

10) Difference-in-Differences

Vitor Kamada

August 2018

Tables, Graphics, and Figures from

**Mastering 'Metrics: The Path from Cause
to Effect**

Angrist & Pischke (2014): Chapter 5.1

“Are Training Subsidies Effective? The Michigan Experience,” *Industrial and Labor Relations Review* 46, 625-636.

$$\log(\text{scrap})_{it} = \beta_0 + \beta_1 \text{grant}_{it} + \alpha_i + u_{it}$$

scrap_{it} : scrap rate of firm i during year t

grant_{it} : 1 if received a job training in 1988

First-Difference Estimator

$$y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it} + \alpha_i + u_{it}$$

$$y_{i2} = (\beta_0 + \delta_0) + \beta_1 x_{i2} + \alpha_i + u_{i2} \quad [t=2]$$

$$y_{i1} = \beta_0 + \beta_1 x_{i1} + \alpha_i + u_{i1} \quad [t=1]$$

$$y_{i2} - y_{i1} = \delta_0 + \beta_1 (x_{i2} - x_{i1}) + u_{i2} - u_{i1}$$

$$\Delta y_i = \delta_0 + \beta_1 \Delta x_i + \Delta u_i$$

Load Data

```
import pandas as pd

import statsmodels.api as sm

file="C:\\Users\\Vitor\\Desktop\\ECO 6100 Introduction to
Econometrics (Fall 2018)\\Lectures\\10) Difference-in-Difference
(DiD)\\JTRAIN.DTA"

df = pd.read_stata(file)

df['const'] = 1

df = df[df['d88'] == 1]
```

OLS vs First-Difference

```
reg1 = sm.OLS(df['lscrap'], df[['const', 'grant']],  
              missing='drop').fit()  
print(reg1.summary())
```

	coef	std err	t	P> t
const	0.4085	0.241	1.698	0.095
grant	0.0566	0.406	0.140	0.890

```
reg2 = sm.OLS(df['clscrap'], df[['const', 'cgrant']],  
              missing='drop').fit()  
print(reg2.summary())
```

	coef	std err	t	P> t
const	-0.0574	0.097	-0.591	0.557
cgrant	-0.3171	0.164	-1.935	0.058

Great Depression

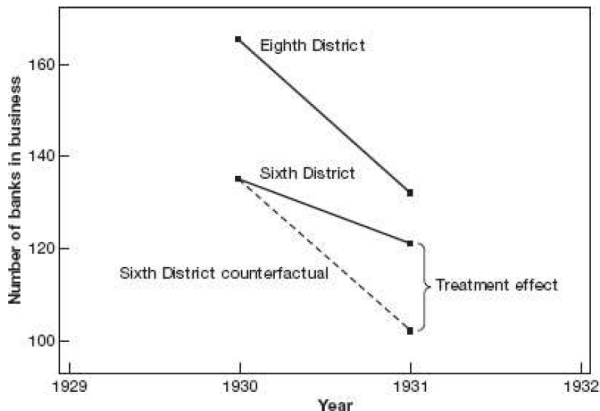
Crisis	Period	Worldwide GDP Fell
Great Depression	1929-1932	15%
Great Recession	2008-2009	1%

Fed	District	Lending (1929-1931)
Atlanta	Sixth	40%
St. Louis	Eighth	-10%

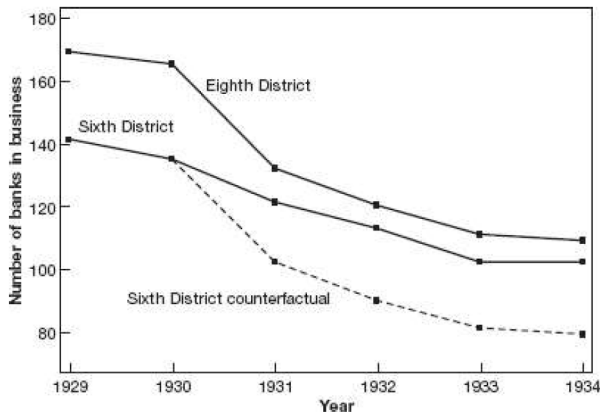
Confidence Crisis vs Real Crisis

Richardson and Troost (2009)

Monetary Intervention Mitigated Banking Panics during the Great Depression: Quasi- Experimental Evidence from a Federal Reserve District Border, 1929–1933. *Journal of Political Economy*, 2009, Vol. 117(6)



Manual Differences-in-Differences Estimator



$$\delta_{DD} = (Y_{6,1931} - Y_{6,1930}) - (Y_{8,1931} - Y_{8,1930})$$
$$(121 - 135) - (132 - 165) = -14 - (-33) = 19$$

Difference-in-Differences Estimator

$$y = \beta_0 + \delta_0 d2 + \beta_1 dT + \delta_1 d2 \cdot dT + u$$

$$\hat{\delta}_1 = (\bar{y}_{2,T} - \bar{y}_{2,C}) - (\bar{y}_{1,T} - \bar{y}_{1,C})$$

$$\hat{\delta}_1 = (\bar{y}_{2,T} - \bar{y}_{1,T}) - (\bar{y}_{2,C} - \bar{y}_{1,C})$$

	Before	After	After-Before
Control	β_0	$\beta_0 + \delta_0$	δ_0
Treatment	$\beta_0 + \beta_1$	$\beta_0 + \delta_0 + \beta_1 + \delta_1$	$\delta_0 + \delta_1$
Treatment-Control	β_1	$\beta_1 + \delta_1$	δ_1

$$\hat{y} = 167 - \underset{(7.6)}{49} d2 + \underset{(8.8)}{29} dT + \underset{(10.7)}{20.5} d2 \cdot dT$$

Wholesale Firm Failures and Sales in 1929 and 1933

	1929	1933	Difference (1933–1929)
Panel A. Number of wholesale firms			
Sixth Federal Reserve District (Atlanta)	783	641	–142
Eighth Federal Reserve District (St. Louis)	930	607	–323
Difference (Sixth–Eighth)	–147	34	181
Panel B. Net wholesale sales (\$ million)			
Sixth District Federal Reserve (Atlanta)	141	60	–81
Eighth District Federal Reserve (St. Louis)	245	83	–162
Difference (Sixth–Eighth)	–104	–23	81

Kiel and McClain (1995)

House Prices during Siting Decision Stages: The Case of an Incinerator from Rumor through Operation. *Journal of Environmental Economics and Management*, Vol 28(2), 241-255

```
import pandas as pd
```

```
import statsmodels.api as sm
```

```
file="C:\\Users\\Vitor\\Desktop\\ECO 6100 Introduction to  
Econometrics (Fall 2018)\\Lectures\\10) Difference-in-Difference  
(DiD)\\KIELMC.DTA"
```

```
df = pd.read_stata(file)
```

```
df['const'] = 1
```

```
df['rprice'] = df['rprice']/1000
```

$$\hat{\delta}_1 = -30,688 - (-18,824) = -11,863$$

```
df81 = df[df['year'] == 1981]
reg1 = sm.OLS(df81['rprice'], df81[['const', 'nearinc']],
              missing='drop').fit()
print(reg1.summary())
```

	coef	std err	t	P> t
const	101.3075	3.093	32.754	0.000
nearinc	-30.6883	5.828	-5.266	0.000

```
df78 = df[df['year'] == 1978]
reg2 = sm.OLS(df78['rprice'], df78[['const', 'nearinc']],
              missing='drop').fit()
print(reg2.summary())
```

	coef	std err	t	P> t
const	82.5172	2.654	31.094	0.000
nearinc	-18.8244	4.745	-3.968	0.000

$$rprice = \beta_0 + \delta_0 y81 + \beta_1 nearinc + \delta_1 y81 \times nearinc + u$$

```
reg3 = sm.OLS(df['rprice'], df[['const', 'y81', 'nearinc',
                                'y81nrinc']], missing='drop').fit()
print(reg3.summary())
```

	coef	std err	t	P> t
const	82.5172	2.727	30.260	0.000
y81	18.7903	4.050	4.640	0.000
nearinc	-18.8244	4.875	-3.861	0.000
y81nrinc	-11.8639	7.457	-1.591	0.113