$  o zmiennej objaśnianej  $Y$  w  $k$ -tej kombinacji

$$h_{kj} = \frac{r_j^2}{1 + \sum_{l=1, l \neq j}^{m_k} |r_{lj}|}, j = 1, 2, \dots, m_k, h_{kj} \in [0, 1]$$

- $k$ -numer kombinacji ( $k = 1, 2, \dots, 2^m - 1$ )
- $m_k$ -liczba zmiennych w  $k$ -tej kombinacji
- $j$ -numer zmiennej w rozpatrywanej kombinacji
- $r_j$ -współczynnik korelacji  $j$  zmiennej objaśniającej ze zmienną objaśnianą
- $r_{lj}$ -współczynnik korelacji między  $l$ -tą i  $j$ -tą zmienną objaśniającą

# Jak wybrać wskaźniki? – Metoda Hellwiga

- Pojemność integralna kombinacji nośników informacji

$$H_k = \sum_{j=1}^{m_k} h_{kj}, k = 1, 2, \dots, 2^m - 1, H_k \in [0, 1]$$

- $k$ -numer kombinacji ( $k = 1, 2, \dots, 2^m - 1$ )
- $m_k$ -liczba zmiennych w  $k$ -tej kombinacji
- $j$ -numer zmiennej w rozpatrywanej kombinacji

- **Wybieramy model z najwyższym  $H_k$**

# Podsumowanie

- ❑ Analiza danych panelowych
  - ❑ Co to są dane panelowe?
- ❑ Jak dobrać dane do modelu ? (metoda Hellwiga)
- ❑ Wybrane metody analizy danych panelowych
  - Pooled OLS
  - Fixed Effect
  - Random Effect
  - First Difference
  - Difference in Difference

# Dziękuję za Państwa czas!



Sylwia Radomska

s.radomska@grape.org.pl

<http://grape.org.pl/sradomska>