



# Semana 5

## Regresión Discontinua y Diferencias en Diferencias

DCDPP - Datos para la evaluación de Políticas Públicas | PUC | 01 de noviembre, 2022

---

 **Pablo A. Celhay** |  [pacelhay@uc.cl](mailto:pacelhay@uc.cl)

# Outline

1. Regresión Discontinua
2. Diferencias en Diferencias

# 1. Regresión Discontinua

# Regresión Discontinua

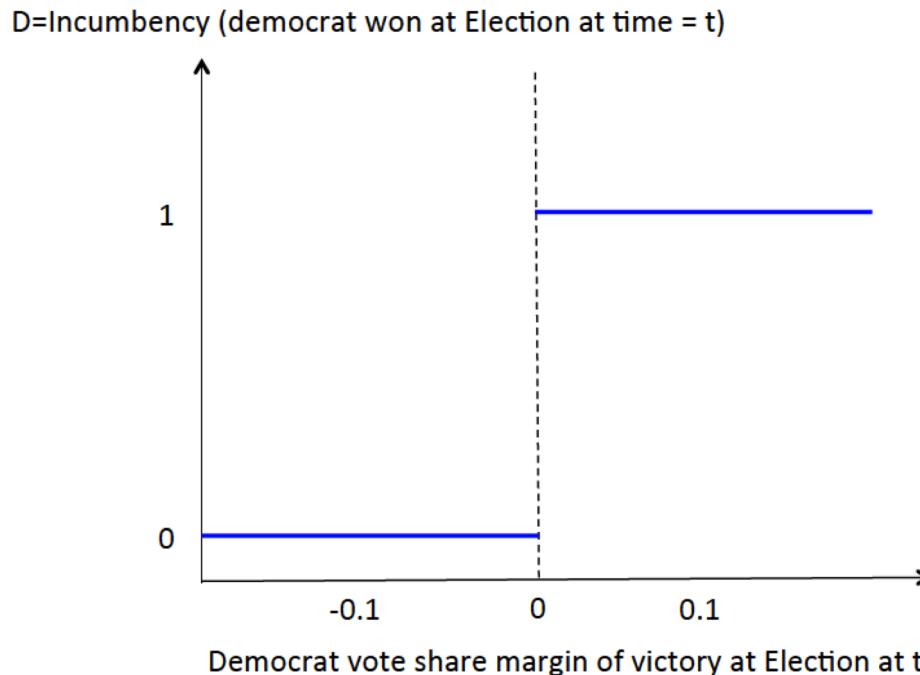
- Un diseño de regresión discontinua es una estrategia de identificación muy poderosa, convincente y aplicable a un gran rango de situaciones
- Es común observar que el acceso o los incentivos a participar en un programa están basados en reglas transparentes de un criterio utiliza puntos de corte ¿Caso más claro cuando uno postula a una carrera en la U?
- Comparar individuos similares pero en distintos lados de este punto de corte puede ser una forma creíble de estimar efectos de un tratamiento.
- RD funciona bien para validez interna ¿Qué es esto? pero no mucho para validez externa

# RD Sharp

## Ejemplo 1: ¿Cuál es el efecto de ser incumbente en la probabilidad de ganar la siguiente elección?

- ¿Cuál es la probabilidad de que una PC/PPD/UDI gane la próxima elección dado que una PC/PPD/UDI gano la última elección?
- Siguiendo nuestra terminología:
  - $Y_i$ : probabilidad de ganar la elección en  $t + 1$  en el distrito  $i$
  - $D_i = 1$  si la última elección la ganó un PC/PPD/UDI
  - ¿Es  $D_i$  asignado de manera aleatoria en la realidad?
  - ¿Cómo podemos estimar el efecto causal de  $D_i$  sobre  $Y_i$  ?
- Lee (2008) compara resultados electorales en casos donde la elección anterior fue ganada por un margen muy pequeño

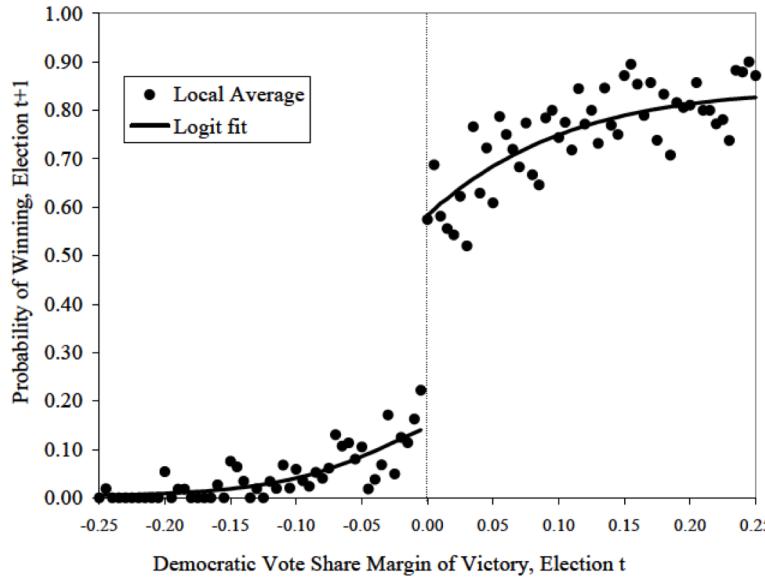
# Visualización: ¿Cómo cambia la probabilidad de ganar?



- $D_i = 1$  si la última elección (tiempo  $t$ ) la ganó un demócrata
- $X_i$ : es la variable de asignación (o **running variable**) en el eje x

# Visualización: ¿Cómo cambia la probabilidad de ganar?

Figure 1: Candidate's Probability of Winning Election  $t+1$ , by Margin of Victory in Election  $t$ : local averages and parametric fit. Source: Lee (2008)



- $Y_i$  : es la probabilidad de que un demócrata gane en  $t + 1$  en el distrito  $i$
- $X_i$ : es la variable de asignación (o **running variable**) en el eje x

# RD Sharp

- $D_i \in \{0, 1\}$ : Tratamiento
- $X_i$ : variable de asignación que determina de manera perfecta el valor de  $D_i$ . Sea  $c$  un punto de corte de esta variable  $X_i \setminus$

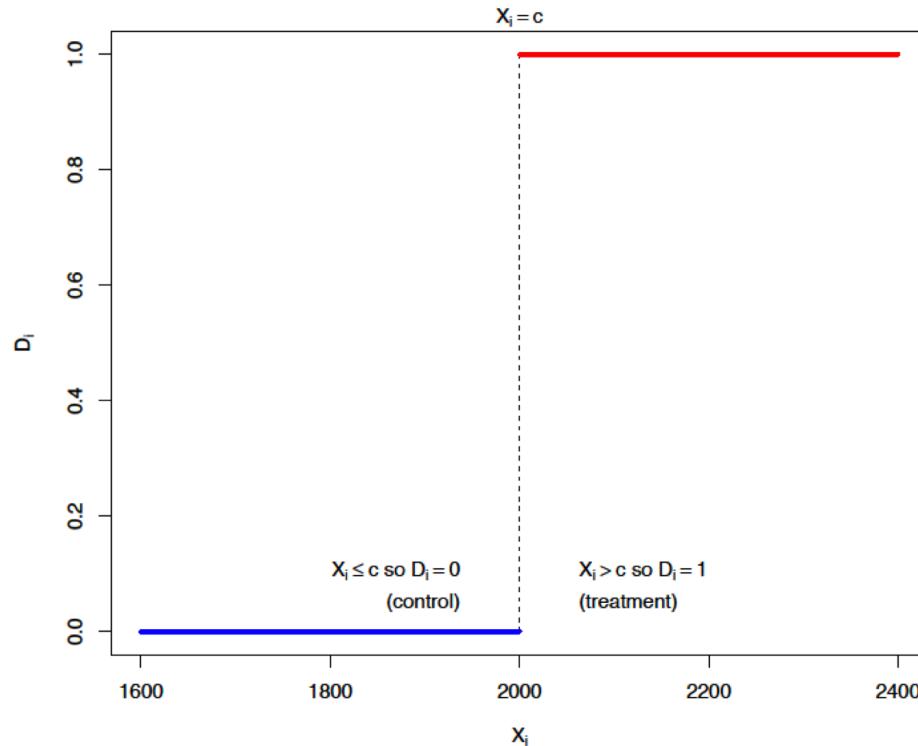
$$D_i = 1\{X_i > c\} \text{ o bien } D_i = 1 \text{ si } X_i > c \text{ y } D_i = 0 \text{ si } X_i \geq 0$$

- $X_i$  puede estar correlacionado con  $Y_{1i}$  o  $Y_{0i}$  de manera directa o a través de variables no observadas.
- La intuición clave y básica es que en el punto de corte  $X_i = c$ , el hecho de que una unidad de análisis sea tratada ( $D_i=1$ ) y otra no ( $D_i=0$ ) es similar a un proceso aleatorio

## Ejemplo 2: ¿Cuál es el efecto de becas para la Universidad?

- Las becas se dan en función de los resultados de un test de admisión a la Universidad (PSU)
  - $Y_i$ : Salarios post universidad
  - $D_i = 1$  si la unidad  $i$  recibe la beca
  - $Y_{1i}$ : Salarios potenciales con beca
  - $Y_{0i}$ : Salarios potenciales sin beca
- $Y_{1i}$  y  $Y_{0i}$  están correlacionados con  $X_i$ : ¿Por qué? en promedio alumn@s con mejor PSU tienen mayores salarios.

# Visualización: ¿Cómo cambia la probabilidad de ganar?



- $D_i = 1$  si obtiene beca
- $X_i$ : es la variable de asignación (o **running variable**) en el eje x: Puntaje PSU

# RD Sharp

- Supuesto clave: Continuidad en los resultados potenciales
  - $E(Y_i|X = x_i)$  es continua en  $x$  alrededor de  $X_i = c$
- El estimador causal: ATE local alrededor del punto de corte:

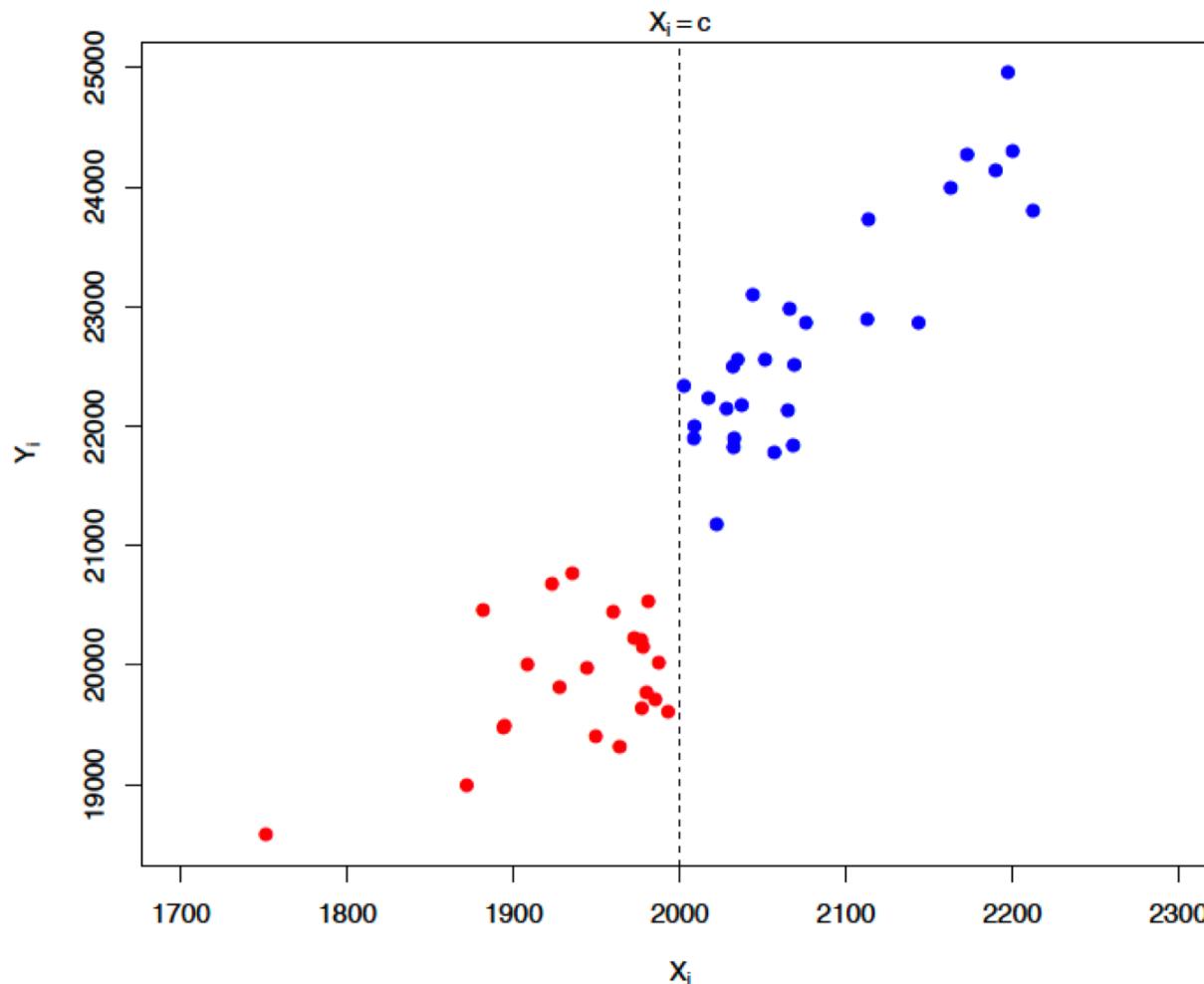
$\tau_{srd} =$  ¿Cómo es ATE? ¿Cómo sería el estimador en el corte?

$\tau_{srd} = E(Y_{1i} - Y_{0i}|X = c)$  - ¿Validez externa?

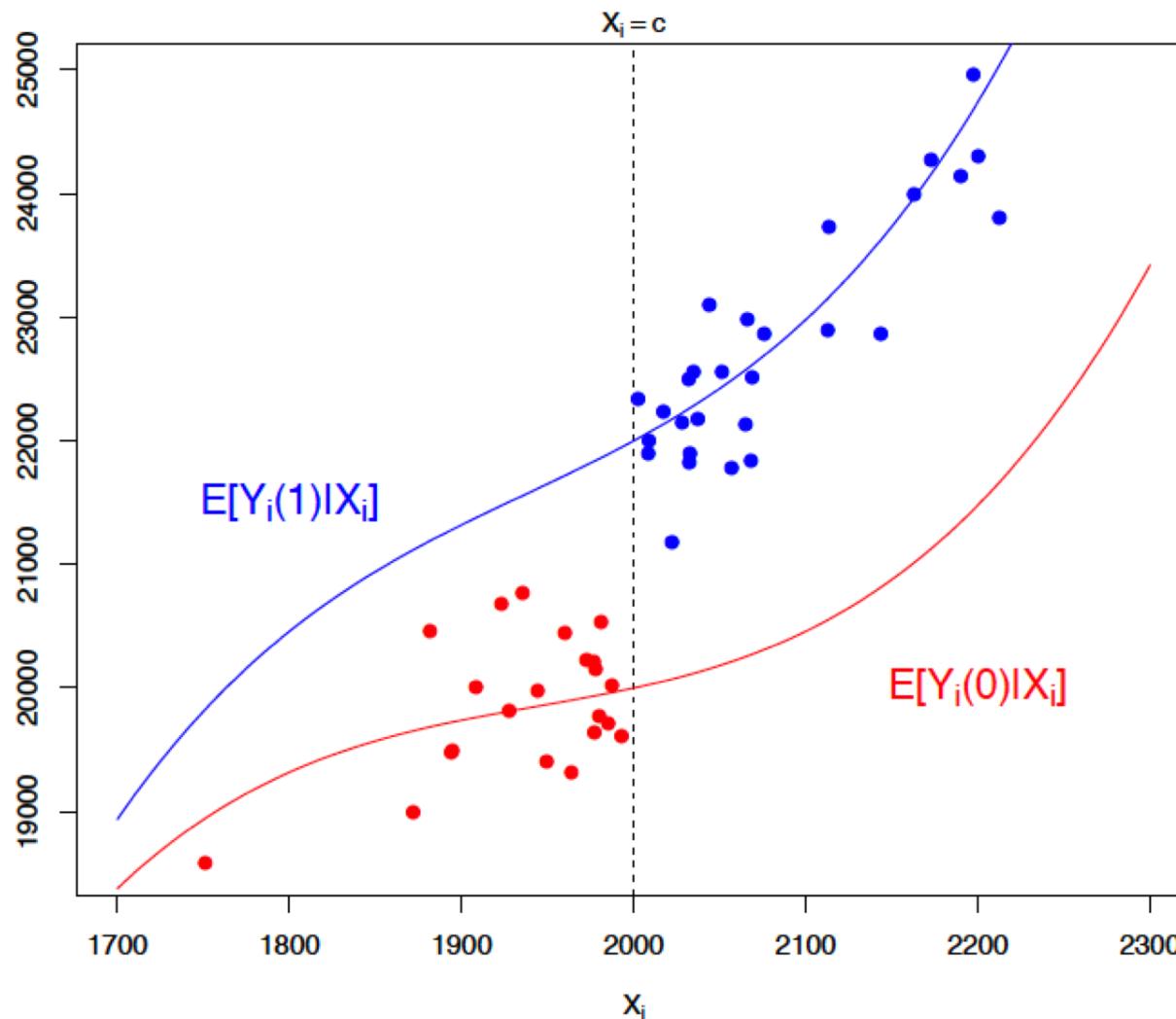
- Si tenemos continuidad en  $X_i$

$$\tau_{srd} = \lim_{x \rightarrow c^-} E(Y_i|X = x) - \lim_{x \rightarrow c^+} E(Y_i|X = x)$$

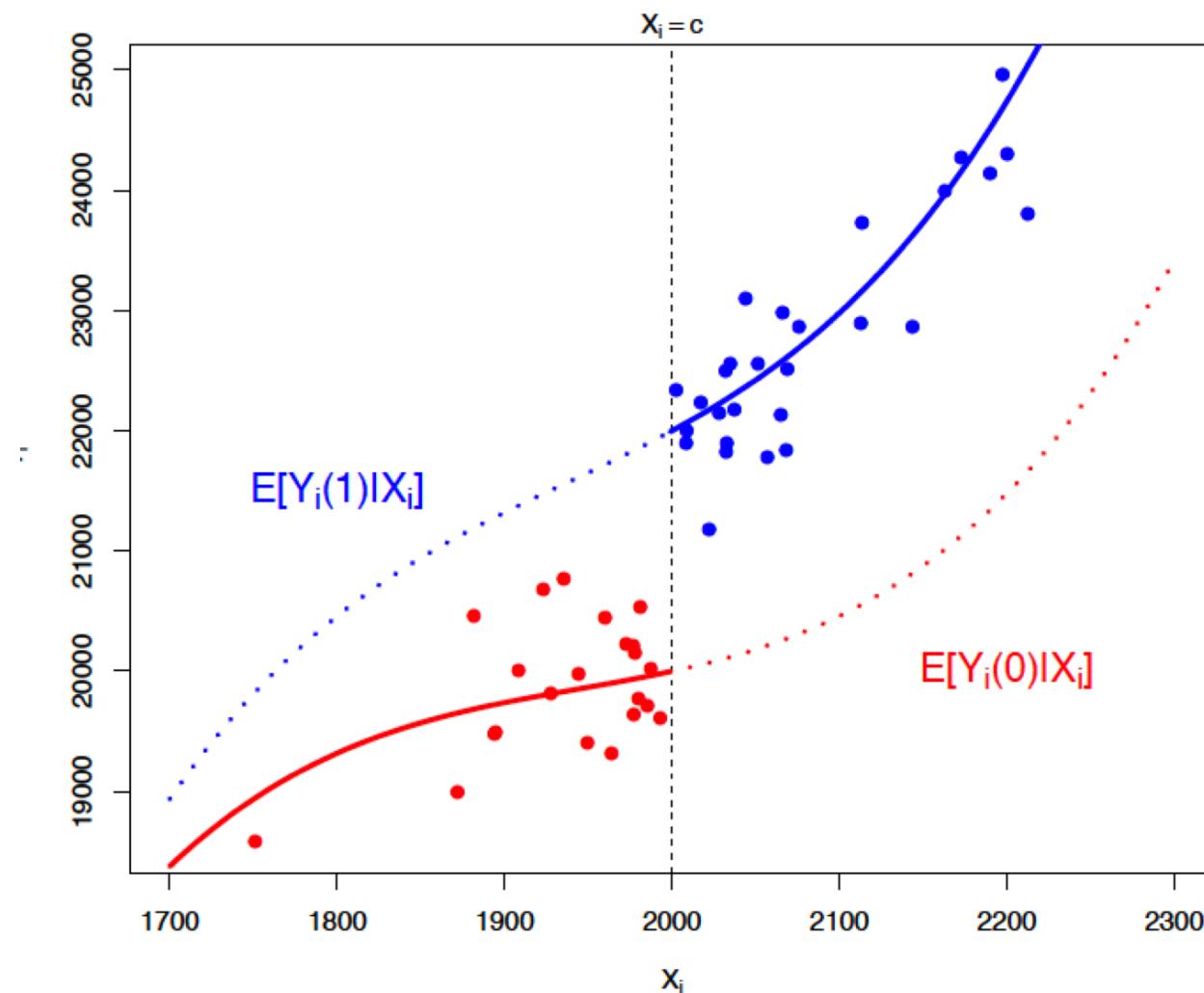
# Visualización: Estimando efectos



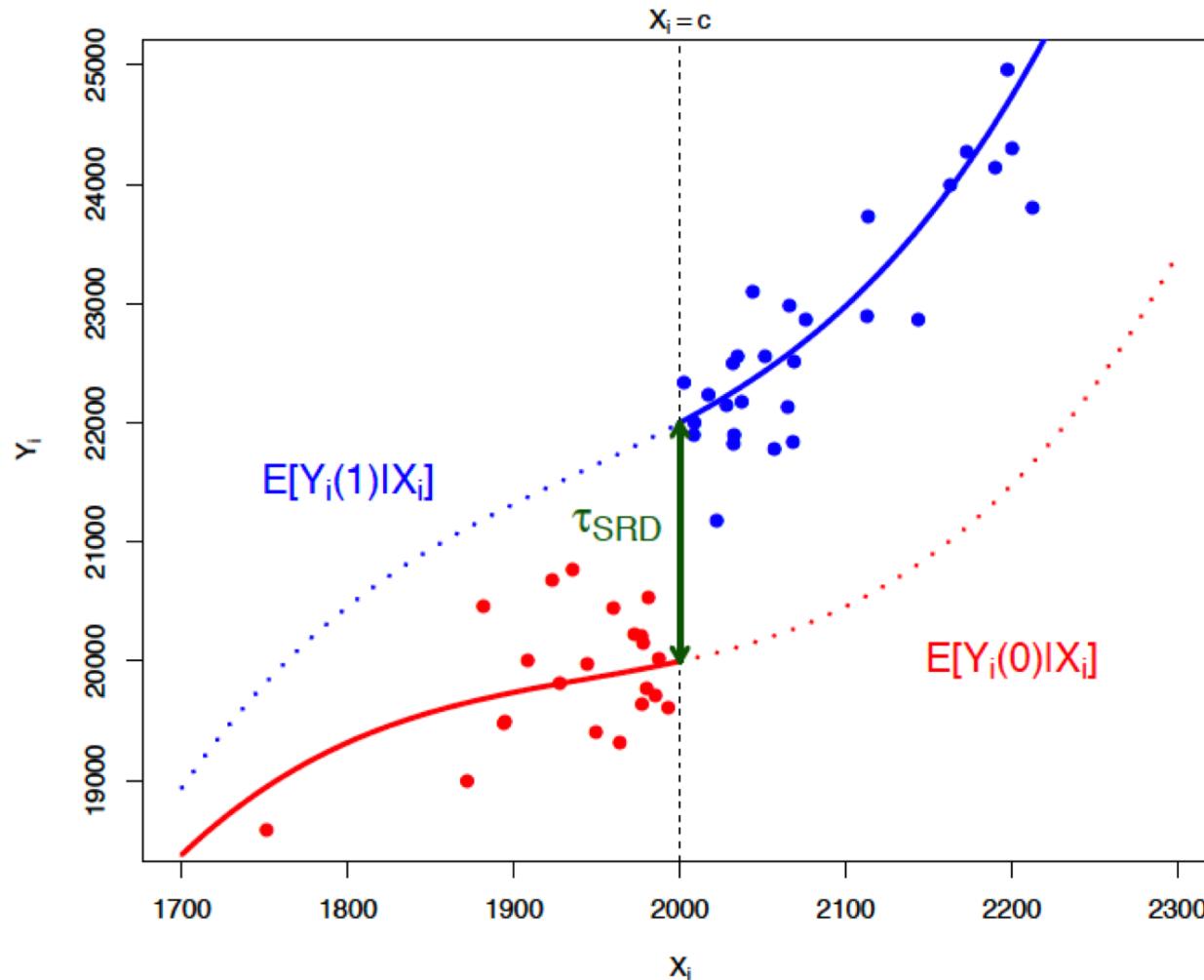
# Visualización: Estimando efectos



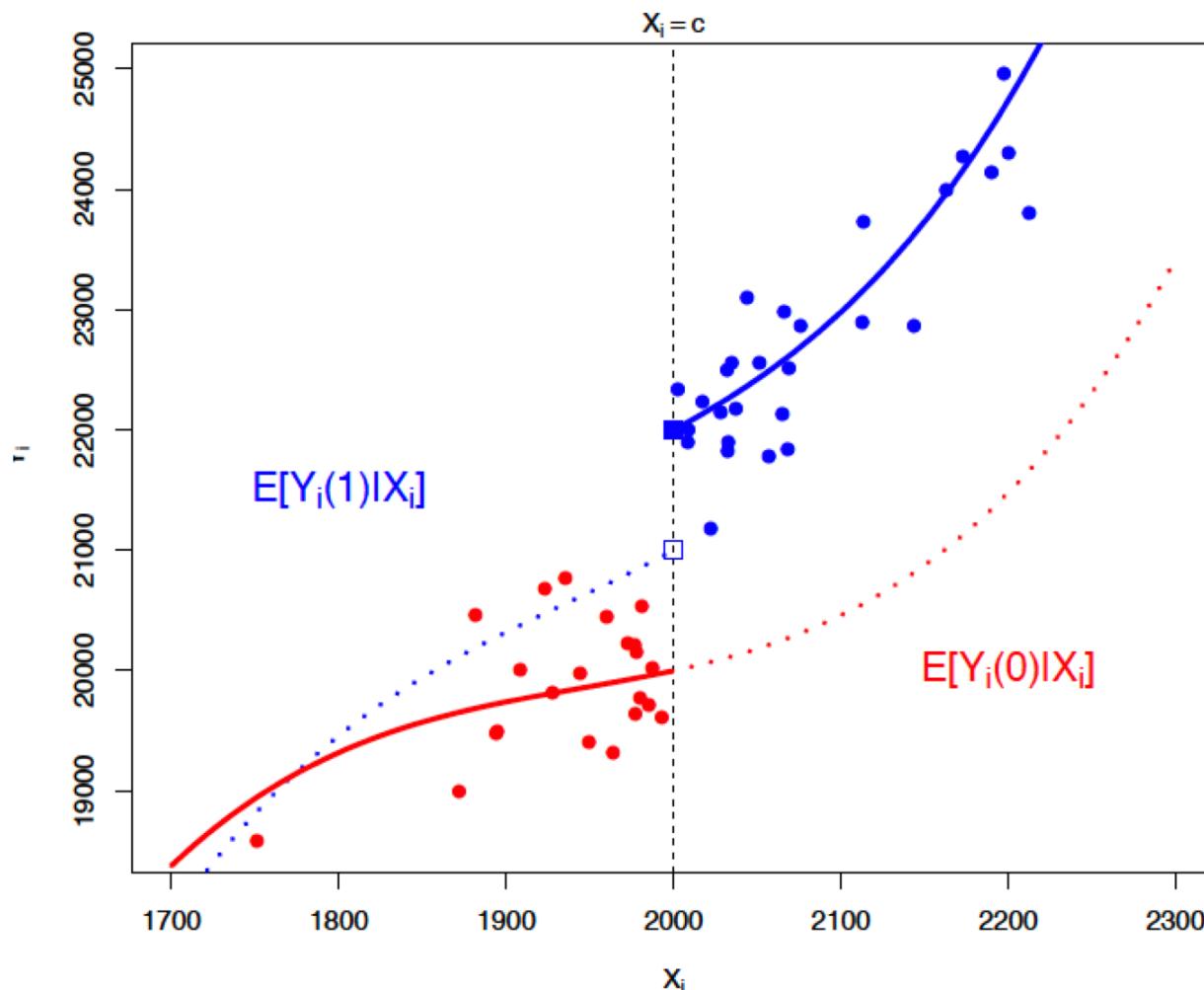
# Visualización: Estimando efectos



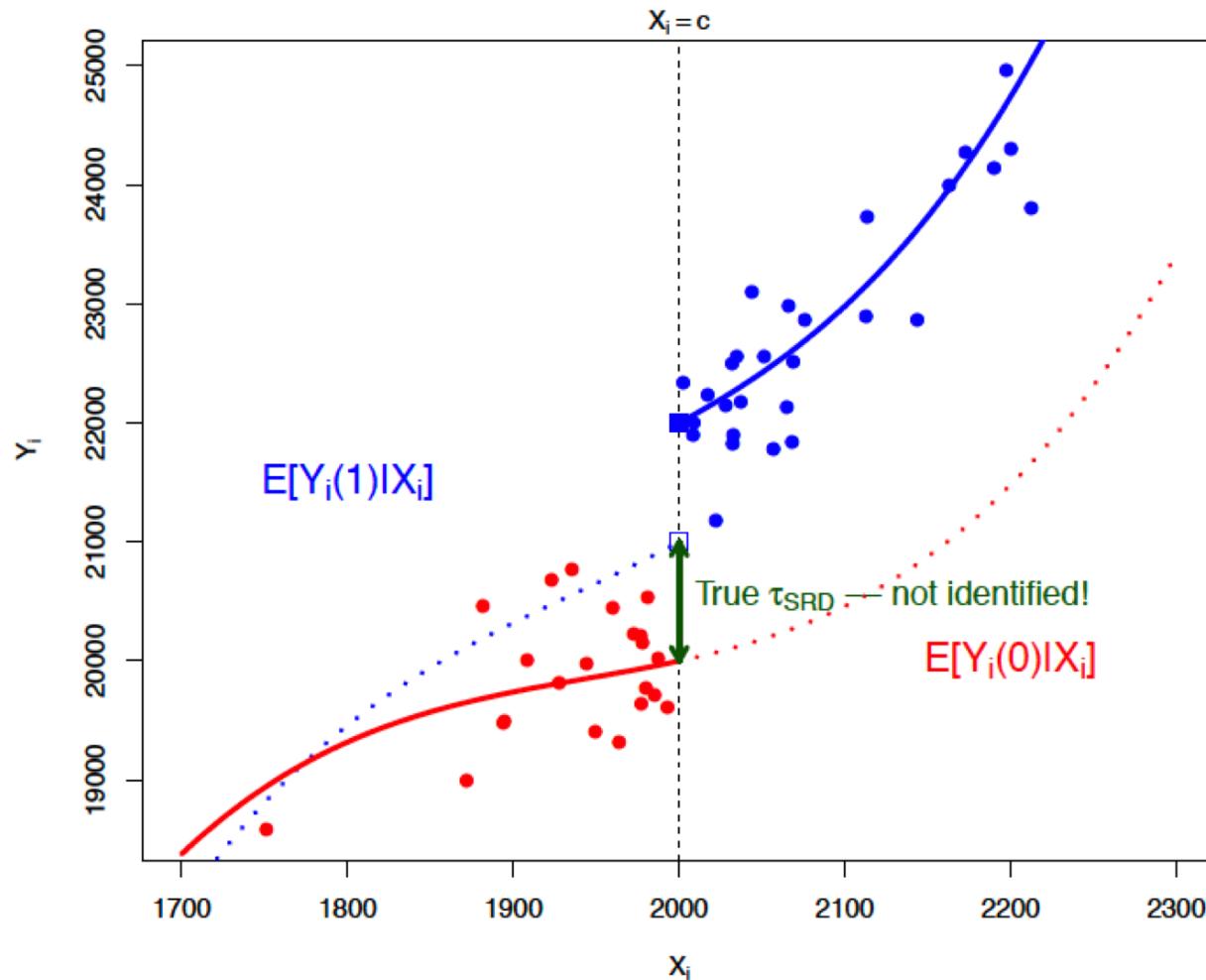
# Visualización: Estimando efectos



# Visualización: Estimando efectos



# Visualización: Estimando efectos



# RD Sharp: Estimación

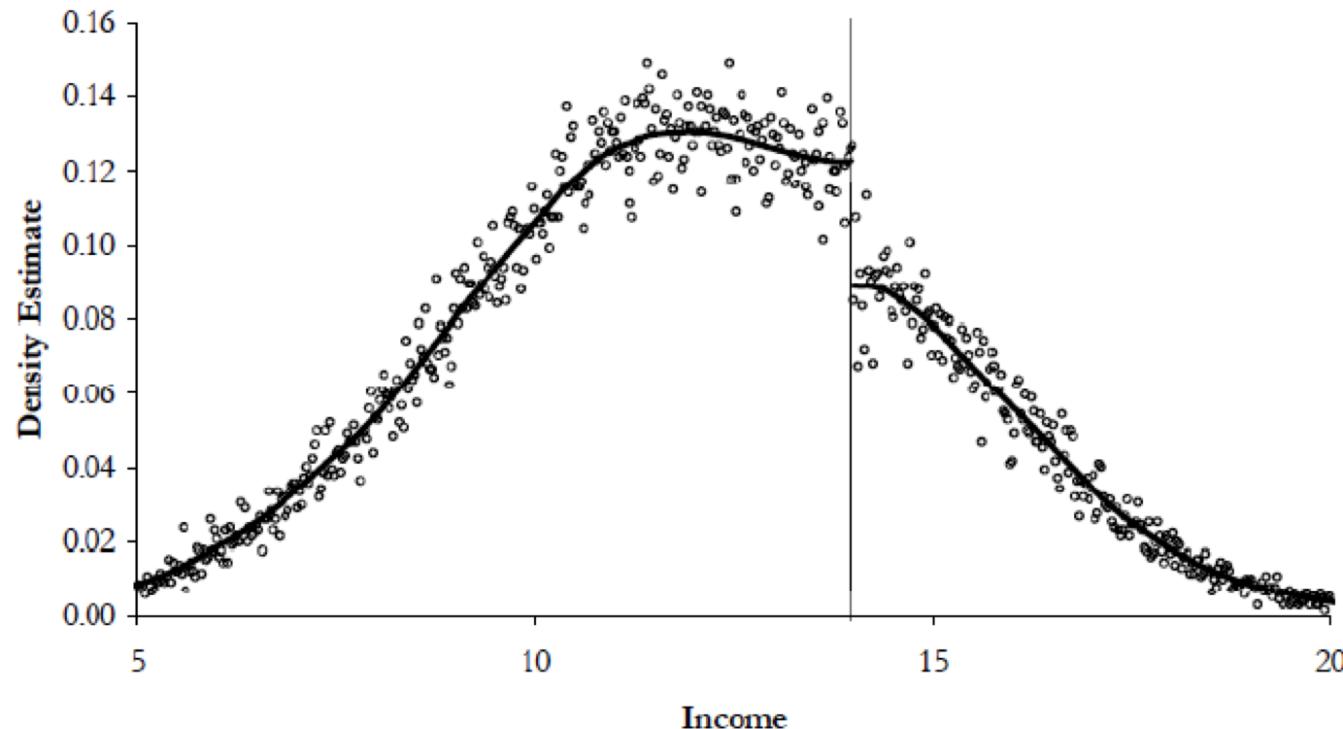
- Concentrarse en una parte de la muestra alrededor del punto de corte
  - $c - h \leq X_i \leq c + h$  donde  $h$  es un ancho de banda (vecindad o ventana)
- Recodificar la variable  $X_i$  en "esviaciones de c":  $X_i = X_i - c$ 
  - $X_i = 0$  si  $X_i = c$
  - $X_i > 0$  si  $X_i > c \rightarrow D_i = 1$
  - $X_i < 0$  si  $X_i < c \rightarrow D_i = 0$
- Decidir en el modelo para  $E[Y_i|X_i]$ :
  - Lineal con pendiente común a cada lado del corte
  - Lineal con pendientes distintas
  - No lineal
  - Cada modelo se construye en base a supuestos sobre los resultados potenciales y los contrafactuales
  - Siempre hay que empezar con una exploración visual (scatter) para ver qué modelo es más plausible

# RD Sharp: Problemas en estimación

1. Otras variables cambian alrededor del punto de corte
  - Revisar si hay saltos en otras variables
2. Hay discontinuidades en otros valores de la variable  $X_i$  distintos de  $c$
3. Manipulación del punto de  $X_i$  alrededor de  $c$ 
  - Verificar continuidad de la densidad (número de obs.) alrededor de  $c$ .

# Visualización: Manipulación

D. Density of Income  
with Pre-Announcement and Manipulation



# Malamud, O., and C. Pop-Eleches. 2011

- ¿Qué pregunta quieren responder?
- ¿Cómo afecta el uso de computadores los resultados escolares?
- ¿Porqué es importante mejorar el acceso a computadores?
- Si hay beneficios importantes de acceder a nuevas tecnologías, esto podría generar brechas socioeconómicas aún mayores **¿Porqué?**
- ¿Cuál es su símil en Chile?

# Malamud, O., and C. Pop-Eleches. 2011



# Malamud, O., and C. Pop-Eleches. 2011

- Buscan estimar el efecto causal de acceder a una computadora en casa en el capital humano de l@s niñ@s que se benefician.
- Programa del Ministerio de Educación de Rumania
- Programa tenía disponible un número fijo de vouchers para comparar computadores
- ¿Cómo se eligieron beneficiarios? En función de un ranking de ingreso familiar
- ¿Porqué es importante este mecanismo de asignación de computadores para la evaluación empírica?

# Malamud, O., and C. Pop-Eleches. 2011

- Yo Elijo mi PC: Teoría de Cambio (Gertler et al. 2010)
  - ¿Cuál es el objetivo de YEPC? (pág. 4)

**Insumos:** Recursos financieros, humanos u otros...

**YEPC** Presupuesto, Número de computadores, hogares con necesidad.

**Actividades:** Acciones emprendidas que transforman insumos en productos...

**YEPC** Entrega de computadores a familias, uso de computadores por integrantes del hogar,

# Malamud, O., and C. Pop-Eleches. 2011

- Yo Elijo mi PC: Teoría de Cambio (Gertler et al. 2010)

**Productos:** Productos resultantes de la transformación de insumos a productos tangibles...

**YEPC** Computadores efectivamente instalados en hogares beneficiarios, computadores en funcionamiento ¿Es esto obvio?

## Junaeb abre investigación por entrega de "notebooks de madera en Punta Arenas

La situación salió a la luz cuando estudiantes beneficiados por el programa "yo elijo mi PC", abrieron las cajas de sus computadores y encontraron pedazos de madera de tamaño y pesos similares a un laptop.

06 de Abril de 2010 | 17:25 | Por Cristina Cáceres, Emol



### EL COMENTARISTA OPINA

Una mujer normal

5 1



### AHORA SE DEBATE



Proponen bajar gastos comunes edificios costeros para usarlos el caso de tsunami. ¿Concuerdas?

# Malamud, O., and C. Pop-Eleches. 2011

- Yo Elijo mi PC: Teoría de Cambio (Gertler et al. 2010)

**Productos:** Productos resultantes de la transformación de insumos a productos tangibles...

**YEPC** Computadores efectivamente instalados en hogares beneficiarios, computadores en funcionamiento

**Resultados:** uso de los productos que realiza la población beneficiaria

**YEPC** Número de computadores en uso, horas de uso, tipos de uso (aprendizaje o recreación)

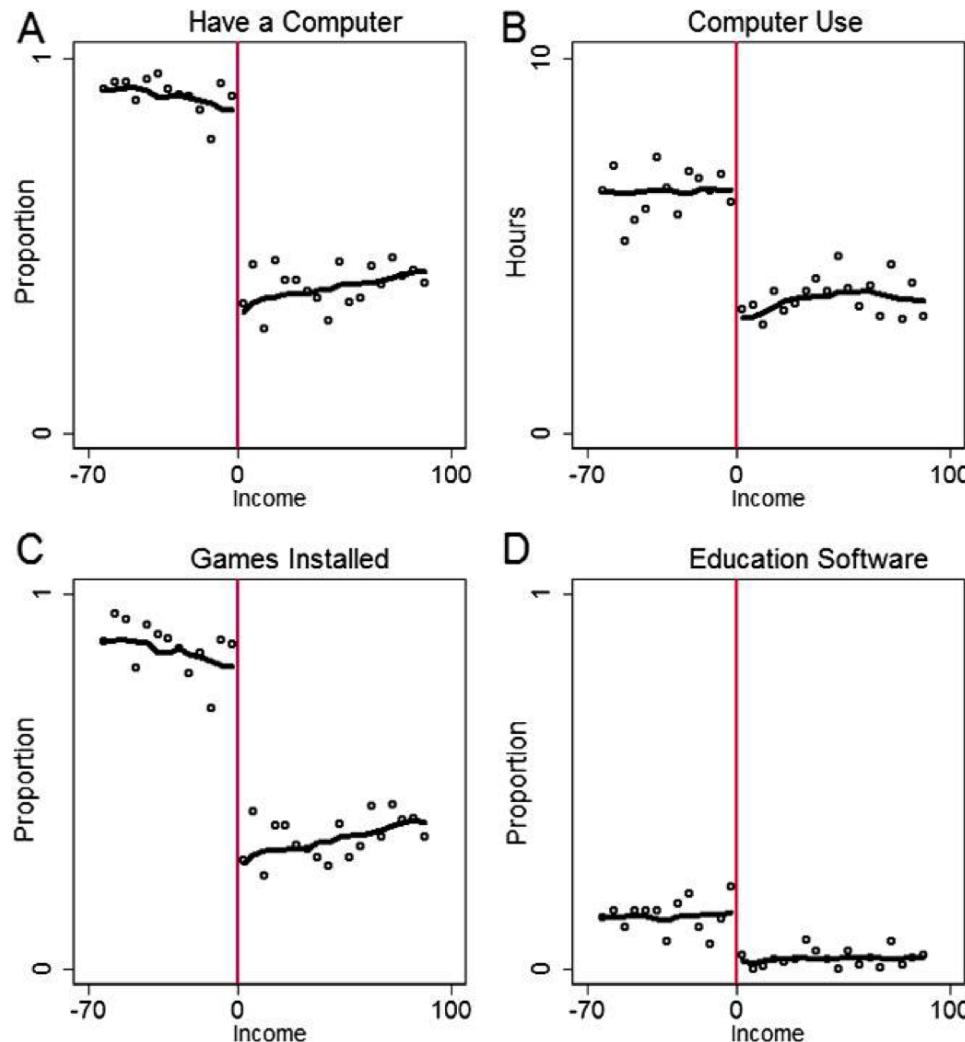
**Resultados finales:** Objetivo final del programa

**YEPC** Mejora puntajes de pruebas estandarizadas y disminuye brecha en test std entre alumn@s de distinto SES, reduce tasa de deserción escolar, mejora empleabilidad, etc.

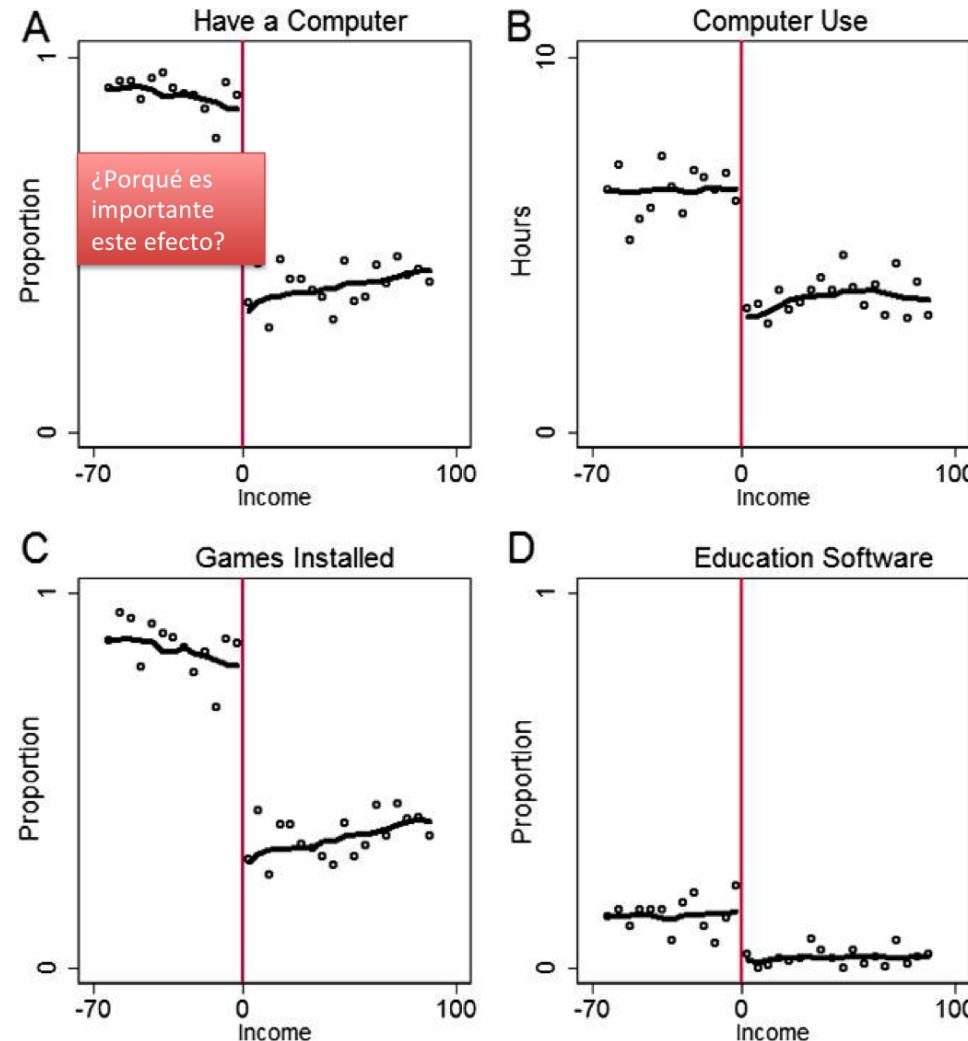
# Antes de ver el detalle

- Pensemos junt@s la regresión que Malamud, O., and C. Pop-Eleches. 2011 deberían correr
  - ¿Cuál es la pregunta de impacto aquí?
  - ¿Cuál es el escenario contrafactual?
  - ¿Qué datos necesitamos?
  - Con estos datos, ¿cómo escribimos la regresión para el análisis de impacto?
  - ¿Cómo estimamos esta regresión? ¿Qué supuestos hay detrás?
- $y_i = \alpha + \delta D_i + f(\text{ingreso}_i) \varepsilon_i$ 
  - ¿Qué es  $y_i$ ?
  - ¿Qué es  $i$ ?
  - ¿Qué es  $D_i$ ?
  - ¿Qué es  $f(\text{ingreso}_i)$ ?

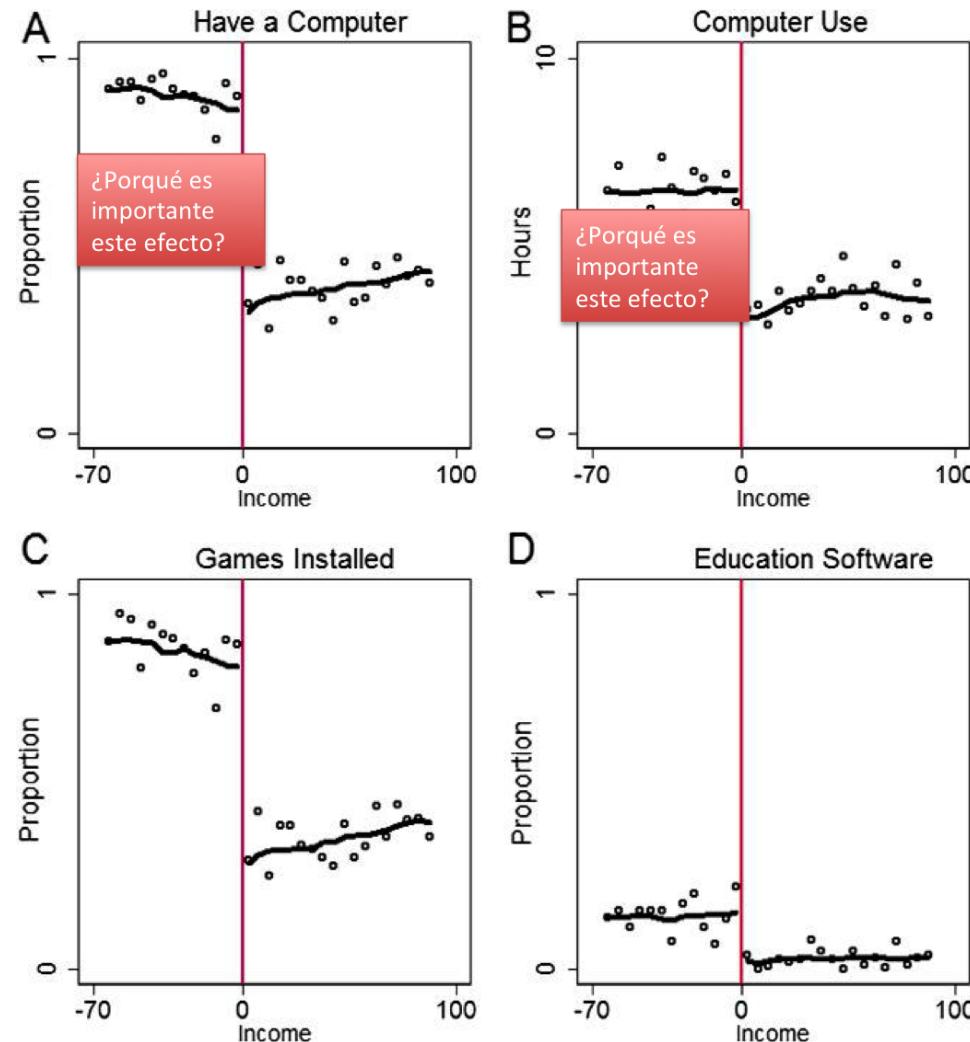
# Inspección visual: Resultados de Corto Plazo



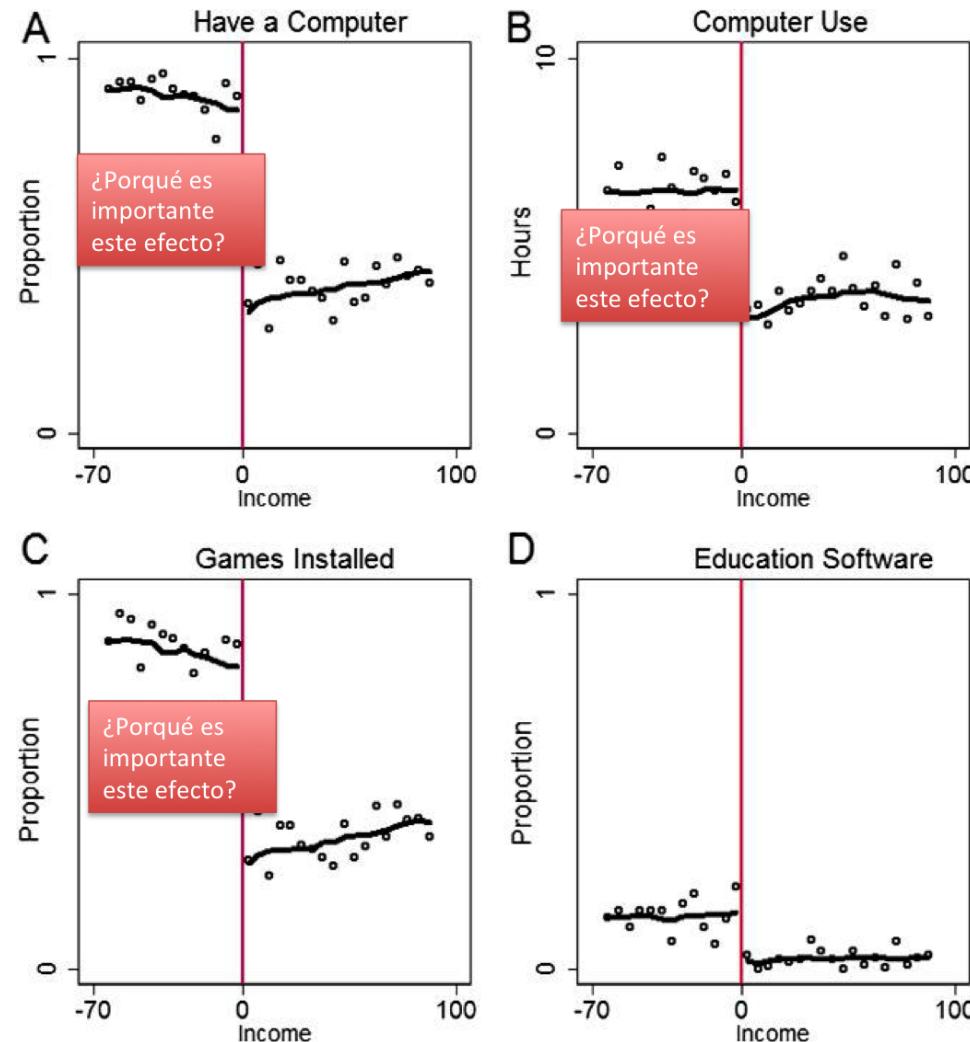
# Inspección visual: Resultados de Corto Plazo



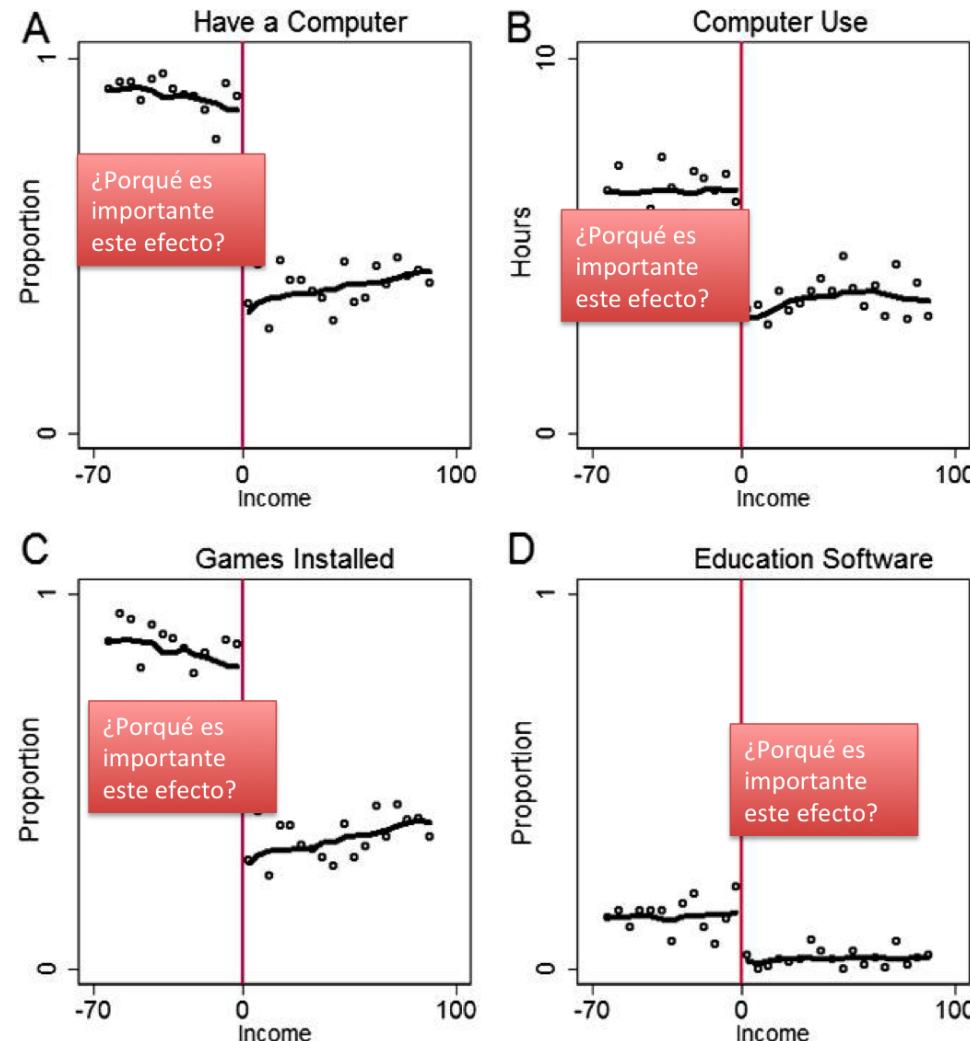
# Inspección visual: Resultados de Corto Plazo



# Inspección visual: Resultados de Corto Plazo



# Inspección visual: Resultados de Corto Plazo



# Estimación

<i>dependent variable</i>	Panel A: Children Survey				
	Have a Computer (1)	Computer Use (2)	Internet (3)	Games Installed (4)	Educational Software (5)
Nonparametric	0.546*** [0.060]	3.407*** [0.754]	0.011 [0.049]	0.497*** [0.058]	0.117** [0.046]
Observations	1,070	1,027	1,069	1,053	975
Nonparametric	0.518*** [0.048]	3.352*** [0.725]	0.029 [0.060]	0.476*** [0.056]	0.187*** [0.066]
CV Bandwidth	1,818	1,138	821	1,169	465
Observations	199	310	113	210	111
Nonparametric	0.699*** [0.119]	4.093*** [1.227]	0.066 [0.094]	0.751*** [0.109]	0.036 [0.089]
IK Bandwidth	Quadratic Spline	Parametric	0.068 [0.044]	0.491*** [0.054]	0.137*** [0.040]
Observations	4,510	4,295	4,503	4,348	4,001

# Inspección visual: Resultados Intermedios

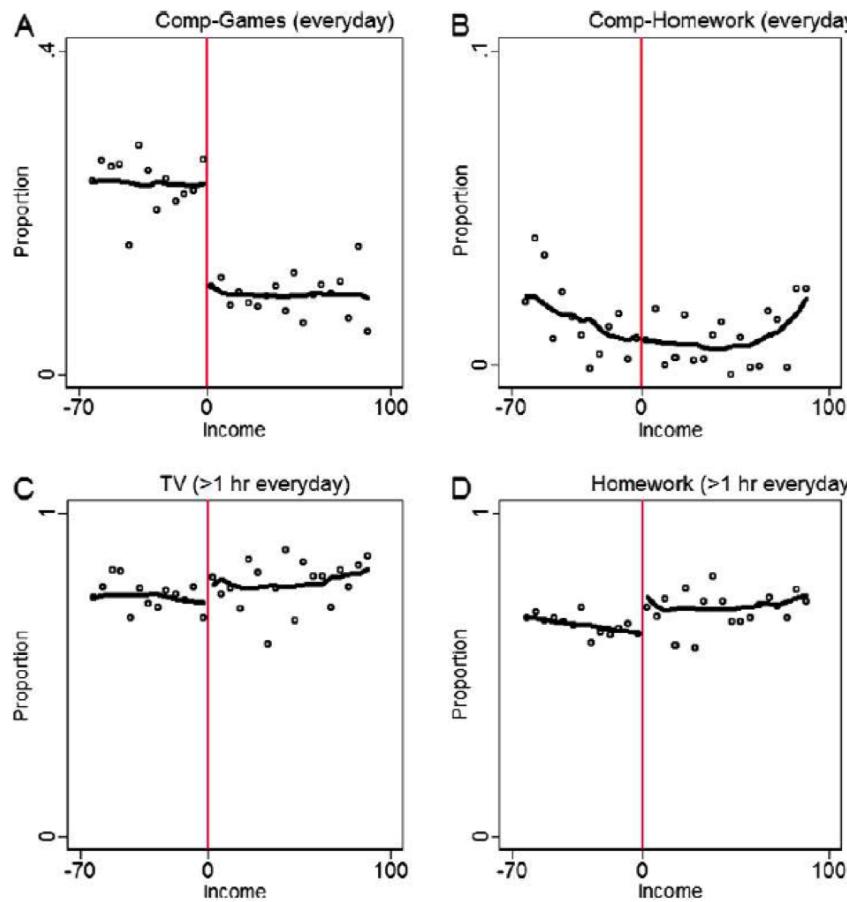


FIGURE II  
Time Use

# Inspección visual: Resultados Finales

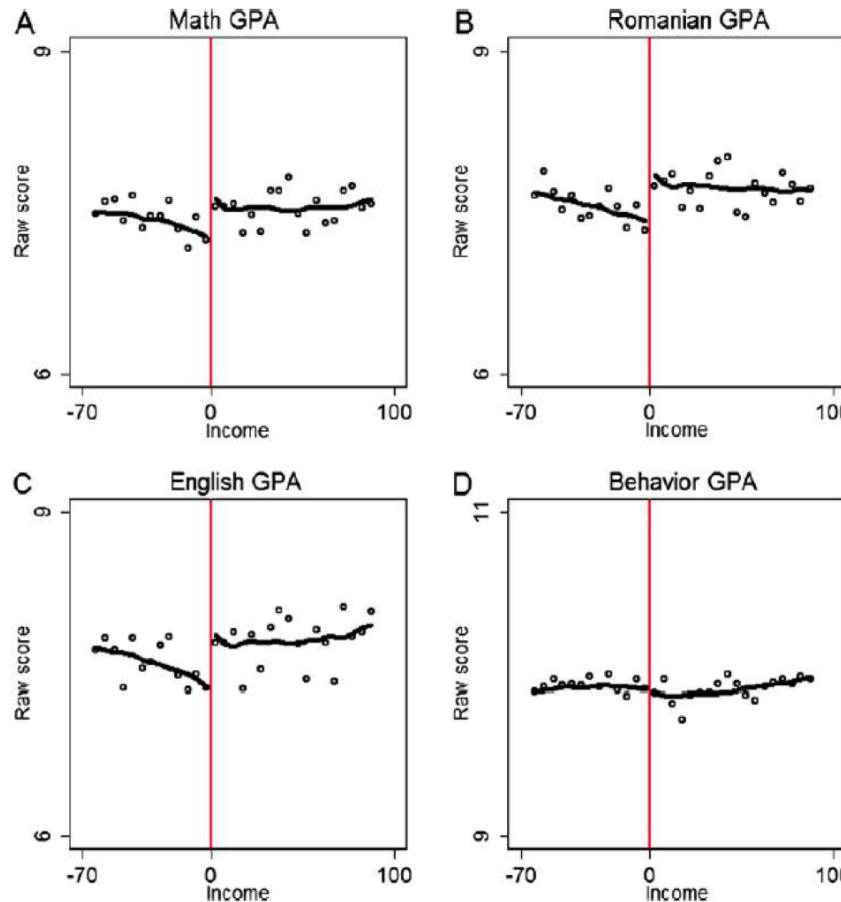


FIGURE III  
Academic Achievement

# Estimación

<i>dependent variable</i>	Panel A: Children Survey			
	Math GPA (1)	Romanian GPA (2)	English GPA (3)	Behavior GPA (4)
Nonparametric	-0.435** [0.171]	-0.562*** [0.181]	-0.634*** [0.225]	0.008 [0.070]
Observations	998	1,002	804	1,018
Nonparametric	-0.411** [0.179]	-0.313** [0.153]	-0.343* [0.193]	-0.047 [0.081]
CV Bandwidth				
Observations	1,077	1,372	1,287	859
Nonparametric	-0.669** [0.329]	-1.090*** [0.320]	-0.683 [0.449]	-0.229 [0.188]
IK Bandwidth				
Observations	263	266	216	207
Parametric	-0.368** [0.158]	-0.392** [0.165]	-0.473** [0.203]	0.014 [0.064]
Quadratic Spline				
Observations	4,179	4,201	3,387	4,266

# Inspección visual: Resultados Finales 2

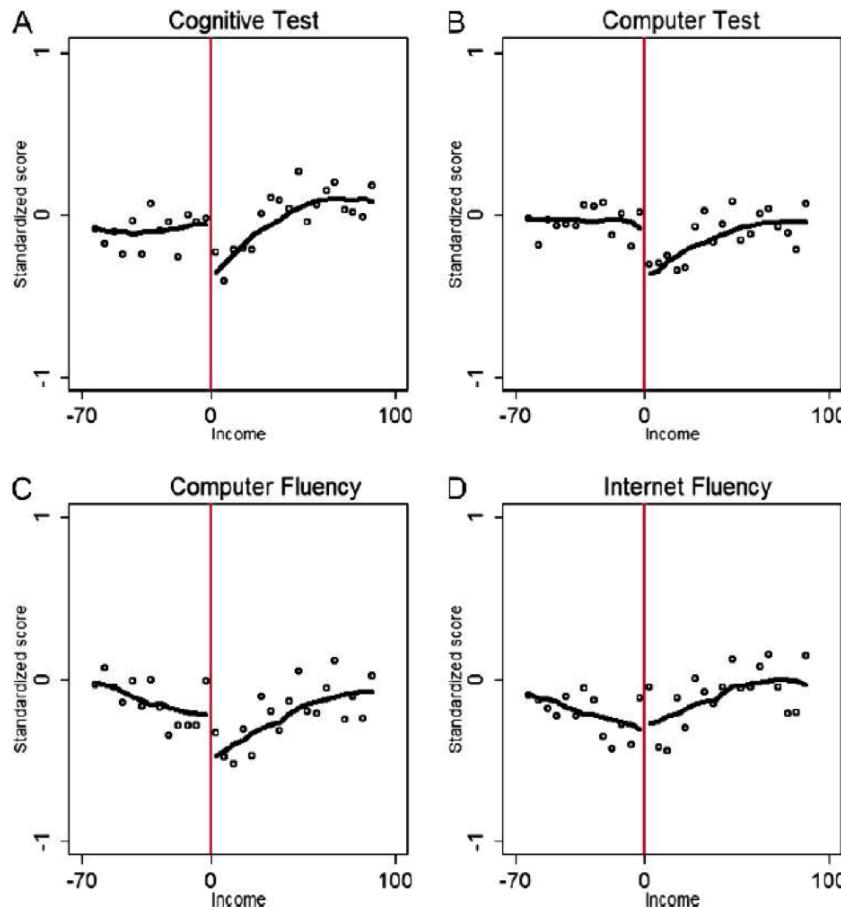


FIGURE IV  
Cognitive and Computer Skills

# Estimación

<i>dependent variable</i>	Raven's Progressive Matrices Test	Computer Test	Children Survey	
	(1)	(2)	Computer Fluency (3)	Applications Fluency (4)
Nonparametric	0.327** [0.134]	0.268** [0.114]	0.212* [0.118]	0.250** [0.118]
Bandwidth - 30				
Observations	1,084	1,027	1,082	1,082
Nonparametric	0.306** [0.125]	0.237* [0.121]	0.214* [0.124]	0.213** [0.104]
CV Bandwidth				
Observations	1,247	887	1,006	1,320
Nonparametric	0.030 [0.302]	0.579*** [0.216]	0.254 [0.272]	0.077 [0.261]
IK Bandwidth				
Observations	265	242	223	254
Parametric	0.381*** [0.120]	0.331*** [0.099]	0.244** [0.103]	0.277*** [0.104]
Quadratic Spline				
Observations	4,532	4,276	4,516	4,516

# Heterogeneidad por reglas sobre uso de PC

<i>dependent variable</i>	Computer Use (1)	Homework (2)	Math GPA (3)	Romanian GPA (4)	English GPA (5)	Raven's Test (6)	Computer Test (7)	Computer Fluency (8)	Applications Fluency (9)
<b>Panel A</b>									
Winner	3.525*** [0.481]	-0.097** [0.038]	-0.258** [0.104]	-0.418*** [0.109]	-0.358*** [0.133]	0.119 [0.081]	0.263*** [0.069]	0.172** [0.069]	-0.028 [0.070]
Computer Rules	4.071*** [0.389]	0.044 [0.028]	0.250*** [0.083]	0.294*** [0.080]	0.320*** [0.092]	0.105 [0.069]	0.322*** [0.056]	0.341*** [0.056]	0.310*** [0.060]
Winner*Computer Rules	-3.231*** [0.471]	0.046 [0.034]	0.003 [0.097]	-0.027 [0.095]	-0.064 [0.113]	-0.014 [0.080]	-0.188*** [0.066]	-0.097 [0.065]	-0.106 [0.070]
Observations	4,268	4,412	4,155	4,176	3,368	4,504	4,255	4,488	4,488
<b>Panel B</b>									
Winner	3.055*** [0.507]	-0.104*** [0.039]	-0.251** [0.107]	-0.443*** [0.111]	-0.418*** [0.137]	0.147* [0.082]	0.244*** [0.071]	0.174** [0.070]	-0.019 [0.071]
Homework Rules	0.644* [0.340]	0.061** [0.027]	0.061 [0.075]	-0.014 [0.077]	-0.090 [0.092]	0.071 [0.059]	0.086 [0.052]	0.085* [0.052]	0.093* [0.053]
Winner*Homework Rules	0.204 [0.434]	0.085*** [0.032]	0.130 [0.090]	0.215** [0.093]	0.340*** [0.113]	-0.016 [0.073]	0.012 [0.065]	0.055 [0.062]	0.011 [0.064]
Observations	4,182	4,335	4,112	4,133	3,335	4,411	4,165	4,395	4,395

# Conclusiones

- Estudiantes que reciben el computador tienen peores notas en el colegio: de casi 1/3 de SD (efecto grande).
- Pero reportan mejores resultados en tests de habilidades para utilizar un computador (computer fluency)
- Algo de evidencia de que los ganadores del voucher mejoran sus habilidades cognitivas (test de Raven)
- Las reglas parentales sirven como mediadores de estos efectos: Ej., reglas sobre tareas median los efectos negativos del computador en las notas
- El efecto neto de entregar computadores es incierto: **¿Porqué?**

## 2. Diferencias en Diferencias

# Diferencias en Diferencias (Experimentos Naturales)

- Cuando no existe un experimento controlado en donde se asignan aleatoriamente personas a un grupo de control o tratamiento, se buscan **experimentos naturales**
  - Por ejemplo, ¿cuál es el efecto de la contaminación sobre mortalidad? No podemos asignar aleatoriamente personas a distintos niveles de contaminación
- Los experimentos naturales son eventos exógenos que cambian el ambiente en que individuos, familias, empresas, ciudades, u otros, operan.
  - ¿Exógeno?
- Típicamente son una ley o política que cambia o fue implementada de manera diferenciada en distintos lugares o para distintas personas.
- La clave está en qué este experimento natural **causa** diferencias en nuestra variable de resultado, ya sea a través del espacio, tiempo, o individuos

# Identificación

- Una **estrategia de identificación** se refiere a cómo un investigador usa datos no experimentales para aproximarse a un experimento real.
- Una evaluación de impacto descansa en los **supuestos de identificación** para aislar el efecto de una política de otras variables que podrían confundir su interpretación (hint: sesgo de selección o sesgo de variable omitida)
  - ¿Cuál es el supuesto de identificación en un modelo de MCO? (hint: MELI)
- La parte difícil en métodos no-experimentales es decidir cuáles supuestos de identificación son posibles y cuáles no.

# Datos de panel

- Muchos métodos no experimentales utilizan Una **datos de panel**
  - Datos de corte transversal: observaciones en un período del tiempo
  - Datos de panel: múltiples observaciones para una misma unidad en el tiempo
  - Datos de corte transversal repetidos: Similar a datos de panel pero se colectan para distintas unidades (personas)
  - ¿Ejemplos en Chile?
- Algunas preguntas que se pueden responder con datos de panel:
  - ¿Cuál es el impacto de la regulación ambiental sobre mortalidad infantil?
  - ¿Cómo afecta una extensión del postnatal las decisiones laborales de las madres?
  - ¿Cuál es el impacto del plan AUGE sobre tasas de utilización de servicios de salud y salud?

# Estimador de diferencias

- Usualmente se observa (se tiene información) de personas antes de que entran a un programa y después de que el programa termina.
- Estos datos pueden ser muy útiles cuando queremos evaluar un programa.
- Un estimador posible del impacto del programa con estos datos es ver el **cambio** antes y después del programa en nuestra variable de resultado
  - Estimador de **diferencias** o el estimador "Antes y Despues"

# Estimador "Antes y Después"

- Supongamos que todas las personas fueron tratadas bajo un programa. El estimador de diferencias se construye como:

$$\Delta^d = \bar{Y}_{1,t_1}^T - \bar{Y}_{0,t_0}^T$$

- Esta notación es la que utilizaremos. Pero lo que se simplifica es:
  - $\bar{Y}_{1,t_1}^T$  es equivalente a  $E[Y_{1i}|D = 1, t = 1]$
  - $\bar{Y}_{0,t_0}^T$  es equivalente a  $E[Y_{0i}|D = 1, t = 0]$

# Estimador "Antes y Después"

- El estimador AD descansa en el **supuesto de identificación** de que el resultado de interés no cambia e la ausencia del programa.
  - ¿Qué concepto esta relacionado a este supuesto?

$$\begin{aligned}\Delta^d &= \bar{Y}_{1,t_1}^T - \bar{Y}_{0,t_0}^T \\ &= \bar{Y}_{1,t_1}^T - \bar{Y}_{0,t_0}^T + \bar{Y}_{0,t_1}^T - \bar{Y}_{0,t_1}^T \\ &= (\bar{Y}_{1,t_1}^T - \bar{Y}_{0,t_1}^T) + (\bar{Y}_{0,t_1}^T - \bar{Y}_{0,t_0}^T) \\ &= \Delta^{ATET} + \text{tendencia contrafactual del grupo de tratamiento}\end{aligned}$$

- El **supuesto de identificación** es que el segundo término es igual a cero.

# Estimador "Antes y Después"

- El segundo término es el contrafactual que no tenemos: ¿Qué hubiese pasado al grupo de tratamiento de no haber sido tratado? → no podemos revisar si el supuesto es válido.
  - Sin evidencia directa sobre la tendencia del contrafactual, uno podría mirar tendencias antes del programa para el grupo de tratamiento. En muchos casos, este supuesto no es razonable. ¿Porqué?
  - El estimador AD no es un buen estimador de impacto: incluye el efecto del programa y una tendencia contrafactual que "contamina" la estimación
  - Para mejorar el AD uno necesita simular esta tendencia contrafactual con un grupo de control: Esto nos lleva al estimador de **Diferencias en Diferencias**

# Estimador de Diferencias en Diferencias

- Necesitamos estimar la tendencia contrafactual para el grupo de tratamiento
- Un estimador posible es la tendencia observada para un grupo que no recibió el tratamiento o programa en un mismo período
  - El estimador de diferencias en diferencias se construye como:

$$\begin{aligned}\Delta^{DID} &= \Delta_{D=1}^d - \Delta_{D=0}^d \\ &= \underbrace{(\bar{Y}_{1,t_1}^T - \bar{Y}_{0,t_0}^T)}_{\text{Diferencias}} \underbrace{-}_{\text{en}} \underbrace{(\bar{Y}_{0,t_1}^C - \bar{Y}_{0,t_0}^C)}_{\text{Diferencias}}\end{aligned}$$

- El trabajo esta en encontrar un grupo de personas no tratadas y datos en los mismos períodos

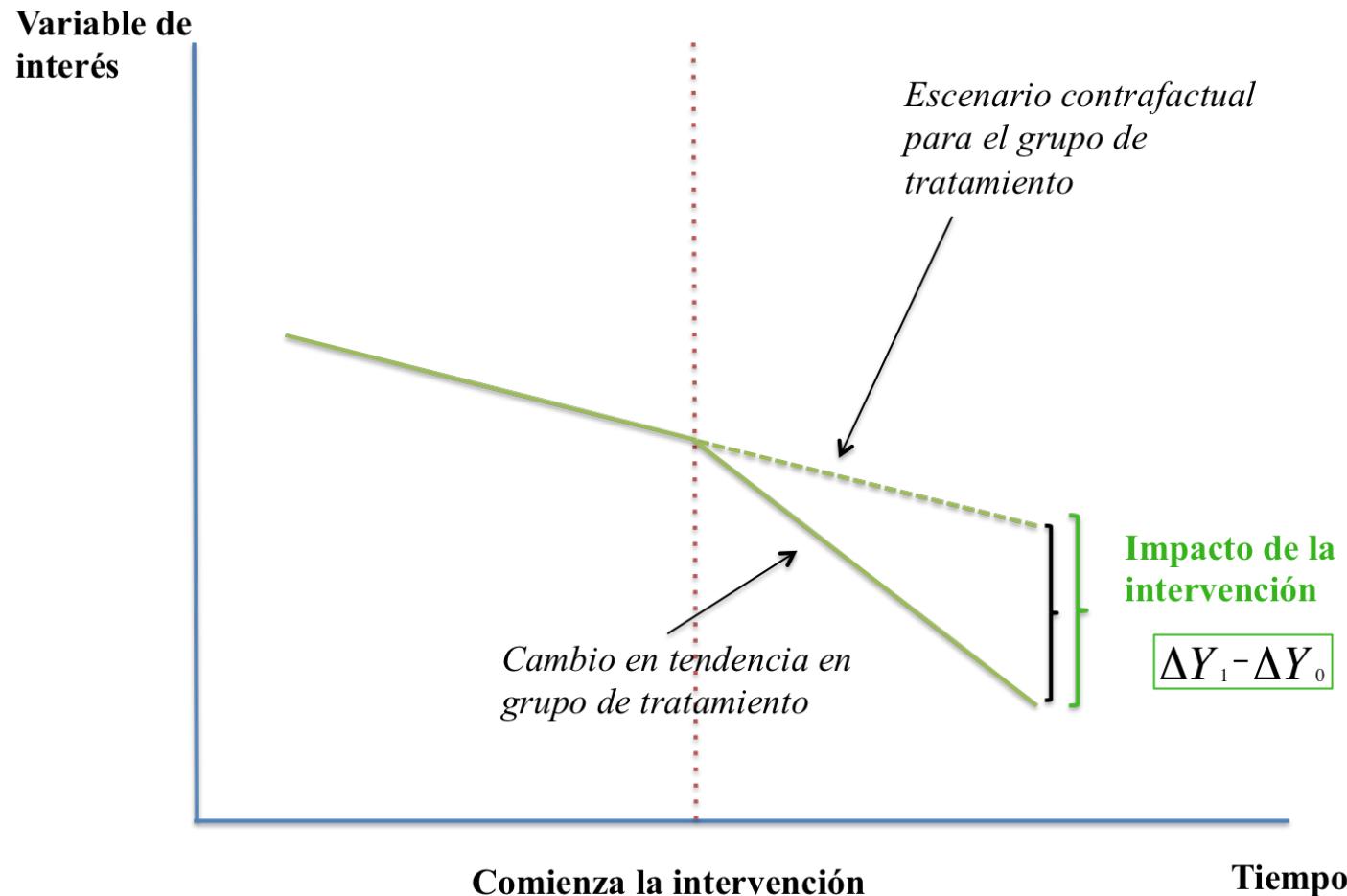
# Estimador de Diferencias en Diferencias

- El estimador DID descansa en el **supuesto de identificación** de que la tendencia observada del grupo de control es la misma que la tendencia contrafactual del tratamiento. ¿Qué tan bueno es mi grupo de control?

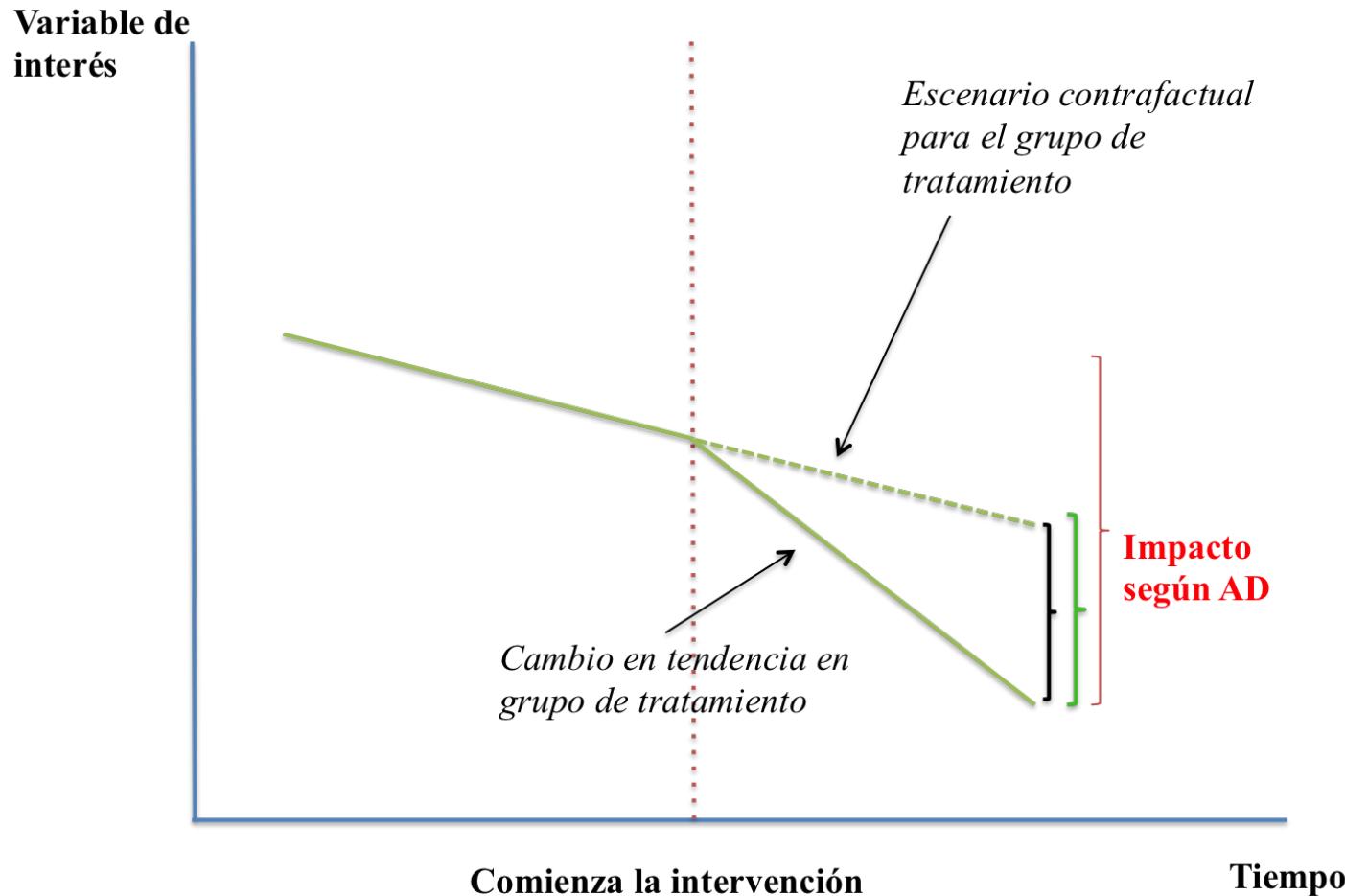
$$\begin{aligned}\Delta^{DID} &= (\bar{Y}_{1,t_1}^T - \bar{Y}_{0,t_0}^T) - (\bar{Y}_{0,t_1}^C - \bar{Y}_{0,t_0}^C) \\ &= (\bar{Y}_{1,t_1}^T - \bar{Y}_{0,t_0}^T) - (\bar{Y}_{0,t_1}^C - \bar{Y}_{0,t_0}^C) + \bar{Y}_{0,t_1}^T - \bar{Y}_{0,t_1}^T \\ &= (\bar{Y}_{1,t_1}^T - \bar{Y}_{0,t_1}^T) - (\bar{Y}_{0,t_1}^T - \bar{Y}_{0,t_0}^T) - (\bar{Y}_{0,t_1}^C - \bar{Y}_{0,t_0}^C) \\ &= \Delta^{ATET} + \text{tendencia contrafactual del grupo de tratamiento} \\ &\quad - \text{tendencia del grupo de control}\end{aligned}$$

- El **supuesto de identificación** es que ambos grupos tendrían la misma tendencia en la ausencia del tratamiento. Esto se llama **supuesto de tendencias paralelas**.

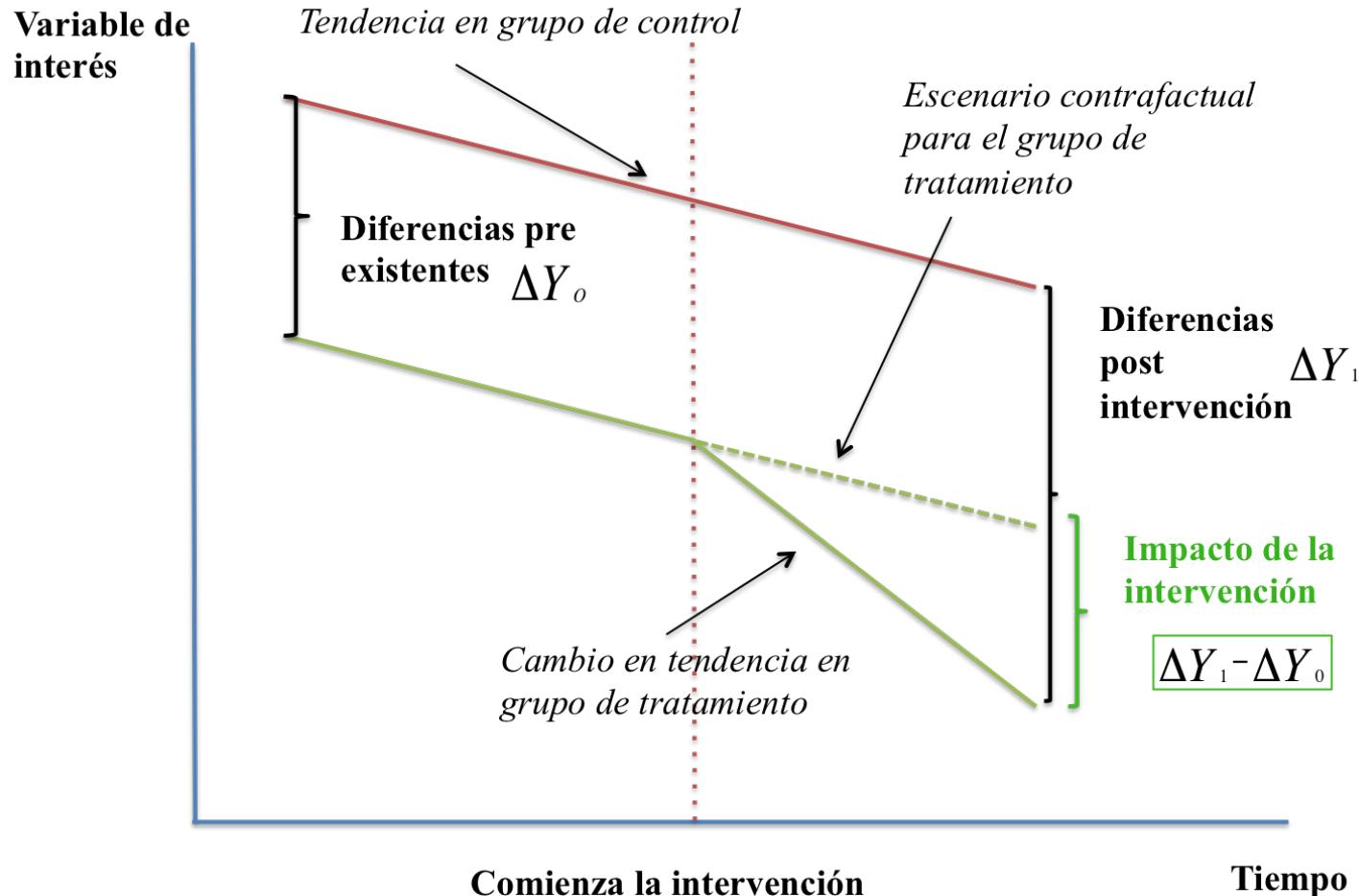
# Visualización: lo que queremos estimar



# Visualización: lo que estima AD



# Visualización: lo que estima DID



# Ejemplo 1: Brote de cólera en Londres en 1854

## John Snow (1854)

- Dos compañías de agua compiten por clientes en algunas regiones del sur del río Thames: Lambeth Company (LC) y Southwark and Vauxhall (SV)
- En 16 sub-distritos (300,000 habitantes), competían de manera directa y en algunos casos entregaban agua a los mismos hogares
- Cada compañía proveía de agua a grupos de hogares muy similares (ricos, pobres, hogares con alto y bajo número de habitantes)
- En 1849 cada compañía sacaba agua del Thames que estaba contaminado. En 1852, LC mueve su infraestructura río arriba en zonas del Thames con agua (más). SV no cambió nada
- Un experimento natural ideal

# Resultados principales

( 19 )

I.—SYNOPSIS OF RESULTS.

Death-Rates per 1,000 of living Population in Two Epidemic Periods.	In Houses enumerated in 1854 as receiving their Water-supply—	
	from the LAMBETH Company.	from the SOUTHWARK and VAUXHALL Company.
CHOLERA .....	1848-9..	12.5
	1853-4..	3.7
DIARRHOEA .....	1848-9..	2.9
	1853-4..	2.1

N.B.—Between the two Epidemic Periods, the Lambeth Water Company had changed its source of supply.

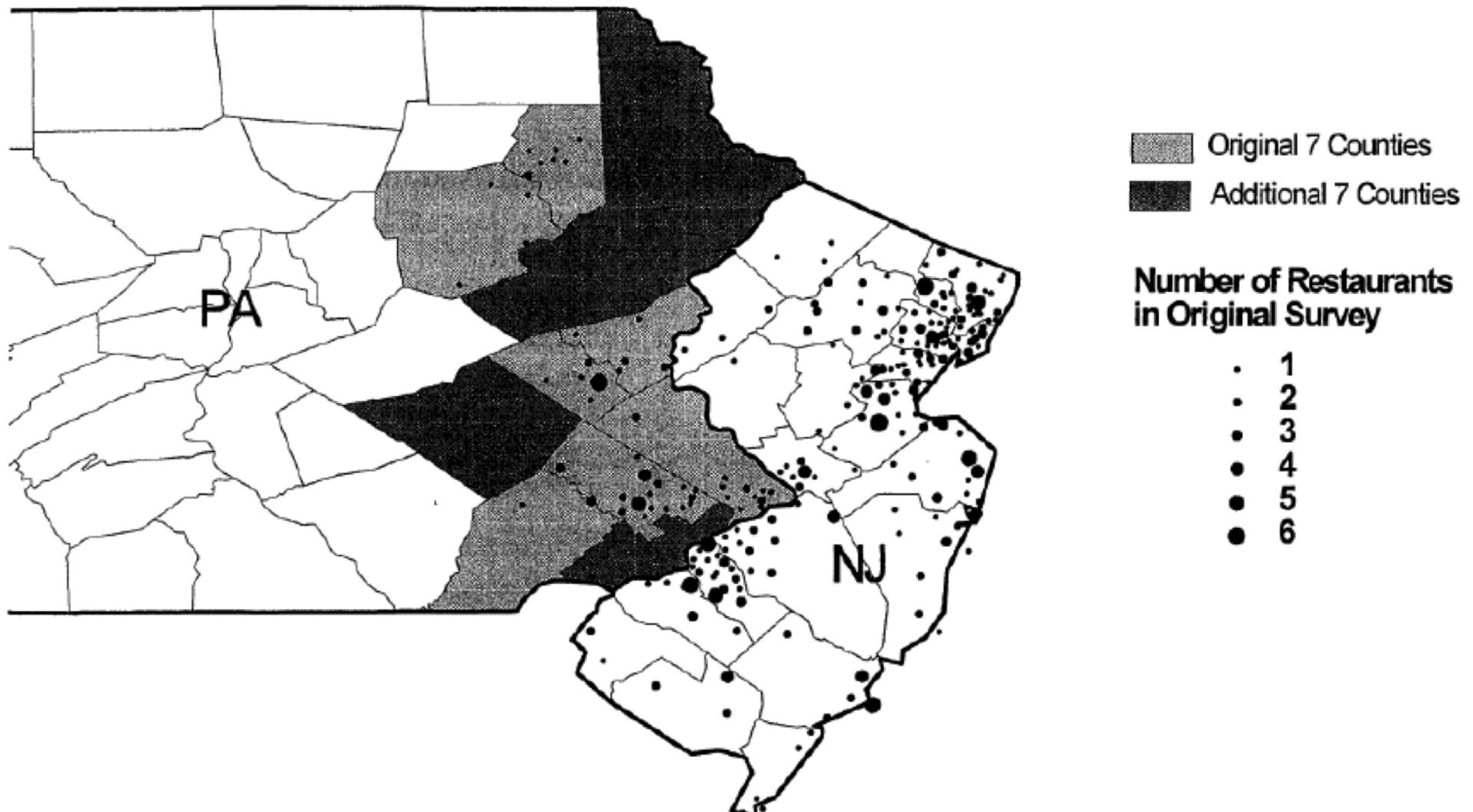
- Tenemos 2 grupos y 2 períodos que registran tasas de mortalidad por cólera y diarrea
  - Lambeth: grupo tratado
  - Southwalk: grupo de control
- El efecto sobre los tratados (ATT) es de 10 muertes menos por cada 1,000 habitantes
- Estimación simple comparando combinaciones de 2x2

# Ejemplo 2: Empleo en industria de comida rápida y salario mínimo

## Card y Krueger (1994)

- La teoría económica predice que subir el salario mínimo reduce el empleo
- Card y Krueger toman un experimento natural para revisar esta predicción
  - Abril 1, 1992: New Jersey sube su salario mínimo de 4.25 a 5.05
  - No hubo cambios en estados vecinos
- Oportunidad para un clásico DID
- Datos: empleo en restaurantes de comida rápida en NJ y PA antes (Feb 1992) y después (Nov 1992) de este aumento en salario mínimo

# Distribución de la muestra



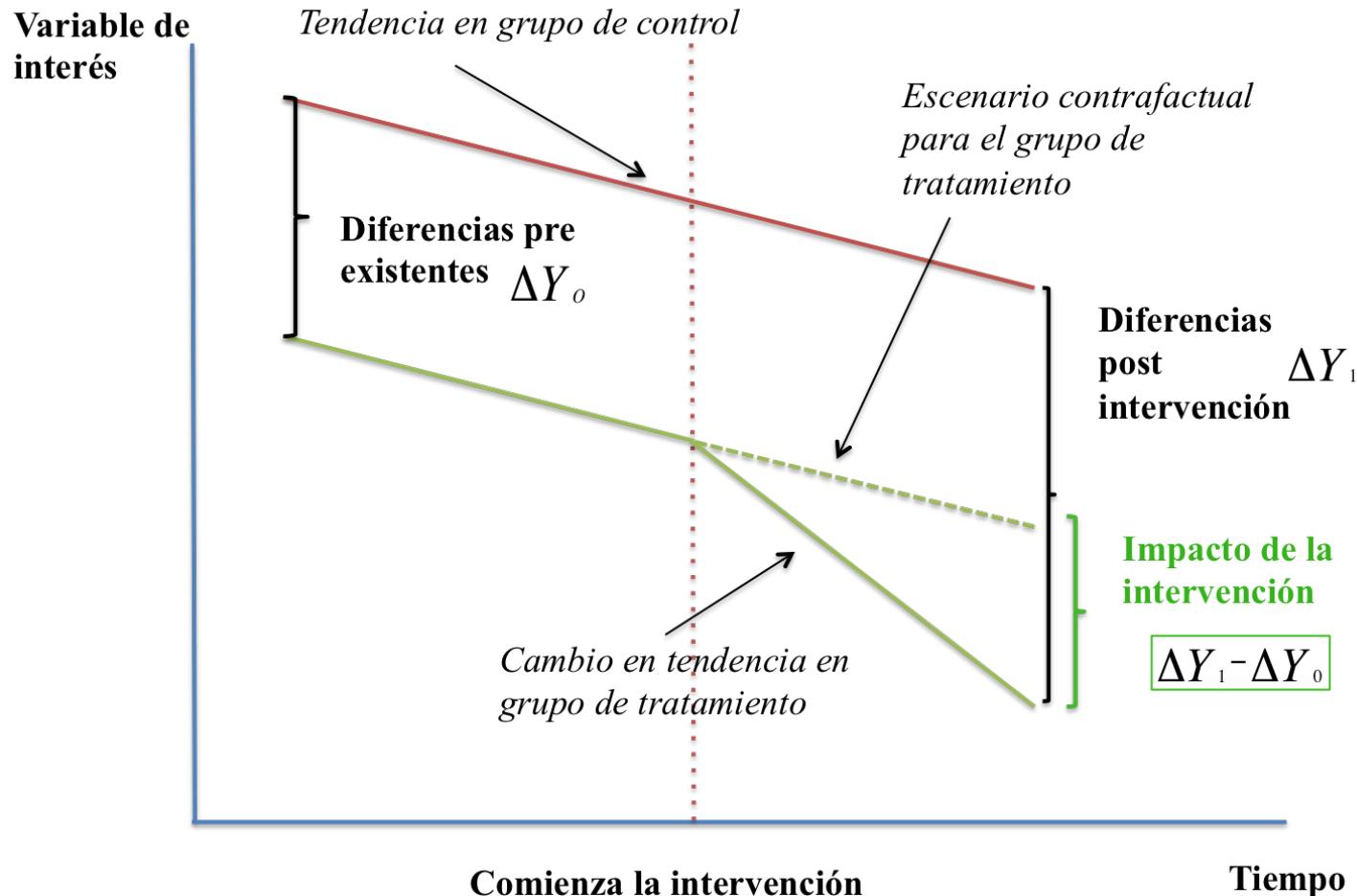
# Resultados principales

Variable	Stores by state		
	PA (i)	NJ (ii)	Difference, NJ – PA (iii)
1. FTE employment before, all available observations	23.33 (1.35)	20.44 (0.51)	-2.89 (1.44)
2. FTE employment after, all available observations	21.17 (0.94)	21.03 (0.52)	-0.14 (1.07)
3. Change in mean FTE employment	-2.16 (1.25)	0.59 (0.54)	2.76 (1.36)

# Supuestos de Card y Krueger

- ¿Cuál es el supuesto detrás de este análisis?
- ¿Porqué podría no cumplirse?

# Visualización: Lo que estima DID



# Supuestos de Card y Krueger

- **Supuesto de identificación:** El empleo en NJ y PA tendría la misma tendencia en la ausencia del aumento en el salario mínimo (tratamiento). ¿Porqué no se cumpliría?

**1** ¿Qué pasa si PA tiene distintas tendencias que NJ desde antes de la política? → no se cumple el supuesto de tendencias paralelas

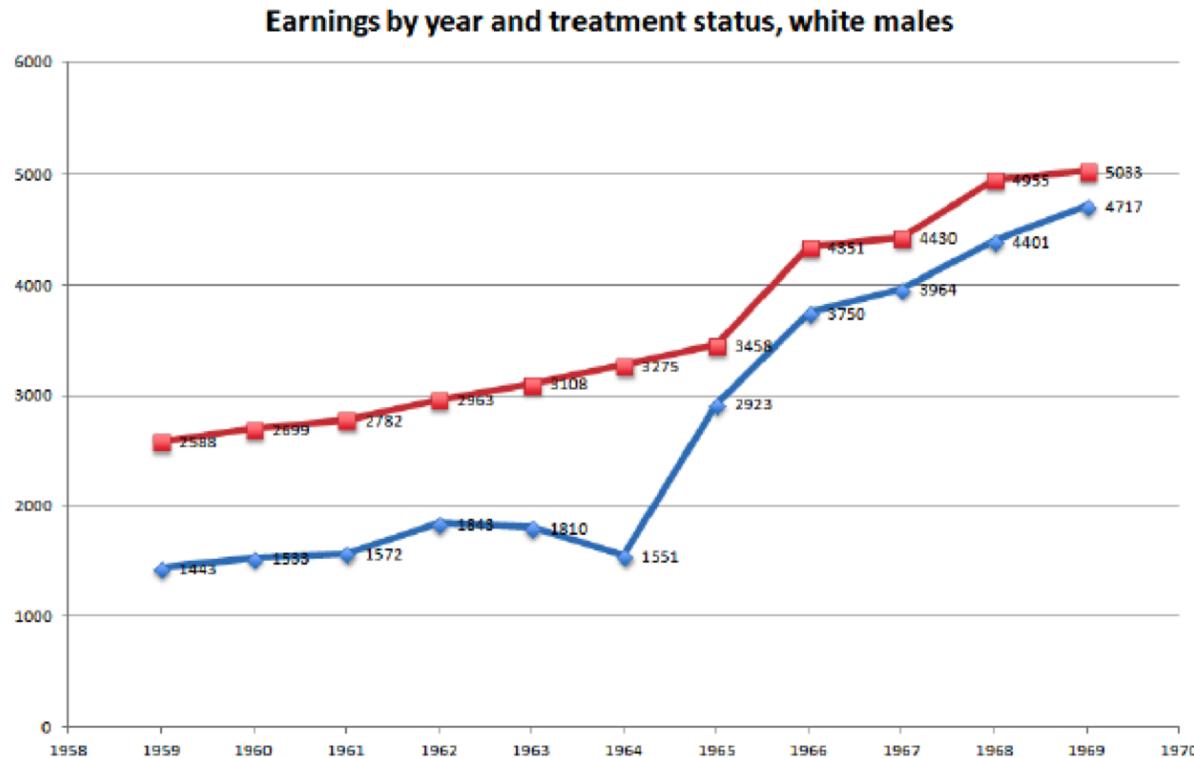
- Revisar si las tendencias son similares antes del programa

**2** ¿Qué pasa si otras cosas pasaron en NJ simultáneamente? → no podemos interpretar el cambio como un efecto del salario mínimo (e.g. programas simultáneos de capacitación laboral)

- Revisar si no hay otros programas en paralelo

# Ejemplo de problemas con tendencias paralelas

- Un ejemplo clásico de no cumplimiento del supuesto de tendencias paralelas se da en programas de capacitación laboral. En el proceso de auto-selección a un programa así:
- ¿cómo se espera que sean los salarios de quienes se inscriben y no se inscriben?



# Efectos de programa de capacitación

- Podemos construir distintos estimadores DID cambiando el año basal: no deberían cambiar

Year	Trainees	Comparisons	$\Delta_{1966}^{DID}$ by base year
1959	1443	2588	544
1960	1533	2699	565
1961	1572	2782	609
1962	1843	2963	519
1963	1810	3108	697
1964	1551	3275	1123
1965	2923	3458	
1966	3750	4351	

- DID son similares para años basales pre-1963.
- El supuesto parece ser más sostenible para años pre 1963. Estos deberían ocuparse para

# Resumen

- Los supuestos de identificación son distintos para distintos estimadores que ocupan datos no experimentales:
- Estimadores de corte transversal: no hay diferencias sistemáticas entre ambos grupos
- Estimadores de diferencia: no hay tendencia contrafactual en la ausencia del tratamiento, o la variable de interés se mantiene constante en el tiempo en ausencia de programa
- Estimadores de diferencias en diferencias: La tendencia observada en el grupo de control es la misma que la tendencia del contrafactual del tratamiento

# Importante

- Estos son supuestos. No podemos (usualmente) probar si son correctos o no
- Esta es la diferencia fundamental entre métodos experimentales y no experimentales
- Para DID debemos argumentar los supuestos de identificación en base a:
- Tendencias paralelas en períodos previos al programa
- No hay otras intervenciones pasando al mismo tiempo
- Sin embargo, esto solo hace el supuesto más creíble ... **no lo prueba**

# Diferencias en Diferencias en formato de regresión

$$Y_{it} = \beta_0 + \beta_1 I_{it1} + \beta_2 D_i + \delta D_i \cdot I_{it1} + \varepsilon_{it}$$

- Dos periodos:  $t = t_0, t_1$
- $I_{it1} = 1$  para  $t = t_1$  y  $I_{it0} = 0$  para  $t = t_0$
- $D_i = 1$  para el grupo de tratamiento y  $D_i = 0$  para el grupo de control
- Entonces  $\delta = \Delta^{DID}$
- Demostración de  $\hat{\delta} = (\bar{Y}_{1,t_1}^T - \bar{Y}_{0,t_0}^T) - (\bar{Y}_{0,t_1}^C - \bar{Y}_{0,t_0}^C)$

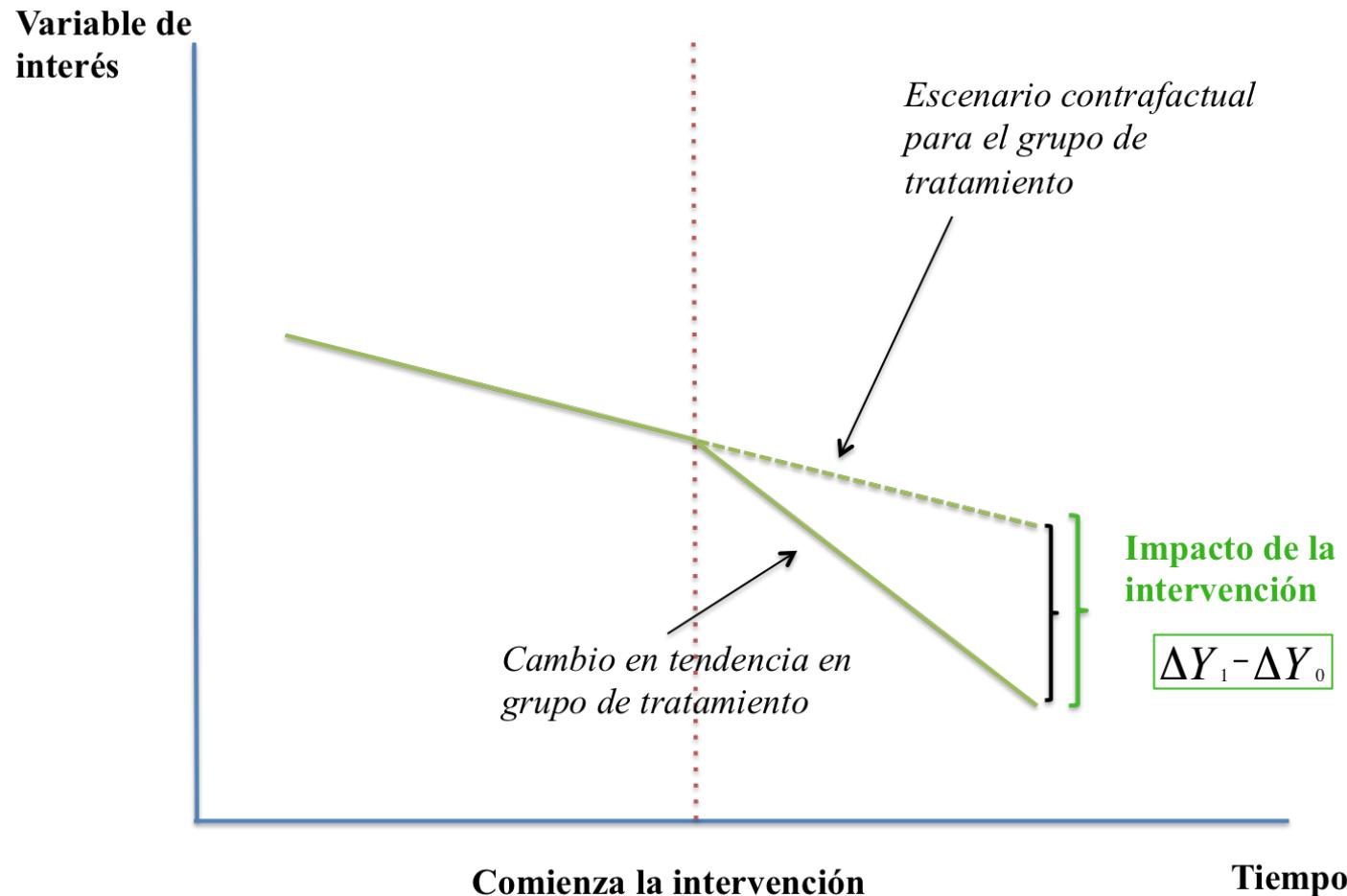
# Diferencias en Diferencias en formato de regresión

- Podemos incluir variables observables y controlar por ellas

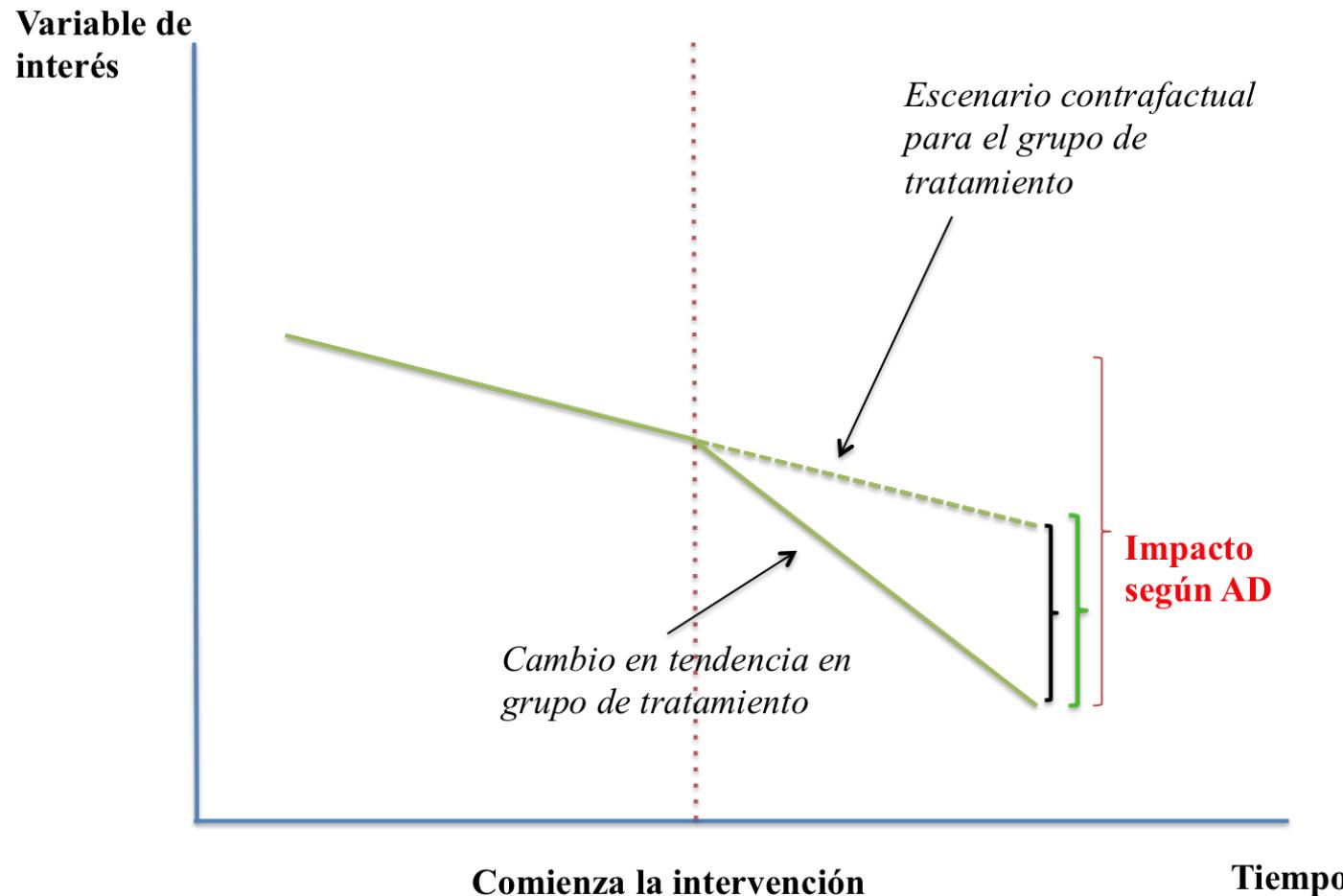
$$Y_{it} = \beta_0 + \beta_1 I_{it1} + \beta_2 D_i + \delta D_i \cdot I_{it1} + X'_{it} \lambda + \varepsilon_{it}$$

- Esto solo se hace para corregir errores estándar o si creemos que hay alguna  $X$  relacionada con  $D_i$
- Se espera que al agregar variables de control no haya cambios significativos en  $\hat{\delta}$

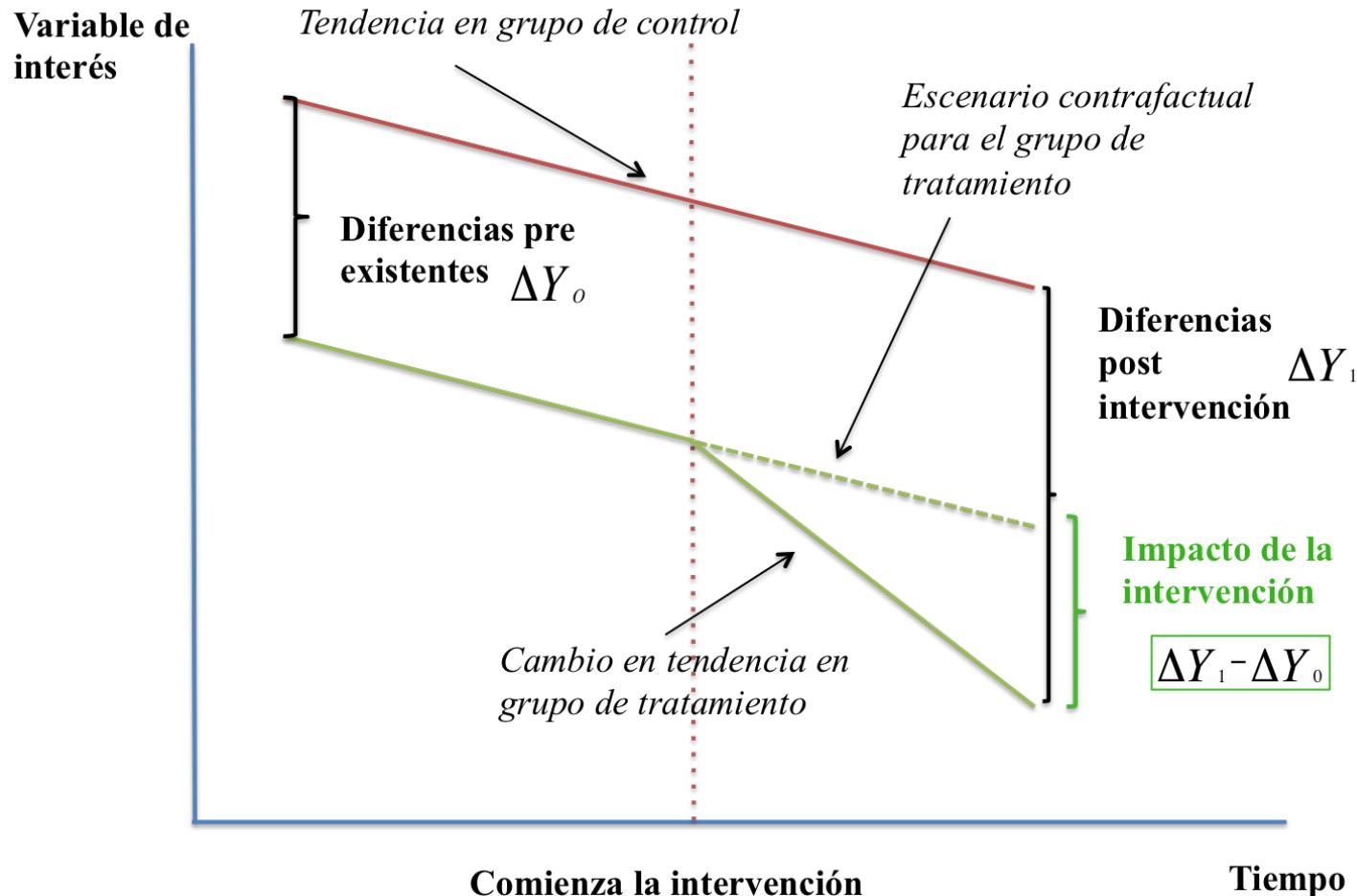
# Visualización: Lo que queremos estimar



# Visualización: Lo que estima AD



# Visualización: Lo que estima DID



# Visualización: Horizonte temporal del problema

## Do Police Reduce Crime? Estimates Using the Allocation of Police Forces After a Terrorist Attack

By RAFAEL DI TELLA AND ERNESTO SCHARGRODSKY\*

*An important challenge in the crime literature is to isolate causal effects of police on crime. Following a terrorist attack on the main Jewish center in Buenos Aires, Argentina, in July 1994, all Jewish institutions received police protection. Thus, this hideous event induced a geographical allocation of police forces that can be presumed exogenous in a crime regression. Using data on the location of car thefts before and after the attack, we find a large deterrent effect of observable police on crime. The effect is local, with no appreciable impact outside the narrow area in which the police are deployed. (JEL K42)*

# Visualización: Horizonte temporal del problema

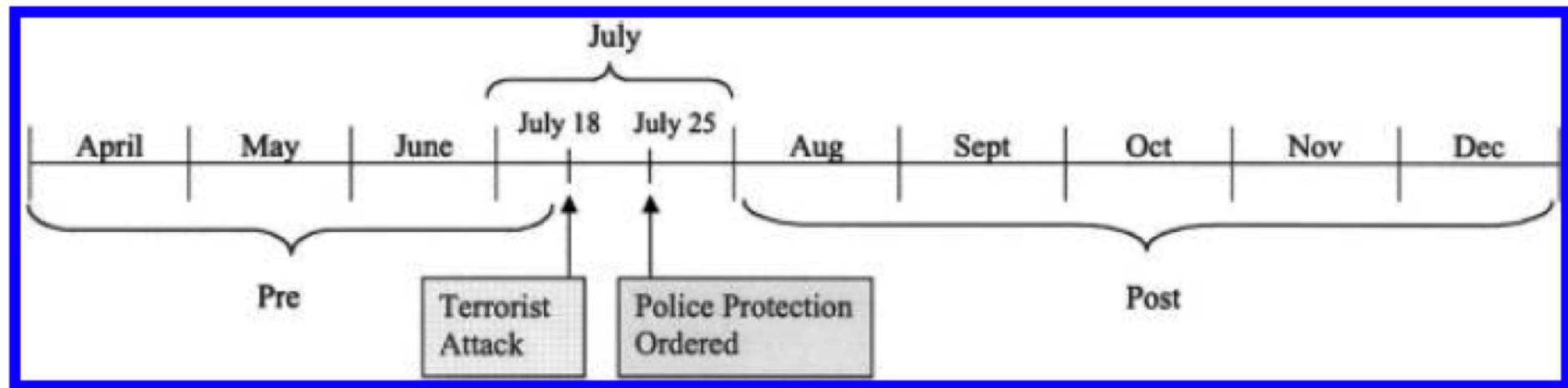
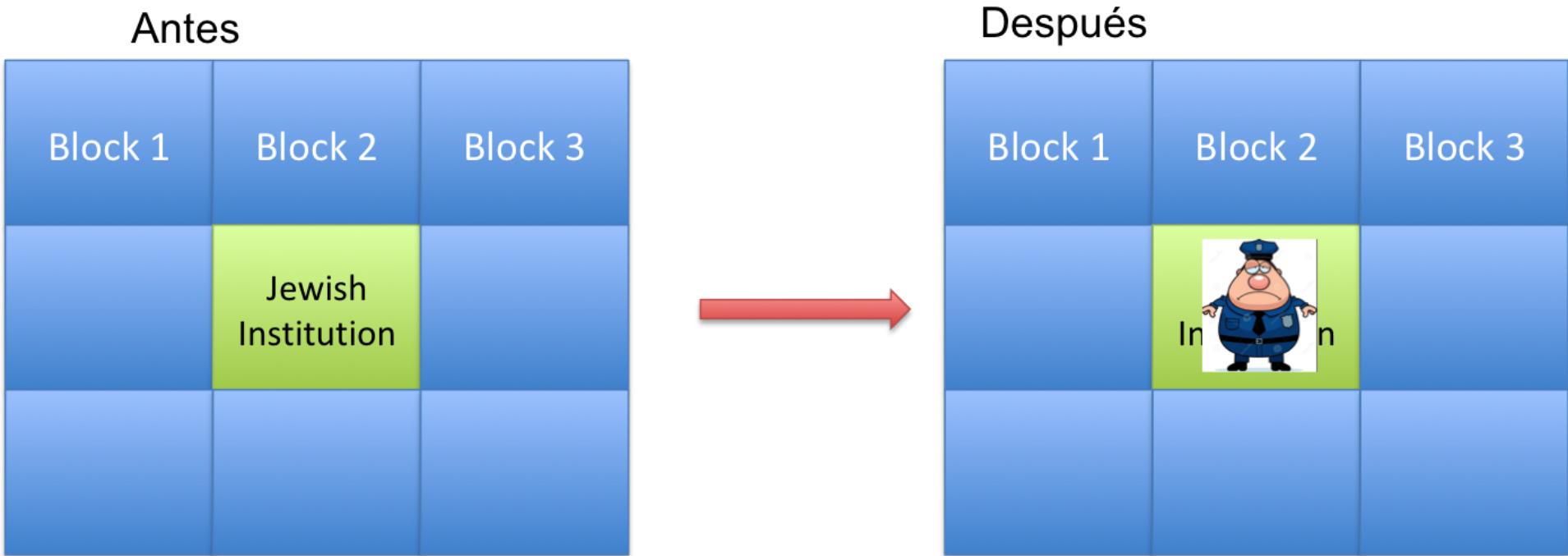


FIGURE 1. TIMELINE OF EVENTS

# Visualización: Horizonte temporal del problema



# Características de los segmentos censales

TABLE 1—DEMOGRAPHIC CHARACTERISTICS OF CONTROL AND TREATMENT AREAS

Demographic characteristics	Census tracts without Jewish institutions (A)	Census tracts with Jewish institutions (B)	Difference (C) = (A) – (B)
Home ownership rate	0.696 (0.008)	0.663 (0.017)	0.032 (0.019)
Overcrowding rate	0.014 (0.001)	0.017 (0.002)	-0.002 (0.003)
Poverty rate	0.042 (0.003)	0.052 (0.008)	-0.010 (0.009)
Education of household head	11.653 (0.147)	11.052 (0.300)	0.600 (0.335)
Number of household members	2.719 (0.023)	2.685 (0.054)	0.034 (0.059)
Female population	0.556 (0.001)	0.552 (0.003)	0.003 (0.003)
Unemployment rate	0.053 (0.001)	0.059 (0.003)	-0.005 (0.003)
Age	38.005 (0.128)	37.690 (0.223)	0.315 (0.258)
Number of census tracts	53	14	

¿Por qué mostrar esta tabla? ¿De qué nos están tratando de convencer los autores?

# Evolución de robo de autos en el tiempo

TABLE 2—MONTHLY EVOLUTION OF CAR THEFT

Month	More than two blocks from nearest Jewish institution (A)	Jewish institution on the block (B)	One block from nearest Jewish institution (C)	Two blocks from nearest Jewish institution (D)	Difference (E) = (B) – (A)	Difference (F) = (C) – (A)	Difference (G) = (D) – (A)
April	0.09955 (0.248)	0.12162 (0.361)	0.12111 (0.287)	0.12278 (0.297)	0.02206 (0.060)	0.02156 (0.025)	0.02323 (0.022)
May	0.10840 (0.235)	0.08783 (0.205)	0.07763 (0.181)	0.09734 (0.259)	-0.02056 (0.035)	-0.03076 (0.018)	-0.01106 (0.020)
June	0.07853 (0.196)	0.12837 (0.286)	0.07763 (0.215)	0.06969 (0.186)	0.04983 (0.047)	-0.00090 (0.019)	-0.00884 (0.015)
July (1–17)	0.03926 (0.145)	0.02027 (0.069)	0.05900 (0.210)	0.03097 (0.141)	-0.01899 (0.013)	0.01973 (0.017)	-0.00829 (0.011)
July (18–31)	0.03926 (0.146)	0.02702 (0.078)	0.07298 (0.217)	0.06858 (0.238)	-0.01224 (0.014)	0.03371 (0.018)	0.02931 (0.017)
August	0.11836 (0.287)	0.04729 (0.175)	0.06677 (0.219)	0.12721 (0.304)	-0.07106 (0.031)	-0.05159 (0.021)	0.00884 (0.024)
September	0.10176 (0.256)	0.01351 (0.057)	0.09006 (0.276)	0.09845 (0.248)	-0.08825 (0.015)	-0.01170 (0.024)	-0.00331 (0.020)
October	0.12112 (0.267)	0.06081 (0.215)	0.09782 (0.260)	0.08849 (0.236)	-0.06031 (0.037)	-0.02330 (0.024)	-0.03263 (0.020)
November	0.09623 (0.240)	0.02702 (0.078)	0.11024 (0.288)	0.10176 (0.217)	-0.06921 (0.017)	0.01400 (0.025)	0.00553 (0.018)
December	0.10176 (0.268)	0.02702 (0.078)	0.11645 (0.278)	0.10619 (0.225)	-0.07474 (0.018)	0.01468 (0.025)	0.00442 (0.019)
Number of blocks	452	37	161	226			

# Visualización de los efectos

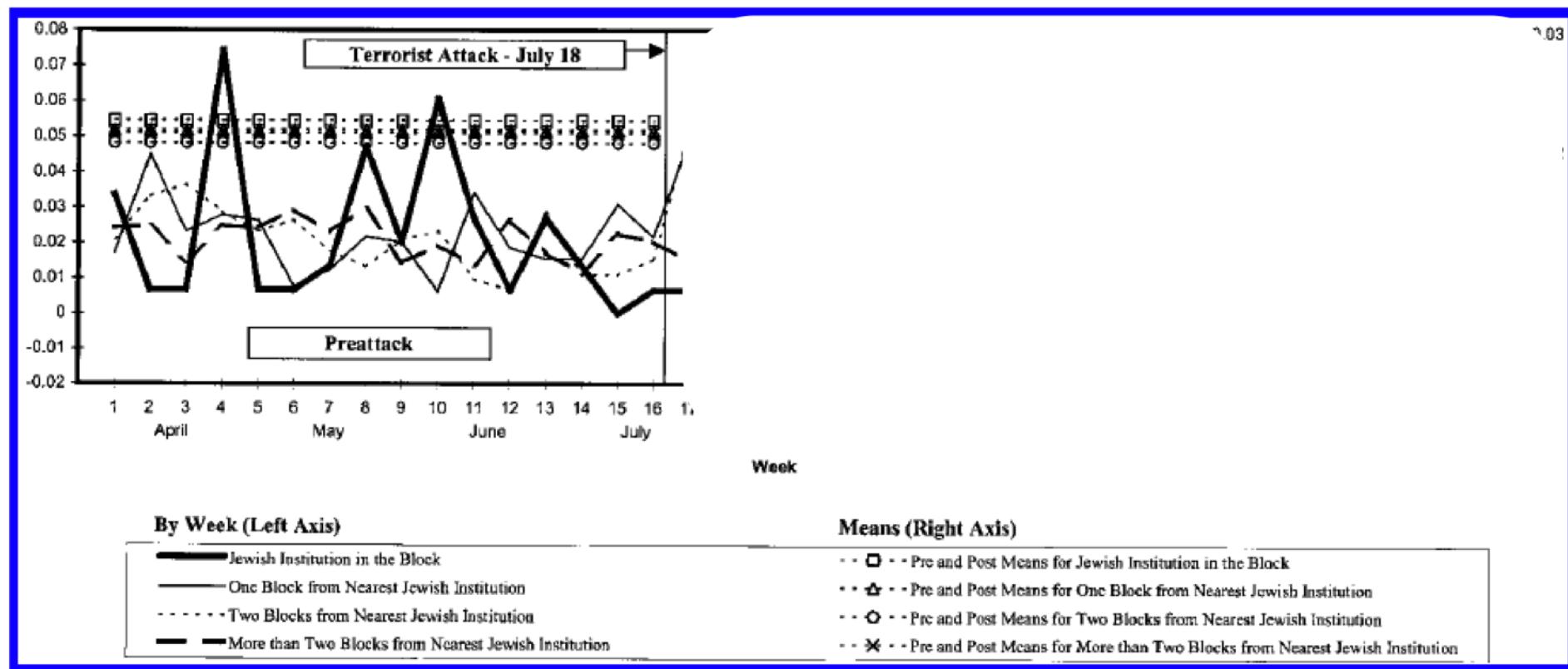


FIGURE 2. WEEKLY EVOLUTION OF CAR THEFTS

# Visualización de los efectos

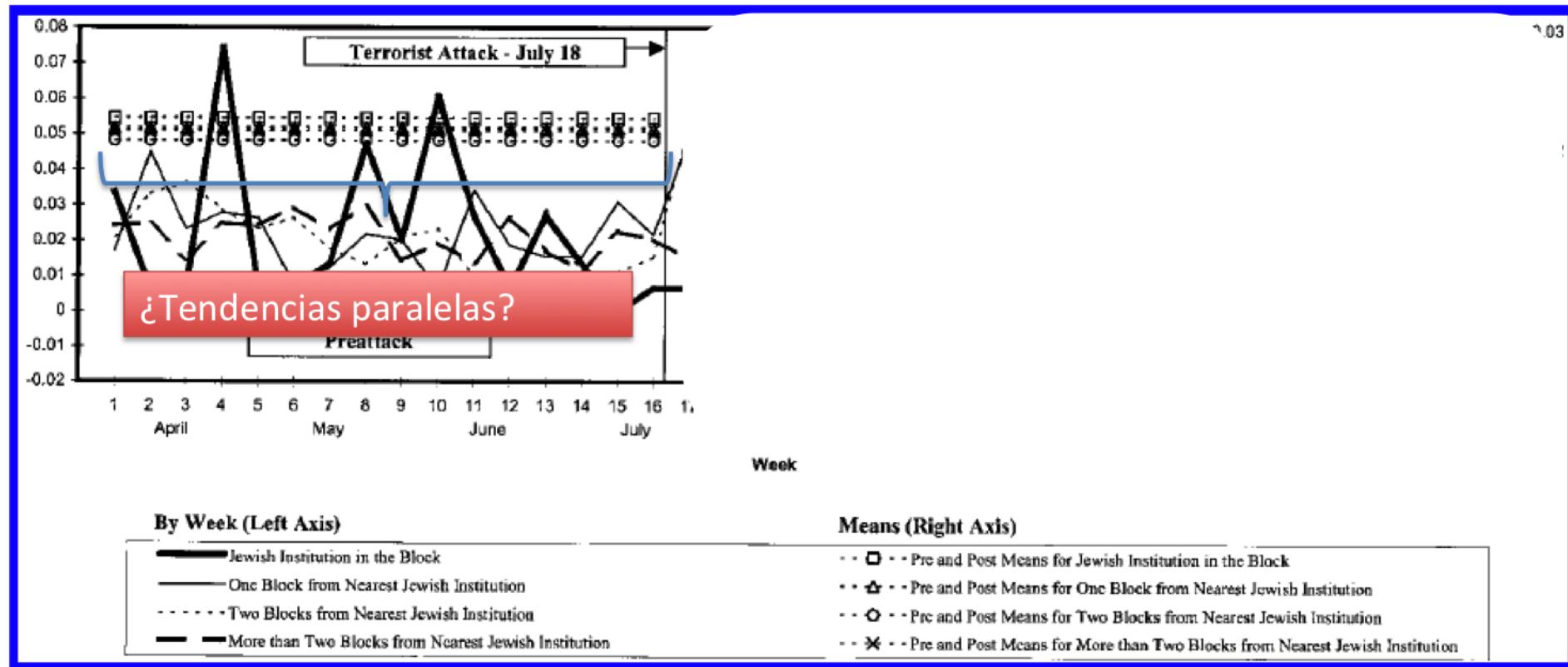


FIGURE 2. WEEKLY EVOLUTION OF CAR THEFTS

# Visualización de los efectos

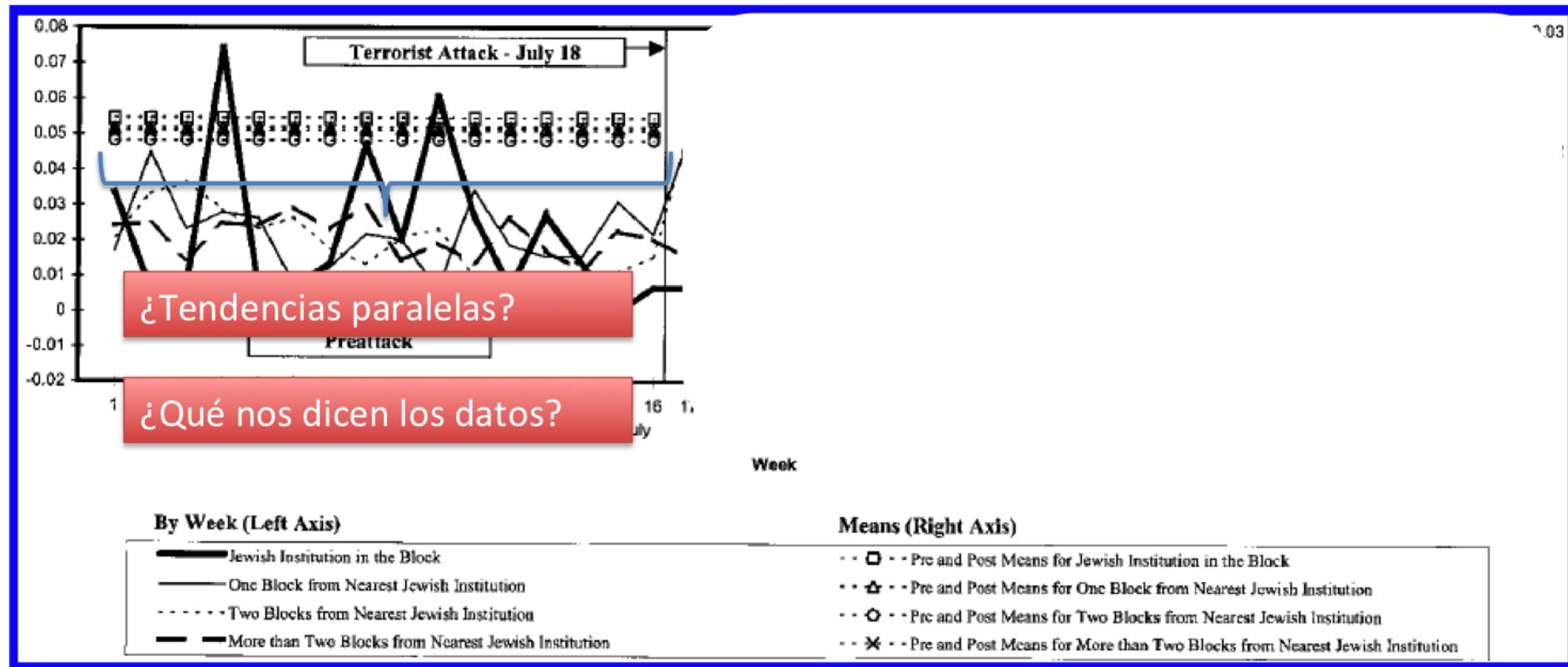


FIGURE 2. WEEKLY EVOLUTION OF CAR THEFTS

# Visualización de los efectos

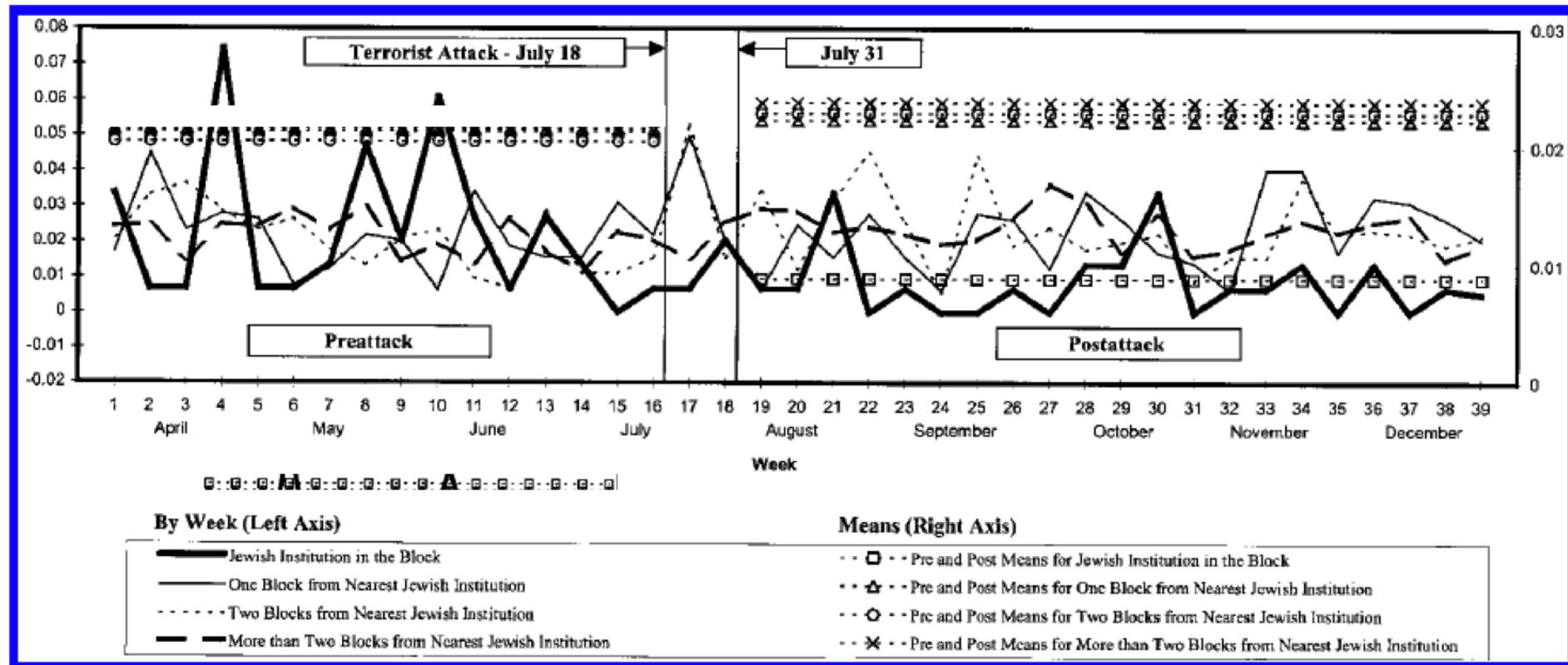


FIGURE 2. WEEKLY EVOLUTION OF CAR THEFTS

# Visualización de los efectos

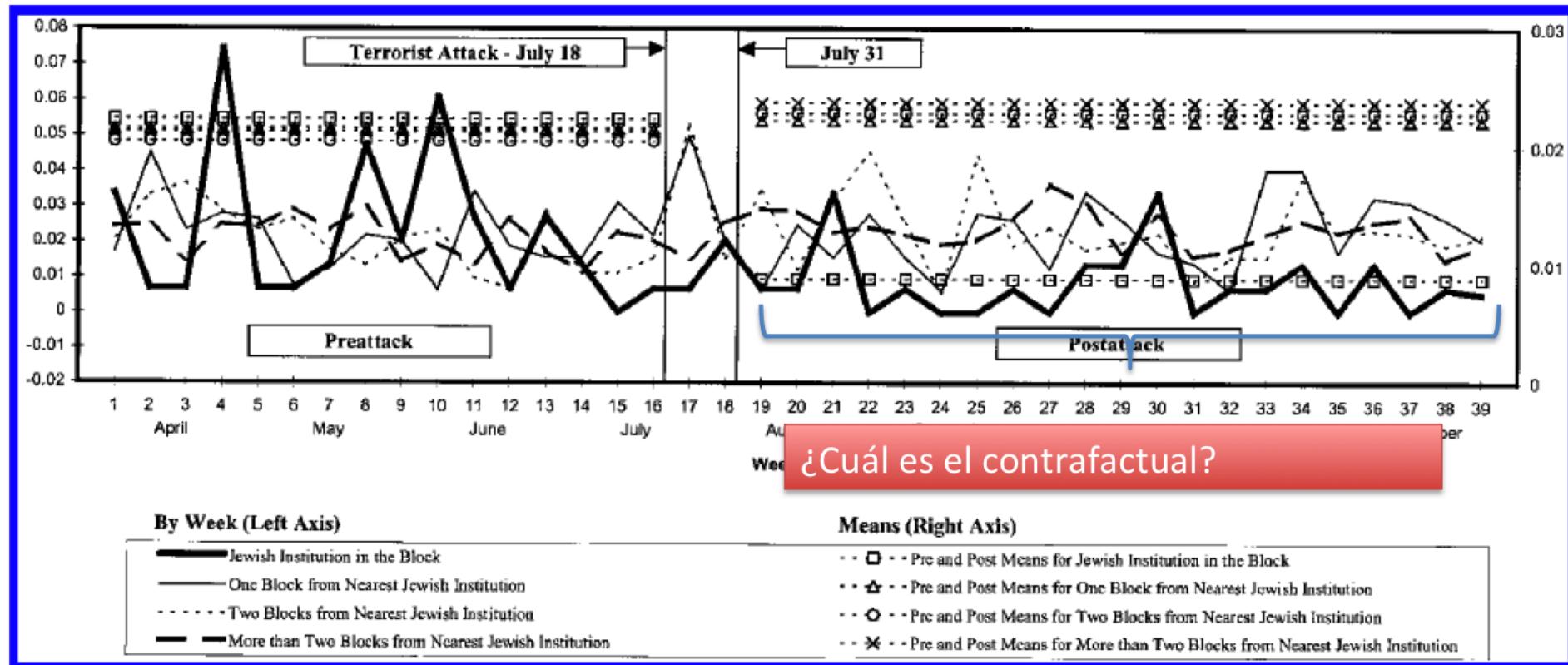


FIGURE 2. WEEKLY EVOLUTION OF CAR THEFTS

# Visualización de los efectos

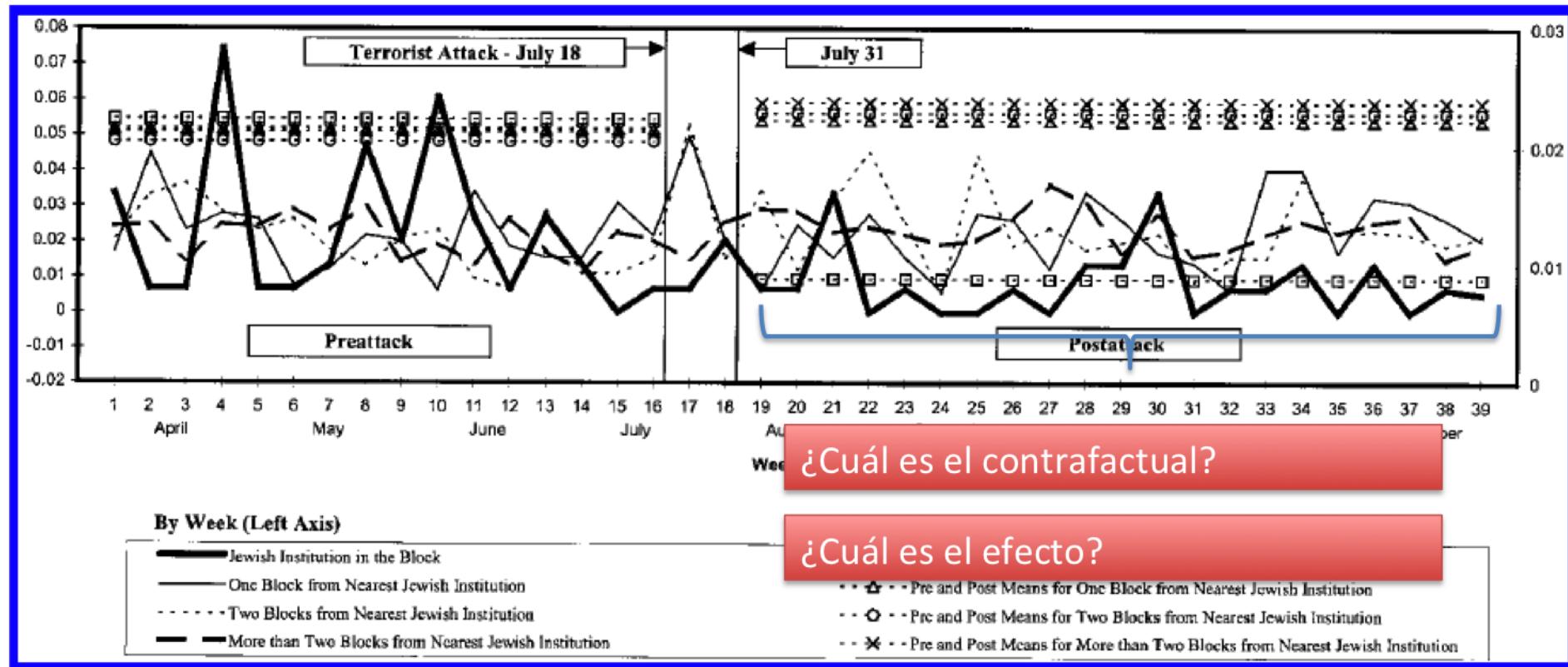


FIGURE 2. WEEKLY EVOLUTION OF CAR THEFTS

# Visualización de los efectos

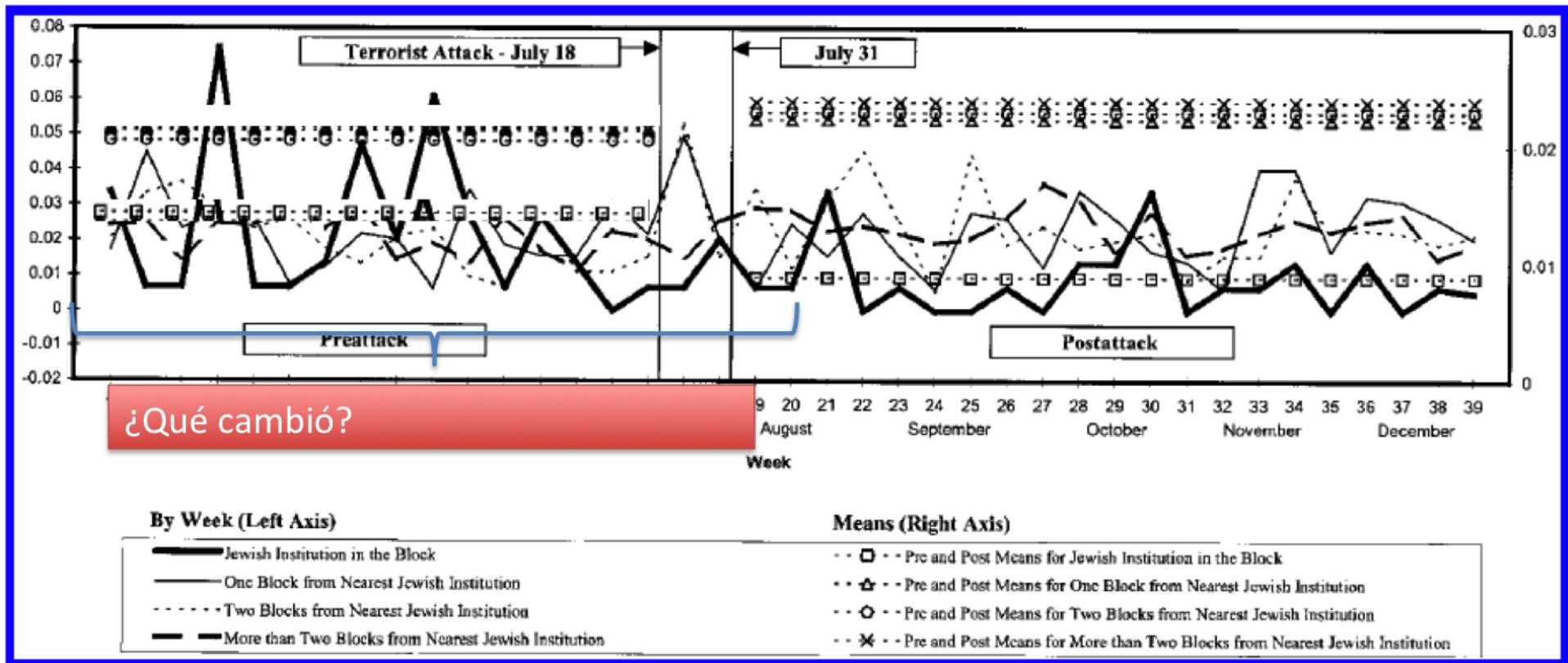


FIGURE 2. WEEKLY EVOLUTION OF CAR THEFTS

# Visualización de los efectos

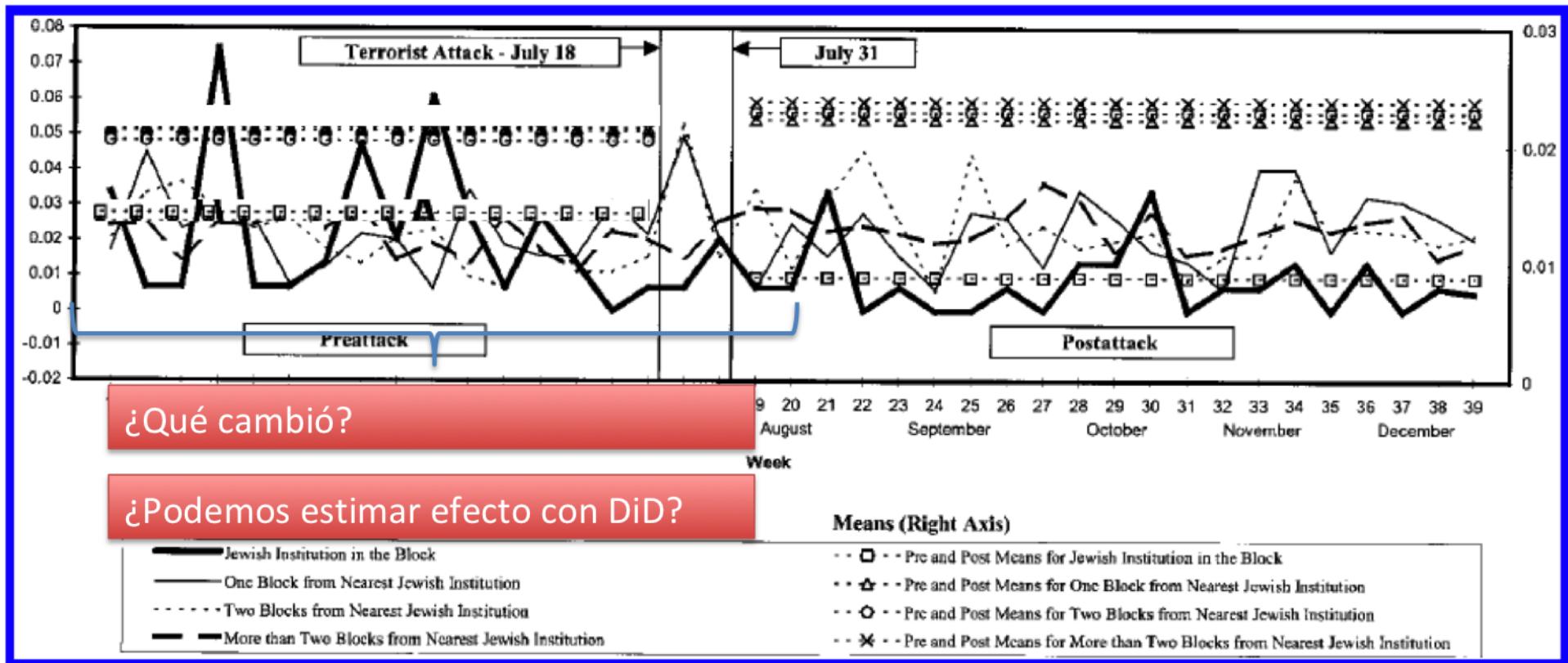


FIGURE 2. WEEKLY EVOLUTION OF CAR THEFTS

# Visualización de los efectos

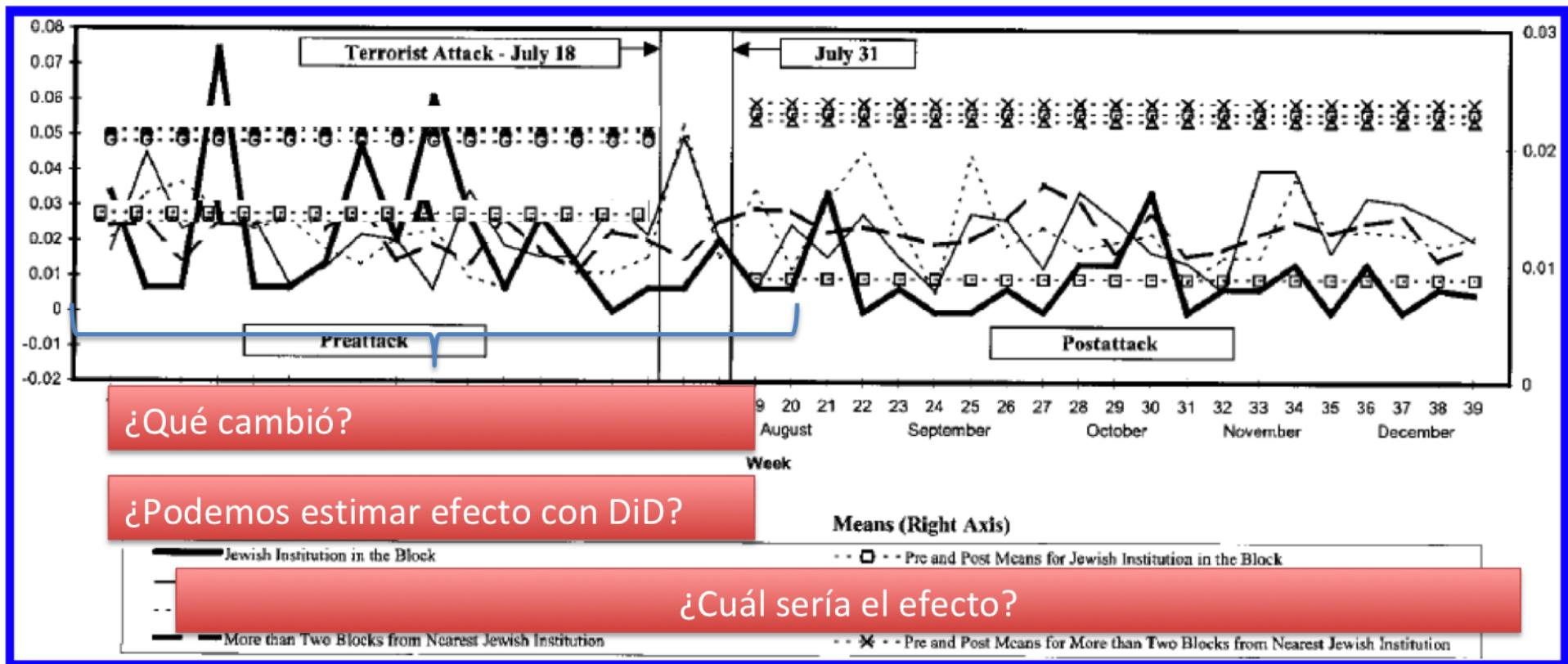


FIGURE 2. WEEKLY EVOLUTION OF CAR THEFTS

# Visualización de los efectos

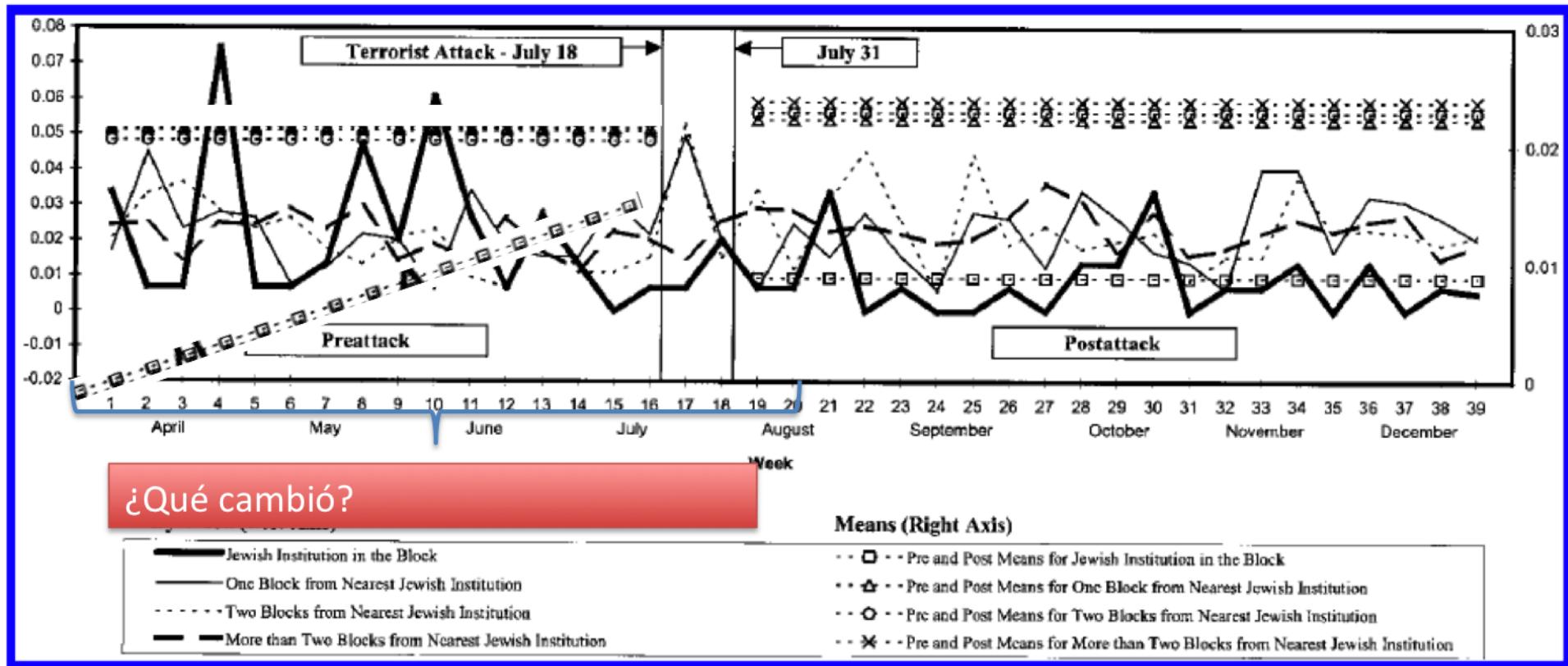


FIGURE 2. WEEKLY EVOLUTION OF CAR THEFTS

# Visualización de los efectos

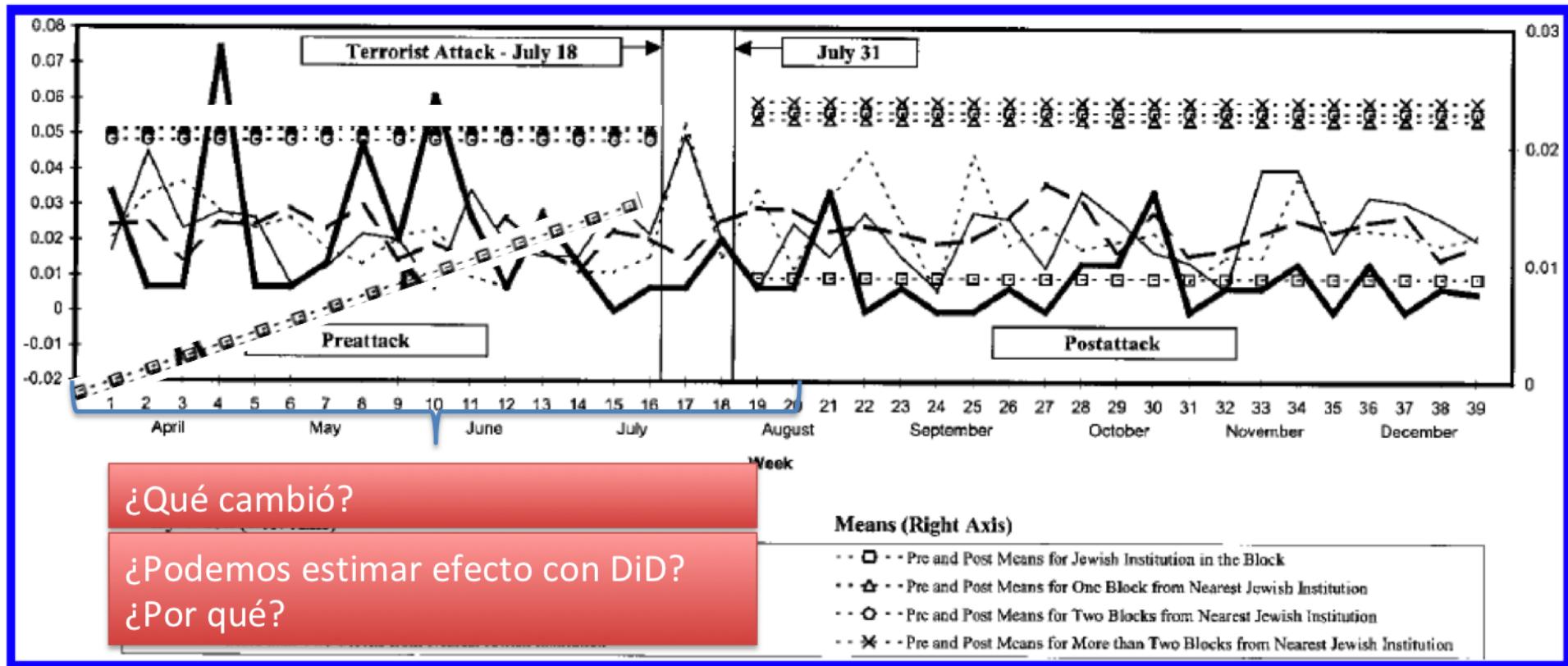


FIGURE 2. WEEKLY EVOLUTION OF CAR THEFTS

# Resultados

TABLE 3—THE EFFECT OF POLICE PRESENCE ON CAR THEFT

	Difference-in-difference			Cross section	Time series
	(A)	(B)	(C)	(D)	(E)
<i>Same-Block Police</i>	-0.07752*** (0.022)	-0.08007*** (0.022)	-0.08080*** (0.022)	-0.07271*** (0.011)	-0.05843*** (0.022)
<i>One-Block Police</i>		-0.01325 (0.013)	-0.01398 (0.014)	-0.01158 (0.010)	-0.00004 (0.013)
<i>Two-Blocks Police</i>			-0.00218 (0.012)	-0.00342 (0.009)	0.01701 (0.010)
Block fixed effect	Yes	Yes	Yes	No	Yes
Month fixed effect	Yes	Yes	Yes	Yes	No
Number of observations	7,884	7,884	7,884	4,380	3,816
R <sup>2</sup>	0.1983	0.1984	0.1984	0.0036	0.1891

# Resultados

TABLE 4—CAR THEFTS BEFORE THE TERRORIST ATTACK

	Police dummies activated on April 30 (A)	Police dummies activated on May 31 (B)	Police dummies activated on June 30 (C)
<i>Same-Block Police</i>	-0.01864 (0.053)	0.01467 (0.040)	-0.03611 (0.038)
<i>One-Block Police</i>	-0.02553 (0.025)	0.01402 (0.019)	0.02310 (0.022)
<i>Two-Blocks Police</i>	-0.03263 (0.022)	-0.01465 (0.017)	-0.00940 (0.016)
Block fixed effect	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes
Number of observations	3,504	3,504	3,504
<i>R</i> <sup>2</sup>	0.3206	0.3202	0.3204

# Resultados

TABLE 6—EXPENSIVE VS. CHEAP, WEEKDAY VS. WEEKEND, AND NIGHT VS. DAY

	Expensive car thefts (A)	Cheap car thefts (B)	Weekday thefts (C)	Weekend thefts (D)	Night thefts (E)	Day thefts (F)
<i>Same-Block Police</i>	-0.02798*** (0.009)	-0.04213** (0.019)	-0.05879*** (0.017)	-0.02201 (0.015)	-0.02922** (0.013)	-0.05157*** (0.018)
<i>One-Block Police</i>	-0.00848 (0.007)	-0.00607 (0.011)	-0.00807 (0.012)	-0.00591 (0.006)	-0.01368 (0.008)	-0.00030 (0.011)
<i>Two-Blocks Police</i>	-0.00763 (0.007)	0.00392 (0.009)	0.00212 (0.010)	-0.00431 (0.006)	-0.00033 (0.006)	-0.00185 (0.010)
Block fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Month fixed effect	Yes	Yes	Yes	Yes	Yes	Yes
Number of observations	7,884	7,884	7,884	7,884	7,884	7,884
R <sup>2</sup>	0.1383	0.1650	0.1629	0.1792	0.1558	0.1737

# Generalizando DiD a Efectos Fijos

- Es común tener datos para más de dos períodos en donde la intervención puede ocurrir en distintos períodos para distintas unidades de análisis
- Esto permite que el efecto del tratamiento varíe en el tiempo

# Retomando la regresión de efectos fijos

$$Y_{it} = \delta D_{it} + \alpha_i + \eta_t + v_{it}$$

- ¿Qué cambio respecto a la notación anterior?
- ¿Qué es  $\alpha_i$ ? ¿Qué es  $\eta_t$ ?
- $\alpha_i$ : componente del error que no varía en el tiempo
- $\eta_t$ : componente del error varía en el tiempo pero es común a todo  $i$
- $v_t$ : factores no obs. que varían en el tiempo y por  $i$

# Retomando la regresión de efectos fijos

$$Y_{it} = \delta D_{it} + \alpha_i + \eta_t + v_{it}$$

- ¿Por qué queremos controlar por  $\alpha_i$  o por  $\eta_t$ ?
- ¿Cuál es el supuesto de identificación de  $\delta$  si no controlamos por  $\alpha_i$  o  $\eta_t$ ?
- ¿Qué pasa si ahora controlamos por  $\alpha_i$  o  $\eta_t$ ?

# Efectos en el tiempo

- Podemos hacer una regresión en donde el efecto cambia en el tiempo:

$$Y_{it} = \sum_{j=0}^J \delta_j D_{i,k+j} + \alpha_i + \eta_t + v_{it}$$

- Supongamos que el tratamiento empieza en  $t = k$
- Definimos periodo después de tratamiento como  $t = k + j$ , donde  $j = 0, \dots, J$
- Entonces,  $\hat{\delta}_j$  es el efecto del tratamiento en  $t = k + j$
- ¿Por qué nos gustaría modelar el efecto de esta forma?

# Ejemplo: impacto de bajas de impuestos a extranjeros sobre migración

## Taxation and International Migration of Superstars Evidence from the European Football Market<sup>†</sup>

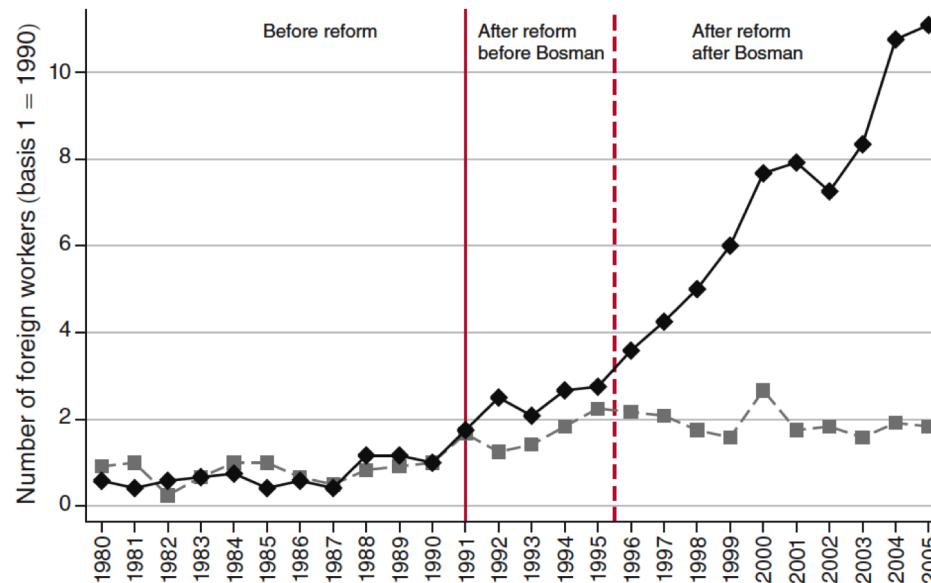
By HENRIK JACOBSEN KLEVEN, CAMILLE LANDAIS,  
AND EMMANUEL SAEZ\*

*We analyze the effects of top tax rates on international migration of football players in 14 European countries since 1985. Both country case studies and multinomial regressions show evidence of strong mobility responses to tax rates, with an elasticity of the number of foreign (domestic) players to the net-of-tax rate around one (around 0.15). We also find evidence of sorting effects (low taxes attract high-ability players who displace low-ability players) and displacement effects (low taxes on foreigners displace domestic players). Those results can be rationalized in a simple model of migration and taxation with rigid labor demand. (JEL F22, H24, H31, J44, J61, L83)*

- Se reducen los impuestos a extranjeros que ganan mucho en 1991
- Grupo de tratamiento: trabajadores extranjeros que ganan justo por sobre el tramo afectado
- Grupo de control: trabajadores extranjeros que ganan justo por debajo del tramo afectado

# Ejemplo: impacto de bajas de impuestos a extranjeros sobre migración

Panel A. Sports and entertainment



- Grupo de tratamiento: trabajadores extranjeros que ganan justo por sobre el tramo afectado
- Grupo de control: trabajadores extranjeros que ganan justo por debajo del tramo afectado
- ¿Cuál es la estrategia de identificación? ¿Es posible?

# Efectos en el tiempo

- En este ejemplo podemos estimar

$$Y_{it} = \sum_{j=0}^J \delta_j D_{i,k+j} + \alpha_i + \eta_t + v_{it}$$

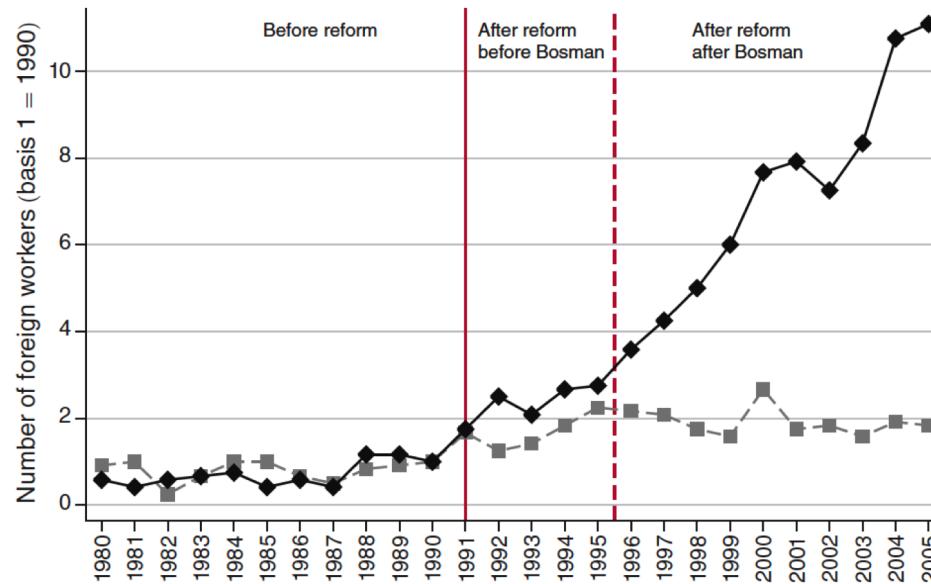
- $Y$ : número de trabajadores extranjeros (cómo base tomamos el año de 1990)
- El tratamiento comienza en  $t = 1991$
- Definimos el periodo después de tratamiento como  $t = 1991 + j$  donde  $j = 0, \dots, J$
- Entonces  $\hat{\delta}_j$  es el efecto del tratamiento en  $t = 1991 + j$
- ¿Por qué nos gustaría modelar el efecto de esta forma?

# Resumen de DiD y EF

- DiD y EF son métodos útiles para análisis de datos de panel
- Estos métodos proveen supuestos de identificación menos exigentes que el estimador de corte transversal (A-D) o de diferencias simple
- En cualquiera de estos dos métodos hay que mostrar que (1) existen tendencias paralelas en períodos anteriores y (2) que no hay cambios en otras políticas al mismo tiempo
- De todas formas los supuestos no son testeables.. Mostrar lo anterior sólo hace que nuestros supuestos sean más creíbles.

# Ejemplo: Impacto de impuesto al tabaco sobre consumo

Panel A. Sports and entertainment



¿Dónde es más probable que haya tendencias paralelas?

# DUDAS



# Semana 5

## Regresión Discontinua y Diferencias en Diferencias

01 de noviembre, 2022

 **Pablo A. Celhay** |  pacelhay@uc.cl

Diseño y formato de la presentación:  José Daniel Conejeros |  jdconejeros@uc.cl |  JDConjeros