Evaluación de Impacto: DiD

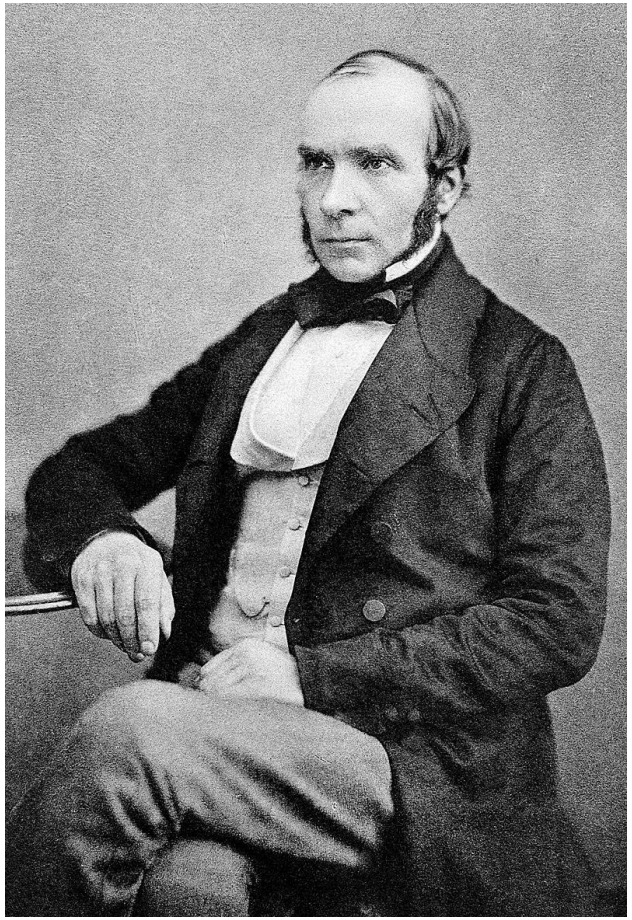
Francesco Bogliacino

Tabla de contenido

La historia de J Snow

- J Snow era un anestesiólogo de Londres del 1800
- A mitad del siglo XIX escribe dos obras sobre el colera
- Su tesis fundamental era: la enfermedad tiene como causa el agua

Fuente: Cunningham (2018)



Fuente: Wikipedia

El colera

- Mortalidad alta (50%)
- Sintomatología es muy dolorosa (diarrea y vómito) e induce deshidratación
- Vibrio cholera

- 1831-1832, 1848-1849, 1853-1854 las tres epidemias en Londres
- En esa época la teoría era la de los miasmas
- No existía teoría de bacterios y microorganismos

Los mapas

“The most terrible outbreak of cholera which ever occurred in this kingdom, is probably that which took place in Broad Street, Golden Square, and the adjoining streets, a few weeks ago. Within two hundred and fifty yards of the spot where Cambridge Street [now Lexington St.] joins Broad Street [now Broadwick], there were upwards of five hundred fatal attacks of cholera in ten days.” (Snow 1855)

Agua vs miasmas

- Micro-organismo que entra en el intestino, se reproduce, infecta el agua e induce la expulsión
- El agua contaminada es reciclada (Londres del 800 no era la de los oligarcas rusos de hoy)
- La gente toma agua y se enferma
- Sigue el proceso

- Miasmas transmiten el “virus” (nadie sabía que era un virus en esa época)
- Más abajo, más transmisión
- Más concentración más transmisión

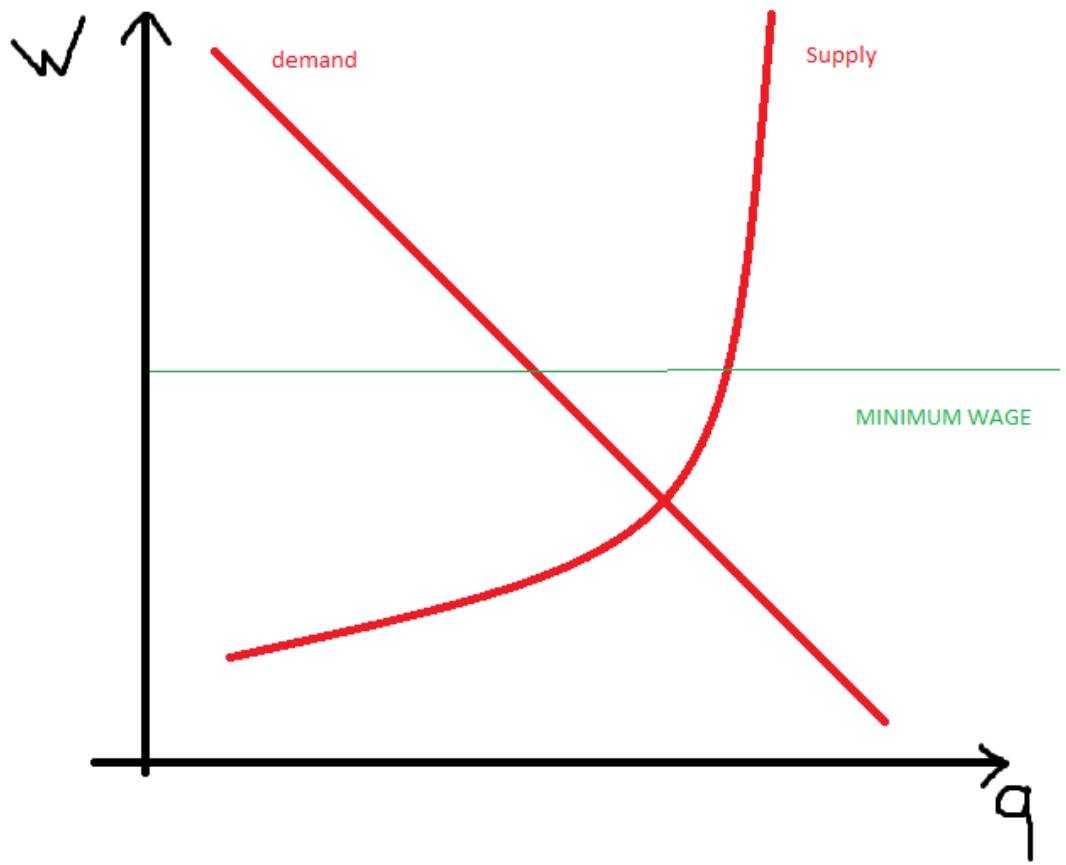
[la teoría es consistente con los hechos estilizados, hay correlación entre los fenómenos]

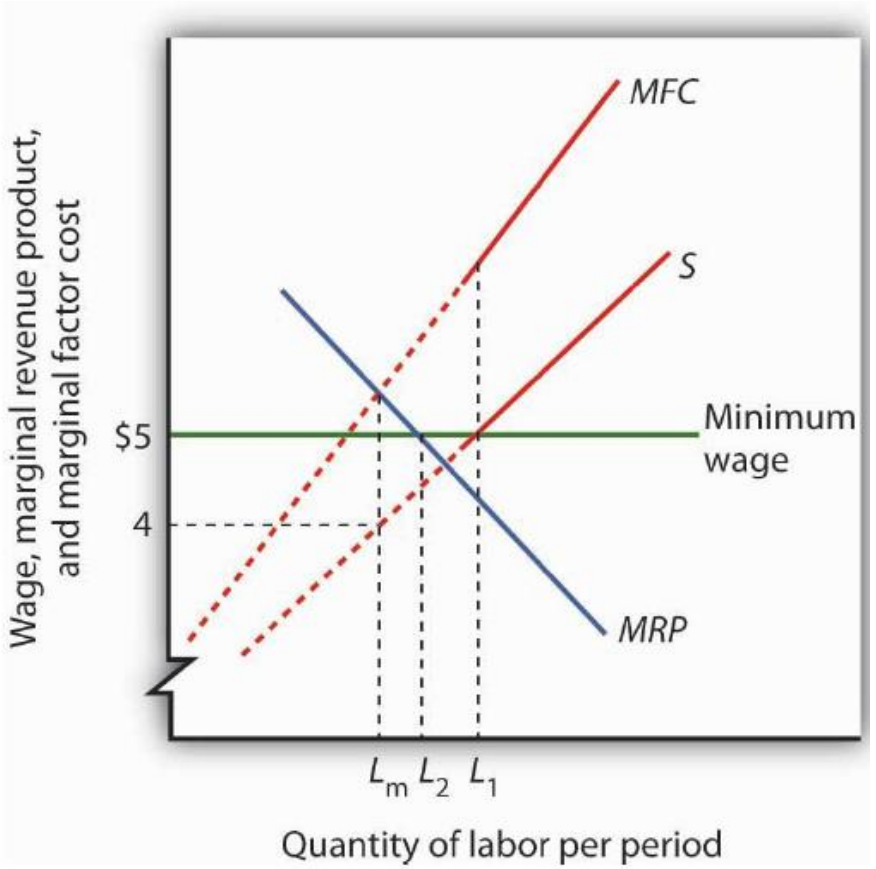
Data	Summary	Statistical Testing	Theory	Refute?	Comment
Albion Terrace, 17 houses & 20-25 deaths, 1849	17 houses infected, surrounding not	None	Water Miasma	NO YES	Sewage leaked into shared water supply after storm. Crucial for *developing* theory
Broad St – Susannah Eley (Hampstead, 1 person)	Single Case, "Far from pump but died"	None	Water Miasma	NO YES	Water bottles shipped to Hampstead by sons
Broad St – St. James workhouse (535 people, 5 deaths)	Counterexample? "Close to pump but survived"	None	Water Miasma	NO YES	In-house well
Broad St 500 residents, categorized by drinking & illness	Infection rates differ by pump drinking	Contingency Table	Water Miasma	NO YES	
S London 480k people, 1849 vs 1854 diff-in-diffs, aggregate sub-district	Mortality rates differ by water supply, not other characteristics	Diff-in-diffs, linear & count regressions, error analysis	Water Miasma	NO YES	Lambeth Water Co changed to clean water 1852, $\Rightarrow$ control / treatment DiD design
S London 480k people, direct District / sub-district comparison, quasi-randomized	Mortality rates differ by water supply company	RCT, Count regressions, detailed error analysis	Water Miasma	NO YES	Mixing of water co customers, control / treatment, effectively randomized

Fuente: Coleman (2019)

Tabla de contenido

Predicción teórica





Source: Cunningham

El estudio de Card and Krueger (1994)

- New Jersey aumenta el salario mínimo de 4.25$ a 5.05$, mientras que Pennsylvania lo deja a 4.25$
- Recolección de datos orientado hacia evaluar el efecto de la política:
  - Fast food: (a) contestan encuesta; (b) tienen muchos trabajadores al mínimo; (c) es difícil que violen la regulación por temas de franquicia; (d) el producto es homogéneo
- Datos: feb 1992, nov 1992
  - Response rate inicial: NJ 91%; PA 72.5%
  - 90% response rate in the second wave
  - Data completed with in site visit (6 cerrados, 2 cerrado temporáneamente, 2 en renovación)-> 99.8% of second wave

El estudio de Card and Krueger (1994)

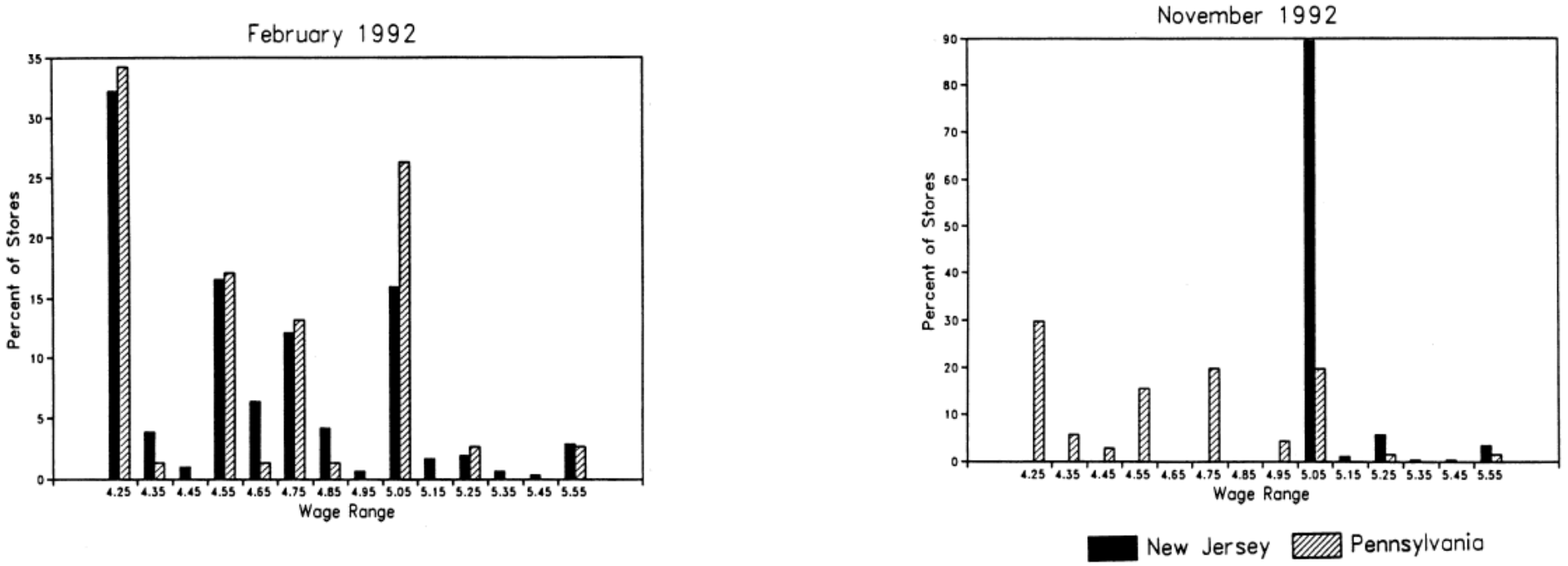


FIGURE 1. DISTRIBUTION OF STARTING WAGE RATES

El estudio de Card and Krueger (1994)

TABLE 2—MEANS OF KEY VARIABLES

Variable	Stores in: NJ	PA	t[a]
1. *Distribution of Store Types (percentages):*			
a. Burger King	41.1	44.3	−0.5
b. KFC	20.5	15.2	1.2
c. Roy Rogers	24.8	21.5	0.6
d. Wendy's	13.6	19.0	−1.1
e. Company-owned	34.1	35.4	−0.2
2. *Means in Wave 1:*			
a. FTE employment	20.4 (0.51)	23.3 (1.35)	−2.0
b. Percentage full-time employees	32.8 (1.3)	35.0 (2.7)	−0.7
c. Starting wage	4.61 (0.02)	4.63 (0.04)	−0.4
d. Wage = $4.25 (percentage)	30.5 (2.5)	32.9 (5.3)	−0.4
e. Price of full meal	3.35 (0.04)	3.04 (0.07)	4.0
f. Hours open (weekday)	14.4 (0.2)	14.5 (0.3)	−0.3
g. Recruiting bonus	23.6 (2.3)	29.1 (5.1)	−1.0

TABLE 3—AVERAGE EMPLOYMENT PER STORE BEFORE AND AFTER THE RISE IN NEW JERSEY MINIMUM WAGE

Variable	Stores by state: PA (i)	NJ (ii)	Difference, NJ − PA (iii)	Stores in New Jersey[a]: Wage = $4.25 (iv)	Wage = $4.26–$4.99 (v)	Wage ≥ $5.00 (vi)	Differences within NJ[b]: Low–high (vii)	Midrange–high (viii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	−2.89 (1.44)	19.56 (0.77)	20.08 (0.84)	22.25 (1.14)	−2.69 (1.37)	−2.17 (1.41)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	−0.14 (1.07)	20.88 (1.01)	20.96 (0.76)	20.21 (1.03)	0.67 (1.44)	0.75 (1.27)
3. Change in mean FTE employment	−2.16 (1.25)	0.59 (0.54)	2.76 (1.36)	1.32 (0.95)	0.87 (0.84)	−2.04 (1.14)	3.36 (1.48)	2.91 (1.41)
4. Change in mean FTE employment, balanced sample of stores[c]	−2.28 (1.25)	0.47 (0.48)	2.75 (1.34)	1.21 (0.82)	0.71 (0.69)	−2.16 (1.01)	3.36 (1.30)	2.87 (1.22)
5. Change in mean FTE employment, setting FTE at temporarily closed stores to 0[d]	−2.28 (1.25)	0.23 (0.49)	2.51 (1.35)	0.90 (0.87)	0.49 (0.69)	−2.39 (1.02)	3.29 (1.34)	2.88 (1.23)

El estudio de Card and Krueger (1994)

TABLE 4—REDUCED-FORM MODELS FOR CHANGE IN EMPLOYMENT

Independent variable	Model (i)	(ii)	(iii)	(iv)	(v)
1. New Jersey dummy	2.33 (1.19)	2.30 (1.20)	—	—	—
2. Initial wage gap[a]	—	—	15.65 (6.08)	14.92 (6.21)	11.91 (7.39)
3. Controls for chain and ownership[b]	no	yes	no	yes	yes
4. Controls for region[c]	no	no	no	no	yes
5. Standard error of regression	8.79	8.78	8.76	8.76	8.75
6. Probability value for controls[d]	—	0.34	—	0.44	0.40

TABLE 6—EFFECTS OF MINIMUM-WAGE INCREASE ON OTHER OUTCOMES

Outcome measure	Mean change in outcome: NJ (i)	PA (ii)	NJ − PA (iii)	Regression of change in outcome variable on: NJ dummy (iv)	Wage gap[a] (v)	Wage gap[b] (vi)
Store Characteristics:						
1. Fraction full-time workers[c] (percentage)	2.64 (1.71)	−4.65 (3.80)	7.29 (4.17)	7.30 (3.96)	33.64 (20.95)	20.28 (24.34)
2. Number of hours open per weekday	−0.00 (0.06)	0.11 (0.08)	−0.11 (0.10)	−0.11 (0.12)	−0.24 (0.65)	0.04 (0.76)
3. Number of cash registers	−0.04 (0.04)	0.13 (0.10)	−0.17 (0.11)	−0.18 (0.10)	−0.31 (0.53)	0.29 (0.62)
4. Number of cash registers open at 11:00 A.M.	−0.03 (0.05)	−0.20 (0.08)	0.17 (0.10)	0.17 (0.12)	0.15 (0.62)	−0.47 (0.74)
Employee Meal Programs:						
5. Low-price meal program (percentage)	−4.67 (2.65)	−1.28 (3.86)	−3.39 (4.68)	−2.01 (5.63)	−30.31 (29.80)	−33.15 (35.04)
6. Free meal program (percentage)	8.41 (2.17)	6.41 (3.33)	2.00 (3.97)	0.49 (4.50)	29.90 (23.75)	36.91 (27.90)
7. Combination of low-price and free meals (percentage)	−4.04 (1.98)	−5.13 (3.11)	1.09 (3.69)	1.20 (4.32)	−11.87 (22.87)	−19.19 (26.81)
Wage Profile:						
8. Time to first raise (weeks)	3.77 (0.89)	1.26 (1.97)	2.51 (2.16)	2.21 (2.03)	4.02 (10.81)	−5.10 (12.74)
9. Usual amount of first raise (cents)	−0.01 (0.01)	−0.02 (0.02)	0.01 (0.02)	0.01 (0.02)	0.03 (0.11)	0.03 (0.11)
10. Slope of wage profile (percent per week)	−0.10 (0.04)	−0.11 (0.09)	0.01 (0.10)	0.01 (0.10)	−0.09 (0.56)	−0.08 (0.57)

El estudio de Card and Krueger (1994)

TABLE 7—REDUCED-FORM MODELS FOR CHANGE IN THE PRICE OF A FULL MEAL

Independent variable	Dependent variable: change in the log price of a full meal				
	(i)	(ii)	(iii)	(iv)	(v)
1. New Jersey dummy	0.033 (0.014)	0.037 (0.014)	—	—	—
2. Initial wage gap[a]	—	—	0.077 (0.075)	0.146 (0.074)	0.063 (0.089)
3. Controls for chain and ownership[b]	no	yes	no	yes	yes
4. Controls for region[c]	no	no	no	no	yes
5. Standard error of regression	0.101	0.097	0.102	0.098	0.097

- Ejemplo de studio que es transparente, bien diseñado desde la recolección
- Todo el ejercicio es informado por la teoría
- Fue muy influyente y generó debats muy encendidos

Tabla de contenido

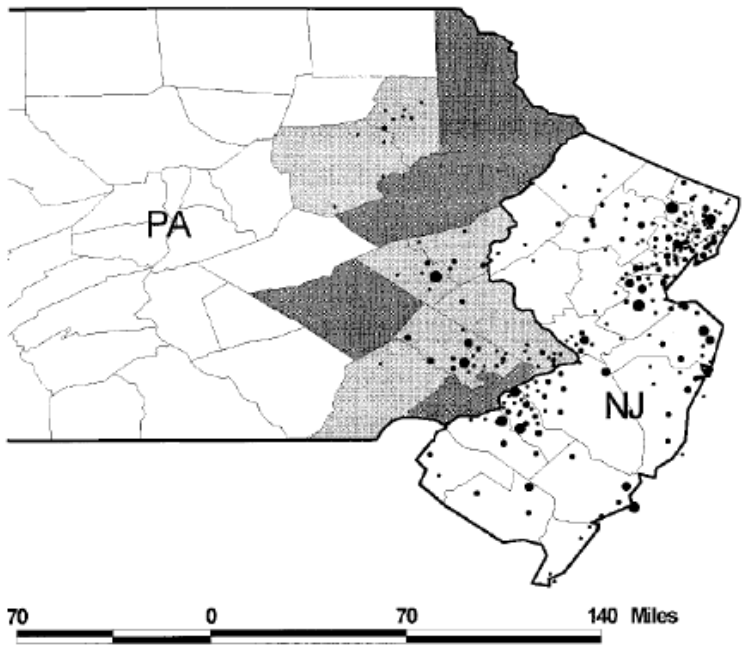
El salario mínimo y el empleo

- New Jersey aumenta el salario mínimo de 4.25$ a 5.05$, mientras que Pennsylvania lo deja a 4.25$
- Para el estado s al tiempo t el empleo con salario mínimo (k) bajo/alto [k=0,1] es

$$E\left[Y_{ist}^{k}\middle|s,t\right]=\gamma_s+\tau_t$$

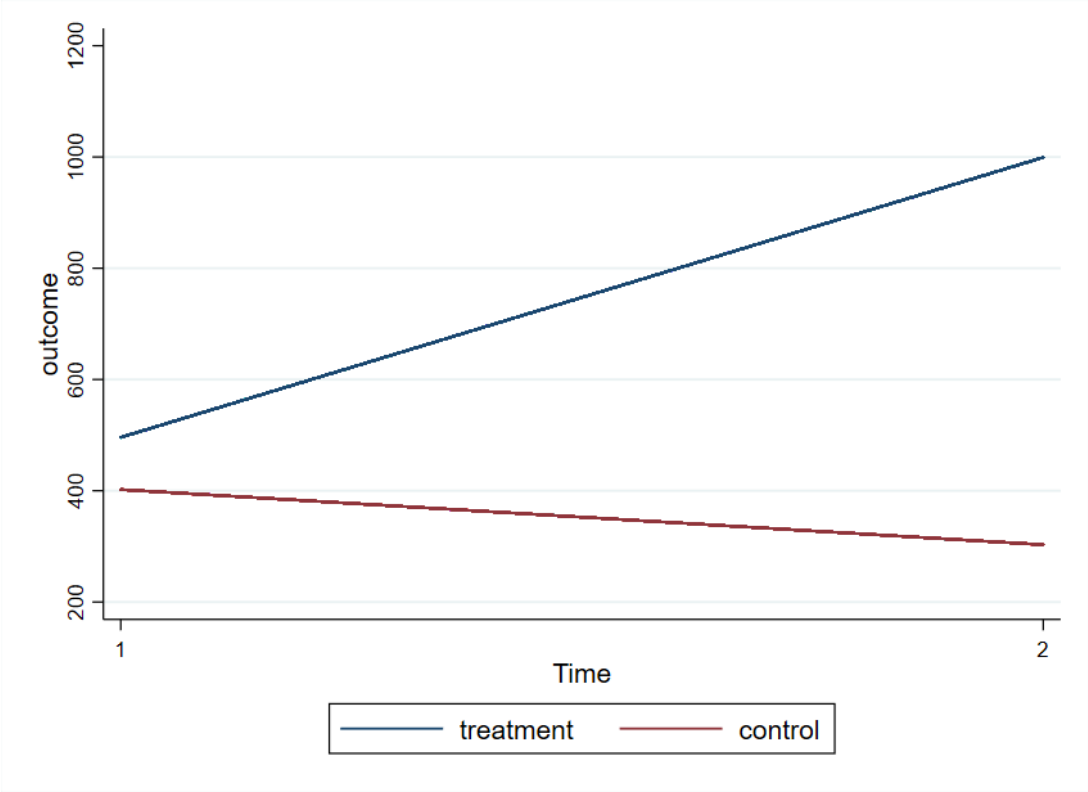
Y

$$E\left[Y_{ist}^{1}-Y_{ist}^{0}\middle|s,t\right]=\delta$$



The switching regression

$$
\begin{aligned}
&Y_{it} \\
&= D_i^{treat}\left[D_i^{t1}\left(Y_{i,t1}^1\right) + (1 \right. \\
&\left. - D_i^{t1})(Y_{i,t0}^1)\right] \\
&+ \left(1 - D_i^{treat}\right)\left[D_i^{t1}\left(Y_{i,t1}^0\right)\right. \\
&\left. + \left(1 - D_i^{t1}\right)\left(Y_{i,t0}^0\right)\right] = \\
&\ Y_{i,t0}^0 + D_i^{t1}\left(Y_{i,t1}^0 - Y_{i,t0}^0\right) \\
&\ + D_i^{treat}\left(Y_{i,t0}^1 - Y_{i,t0}^0\right) + \\
&+ D_i^{treat} D_i^{t1}\left[\left(Y_{i,t1}^1 - Y_{i,t0}^1\right)\right. \\
&\left. - \left(Y_{i,t0}^0 - Y_{i,t0}^0\right)\right]
\end{aligned}
$$



La lógica de Dif in Dif

$$E\left[Y_{ist}^{k}\middle|s,t\right] = \gamma_s + \tau_t$$

Y

$$E\left[Y_{ist}^{1} - Y_{ist}^{0}\middle|s,t\right] = \delta$$

$$Y_{ist} = \gamma_s + \tau_t + \delta D_{st} + \varepsilon_{ist}$$

$$E[Y_{ist}|NJ, post] = \gamma_{NJ} + \tau_t + \delta$$

$$E[Y_{ist}|NJ, pre] = \gamma_{NJ}$$

$$E[Y_{ist}|PA, post] = \gamma_{PA} + \tau_t + \delta$$

$$E[Y_{ist}|PA, pre] = \gamma_{PA}$$

$$E[Y_{ist}|NJ, post] - E[Y_{ist}|NJ, pre] = \tau_t + \delta$$

No es suficiente

$$E[Y_{ist}|PA, post] - E[Y_{ist}|PA, pre] = \tau_t$$

No es suficiente

$$E[Y_{ist}|NJ, post] - E[Y_{ist}|NJ, pre] - \left[E[Y_{ist}|PA, post] - E[Y_{ist}|PA, pre]\right] = \delta$$

Si las tendencias son paralelas

Potential Outcome & Switching Equation

$$\left[E\left(Y^1_{T,post} - Y^1_{T,pre}\right) - E\left(Y^0_{C,post} - Y^0_{C,pre}\right)\right] =$$

$$\left[E\left(Y^1_{T,post} - Y^1_{T,pre}\right) - E\left(Y^0_{C,post} - Y^0_{C,pre}\right)\right] + \mathrm{E}\left[Y^0_{T,post}\right] - \mathrm{E}\left[Y^0_{T,post}\right] =$$

$$E\left(Y^1_{T,post} - Y^0_{T,post}\right) =$$

ATT

$$\mathrm{E}\left[Y^0_{T,post} - Y^1_{T,pre}\right] - E\left(Y^0_{C,post} - Y^0_{C,pre}\right)$$

La diferencia entre "trends": si las tendencias son paralelas es igual a cero

Parallel trends

- El DiD se puede usar con *repeated cross section*, por eso es tal vez el más común de los diseños de investigación
- Sin parallel trends estoy violando identif
- Si cambia la composición de los grupos

  puede ser muy común la variación de tendencias paralelas

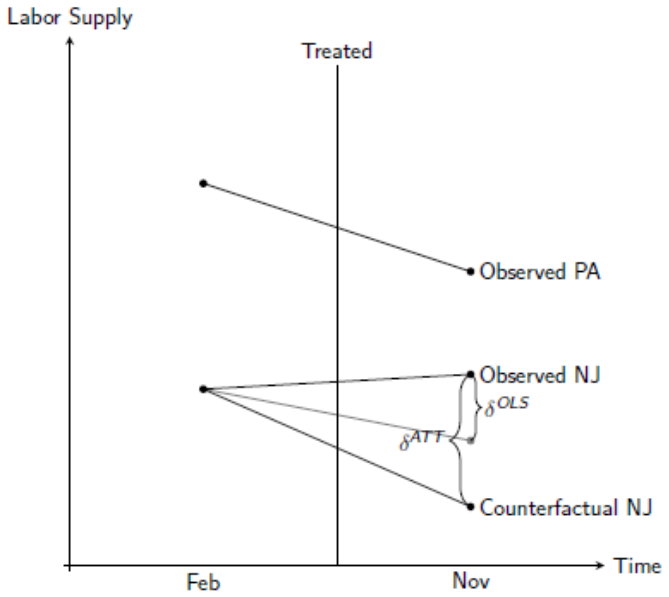


Figure: DD regression diagram without parallel trends

Table 1: Descriptive Statistics for Internet User and Non-user Groups[a]

Year	1997		1998		1999	
	Internet User	Non-user	Internet User	Non-user	Internet User	Non-user
Average Expenditure						
Recorded Music	$25.73	$10.90	$24.18	$9.97	$20.92	$9.37
Entertainment	$195.03	$96.71	$193.38	$84.92	$182.42	$80.19
Zero Expenditure						
Recorded Music	.56	.79	.60	.80	.64	.81
Entertainment	.08	.32	.09	.36	.14	.39
Demographics						
Age	40.2	49.0	42.3	49.0	44.1	49.4
Income	$52,887	$30,459	$51,996	$28,169	$49,970	$26,649
High School Grad.	.18	.31	.17	.32	.21	.32
Some College	.37	.28	.35	.27	.34	.27
College Grad.	.43	.21	.45	.21	.42	.20
Manager	.16	.08	.16	.08	.14	.08

Figure 1: Internet Diffusion and Average Quarterly Music Expenditure in the CEX

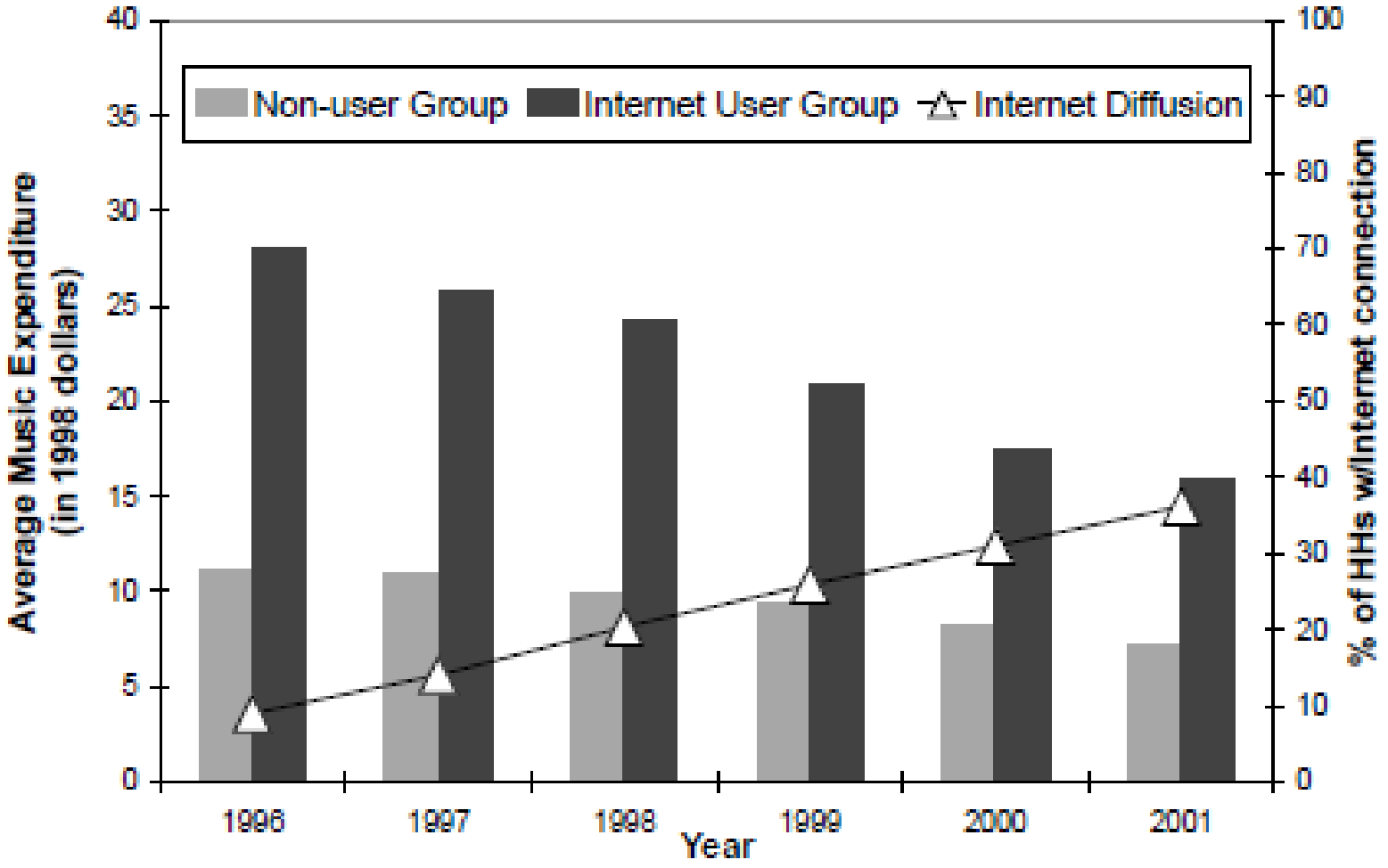


Tabla de contenido

649	325	0	NJ	19.16265	1	0	0
650	325	1	NJ	19.83828	1	1	1
651	326	0	NJ	21.34171	1	0	0
652	326	1	NJ	20.64373	1	1	1
653	327	0	NJ	20.80775	1	0	0
654	327	1	NJ	20.96773	1	1	1
655	328	0	NJ	19.56505	1	0	0
656	328	1	NJ	20.39766	1	1	1
657	329	0	NJ	20.68392	1	0	0
658	329	1	NJ	22.0185	1	1	1
659	330	0	NJ	20.36015	1	0	0
660	330	1	NJ	21.88702	1	1	1
661	331	0	NJ	21.43298	1	0	0
662	331	1	NJ	21.13688	1	1	1
663	332	0	PA	23.62994	0	0	0
664	332	1	PA	19.59755	0	1	0
665	333	0	PA	23.68527	0	0	0
666	333	1	PA	20.46726	0	1	0
667	334	0	PA	21.45385	0	0	0
668	334	1	PA	22.15355	0	1	0

La estructura de la base de datos

OLS

$$y_{ist} = \alpha + \beta NJ_i + \gamma d_t + \delta(NJ_i\, d_t) + \varepsilon_{ist}$$

- NJ es igual a uno para NJ y 0 para PA
- Bajo parallel trends
  - Para PA, pre= α
  - Para PA, post= $\alpha + \gamma$
  - Para NJ, pre= $\alpha + \beta$
  - Para NJ, post= $\alpha + \beta + \gamma + \delta$

```
. reg empl NJ expost NJ_expost, rob

Linear regression                               Number of obs     =        820
                                                F(3, 816)         =     164.53
                                                Prob > F          =     0.0000
                                                R-squared         =     0.5734
                                                Root MSE          =      .6812

------------------------------------------------------------------------------
             |               Robust
        empl |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          NJ |  -2.822551    .159201   -17.73   0.000    -3.135043   -2.510059
      expost |  -2.252137   .1825651   -12.34   0.000     -2.61049   -1.893785
   NJ_expost |   2.861982   .1868274    15.32   0.000     2.495263    3.228701
       _cons |   23.23443   .1566275   148.34   0.000     22.92699    23.54187
------------------------------------------------------------------------------
```

- DiD Estimate: (NJ, post-NJ, pre)-(NJ, post-NJ, pre)=

$$\alpha + \beta + \gamma + \delta - \alpha - \beta - (\alpha + \gamma - \alpha) = \delta$$

TWFE

$$y_{ist} = \alpha + \gamma d_t + \delta(NJ_i\, d_t) + \mu_i + \varepsilon_{ist}$$

```
. xtreg empl expost NJ_expost, fe vce(rob)

Fixed-effects (within) regression               Number of obs     =       820
Group variable: iid                             Number of groups  =       410

R-sq:                                           Obs per group:
     within  = 0.5741                                         min =         2
     between = 0.5726                                         avg =       2.0
     overall = 0.0001                                         max =         2

                                                F(2,409)          =    190.01
corr(u_i, Xb)  = -0.6501                        Prob > F          =    0.0000

                                  (Std. Err. adjusted for 410 clusters in iid)
------------------------------------------------------------------------------
             |               Robust
        empl |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      expost |  -2.252137    .186049   -12.11   0.000    -2.617869   -1.886406
   NJ_expost |   2.861982   .1902816    15.04   0.000      2.48793    3.236034
       _cons |   20.95573   .0241002   869.53   0.000     20.90836    21.00311
-------------+----------------------------------------------------------------
     sigma_u |  1.2114728
     sigma_e |  .69012486
         rho |  .75499622   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Tabla de contenido

Los diseños DiD-> State level variation

BDM

TABLE II
DD Rejection Rates for Placebo Laws

A. CPS DATA

Data	$\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3$	Modifications	Rejection rate: No effect	Rejection rate: 2% effect
1) CPS micro, log wage			.675 (.027)	.855 (.020)
2) CPS micro, log wage		Cluster at state-year level	.44 (.029)	.74 (.025)
3) CPS agg, log wage	.509, .440, .332		.435 (.029)	.72 (.026)
4) CPS agg, log wage	.509, .440, .332	Sampling w/replacement	.49 (.025)	.663 (.024)
5) CPS agg, log wage	.509, .440, .332	Serially uncorrelated laws	.05 (.011)	.988 (.006)
6) CPS agg, employment	.470, .418, .367		.46 (.025)	.88 (.016)
7) CPS agg, hours worked	.151, .114, .063		.265 (.022)	.280 (.022)
8) CPS agg, changes in log wage	−.046, .032, .002		0	.978 (.007)

B. MONTE CARLO SIMULATIONS WITH SAMPLING FROM AR(1) DISTRIBUTION

Data	ρ	Modifications	Rejection rate: No effect	Rejection rate: 2% effect
9) AR(1)	.8		.373 (.028)	.725 (.026)
10) AR(1)	0		.053 (.013)	.783 (.024)
11) AR(1)	.2		.123 (.019)	.738 (.025)
12) AR(1)	.4		.19 (.023)	.713 (.026)
13) AR(1)	.6		.333 (.027)	.700 (.026)
14) AR(1)	−.4		.008 (.005)	.7 (.026)

BDM

- <DiD.do>

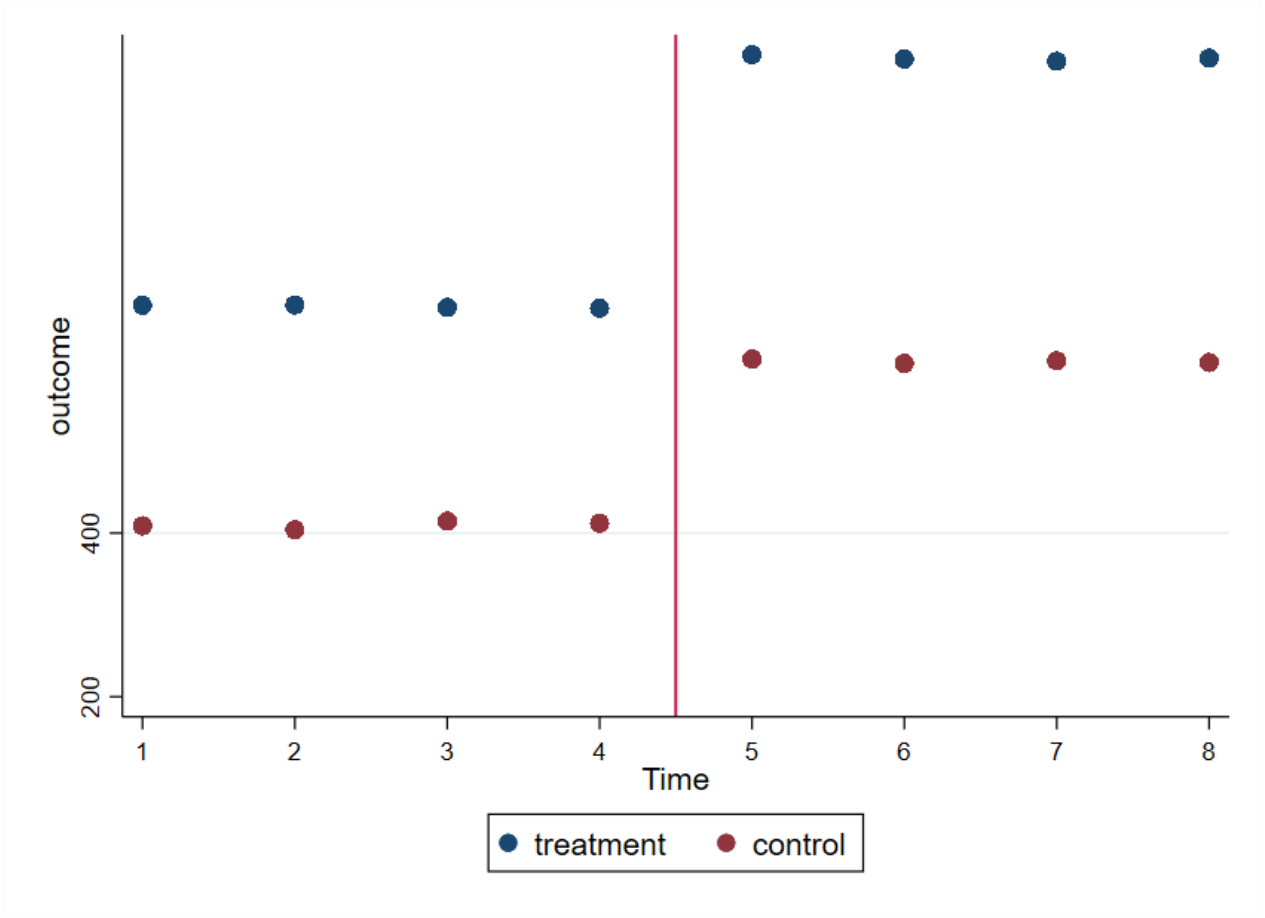
- BDM no sugieren modelar el término de error
- Sugieren:
  - Hacer block randomization a nivel de estado
  - Clustering a nivel de estado
  - Collapsar la info a nivel de estado antes y después y correr la regresión

Tabla de contenido

Ejemplo

- 50 estados
- La ley se pasa entre t=4 y t=5
- T=1, …, 8
- Queremos ver el impacto sobre ingresos

(1) Plot the data to see parallel trend



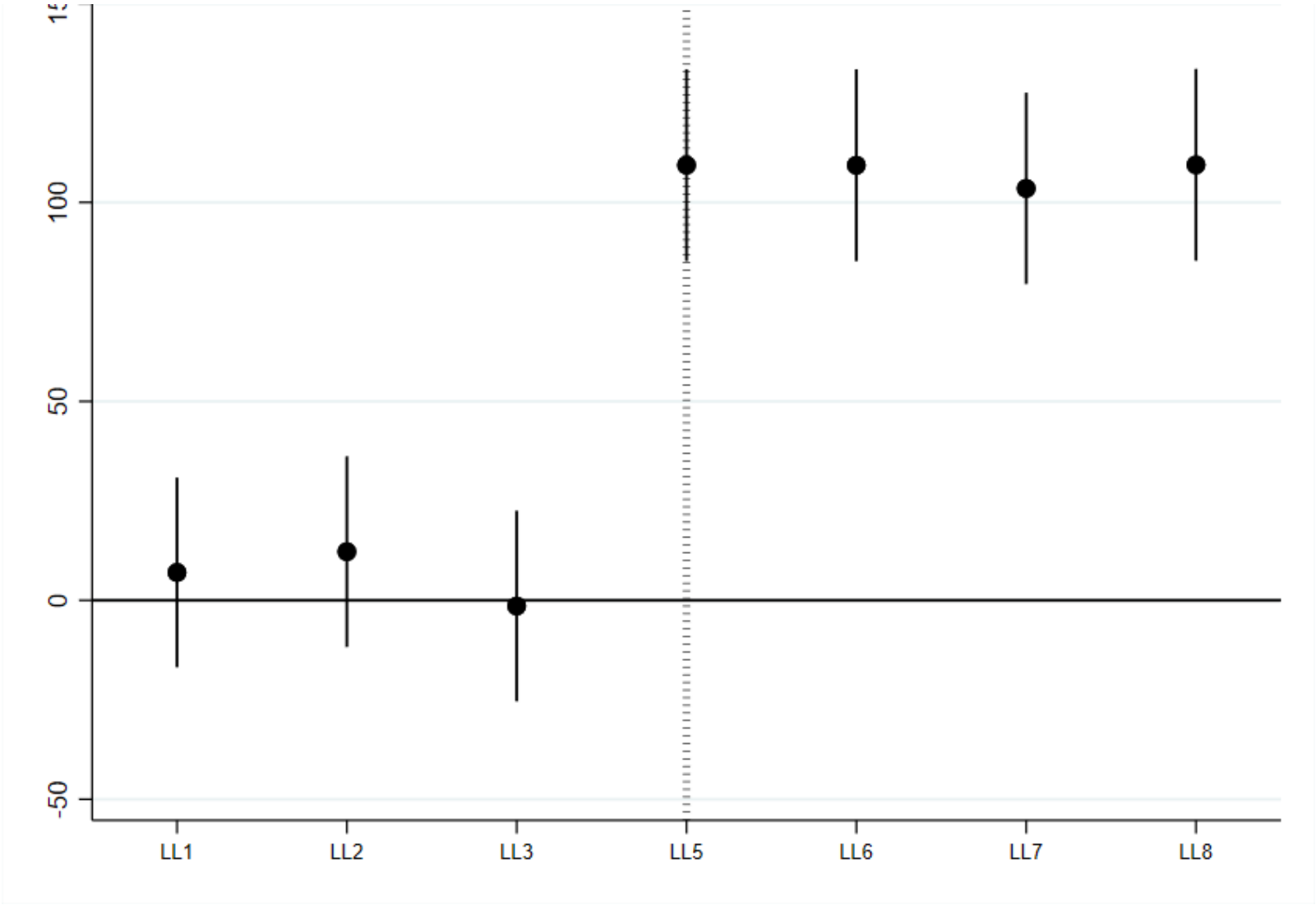
(2) Baseline Results

Dependent Variable: Outcome	(1) OLS	(2) OLS	(3) FE
DiD	96.37***	96.37***	96.37***
	(7.38)	(6.05)	(5.10)
Constant	366.67***	379.71***	514.68***
	(3.73)	(11.50)	(2.77)
Observations	80,000	80,000	80,000
R-squared	0.14	0.42	0.18
Treatment Dummy	Yes		
Expost Dummy	Yes		
Standard errors	Rob	Rob	Cluster(iid)
Year FE		Yes	Yes
State FE		Yes	
Number of iid			10,000
Individual FE			Yes

(3) BDM

	(1)	(2)	(3)
Outcome	FE	FE	OLS
DiD	95.34***	95.34***	95.34***
	(4.82)	(4.02)	(0.24)
Constant	548.88***	548.88***	289.77***
	(2.61)	(4.38)	(0.43)
Observations	80,000	80,000	20,000
R-squared	0.18	0.18	1.00
Number of iid	10,000	10,000	
Year FE	Yes	Yes	Yes
Individual FE	Yes	Yes	
Standard errors	Cluster(state)	Block Boots	Averaged Pre and Post
State FE			Yes

$$y_{ist} = \sum_{\tau=-2}^{-q} \gamma_\tau D_i + \sum_{\tau=0}^{m} \delta_\tau D_i$$
$$\mu_s + \lambda_t + \zeta_i + \varepsilon_{ist}$$



(4) Event study

(5) Placebo

	(1) FE Outcome	(2) FE Placebo Outcome
DiD (placebo)	2.13	
	(10.29)	
DiD		-6.70
		(5.73)
Constant	525.01***	1,624.22***
	(4.60)	(3.80)
Observations	80,000	80,000
R-squared	0.17	0.21
Number of iid	10,000	10,000
Year FE	Yes	Yes
Individual FE	Yes	Yes
Standard errors	Cluster state	Cluster state

Tabla de contenido