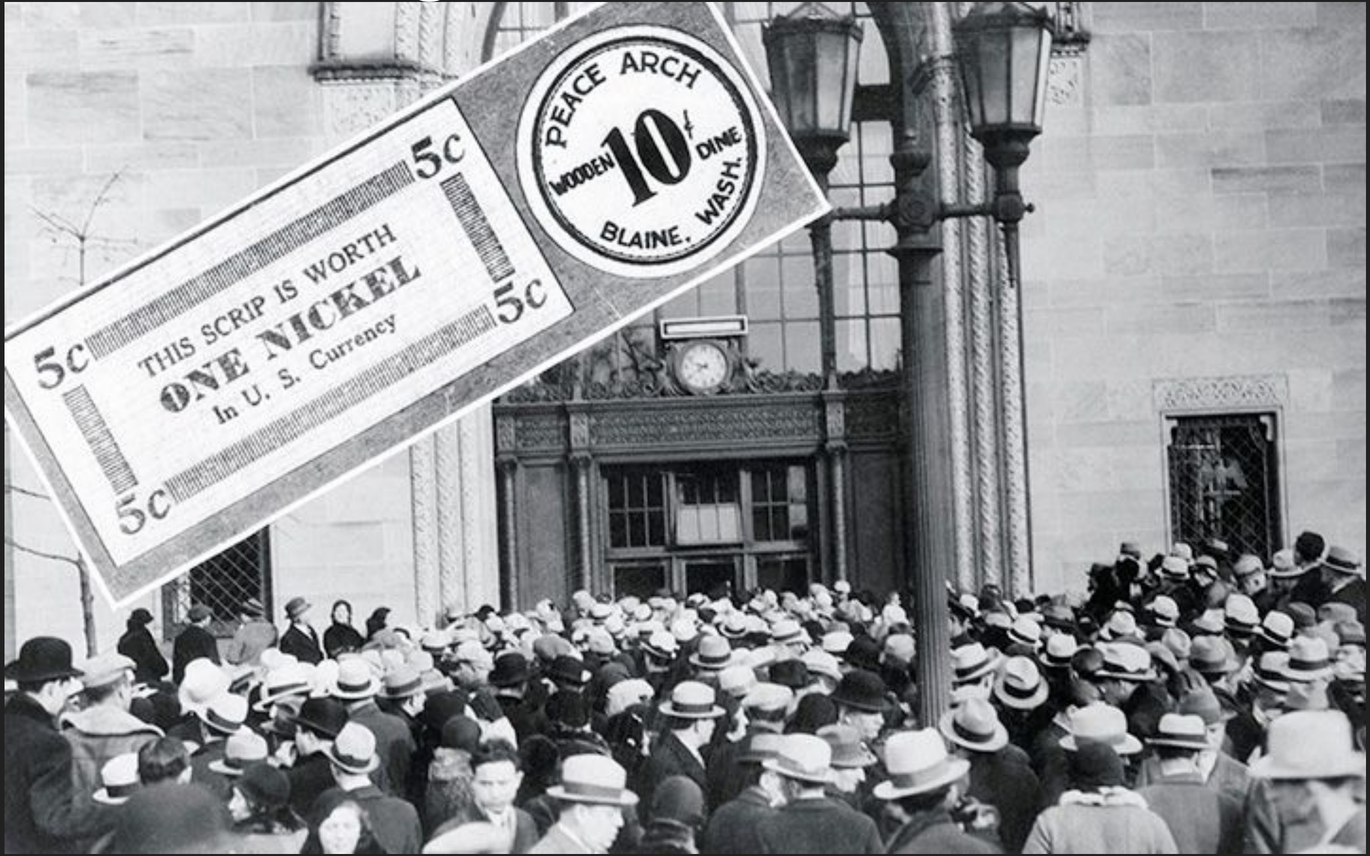
Diff-in-diff and fixed effects

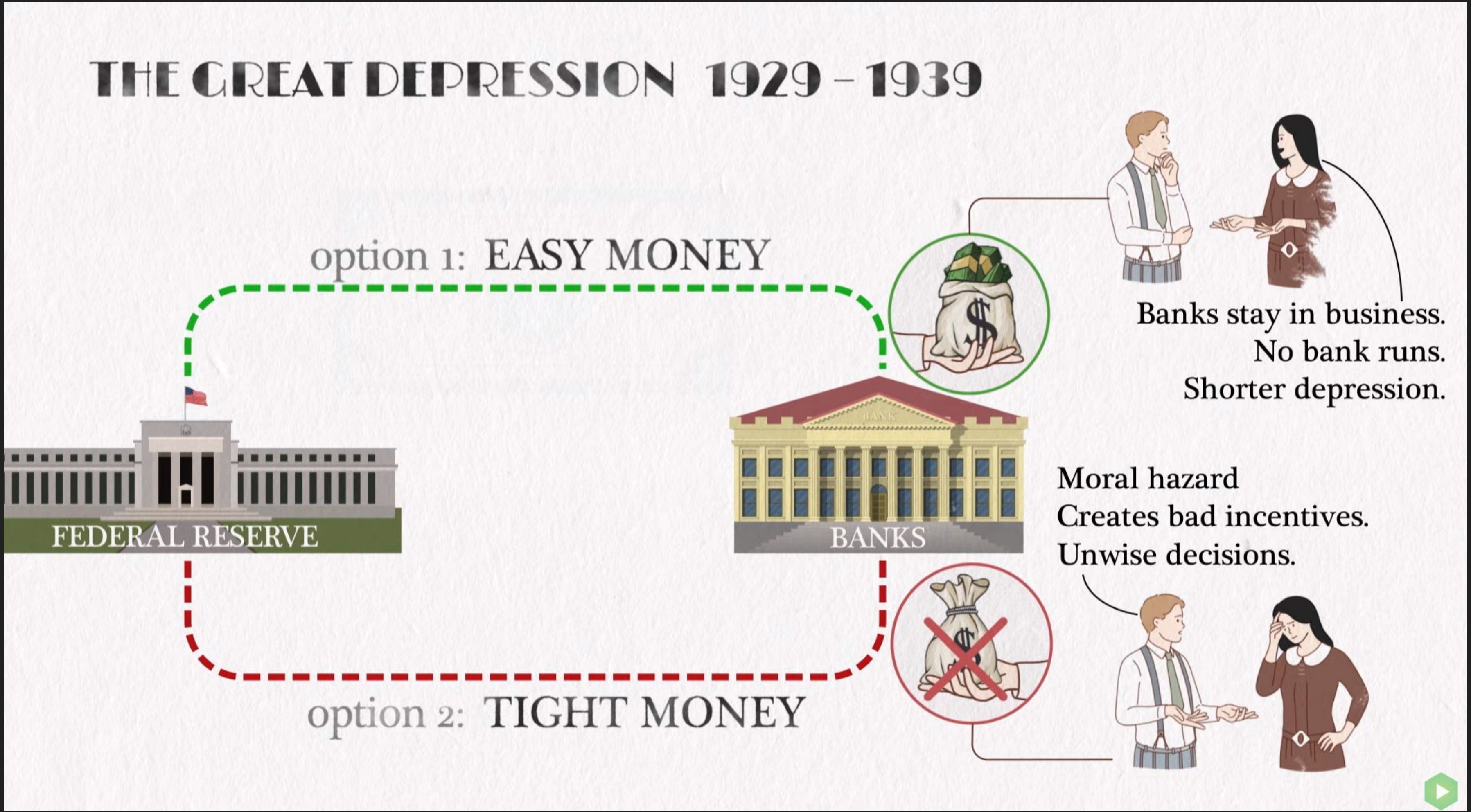
IN4143: Data Analysis and Causal Inference

Quiz about the video: “Introduction to differences in differences”

(U-cursos)

Banking Panics of 1931-33

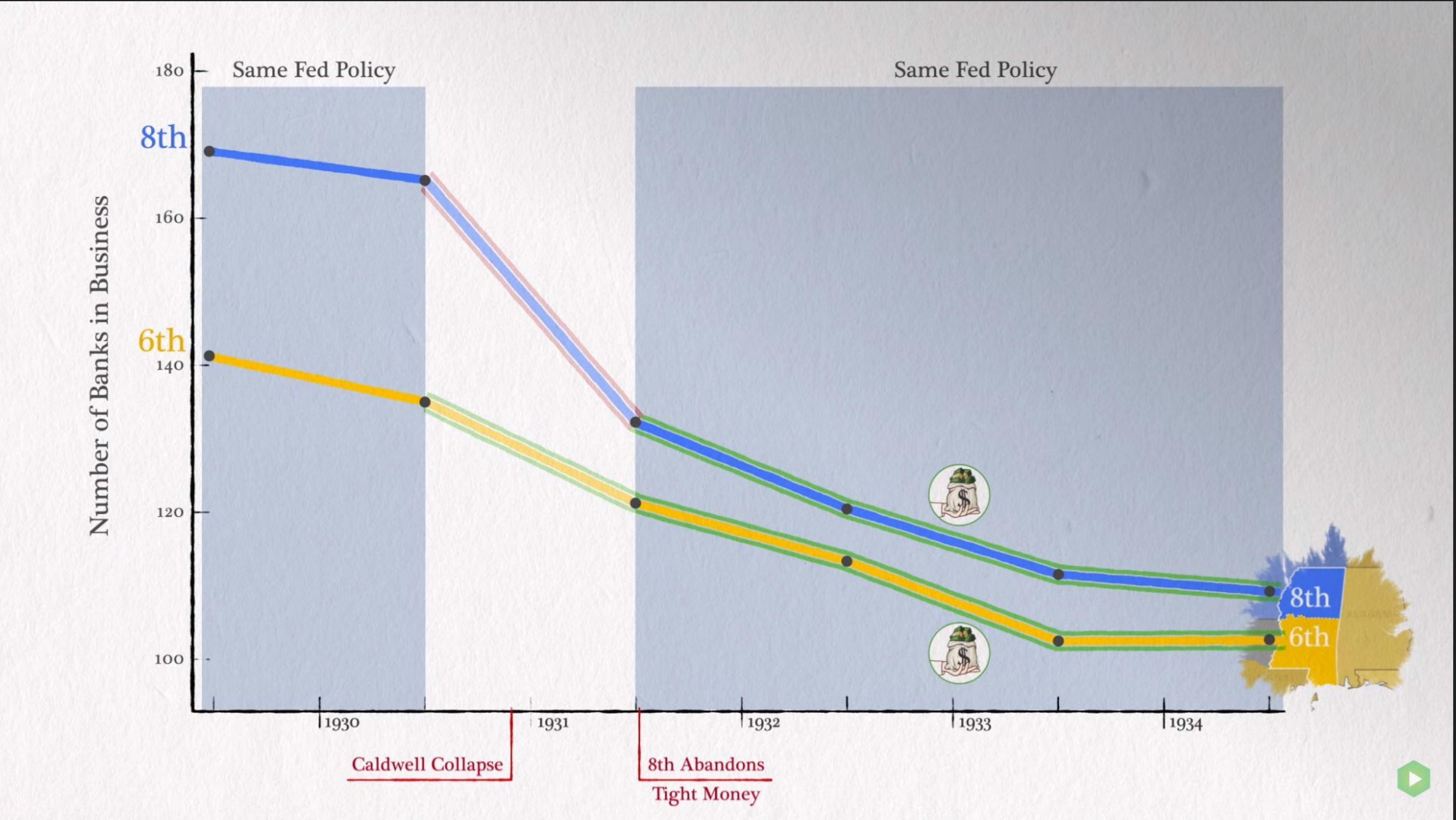




Source: Video Introduction to Differences-In-Differences (Marginal Revolution University Youtube Channel)

Source: Video Introduction to Differences-In-Differences (Marginal Revolution University Youtube Channel)
)

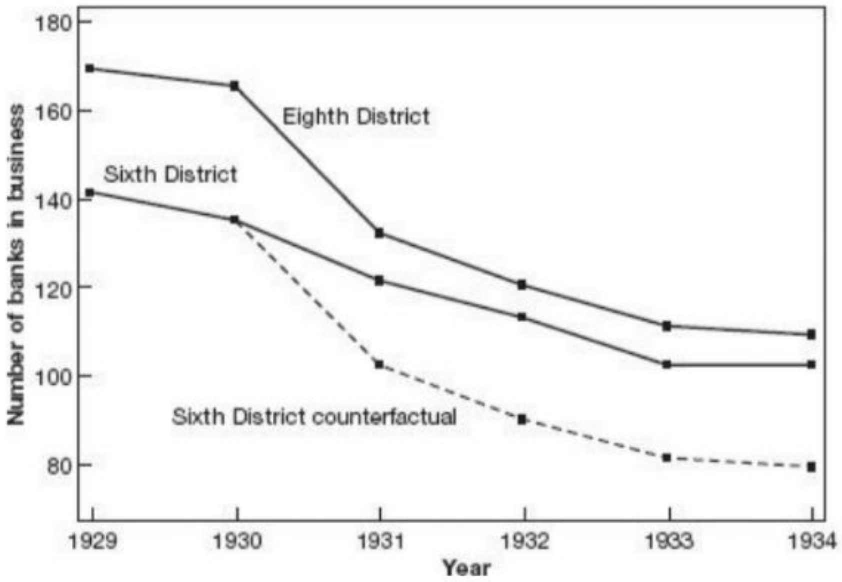
Source: Angrist and Pischke 2014



Source: Video Introduction to Differences-In-Differences (Marginal Revolution University Youtube Channel)

FIGURE 5.3

Trends in bank failures in the Sixth and Eighth Federal Reserve Districts, and the Sixth District's DD counterfactual



Notes: This figure adds DD counterfactual outcomes to the banking data plotted in Figure 5.2. The dashed line depicts the counterfactual evolution of the number of banks in the Sixth District if the same number of banks had failed in that district after 1930 as did in the Eighth.

Source: Angrist and Pischke 2014

Econometric model

$$Y_{dt} = \beta\, TREAT_d + \gamma\, POST_t + \square_{rDD}\, (TREAT_d \times POST_t) + e_{dt}$$

**d stands for district d*

Econometric model

$$Y_{dt} = \beta\ TREAT_d + \gamma\ POST_t + \square_{rDD}\ (TREAT_d \times POST_t) + e_{dt}$$

$$\boxed{Y_{dt} = 167 - \underset{(8.8)}{29}\ TREAT_d + \underset{(7.6)}{49}\ POST_t + \underset{(10.7)}{20.5}\ (TREAT_d \times POST_t) + e_{dt}}$$

**d stands for district d*

Econometric model

$$Y_{dt} = 167 - \underset{(8.8)}{29}\ TREAT_d + \underset{(7.6)}{49}\ POST_t + \underset{(10.7)}{20.5}\ (TREAT_d \times POST_t) + e_{dt}$$

1. What is the difference in the initial period between treated and untreated distr.?
2. What is the base assumption of diff-in-diff? Does it require pre-treatment balance?
3. Do they differ significantly in the dependent variable before treatment?
4. Does the control group change the number of operating banks between the pre and post-treatment periods?
5. What can you conclude about the easy money policy? How many banks were saved because of the easy money policy?

**peer tutoring*

R code

```
> summary(lm(Y ~ TREAT*POST, df))

Call:
lm(formula = Y ~ TREAT * POST, data = df)

Residuals:
   Min     1Q Median     3Q    Max
  -1.0   -0.5    0.0    0.5    1.0

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.000e+00  5.270e-01   3.795 0.009023 **
TREAT       5.000e+00  7.454e-01   6.708 0.000533 ***
POST        2.500e+00  8.333e-01   3.000 0.024008 *
TREAT:POST  4.300e-16  1.179e+00   0.000 1.000000
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9129 on 6 degrees of freedom
Multiple R-squared:  0.9394,    Adjusted R-squared:  0.9091
F-statistic:    31 on 3 and 6 DF,  p-value: 0.0004758
```

Summary of diff-in-diff

- Diff-in-diff (DD) estimator is good to provide causal explanations because reduces unobservable differences between groups

- It needs to assume that growing trends are parallel
    - This can come from balanced or unbalanced groups
    - It can be assumed (by random assignment) or checked with data

- Compares times differences assuming a linear behavior

- Adding covariates help lowering standard errors

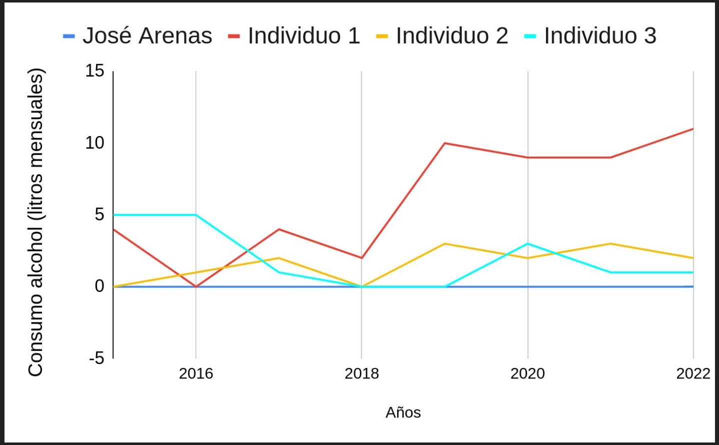
Fixed Effects Models

Panel Data

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

- We follow subject *i* for several *t* times
- Example:
  - y_{it} is a test result of the student *i* in the quarter *t*

- Sample can be **balanced** or **unbalanced**
- Puts some difficulties because errors (ε_{it}) change across subjects and across time periods.

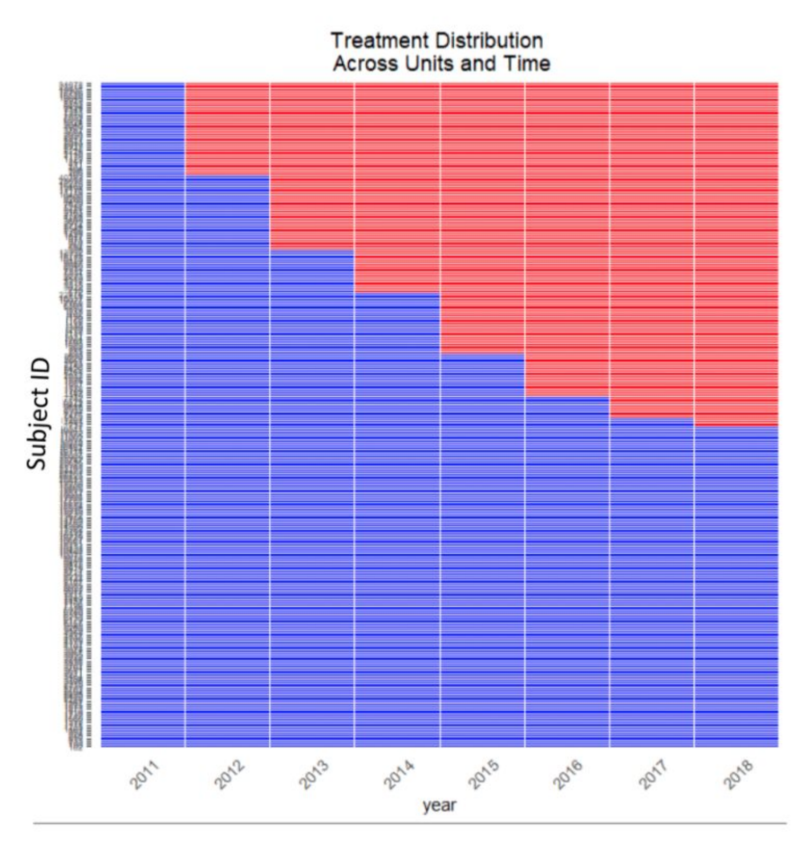


Panel Data

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

- We follow subject *i* for several *t* times
- Example:
  - y_{it} is a test result of the student *i* in the quarter *t*

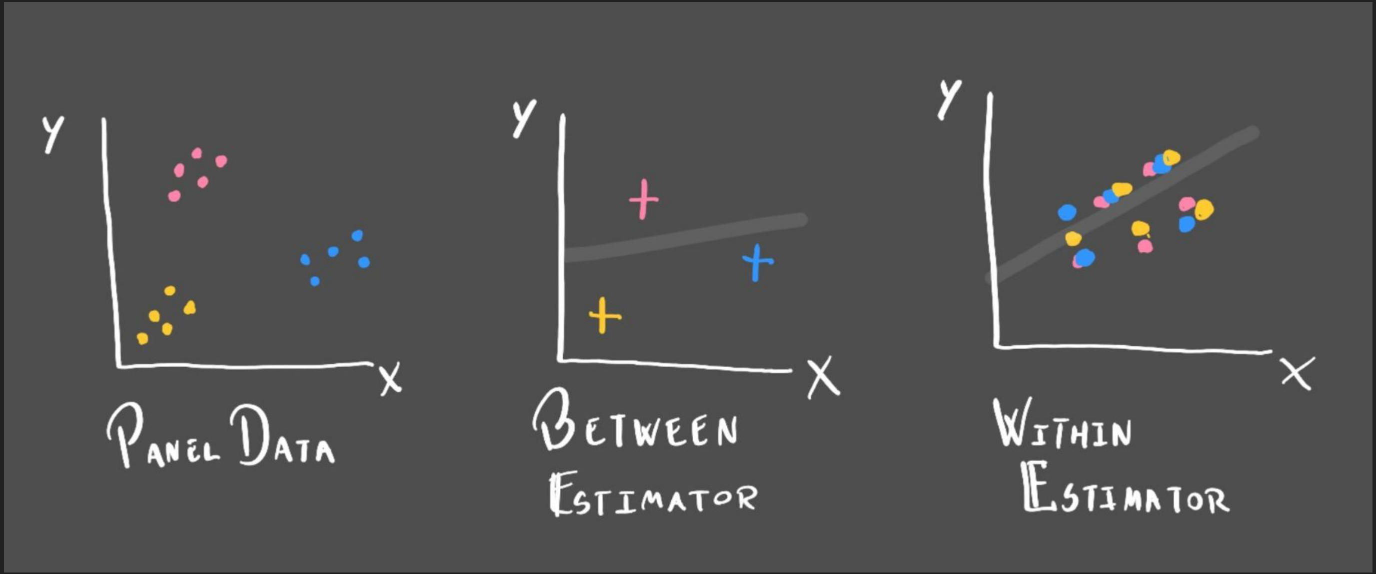
- Sample can be **balanced** or **unbalanced**
- Puts some difficulties because errors (ε_{it}) change across subjects and across time periods.



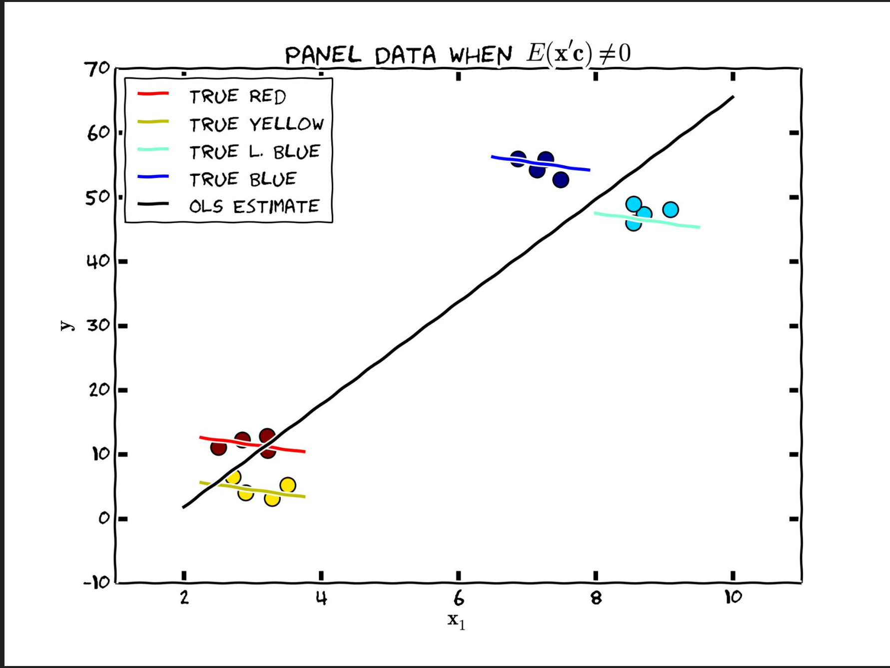
Panel data can use time (in)variance

- In panel data we have:
  - "**between**" information
    - How the variables change across subjects
    - i.e.: Comparing results from the student i with student j because they went to different classes
  - "**within**" information
    - How the variables change for a particular subject
    - i.e.: Comparing student i's results in time t with time t+1 because he attended classes
- Also we could compare *growth* (t and t+1) between subjects:
  - Does student *i* improved his results from *t* to *t+1* more than student *j*?

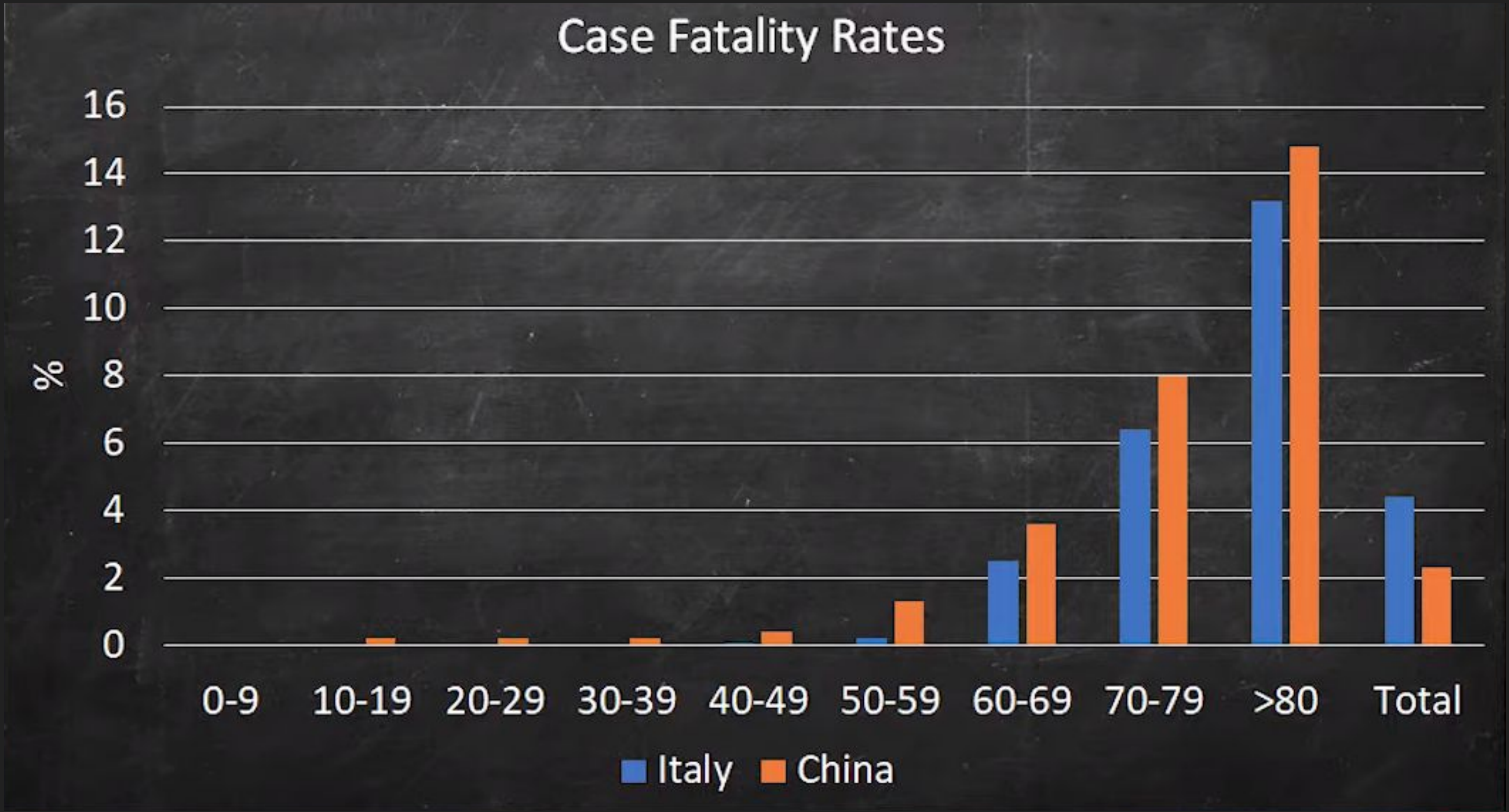
Within v/s Between



Within estimator



Source: Cross Section Econometrics Notes (Rob Hicks 2021)



Source: Video How Simpson's Paradox Explains Weird Covid19 Statistics (Dr. Trefor Bazett Youtube Channel)

Fixed effects or the within estimator

- **Main assumption** → What if we could “erase” the constant features of a student (i.e., gender, ethnicity, place) and focus only in the “growing” part

- Let’s think about the model:

$$y_{it} = \beta_0 + \beta_X X_{it} + \overbrace{\beta_i Z_{it} + \varepsilon_{it}}^{u_{it}}$$

$$y_{it} = \boldsymbol{\alpha_i} + \beta_X X_{it} + \varepsilon_{it}$$

“unobserved heterogeneity”

**what if we use OLS w/out accounting for this unobserved heterogeneity?*

Fixed effects or the within estimator

- **Demeaning** interpretation:

$$y_{it} = \boldsymbol{\alpha_i} + \beta_X X_{it} + \varepsilon_{it}$$

$$\Big\downarrow \text{ demeaning}$$

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

**interpretation → deviations from the mean*

R code

```r
# Demeaning data

df.demeaned <- with(df,
                     data.frame(Y = Y - ave(Y, id),
                                X = X - ave(X, id))
                    )

# OLS with deameaned variables
fixed_effects.demeaned_model = lm(Y ~ X - 1, data = df.demeaned)

# print summary using robust standard errors for the f.e. model
coeftest(fixed_effects.demeaned_model, vcov. = vcovHC, type = "HC1")
```

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

Remarks

1. F.E. estimator is equivalent to an OLS estimation on the demeaned variables model

2. All time-invariant unobserved variables are removed from the estimation

3. All time-invariant observed variables are also removed from the estimation (perfect collinearity)

4. Identification of the f.e. coefficients comes from temporal variation of Xs
   - Efficiency → Statistical precision of the coefficients is related to the temporal variance of X
   - Unbiased estimation → An unbiased estimation requires that temporal variations in X don't correlate with temporal variations in the error term

Fixed effects or the within estimator

- This model is equivalent to add a dummy variable for each individual subject *i* to the original model

$$y_{it} = \beta_0 + \beta_1 X_{it} + \square_1 D_{i=1} + \dots + \square_N D_{i=N} + \varepsilon_{it}$$

**interpretation → "fixed effects"*

R code

```r
library(plm)

# estimate the fixed effects regression with plm()
fixed_effects.model <- plm(Y ~ X,
                           data = df,
                           index = c("id", "month"),
                           model = "within")

# print summary using robust standard errors for the f.e. model
coeftest(fixed_effects.model, vcov. = vcovHC, type = "HC1")
```

$$y_{it} = \quad \beta_0 + \beta_1 X_{it} + \square_1 D_{i=1} + \ldots + \square_N D_{i=N} + \varepsilon_{it}$$

Demeaning v/s Dummy for each *i*

Model 1

OLS

Model 2

OLS demeaned variables

Model 3

f.e. regression

	Model 1	Model 2	Model 3
prbarr	0.05 **		-0.03 *
	(0.02)		(0.01)
demeaned_prob		-0.03 *	
		(0.01)	
N	27	27	27
R2	0.25	0.21	0.94
*** p < 0.001; ** p < 0.01; * p < 0.05.			

Source: Video Econometrics - Within Variation And Fixed Effects (Econometrics, Causality, and Coding with Dr. HK Youtube Channel)

Dirección Nacional del Servicio Civil → fortalecer función pública

Servicio Civil
Ministerio de Hacienda
Gobierno de Chile

Secciones biobiochile.cl

Jueves 16 enero de 2020 | Publicado a las 08:17

Uno de cada tres funcionarios públicos reconoce haber conseguido el empleo con "pitutos"

Por Luciano Veloso

Líbero TV Podcast Radio Alerta Líbero! Buscar

Publicado el 28 de diciembre, 2018

Presidente de Alta Dirección Pública: “Hoy todos los directores de servicio pasaron por un filtro de mérito»

Dirección Nacional del Servicio Civil → fortalecer función pública

- Generación de concursos públicos para los directores de instituciones
  - Sistema de Alta Dirección Pública

- ¿Qué ocurre en el tiempo entre que un directivo/a deja su cargo y se debe concursar otro? → Designación de un TyP (Provisional y Transitorio) por parte de la autoridad
  - Incentivos Perversos
    - Quitar a directivos para poder asignar a alguien de confianza
    - El TyP podía participar de los concursos con “información adicional” y experiencia en el cargo

- 2016 → Reforma que elimina TyP
  - ¿Cuál fue su efecto sobre # de postulantes?

Panel de datos

N° Concursos	N° Cargos	%	N° Datos
1	228	20.6	228
2	251	22.6	502
3	242	21.8	726
4	198	17.9	792
5	110	9.9	550
6	60	5.4	360
7	18	1.6	126
8	1	0.1	8
9	1	0.1	9
Total	1109	100%	3301

**desbalanceado*

- Variable dependiente:
  - Número de postulantes para el cargo *i* en el concurso *t*
- Ejemplo
  - El quinto concurso para director del SII convocado en mayo del 2015 tuvo 77 postulantes

Fuente: Fernández Vergara, A., Fuenzalida , J. & Castro, C. (2019). Impacto de sistemas de selección por mérito , el caso de Chile postreforma (ATN O/C 16465 - RG). Santiago: Latin American Laboratory of Probity and Transparency Policies, Inter - American Development Bank
http://www.sistemaspublicos.cl/publicacion/documentos-de-politica-el-caso-de-chile-post-reforma-a-la-alta-direccion-publica/

Resultados primer análisis (# postulantes vs TyP)

1. Efecto de los TyP en postulaciones (pre-reforma)
   - Los efectos fijos “limpian” el sesgo por variable omitida (suponemos, invariante en el tiempo)

Variable tratamiento	TYP en concurso		
VD: Núm. Post	Sin controles	Con controles	FE: con controles
Figura TYP	-1.186	4.362	-4.386
	(3.283)	(3.206)	(1.880)*
Constante	110.210	225.036	204.694
	(2.317)**	(161.966)	(92.822)*
Controles+	No	Sí	Sí
Efectos temporales	No	No	Si
Efectos fijos	No	No	Si
R2	0.00	0.09	0.37
N	1,971	1,971	1,971

* p<0.05; ** p<0.01 ; errores estándar entre paréntesis

+Controles: coalición de gobierno, estacionalidad por meses, notas obtenidas en los concursos

Fuente: Fernández Vergara, A., Fuenzalida , J. & Castro, C. (2019). Impacto de sistemas de selección por mérito , el caso de Chile postreforma (ATN O/C 16465 - RG). Santiago: Latin American Laboratory of Probity and Transparency Policies, Inter - American Development Bank
http://www.sistemaspublicos.cl/publicacion/documentos-de-politica-el-caso-de-chile-post-reforma-a-la-alta-direccion-publica/

Resultados segundo análisis (# postulantes vs Reforma)

2. Efecto de la reforma en # postulaciones (pre vs post reforma)

VD: Núm. Post	Sin controles	Con controles	FE: con controles
REFORMA	26.675	26.070	30.471
	(3.117)**	(7.936)**	(5.452)**
Constante	111.532	49.960	175.455
	(1.677)**	(132.207)	(93.964)
Controles+	No	Sí	Sí
Efectos temporales	No	No	Sí
Efectos fijos	No	No	Sí
R2	0.02	0.07	0.22
N	3,285	3,285	3,285

* p<0.05; ** p<0.01

+Controles: coalición de gobierno, estacionalidad por meses y desempleo mensual

Fuente: Fernández Vergara, A., Fuenzalida , J. & Castro, C. (2019). Impacto de sistemas de selección por mérito , el caso de Chile postreforma (ATN O/C 16465 - RG). Santiago: Latin American Laboratory of Probity and Transparency Policies, Inter - American Development Bank
http://www.sistemaspublicos.cl/publicacion/documentos-de-politica-el-caso-de-chile-post-reforma-a-la-alta-direccion-publica/

Conclusiones

- La designación de TyP reduce el número de postulantes (4 aprox. en promedio, por concurso)

- La reforma que eliminó los TyP es consistente con un aumento en el número de postulantes

- El efecto de la reforma es de 30 postulantes adicionales (aprox.), en promedio, por concurso

Any question?