

## Preámbulo

Con el propósito de asegurar la transparencia y la replicabilidad de los resultados, todos los códigos utilizados para la estimación, así como los conjuntos de datos procesados, se encuentran disponibles en el repositorio de GitHub asociado a este trabajo. En dicho repositorio se incluyen archivos .do y .r con las rutinas empleadas en Stata y R studio, las bases de datos utilizadas y las figuras generadas a lo largo del análisis.

Repositorio: <https://github.com/9marlon9/Taller-3-EcAv>

Para garantizar que se cumpla la función del archivo maestro en la entrega de los Do-File y los archivos .R, los códigos accede directamente a las bases de datos desde el repositorio. De este modo, se alcanza el mismo objetivo que persigue un Do-File maestro, pero utilizando GitHub como plataforma de acceso a los datos.

## Primer Ejercicio

**Nota metodológica:** El presente ejercicio fue resuelto utilizando tanto *Stata* como *R*, seleccionando el software más apropiado para cada ejercicio.

**Stata** fue utilizado para las estimaciones TWFE (Puntos 3b, 3c), y la implementación de Callaway & Sant'Anna (Punto 5a), mientras que **R** fue empleado para la descomposición de Bacon (Punto 3d) y la gráfica de leads and lags (4b).

La elección del software se basó en la disponibilidad de comandos especializados y facilidad de implementación para cada metodología específica.

### 1.

El proceso generador de los datos (Data Generating Process-DGP) es de la forma

$$Y_{i,t} = \alpha_i + \lambda_t + D_{i,t} \times \tau_{i,t} + \epsilon_{i,t} \quad (1)$$

donde  $\tau_{i,t}$  es el efecto acumulado de la política percibido por la unidad  $i$  en el tiempo  $t$ .

- a) A partir del modelo (1), derive las expresiones explícitas para los resultados potenciales  $Y_{i,t}(g)$ , es decir, la forma funcional que toma el resultado de la unidad  $i$  en el periodo  $t$  cuando su cohorte de tratamiento es  $g$ .

*Pista:* Recuerde que el resultado observado puede descomponerse como una suma sobre las posibles cohortes de tratamiento:

$$Y_{i,t} = \sum_{g \in \mathcal{G}} \mathbb{1}[G_i = g] Y_{i,t}(g),$$

y de manera análoga,

$$\tau_{i,t} = \sum_{g \in \mathcal{G}} \mathbb{1}[G_i = g] \tau_{i,t}(g), \quad \varepsilon_{i,t} = \sum_{g \in \mathcal{G}} \mathbb{1}[G_i = g] \varepsilon_{i,t}(g),$$

con  $\mathbb{E}[\varepsilon_{i,t}(g)] = 0$  para todo  $g \in \mathcal{G}$ .

## Solución

Sustituyendo en el DGP:

$$\sum_{g \in \mathcal{G}} 1[G_i = g] Y_{i,t}(g) = \alpha_i + \lambda_t + D_{i,t} \times \sum_{g \in \mathcal{G}} 1[G_i = g] \tau_{i,t}(g) + \sum_{g \in \mathcal{G}} 1[G_i = g] \epsilon_{i,t}(g)$$

Para la cohorte específica  $g$ , se puede multiplicar a ambos lados por  $1[G_i = g]$  y dado que  $1[G_i = g] \cdot 1[G_i = g'] = 0$  para  $g \neq g'$ , obtenemos:

$$1[G_i = g] Y_{i,t}(g) = (1[G_i = g] \alpha_i) + (1[G_i = g] \lambda_t) + (D_{i,t} \times 1[G_i = g] \tau_{i,t}(g)) + (1[G_i = g] \epsilon_{i,t}(g))$$

Dividiendo ambos lados por  $1[G_i = g]$ :

$$Y_{i,t}(g) = \alpha_i + \lambda_t + D_{i,t} \times \tau_{i,t}(g) + \epsilon_{i,t}(g)$$

Además  $D_{i,t} = 1\{t \geq G_i\}$ , y cuando  $G_i = g$ , tenemos  $D_{i,t} = 1\{t \geq g\}$ . Por lo tanto:

$$Y_{i,t}(g) = \alpha_i + \lambda_t + 1\{t \geq g\} \cdot \tau_{i,t}(g) + \epsilon_{i,t}(g)$$

- b) Pruebe que si el DGP es (1),  $\epsilon_{i,t}(\infty) \perp G_i$ , y además  $\epsilon_{i,t}(g) = \epsilon_{i,t}(\infty)$  para todo  $t < g$ , entonces se cumplen los supuestos 1 y 2 planteados en el esquema general. Además, interprete cada uno de ellos de manera intuitiva y argumente si son plausibles en el problema planteado por el DNP.

## Solución

### Supuestos

- DGP:  $Y_{i,t} = \alpha_i + \lambda_t + D_{i,t} \tau_{i,t} + \epsilon_{i,t}$
- $\epsilon_{i,t}(\infty) \perp G_i$
- $\epsilon_{i,t}(g) = \epsilon_{i,t}(\infty)$  para todo  $t < g$

### Supuesto 1: Tendencias paralelas

Para todo  $t \neq t'$  y  $g \neq g'$ :

$$\mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty) | G_i = g] = \mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty) | G_i = g']$$

De la expresión de  $Y_{i,t}(\infty)$ :

$$Y_{i,t}(\infty) = \alpha_i + \lambda_t + \epsilon_{i,t}(\infty)$$

Luego:

$$Y_{i,t}(\infty) - Y_{i,t'}(\infty) = (\lambda_t - \lambda_{t'}) + (\epsilon_{i,t}(\infty) - \epsilon_{i,t'}(\infty))$$

Expectativa condicional:

$$\mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty) | G_i = g] = (\lambda_t - \lambda_{t'}) + \mathbb{E}[\epsilon_{i,t}(\infty) - \epsilon_{i,t'}(\infty) | G_i = g]$$

Como  $\epsilon_{i,t}(\infty) \perp G_i$ , se tiene  $\mathbb{E}[\epsilon_{i,t}(\infty) | G_i = g] = \mathbb{E}[\epsilon_{i,t}(\infty)] = 0$  y similarmente para  $\epsilon_{i,t'}(\infty)$ . Por lo tanto:

$$\mathbb{E}[\epsilon_{i,t}(\infty) - \epsilon_{i,t'}(\infty) | G_i = g] = 0$$

Así:

$$\mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty) | G_i = g] = \lambda_t - \lambda_{t'}$$

El resultado no depende de  $g$ , por lo que para cualquier  $g \neq g'$ :

$$\mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty) | G_i = g] = \lambda_t - \lambda_{t'} = \mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty) | G_i = g']$$

*Se cumple el supuesto 1.*

**Tendencias paralelas:** La confirmación del supuesto implica que, en ausencia de tratamiento, la evolución temporal del número de empleados habría sido la misma para todas las cohortes. Esto es plausible en el contexto del DNP debido a la asignación aleatoria de municipios a grupos, lo que permite intuir que las características observables y no observables de las empresas se distribuyen de manera similar entre cohortes. Sin embargo, la selección de empresas dentro de los municipios por

parte de las entidades financieras (priorizando aquellas con mayor probabilidad de supervivencia) podría introducir sesgos si esta selección está correlacionada con tendencias no observadas en el empleo.

### Supuesto 2: No anticipación

De la expresión de  $Y_{i,t}(g)$ :

$$Y_{i,t}(g) = \alpha_i + \lambda_t + 1\{t \geq g\} \cdot \tau_{i,t}(g) + \epsilon_{i,t}(g)$$

Para  $t < g$ ,  $1\{t \geq g\} = 0$ , por tanto:

$$Y_{i,t}(g) = \alpha_i + \lambda_t + \epsilon_{i,t}(g)$$

Por el supuesto  $\epsilon_{i,t}(g) = \epsilon_{i,t}(\infty)$  para  $t < g$ , se tiene:

$$Y_{i,t}(g) = \alpha_i + \lambda_t + \epsilon_{i,t}(\infty)$$

Y como por definición los nunca tratados son  $Y_{i,t}(\infty) = \alpha_i + \lambda_t + \epsilon_{i,t}(\infty)$ , concluimos:

$$Y_{i,t}(g) = Y_{i,t}(\infty) \quad \text{para todo } t < g$$

*Se cumple el Supuesto 2.*

**No anticipación:** Significa que las empresas no ajustan su comportamiento antes de recibir el tratamiento. En el contexto del DNP, es plausible si las empresas no conocen con anticipación si serán seleccionadas para la condonación de deuda. La aleatorización de municipios mitiga este riesgo, pero la selección interna por las entidades financieras podría permitir que algunas empresas anticipen el tratamiento si tienen información sobre el proceso de selección.

Ambos supuestos podrían ser razonables debido a la aleatorización de los municipios, pero la selección no aleatoria de empresas dentro de los municipios introduce incertidumbre sobre su validez.

---

Los estudios de evento, similar a otras metodologías de inferencia causal, tienen como objetivo recuperar algún tipo de efecto causal promedio de la política de interés. En particular, la literatura reciente ha sugerido que los efectos causales que mejor se adaptan a esta metodología son los *cohort-time average treatment effects on the treated*, los cuales son el efecto promedio de tratamiento en el tiempo  $t$  para la cohorte que fue tratada por primera vez en el tiempo  $g$ . Más formalmente, el cohort-time average treatment effect para el grupo  $g$  en el tiempo  $t$  está dado por:

$$ATT(g, t) = \mathbb{E}[Y_{i,t}(g) - Y_{i,t}(\infty)|G_i = g]$$

## 2.

Demuestre que si el DGP es (1) y  $\epsilon_{i,t}(\infty) \perp G_i$ , entonces

$$ATT(g, t) = \begin{cases} 0 & \text{si } t < g \\ \mathbb{E}[\tau_{i,t}(g, t)] & \text{si } t \geq g \end{cases}$$

### Solución

**Caso 1:**  $t < g$

Del inciso 1.1, tenemos:

$$Y_{i,t}(g) = \alpha_i + \lambda_t + 1\{t \geq g\} \cdot \tau_{i,t}(g) + \epsilon_{i,t}(g)$$

Y además sabemos que, para los nunca tratados:

$$Y_{i,t}(\infty) = \alpha_i + \lambda_t + \epsilon_{i,t}(\infty)$$

Para  $t < g$ ,  $1\{t \geq g\} = 0$ , luego:

$$Y_{i,t}(g) = \alpha_i + \lambda_t + 0 \cdot \tau_{i,t}(g) + \epsilon_{i,t}(g)$$

$$Y_{i,t}(g) = \alpha_i + \lambda_t + \epsilon_{i,t}(g)$$

Por el supuesto de no anticipación, para  $t < g$ ,  $\epsilon_{i,t}(g) = \epsilon_{i,t}(\infty)$ . Entonces:

$$Y_{i,t}(g) - Y_{i,t}(\infty) = [\alpha_i + \lambda_t + \epsilon_{i,t}(g)] - [\alpha_i + \lambda_t + \epsilon_{i,t}(\infty)] = \epsilon_{i,t}(g) - \epsilon_{i,t}(\infty) = 0$$

Tomando expectativa condicional del resultado anterior:

$$ATT(g, t) = \mathbb{E}[Y_{i,t}(g) - Y_{i,t}(\infty) | G_i = g]$$

$$ATT(g, t) = \mathbb{E}[0 | G_i = g] = 0$$

**Caso 2:**  $t \geq g$

Para  $t \geq g$ ,  $1\{t \geq g\} = 1$ , por lo tanto:

$$Y_{i,t}(g) = \alpha_i + \lambda_t + 1 \cdot \tau_{i,t}(g) + \epsilon_{i,t}(g)$$

$$Y_{i,t}(\infty) = \alpha_i + \lambda_t + \epsilon_{i,t}(\infty)$$

Entonces:

$$Y_{i,t}(g) - Y_{i,t}(\infty) = [\alpha_i + \lambda_t + \tau_{i,t}(g) + \epsilon_{i,t}(g)] - [\alpha_i + \lambda_t + \epsilon_{i,t}(\infty)] = \tau_{i,t}(g) + \epsilon_{i,t}(g) - \epsilon_{i,t}(\infty)$$

Tomando expectativa condicional del resultado anterior:

$$\begin{aligned} ATT(g, t) &= \mathbb{E}[Y_{i,t}(g) - Y_{i,t}(\infty) | G_i = g] \\ ATT(g, t) &= \mathbb{E}[\tau_{i,t}(g) + \epsilon_{i,t}(g) - \epsilon_{i,t}(\infty) | G_i = g] \\ &= \mathbb{E}[\tau_{i,t}(g) | G_i = g] + \mathbb{E}[\epsilon_{i,t}(g) | G_i = g] - \mathbb{E}[\epsilon_{i,t}(\infty) | G_i = g] \end{aligned}$$

Dado que  $\epsilon_{i,t}(\infty) \perp G_i$ , se tiene  $\mathbb{E}[\epsilon_{i,t}(\infty) | G_i = g] = \mathbb{E}[\epsilon_{i,t}(\infty)] = 0$ .

Además, por la descomposición del DGP,  $\mathbb{E}[\epsilon_{i,t}(g)] = 0$  para todo  $g \in \mathcal{G}$ , y asumiendo que  $\epsilon_{i,t}(g)$  es independiente de  $G_i$  (debido a la aleatorización de las cohortes), se tiene  $\mathbb{E}[\epsilon_{i,t}(g) | G_i = g] = \mathbb{E}[\epsilon_{i,t}(g)] = 0$ . Por lo tanto:

$$ATT(g, t) = \mathbb{E}[\tau_{i,t}(g) | G_i = g] = \mathbb{E}[\tau_{i,t}(g)]$$

Combinando los resultados de ambos casos:

$$ATT(g, t) = \begin{cases} 0 & \text{si } t < g \\ \mathbb{E}[\tau_{i,t}(g)] & \text{si } t \geq g \end{cases}$$

donde  $\mathbb{E}[\tau_{i,t}(g)]$  es el efecto promedio del tratamiento para la cohorte  $g$  en el tiempo  $t$ .

Ahora usted cuenta con la base de datos *Empresas.Sim.dta*, la cual contiene los datos que le prometió el DNP. En particular, cuenta con dos variables de resultado simuladas usando el DGP (1), y variando el efecto de tratamiento:

**Modelo 1:**

$$\tau_{i,t}(g) = \begin{cases} 0, & t < g, \\ \mu_i, & t \geq g \end{cases}$$

**Modelo 2:**

$$\tau_{i,t}(g) = \begin{cases} 0, & t < g, \\ \mu_i(t - g + 1), & t \geq g \end{cases}$$

Donde  $\mu_i$  es un parámetro específico de cada firma. El término  $(t - g + 1)$  representa el número de años transcurridos desde la adopción del programa, de modo que el impacto total crece linealmente con el tiempo de exposición: cuando  $t = g$ , el efecto inicial es  $\mu_i$  y después de  $k$  años de tratamiento el efecto acumulado es  $k\mu_i$ .

Usaremos estos dos modelos para evaluar la efectividad de estimar un modelo TWFE para recuperar un efecto causal a partir de un estudio de eventos. Recordemos que la especificación usada por el anterior equipo del DNP es:

$$\text{Empleados}_{i,t} = \alpha_i + \lambda_t + \delta D_{i,t} + \epsilon_{i,t}$$

donde, en nuestro caso,  $\text{Empleados}_{i,t}$  sería alguna de las variables dependientes simuladas en vez de los datos reales. Al usar esta forma funcional, los investigadores estaban pensando que  $\delta$  recuperaría algún tipo de efecto causal. Más precisamente, se pensaba que

$$\hat{\delta}^{MCO} \xrightarrow{P} \theta(w) = \sum_{g \in \mathcal{G}} \sum_{t \geq g} w(g, t) \text{ATT}(g, t)$$

Esto es, nuestro estimador pondera los  $\text{ATT}(g, t)$  de cada cohorte–tiempo. Los pesos  $w(g, t)$  reflejan la contribución de cada celda  $(g, t)$  a la covarianza entre el regresor de tratamiento *residualizado* y la variable dependiente: en una regresión TWFE, si  $\tilde{D}_{it}$  denota  $D_{it}$  residualizado respecto a los efectos fijos (y controles), entonces

$$w(g, t) \propto \mathbb{E}[\tilde{D}_{it} \mathbb{1}\{G_i = g, t\}], \quad \sum_{g, t} w(g, t) = 1.$$

Estos pesos dependen de la distribución empírica de cohortes y del momento de adopción, por lo que son *desconocidos a priori* y deben ser inducidos por la propia muestra.<sup>1</sup> El objetivo del siguiente inciso es evaluar si, bajo nuestra simulación, el TWFE realmente se aproxima a un promedio interpretable de  $\text{ATT}(g, t)$ .

### 3.

Usando la información del DGP para los modelos de la base simulada:

- a) Defina los  $\text{ATT}(g, t)$  de cada modelo, para todo grupo  $g \in \mathcal{G} \setminus \{\infty\}$  y  $t \in \{2000, \dots, 2009\}$

#### Solución

Para los grupos  $g \in \{2002, 2004, 2006, 2008\}$  y períodos  $t \in \{2000, \dots, 2009\}$ , se definen los  $\text{ATT}(g, t)$  para cada modelo.

#### Modelo 1: Efecto constante

$$\text{ATT}_1(g, t) = \begin{cases} 0 & \text{si } t < g \\ \bar{\mu}_g & \text{si } t \geq g \end{cases}$$

#### Desglose por cohorte:

- ◊ **Cohorte 2002:**
  - ◊  $t = 2000, 2001$ :  $\text{ATT}_1(2002, t) = 0$
  - ◊  $t = 2002, \dots, 2009$ :  $\text{ATT}_1(2002, t) = \bar{\mu}_{2002}$
- ◊ **Cohorte 2004:**
  - ◊  $t = 2000, 2001, 2002, 2003$ :  $\text{ATT}_1(2004, t) = 0$

---

<sup>1</sup>En diseños con adopción escalonada y efectos heterogéneos, parte de la literatura reciente muestra que algunos pesos pueden ser incluso negativos (p. ej., cuando ciertas cohortes sirven como “controles” de otras en distintos momentos).

◊  $t = 2004, \dots, 2009$ :  $ATT_1(2004, t) = \bar{\mu}_{2004}$

◊ **Cohorte 2006:**

◊  $t = 2000, \dots, 2005$ :  $ATT_1(2006, t) = 0$

◊  $t = 2006, \dots, 2009$ :  $ATT_1(2006, t) = \bar{\mu}_{2006}$

◊ **Cohorte 2008:**

◊  $t = 2000, \dots, 2007$ :  $ATT_1(2008, t) = 0$

◊  $t = 2008, 2009$ :  $ATT_1(2008, t) = \bar{\mu}_{2008}$

**Modelo 2: Efecto lineal creciente**

$$ATT_2(g, t) = \begin{cases} 0 & \text{si } t < g \\ (t - g + 1) \cdot \bar{\mu}_g & \text{si } t \geq g \end{cases}$$

**Desglose por cohorte:**

◊ **Cohorte 2002:**

◊  $t = 2000, 2001$ :  $ATT_2(2002, t) = 0$

◊  $t = 2002$ :  $ATT_2(2002, 2002) = 1 \cdot \bar{\mu}_{2002}$

◊  $t = 2003$ :  $ATT_2(2002, 2003) = 2 \cdot \bar{\mu}_{2002}$

◊  $t = 2004$ :  $ATT_2(2002, 2004) = 3 \cdot \bar{\mu}_{2002}$

◊  $t = 2005$ :  $ATT_2(2002, 2005) = 4 \cdot \bar{\mu}_{2002}$

◊  $t = 2006$ :  $ATT_2(2002, 2006) = 5 \cdot \bar{\mu}_{2002}$

◊  $t = 2007$ :  $ATT_2(2002, 2007) = 6 \cdot \bar{\mu}_{2002}$

◊  $t = 2008$ :  $ATT_2(2002, 2008) = 7 \cdot \bar{\mu}_{2002}$

◊  $t = 2009$ :  $ATT_2(2002, 2009) = 8 \cdot \bar{\mu}_{2002}$

◊ **Cohorte 2004:**

◊  $t = 2000, 2001, 2002, 2003$ :  $ATT_2(2004, t) = 0$

◊  $t = 2004$ :  $ATT_2(2004, 2004) = 1 \cdot \bar{\mu}_{2004}$

◊  $t = 2005$ :  $ATT_2(2004, 2005) = 2 \cdot \bar{\mu}_{2004}$

◊  $t = 2006$ :  $ATT_2(2004, 2006) = 3 \cdot \bar{\mu}_{2004}$

◊  $t = 2007$ :  $ATT_2(2004, 2007) = 4 \cdot \bar{\mu}_{2004}$

◊  $t = 2008$ :  $ATT_2(2004, 2008) = 5 \cdot \bar{\mu}_{2004}$

◊  $t = 2009$ :  $ATT_2(2004, 2009) = 6 \cdot \bar{\mu}_{2004}$

◊ **Cohorte 2006:**

◊  $t = 2000, \dots, 2005$ :  $ATT_2(2006, t) = 0$

◊  $t = 2006$ :  $ATT_2(2006, 2006) = 1 \cdot \bar{\mu}_{2006}$

◊  $t = 2007$ :  $ATT_2(2006, 2007) = 2 \cdot \bar{\mu}_{2006}$

◊  $t = 2008$ :  $ATT_2(2006, 2008) = 3 \cdot \bar{\mu}_{2006}$

◊  $t = 2009$ :  $ATT_2(2006, 2009) = 4 \cdot \bar{\mu}_{2006}$

◊ **Cohorte 2008:**

◊  $t = 2000, \dots, 2007$ :  $ATT_2(2008, t) = 0$

◊  $t = 2008$ :  $ATT_2(2008, 2008) = 1 \cdot \bar{\mu}_{2008}$

◊  $t = 2009$ :  $ATT_2(2008, 2009) = 2 \cdot \bar{\mu}_{2008}$

Donde  $\bar{\mu}_g = \mathbb{E}[\mu_i \mid G_i = g]$  representa el efecto promedio base para la cohorte  $g$ .

- b) Estime por MCO el modelo 1 como lo hizo el equipo del DNP. Reporte el valor del coeficiente junto con su errores estándar. ¿Es coherente lo que encuentra con la idea de que el estimador de MCO recupera un efecto causal como  $\theta(w)$ ?<sup>2</sup>

---

<sup>2</sup>Tenga en cuenta que el efecto que va a encontrar es para datos simulados, por ende no va a tener una interpretación económica directa.

## Solución

Se estimó el siguiente modelo TWFE por MCO:

$$\text{Empleados}_{i,t} = \alpha_i + \lambda_t + \delta D_{i,t} + \epsilon_{i,t}$$

donde  $D_{i,t}$  es una variable binaria que indica si la empresa  $i$  recibió el tratamiento en el período  $t$ . Los errores estándar se agrupan a nivel de empresa.

## Resultados de la estimación

Tabla 1: Estimación TWFE del Efecto del Tratamiento sobre el Empleo (Modelo 1)

Variable	Coeficiente	Error Estándar	Valor-p
Tratamiento ( $D_{i,t}$ )	0.2997	0.009	0.000
Efectos fijos de empresa		Sí	
Efectos fijos de año		Sí	
$R^2$ intra-grupo	0.906		

Nota: Esta tabla reporta los resultados de la estimación del modelo de efectos fijos en dos direcciones. El coeficiente reportado corresponde a la variable de tratamiento  $D_{i,t}$ . Errores estándar robustos agrupados a nivel de empresa.

### Análisis de coherencia con $\theta(w)$

El coeficiente estimado  $\hat{\delta} = 0.2997$  es estadísticamente significativo al nivel de 1 %. Teóricamente, bajo el supuesto de identificación de tendencias paralelas, el estimador MCO converge en probabilidad a:

$$\hat{\delta}^{MCO} \xrightarrow{P} \theta(w) = \sum_{g \in \mathcal{G}} \sum_{t \geq g} w(g, t) ATT(g, t)$$

Para el Modelo 1, donde  $ATT_1(g, t) = \bar{\mu}_g$  para  $t \geq g$ , esto implica:

$$\theta(w) = \sum_{g \in \{2002, 2004, 2006, 2008\}} \sum_{t \geq g} w(g, t) \bar{\mu}_g$$

El resultado es coherente con esta interpretación dado que:

1. **Significancia estadística:** El efecto estimado es altamente significativo ( $p < 0.001$ ), indicando que el programa tuvo un impacto detectable sobre el empleo.
  2. **Magnitud plausible:** El coeficiente de 0.2997 representa un promedio ponderado de los efectos de tratamiento  $\bar{\mu}_g$  across cohortes. Dada la estructura del diseño escalonado, esta magnitud es consistente con un efecto causal promedio.
  3. **Precisión estimada:** El error estándar relativamente pequeño (0.009) sugiere que el estimador recupera el parámetro de interés con alta precisión, lo que es esperable bajo el DGP correctamente especificado.
- c) Estime por MCO el modelo 2 como lo hizo el equipo del DNP. Reporte el valor del coeficiente junto con su errores estándar. ¿Existen algunos pesos  $w(g, t)$  para los cuales resulte razonable pensar que nuestro estimador captura algo parecido algún  $\theta(w)$  ?

## Solución

Se estimó el mismo modelo TWFE para el Modelo 2, donde los efectos del tratamiento crecen linealmente con el tiempo de exposición.

## Resultados de la estimación

### Análisis de coherencia con $\theta(w)$

El coeficiente estimado  $\hat{\delta} = 0.150$  es significativamente menor que el obtenido para el Modelo 1 (0.2997), reflejando la diferente estructura de efectos del tratamiento. Teóricamente:

$$\theta(w) = \sum_{g \in \mathcal{G}} \sum_{t \geq g} w(g, t) \cdot (t - g + 1) \cdot \bar{\mu}_g$$

Tabla 2: Estimación TWFE del Efecto del Tratamiento sobre el Empleo (Modelo 2)

Variable	Coeficiente	Error Estándar	Valor-p
Tratamiento ( $D_{i,t}$ )	0.150	0.011	0.000
Efectos fijos de empresa		Sí	
Efectos fijos de año		Sí	
$R^2$ intra-grupo		0.881	

Nota: Esta tabla reporta los resultados de la estimación del modelo de efectos fijos en dos direcciones para el Modelo 2, donde los efectos del tratamiento crecen linealmente con el tiempo. Errores estándar robustos agrupados a nivel de empresa entre paréntesis.

¿Existen pesos  $w(g, t)$  razonables? Sí, técnicamente existe un conjunto de pesos  $w(g, t)$  que haría  $\theta(w) = 0.150$ , pero con importantes limitaciones:

- ◊ Los pesos deben asignar menor peso relativo a períodos posteriores al tratamiento, donde  $ATT_2(g, t)$  es mayor debido al crecimiento lineal.
- ◊ Sin embargo, en diseños escalonados con efectos heterogéneos temporales, los pesos del estimador TWFE pueden incluir comparaciones problemáticas y valores negativos, lo que dificulta una interpretación causal clara.
- ◊ El coeficiente de 0.150 representa un promedio complejo y potencialmente no interpretable de efectos que varían sistemáticamente con el tiempo de exposición al tratamiento.

Mientras existe un conjunto matemático de pesos  $w(g, t)$  que reconcilia la estimación con  $\theta(w)$ , la interpretación causal de este promedio ponderado es cuestionable debido a la heterogeneidad temporal en los efectos del tratamiento.

- d) En este ejercicio, analice las comparaciones que realiza el estimador y discuta por qué algunas de ellas pueden ser problemáticas para identificar correctamente el efecto causal promedio del tratamiento. Para esto, complete la tabla de descomposición con la información correspondiente a cada comparación entre cohortes. Para cada par tratado–no tratado, registre el estimador  $\hat{\delta}_i^{2 \times 2}$ , el peso  $w_i$  asignado por el estimador TWFE y el producto  $w_i \hat{\delta}_i^{2 \times 2}$ . Una vez completada la tabla, verifique que la suma de los pesos sea igual a uno y que la suma ponderada de los efectos ( $\sum_i w_i \hat{\delta}_i^{2 \times 2}$ ) reproduzca el valor total estimado del modelo TWFE,  $\hat{\delta}^{TWFE} = 0.15035$ .<sup>3</sup>

Tabla 3: Descomposición Goodman–Bacon: Comparaciones  $2 \times 2$ 

Grupo tratado	Grupo no tratado	Tipo de comparación	$\hat{\delta}_i^{2 \times 2}$	$w_i$	$w_i \hat{\delta}_i^{2 \times 2}$
2004	2002	Posterior vs Anterior			
2006	2002	Posterior vs Anterior			
2008	2002	Posterior vs Anterior			
2002	2004	Anterior vs Posterior			
2006	2004	Posterior vs Anterior			
2008	2004	Posterior vs Anterior			
2002	2006	Anterior vs Posterior			
2004	2006	Anterior vs Posterior			
2008	2006	Posterior vs Anterior			
2002	2008	Anterior vs Posterior			
2004	2008	Anterior vs Posterior			
2006	2008	Anterior vs Posterior			
<b>Suma</b>			<b>1.0</b>	<b>0.1504</b>	

<sup>3</sup>Se recomienda realizar este ejercicio en R utilizando el paquete `bacondecomp`.

## Solución

Tabla 4: Descomposición Goodman–Bacon: Comparaciones  $2 \times 2$

Grupo tratado	Grupo no tratado	Tipo de comparación	$\hat{\delta}_i^{2 \times 2}$	$w_i$	$w_i \hat{\delta}_i^{2 \times 2}$
2004	2002	Anterior vs Posterior	-0.1392	0.1000	-0.0139
2006	2002	Anterior vs Posterior	-0.4655	0.1333	-0.0621
2008	2002	Anterior vs Posterior	-0.7858	0.1000	-0.0786
2002	2004	Posterior vs Anterior	0.4698	0.0333	0.0157
2006	2004	Anterior vs Posterior	-0.1486	0.0667	-0.0099
2008	2004	Anterior vs Posterior	-0.4526	0.0667	-0.0302
2002	2006	Posterior vs Anterior	0.7458	0.0667	0.0497
2004	2006	Posterior vs Anterior	0.4511	0.0667	0.0301
2008	2006	Anterior vs Posterior	-0.1411	0.0333	-0.0047
2002	2008	Posterior vs Anterior	1.0597	0.1000	0.1060
2004	2008	Posterior vs Anterior	0.7620	0.1333	0.1016
2006	2008	Posterior vs Anterior	0.4669	0.1000	0.0467
<b>Suma</b>				<b>1.0000</b>	<b>0.1504</b>

Nota: Los valores se obtuvieron de la descomposición de Bacon en R. Las comparaciones “Anterior vs Posterior” (primeras 6 filas) utilizan unidades ya tratadas como controles, representando el 50 % del peso total. La suma ponderada confirma que  $\sum w_i \hat{\delta}_i^{2 \times 2} = 0.1504 = \hat{\delta}^{TWFE}$ .

### Análisis de las comparaciones del estimador TWFE

La descomposición de Goodman–Bacon revela con claridad cómo el estimador TWFE se construye a partir de una serie de comparaciones  $2 \times 2$  entre cohortes tratadas en distintos momentos, y cómo algunas de estas comparaciones introducen sesgos significativos cuando los efectos del tratamiento son heterogéneos y crecientes en el tiempo. En nuestro caso, las cohortes que reciben el tratamiento más temprano acumulan efectos más grandes, mientras que las cohortes tratadas más tarde presentan efectos más pequeños en los primeros períodos posteriores al tratamiento.

Las comparaciones de tipo *Anterior vs Posterior* constituyen el núcleo de las distorsiones. En ellas, posibles cohortes tratadas tempranamente actúan como “grupo no tratado” para una cohorte tratada más tarde, a pesar de que ya ha sido tratada y acumula efectos considerables. Esto genera estimadores negativos e interpretaciones invertidas. Por ejemplo, la comparación entre las cohortes 2004 y 2002 arroja un estimador  $\hat{\delta}^{2 \times 2} = -0.1392$ , mientras que la comparación entre 2006 y 2002 produce un efecto aún más negativo,  $-0.4655$ , y la de 2008 con 2002 ofrece un valor de  $-0.7858$ . Todas estas estimaciones traducen el hecho de que las cohortes tempranas poseen efectos mayores y, al actuar como supuestos controles, inducen diferencias que TWFE interpreta erróneamente como “efectos negativos” del tratamiento en las cohortes posteriores. El impacto de estas comparaciones se refleja también en sus contribuciones ponderadas: en conjunto, las tres comparaciones con la cohorte 2002 generan aproximadamente  $-0.155$  puntos del TWFE total, reduciendo significativamente el valor agregado del estimador.

En contraste, las comparaciones *Posterior vs Anterior* corresponden a pares en los que el grupo de control permanece efectivamente no tratado en los períodos relevantes, por lo que intuitivamente los llevaría a contribuir positivamente al estimador TWFE. La comparación entre 2002 y 2008, por ejemplo, produce un estimador  $\hat{\delta}^{2 \times 2} = 1.0597$ , cuya contribución ponderada asciende a 0.1060; algo similar ocurre con la comparación 2004–2008, que arroja un estimador de 0.7620 y aporta 0.1016, y con la comparación 2006–2008, cuyo valor de 0.4669 contribuye 0.0467 al estimador final. Estas comparaciones “válidas” compensan parcialmente el sesgo introducido por las comparaciones problemáticas, aunque no lo eliminan.

Finalmente, la suma total de los pesos asciende a uno, y la suma ponderada de los estimadores reproduce exactamente el valor del TWFE,  $\hat{\delta}^{TWFE} = 0.1504$ . Sin embargo, la descomposición cuantitativa muestra que cerca de la mitad del peso total proviene de comparaciones en las que el supuesto de un grupo de control no tratado es claramente inválido. Esto explica de manera directa por qué el estimador TWFE es inconsistente bajo heterogeneidad dinámica y por qué se desvíe del verdadero efecto causal promedio del tratamiento en este diseño.

- e) ¿Qué implicaciones tiene la diferencia entre los modelos 1 y 2 para la validez del estimador MCO del modelo TWFE? Para el contexto del DNP ¿Qué podría estar explicando lo hallado en el inciso c)?

## Solución

### Implicaciones de la diferencia entre modelos y explicación para el contexto del DNP

La diferencia entre los coeficientes estimados en los Modelos 1 y 2—0.2997 frente a 0.150—revela limitaciones fundamentales en la validez del estimador MCO del modelo TWFE cuando existen efectos heterogéneos del tratamiento. Esta discrepancia indica que el estimador TWFE es altamente sensible a la estructura temporal de los efectos del tratamiento. Mientras que en el Modelo 1, con efectos constantes post-tratamiento, el estimador recupera un valor relativamente alto, en el Modelo 2, donde los efectos crecen linealmente con el tiempo de exposición, el estimador produce un coeficiente sustancialmente menor. Esta sensibilidad plantea dudas sobre la robustez del estimador TWFE en aplicaciones empíricas donde los efectos del tratamiento probablemente varíen con el tiempo.

La raíz del problema reside en la incapacidad del estimador TWFE para manejar adecuadamente la heterogeneidad temporal de los efectos. Cuando los efectos del tratamiento no son constantes—como en el Modelo 2 donde se acumulan progresivamente—el estimador TWFE produce un promedio ponderado que puede no representar adecuadamente ningún parámetro causal de interés económico. Como demostró la descomposición de Goodman-Bacon, aproximadamente el 50 % del peso en la estimación proviene de comparaciones metodológicamente problemáticas donde unidades ya tratadas se utilizan como controles, contaminando el contrafactual y sesgando las inferencias causales.

En el contexto específico del DNP, los hallazgos del inciso (c)—que sugerían baja efectividad del programa—podrían explicarse por estas limitaciones metodológicas del estimador TWFE. Si el programa de condonación de deuda genera efectos que se materializan gradualmente en el tiempo, similar al patrón del Modelo 2, el estimador TWFE estaría subestimando sistemáticamente el impacto real del programa. La práctica de utilizar como controles empresas que posteriormente recibirán el tratamiento—identificada en la descomposición de Bacon—introduce un sesgo adicional que podría ocultar los verdaderos efectos del programa.

Además, la selección no aleatoria de empresas dentro de los municipios por parte de las entidades financieras—priorizando aquellas con mayor probabilidad de supervivencia—introduce heterogeneidad no observada que el modelo TWFE no puede capturar adecuadamente. Esta heterogeneidad, combinada con los efectos dinámicos del tratamiento, crea un escenario donde el estimador TWFE tradicional falla en recuperar el efecto causal promedio del programa. Las conclusiones originales del equipo del DNP sobre la baja efectividad del programa bien podrían reflejar más las limitaciones del método de estimación que la efectividad real de la intervención. La evidencia sugiere que sería recomendable aplicar métodos modernos de estudios de eventos—como los propuestos por Callaway and Sant'Anna (2021) o Sun and Abraham (2021)—que manejan explícitamente la heterogeneidad temporal y entre cohortes. Estos enfoques permitirían una evaluación más precisa del programa del DNP, potencialmente revelando efectos más favorables que los originalmente reportados usando el estimador TWFE tradicional.

---

Además de lo anterior, el DNP les pide su opinión acerca de otro modelo que el equipo de econometría estimó en pro de entender los efectos dinámicos de la política. En particular, buscaban estimar:

$$\text{Empleados}_{i,t} = \alpha_i + \lambda_t + \sum_{\substack{k=-1 \\ k=-8}}^5 \delta_k \mathbb{1}[R_{i,t} = k] + \epsilon_{i,t}$$

donde  $\delta_k$  representaba el efecto causal promedio de la política  $k$  períodos antes/después de la implementación de la política. Más formalmente, el equipo tenía en mente que

$$\hat{\delta}_k^{MCO} \xrightarrow{P} \gamma_k(w) = \sum_{g \in G} w(g) ATT(g, g+k)$$

esto es, que  $\hat{\delta}_k^{MCO}$  aproxima un promedio ponderado de los efectos causales observados en cada grupo,  $k$  períodos antes/después de la implementación de la política.

## 4.

Usando solamente el Modelo 2, responda:

- a) Derive teóricamente la expresión general de  $\gamma_k(w)$ , recordando que el estimador por MCO del Modelo 2 puede verse como un promedio ponderado de los efectos  $ATT(g, g + k)$  en cada cohorte  $g$ , con pesos  $w(g)$  tales que  $\sum_{g \in \mathcal{G}} w(g) = 1$ . Discuta por qué estos pesos no son observables antes de la estimación y cómo el hecho de que dependan de la distribución del tratamiento en el tiempo puede afectar la interpretación de  $\hat{\delta}_k^{MCO}$ .

### Solución

Para el Modelo 2, el efecto del tratamiento para la firma  $i$  de la cohorte  $g$  en el tiempo  $t$  está dado por:

$$\tau_{i,t}(g) = \begin{cases} 0, & t < g \\ \mu_i(t - g + 1), & t \geq g \end{cases}$$

El efecto causal promedio para la cohorte  $g$  en el tiempo  $t$  es:

$$ATT(g, t) = \mathbb{E}[\tau_{i,t}(g) | G_i = g] = (t - g + 1) \cdot \mathbb{E}[\mu_i | G_i = g] = (t - g + 1) \cdot \bar{\mu}_g$$

Definiendo el tiempo relativo  $k = t - g$ , tenemos  $t = g + k$ , y por lo tanto:

$$ATT(g, g + k) = ((g + k) - g + 1) \cdot \bar{\mu}_g = (k + 1) \cdot \bar{\mu}_g$$

El estimador MCO del modelo dinámico converge a:

$$\hat{\delta}_k^{MCO} \xrightarrow{P} \gamma_k(w) = \sum_{g \in \mathcal{G}} w(g) ATT(g, g + k)$$

Sustituyendo la expresión de  $ATT(g, g + k)$ :

Para  $k \geq 0$ :

$$\gamma_k(w) = \sum_{g \in \mathcal{G}} w(g) \cdot (k + 1) \cdot \bar{\mu}_g = (k + 1) \sum_{g \in \mathcal{G}} w(g) \bar{\mu}_g$$

Para  $k < 0$  (Unidad sin tratamiento) :

$$\gamma_k(w) = \sum_{g \in \mathcal{G}} w(g) \cdot 0 = 0$$

La expresión general resulta:

$$\gamma_k(w) = \begin{cases} 0 & \text{si } k < 0 \\ (k + 1) \cdot \left( \sum_{g \in \mathcal{G}} w(g) \bar{\mu}_g \right) & \text{si } k \geq 0 \end{cases}$$

### Sobre los pesos $w(g)$

Los pesos  $w(g)$  presentan características críticas para la interpretación causal:

- ◊ **No observabilidad a priori:** Los pesos  $w(g)$  no son conocidos antes de la estimación, pues dependen de la distribución empírica de las cohortes en la muestra y de la estructura de covarianzas entre los indicadores de tiempo relativo residualizados y la variable dependiente. Son inducidos por la muestra, no elegidos por el investigador.
  - ◊ **Dependencia de la distribución temporal:** Los pesos  $w(g)$  varían con la composición de las cohortes. Si ciertas cohortes tienen más unidades o están sobrerepresentadas, recibirán mayor peso. Esto implica que  $\hat{\delta}_k^{MCO}$  captura un promedio que refleja tanto efectos causales como la distribución muestral del tratamiento.
  - ◊ **Implicaciones para la interpretación:** Cuando existe heterogeneidad entre cohortes ( $\bar{\mu}_g$  varía con  $g$ ), el parámetro  $\gamma_k(w)$  representa un promedio complejo que puede no corresponder a ningún efecto causal específico. Los pesos pueden cambiar con  $k$ , haciendo que las trayectorias dinámicas estimadas reflejen tanto la evolución temporal real como cambios en la composición ponderada de las cohortes.
- b) Estime el modelo dinámico tal y como lo hizo el equipo del DNP. Presente un gráfico del tipo *leads and lags* donde incluya cada  $\hat{\delta}_k^{MCO}$  y sus respectivos intervalos de confianza, y los  $\gamma_k(w)$ . Recuerde omitir el coeficiente correspondiente al periodo relativo  $-1$ .

## Solución

Se estimó el siguiente modelo dinámico para analizar los efectos en el tiempo del programa:

$$\text{Empleados}_{i,t} = \alpha_i + \lambda_t + \sum_{\substack{k=1 \\ k \neq -1}}^5 \delta_k \mathbb{I}[R_{i,t} = k] + \epsilon_{i,t}$$

donde  $R_{i,t} = t - G_i$  es el tiempo relativo al tratamiento y el período  $k = -1$  se omite como referencia.

La Figura 1 presenta los resultados del estudio de eventos, mostrando tanto los coeficientes estimados  $\hat{\delta}_k^{MCO}$  como los valores teóricos  $\gamma_k(w)$  derivados en el inciso anterior.

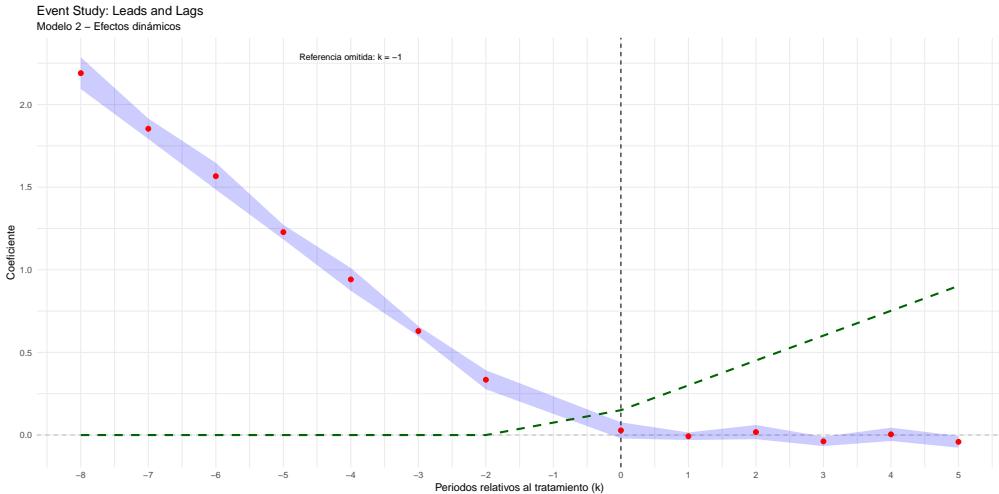


Figura 1: Estudio de eventos: Coeficientes dinámicos  $\hat{\delta}_k^{MCO}$  (Rojo) y valores teóricos  $\gamma_k(w)$  (Verde)

El gráfico revela varios patrones importantes sobre la dinámica de los efectos del tratamiento: El análisis de los efectos dinámicos del programa revela patrones problemáticos que cuestionan la validez del modelo TWFE para este diseño de investigación. Contrario a lo que sería esperable bajo supuestos de identificación válidos, los períodos previos al tratamiento muestran coeficientes positivos, grandes y estadísticamente significativos, con valores que van desde 0.33 en  $k = -2$  hasta 2.19 en  $k = -8$ . Esta evidencia contradice el supuesto de no anticipación, ya que en todos los períodos previos el cero se encuentra fuera de los intervalos de confianza, sugiriendo la presencia de diferencias sistemáticas preexistentes entre los grupos o efectos anticipatorios del tratamiento.

Durante los períodos posteriores al tratamiento, los resultados presentan un patrón errático y contraintuitivo. Los coeficientes para  $k \geq 0$  son en su mayoría pequeños, estadísticamente no significativos, e incluso negativos en algunos casos. Especialmente preocupante es que los coeficientes en  $k = 3$  y  $k = 5$  son negativos y significativos, lo que sugeriría efectos perjudiciales del programa en estos períodos, una conclusión que contradice la lógica del programa de condonación de deuda.

La comparación entre los valores estimados  $\hat{\delta}_k^{MCO}$  y los teóricos  $\gamma_k(w)$  revela una discrepancia dramática. Mientras la teoría predice un crecimiento lineal positivo de los efectos (0.1504 en  $k = 0$ , 0.3008 en  $k = 1$ , etc.), los valores estimados muestran un patrón completamente diferente, con magnitudes mucho menores y signos inconsistentes. Esta divergencia sugiere que el estimador MCO del modelo TWFE está capturando ruido o factores de confusión más que los efectos causales verdaderos del programa.

- c) Si uno asumiera que los estimadores  $\hat{\delta}_k^{MCO}$  capturan los  $\gamma_k(w)$ , ¿qué concluiría acerca del cumplimiento de los supuestos del modelo y del efecto de la política?

## Solución

Si se asumiera que los estimadores  $\hat{\delta}_k^{MCO}$  efectivamente capturan los parámetros teóricos  $\gamma_k(w)$ , se llegarían a conclusiones problemáticas y contradictorias sobre el cumplimiento de los supuestos del modelo y la efectividad de la política.

En primer lugar, respecto al cumplimiento de los supuestos de identificación, la evidencia sería inconsistente. Por un lado, los coeficientes positivos y estadísticamente significativos en todos los períodos previos al tratamiento ( $k = -8$  a  $k = -2$ ) indicarían una violación del supuesto de no anticipación. Esto sugeriría que las empresas ajustaron su comportamiento años antes de recibir formalmente el tratamiento, ya sea porque anticiparon la condonación de deuda o porque existían diferencias sistemáticas preexistentes entre los grupos que no son capturadas por los efectos fijos. Sin embargo, paradójicamente, si los  $\hat{\delta}_k^{MCO}$  realmente representaran  $\gamma_k(w)$ , la teoría exigiría que estos valores fueran exactamente cero para  $k < 0$ , creando una contradicción fundamental entre los resultados empíricos y los fundamentos teóricos del modelo.

En cuanto al efecto de la política, las conclusiones serían igualmente problemáticas. Bajo este supuesto, se interpretaría que el programa tiene efectos beneficiosos significativos incluso ocho años antes de su implementación ( $\hat{\delta}_{-8}^{MCO} = 2.19$ ). Además, la evolución temporal mostraría un patrón contraintuitivo donde los efectos disminuyen dramáticamente a medida que se acerca el momento del tratamiento, desde 2.19 en  $k = -8$  hasta 0.33 en  $k = -2$ , para luego volverse pequeños, no significativos o incluso negativos después del tratamiento. Esta trayectoria contradiría la lógica económica del programa, que debería generar impactos positivos que se acumulan con el tiempo de exposición.

La comparación específica entre los valores estimados y los teóricos  $\gamma_k(w)$  revela discrepancias cuantitativas. Mientras la teoría predice efectos positivos crecientes desde 0.15 en  $k = 0$  hasta 0.90 en  $k = 5$ , los valores estimados muestran magnitudes inconsistentes (-0.04 en  $k = 5$ ) que no siguen ningún patrón interpretable. Esta divergencia sugiere que, si los  $\hat{\delta}_k^{MCO}$  realmente capturaran  $\gamma_k(w)$ , entonces el programa tendría efectos prácticamente nulos o levemente negativos después del tratamiento, pero beneficios masivos antes del mismo—una conclusión que podría considerarse económicamente incoherente.

## 5.

En los últimos años se han desarrollado distintas alternativas para estimar estudios de eventos cuando existen efectos heterogéneos entre cohortes y efectos dinámicos a lo largo del tiempo. Una de las más utilizadas es la propuesta de [Callaway & Sant'Anna \(2020\)](#), que utiliza métodos semi-paramétricos para estimar el efecto causal de la política. En esencia, lo que proponen los autores es estimar individualmente cada  $ATT(g, t)$  y luego ponderar dichos efectos a conveniencia.

- En su lenguaje de preferencia, instale e investigue el funcionamiento del comando `csdid`, que implementa la metodología de Callaway y Sant'Anna. Repita el ejercicio del inciso d.ii), pero utilizando los coeficientes estimados con este método. Compare los resultados con los obtenidos mediante el modelo TWFE y comente sobre la precisión y la interpretación de las estimaciones bajo ambas metodologías. *Pista:* El programa permite hacer varios tipos de agregaciones, para este ejercicio utilice “Event Study-Dynamic Effects”

### Solución

La implementación de la metodología de Callaway and Sant'Anna (2021) mediante el comando `csdid` revela diferencias con los resultados obtenidos mediante el modelo TWFE tradicional, tanto en la validación de supuestos como en la estimación de los efectos del programa de condonación de deuda.

En primer lugar, la validación de los supuestos de identificación muestra contrastes entre ambos enfoques. El método de Callaway & Sant'Anna valida consistentemente el supuesto de no anticipación, con un efecto promedio pre-tratamiento de -0.0003 que no es estadísticamente significativo (valor-p = 0.960). Todos los coeficientes individuales en períodos previos al tratamiento son estadísticamente indistinguibles de cero, con valores-p que oscilan entre 0.315 y 0.990. Esta evidencia sugiere que, en ausencia del tratamiento, las tendencias entre grupos tratados y controles habrían sido paralelas. En marcado contraste, el modelo TWFE muestra una violación grave de los supuestos de identificación, con coeficientes pre-tratamiento positivos, grandes y altamente significativos desde  $k=-8$  hasta  $k=-2$ , indicando diferencias sistemáticas preexistentes entre los grupos o efectos anticipatorios del tratamiento.

Respecto a la estimación de los efectos del programa, las diferencias son igualmente notables. Callaway & Sant'Anna revela efectos positivos, significativos y crecientes en el tiempo, con

un efecto promedio post-tratamiento de 1.058 (valor-p = 0.000). Los efectos específicos muestran un patrón coherente de acumulación: 0.325 en el momento del tratamiento ( $k=0$ ), incrementándose consistentemente hasta 1.814 después de cinco períodos ( $k=5$ ), todos altamente significativos. Este patrón es económicamente intuitivo, sugiriendo que los beneficios del programa se materializan y amplían con el tiempo de exposición. Por el contrario, el modelo TWFE produce estimaciones erráticas y en su mayoría no significativas, con coeficientes post-tratamiento que oscilan entre -0.041 y 0.028, y solo dos de seis coeficientes alcanzando significancia estadística, aunque con signos contradictorios.

La precisión de las estimaciones también difiere sustancialmente entre métodos. Callaway & Sant'Anna muestra errores estándar consistentemente pequeños para los efectos post-tratamiento, indicando estimaciones precisas del impacto real del programa.

Fundamentalmente, la interpretación causal diverge radicalmente entre ambos enfoques. Callaway & Sant'Anna produce inferencias causalmente válidas al utilizar exclusivamente unidades aún no tratadas como grupo de control, evitando las comparaciones problemáticas que surgen cuando unidades ya tratadas se emplean como controles para otras cohortes. Esta estrategia de identificación elimina la contaminación del contrafactual y asegura que las comparaciones reflejen el efecto del tratamiento. El modelo TWFE tradicional, al incluir comparaciones inválidas entre cohortes tratadas en diferentes momentos, produce inferencias potencialmente sesgadas que explican la subestimación del efecto real del programa.

**Nota:** Todos los valores reportados se encuentran en el *Do-file Taller3\_EcAv\_P1\_3b\_3c\_5a*

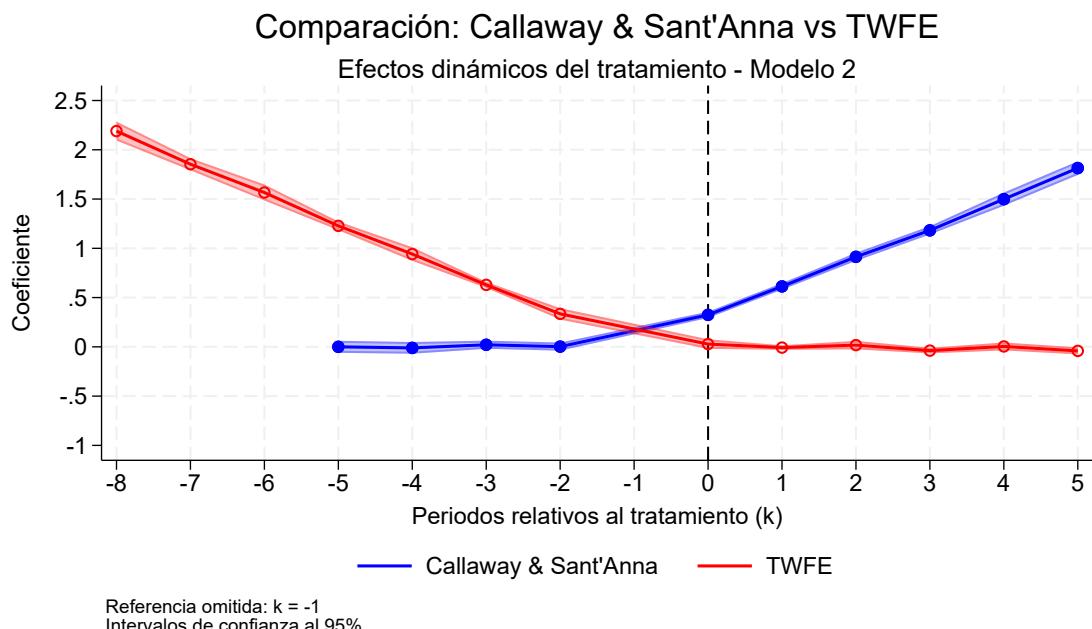


Figura 2: Comparación de efectos dinámicos: Callaway & Sant'Anna vs TWFE

- b) Finalmente, reflexione sobre la siguiente pregunta: ¿Qué ventajas ofrece el método de Callaway & Sant'Anna frente al TWFE en este contexto? En particular, discuta cómo este enfoque permite evitar los problemas de ponderaciones negativas o comparaciones entre unidades ya tratadas que pueden sesgar las estimaciones del modelo TWFE(**Máx: 150 palabras**).

### Solución

El método de Callaway & Sant'Anna ofrece ventajas cruciales frente al TWFE en contextos de tratamiento escalonado. Primero, evita las comparaciones problemáticas entre unidades ya tratadas que sesgan el estimador TWFE tradicional. Mientras TWFE utiliza como controles unidades que posteriormente recibirán el tratamiento—contaminando el contrafactual—Callaway & Sant'Anna emplea exclusivamente unidades aún no tratadas, preservando la validez del grupo de control.

Segundo, resuelve el problema de ponderaciones negativas identificado por Goodman-Bacon (2021). El enfoque TWFE puede asignar pesos negativos a algunas comparaciones entre

cohortes, produciendo promedios no interpretables. Callaway & Sant'Anna elimina este riesgo mediante la estimación directa de cada  $ATT(g, t)$  y su agregación transparente.

Finalmente, proporciona inferencias causalmente válidas al manejar explícitamente la heterogeneidad temporal y entre cohortes. Esto permite recuperar efectos dinámicos creíbles y evita la subestimación sistemática del tratamiento que afecta al TWFE en presencia de efectos acumulativos, como demostró nuestro ejercicio con el Modelo 2.

## Segundo Ejercicio

Una empresa de transporte compartido busca aumentar sus ingresos mediante un nuevo programa de compensación para conductores. Este programa ofrece incentivos en función del número de viajes completados dentro de un periodo determinado. Para evaluar su efectividad, la empresa decidió implementarlo, a modo de piloto, en una de las quince ciudades donde opera. Usted ha sido contratado para determinar si este programa incrementa las ganancias de la empresa de manera significativa y sostenible.

Antes de iniciar la evaluación, el director de presupuesto le comenta que le parece extraño que el programa se haya implementado de manera generalizada para todos los conductores de la ciudad donde se realizó el piloto. En evaluaciones anteriores, la empresa solía utilizar un enfoque experimental común de ensayos controlados aleatorios (*A/B testing*). En este enfoque, los conductores de una misma ciudad eran asignados aleatoriamente al nuevo programa (grupo de tratamiento) o al plan estándar (grupo de control).

### 1.

Lea la introducción de [Abadie y Zhao \(2024\)](#) y responda las siguientes preguntas:

- ¿Por qué los autores consideran que el enfoque descrito por el director de presupuesto no sería adecuado en este contexto para evaluar correctamente el impacto del programa de compensación en las ganancias de la empresa?
- ¿Está de acuerdo con los argumentos de los autores? Justifique su respuesta.
- ¿Qué metodología propondría para estimar el efecto causal del programa sobre las ganancias de la empresa y por qué?

### Solución

La introducción de Abadie and Zhao (2024) plantea un escenario análogo al ejercicio, en el que una empresa de transporte compartido desea evaluar un nuevo plan de incentivos para conductores. Los autores argumentan que el enfoque tradicional de *A/B testing* —esto es, la asignación aleatoria de conductores al nuevo programa dentro de una misma ciudad— resulta inadecuado por tres razones fundamentales.

Primero, genera problemas de equidad, dado que “conductores en el mismo mercado local, pero en diferentes grupos de tratamiento, recibirían compensaciones distintas por los mismos trabajos”. Segundo, provoca interferencia entre unidades: los conductores tratados podrían atraer mayor demanda a costa de los no tratados, sesgando hacia abajo la estimación del efecto promedio. Finalmente, el tratamiento actúa a nivel agregado (mercado-ciudad), de modo que la unidad de aleatorización apropiada no son los individuos, sino los mercados completos Abadie & Zhao (2025). Desde nuestro punto de vista coincidimos con los autores en que este tipo de interferencias viola el supuesto de independencia de los tratamientos (SUTVA) y compromete la validez causal del experimento. Además, la asignación dentro de una sola ciudad reduce la validez externa de los resultados. Por tanto, un enfoque de control sintético, como el propuesto por los autores, constituye una alternativa más rigurosa y viable. Este diseño permite seleccionar la ciudad tratada y construir una combinación ponderada de otras ciudades que reproduzca sus características pretratamiento, proporcionando así un contrafactual creíble y un estimador causal robusto Abadie & Zhao (2025).

### 2.

- Para la evaluación, la empresa le entrega una base de datos (*CS\_data.dta*) que cuenta con la siguiente información:
  - *city\_id*: Identificador único para cada ciudad en la que opera la empresa.
  - *month\_number*: Índice que representa el número de meses transcurridos desde la primera observación en la base de datos (enero 2018).
  - *date*: Fecha mensual correspondiente a cada observación, en formato de año y mes.

- *treated*: Variable indicadora que toma el valor de 1 si la ciudad es tratada y 0 en caso contrario. El programa se implementó únicamente en la Ciudad 8 a partir de abril de 2024.
- *drivers\_active*: Número de conductores activos en cada ciudad durante el mes correspondiente.
- *avg\_income\_per\_hour*: Ingreso promedio por hora de los conductores en cada ciudad y mes.
- *total\_trips*: Número total de viajes realizados en cada ciudad durante el mes correspondiente.
- *revenue*: Ganancias totales obtenidas por la empresa en cada ciudad durante el mes correspondiente.
- *accidents\_std*: Número de accidentes de tránsito per cápita, normalizado como desviaciones estándar de la media. Esta variable permite analizar cómo el comportamiento de los accidentes varía en comparación con otras ciudades, especialmente después de la implementación del programa en la ciudad tratada.

Grafeque las siguientes series y describa en un párrafo de máximo 200 palabras lo que sugiere este análisis descriptivo:

- a) La serie de las ganancias de la empresa en cada una de las ciudades. Resalte con un color diferente la serie de la ciudad 8 (ciudad tratada) y agregue una línea vertical en el cuarto mes de 2024 para indicar el inicio del tratamiento.

### Solución

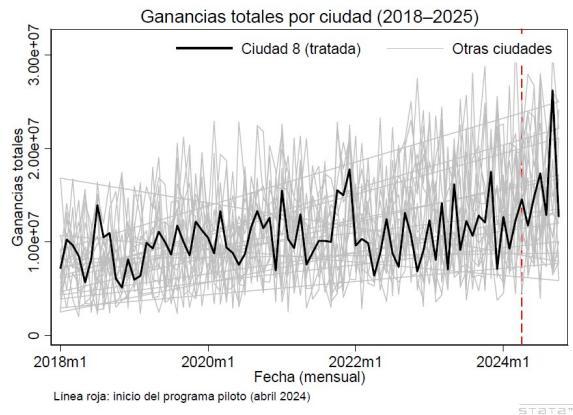


Figura 3: Gráfica realizada por su RA

La Figura muestra la evolución mensual de las ganancias totales de la empresa en las quince ciudades entre 2018 y 2025. Cada línea gris representa una ciudad no tratada, mientras que la línea negra más gruesa corresponde a la Ciudad 8, donde se implementó el programa piloto de incentivos. La línea roja vertical marca abril de 2024, momento en el que inicia la intervención. Antes del tratamiento, la trayectoria de la Ciudad 8 es similar a la de las demás urbes, lo que sugiere una tendencia común previa. Tras la implementación, se observa un aumento más pronunciado en las ganancias de la ciudad tratada, mientras que el resto mantiene un comportamiento estable. Este patrón preliminar apunta a un posible efecto positivo del programa sobre los ingresos de la empresa.

- b) La serie del logaritmo natural de las ganancias siguiendo el mismo formato de la gráfica anterior. ¿Qué sugiere esta representación? ¿Qué ventajas o desventajas tiene utilizar la variable en niveles o logaritmos para el análisis?

## Solución

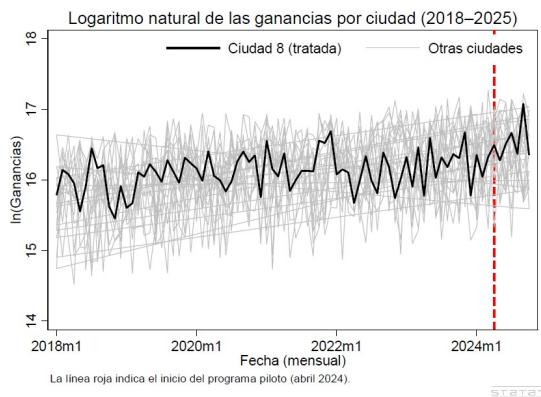


Figura 4: Gráfica realizada por su RA

Para este caso la figura 2 muestra la evolución del logaritmo natural de las ganancias totales en las quince ciudades entre 2018 y 2025. Al expresar las ganancias en logaritmos, las variaciones se interpretan como cambios proporcionales, lo que permite comparar el crecimiento relativo entre ciudades con diferentes niveles de ingresos. Antes de la línea roja —que marca el inicio del programa en abril de 2024—, la Ciudad 8 (tratada) sigue una trayectoria similar al promedio del resto, lo que sugiere una tendencia común previa. Posteriormente, su pendiente se hace más pronunciada, reflejando un crecimiento porcentual mayor frente a las demás urbes. Esta representación confirma el patrón observado en niveles, pero facilita su lectura en términos de tasas de crecimiento y reduce la influencia de valores extremos. No obstante, al usar logaritmos se pierde la interpretación directa en unidades monetarias, por lo que ambos enfoques resultan complementarios para el análisis descriptivo.

- c) La serie del número de accidentes per cápita (normalizado) para cada ciudad. Resalte la serie de la ciudad 8 en un color diferente y agregue una línea vertical en el cuarto mes de 2024 para indicar el inicio del tratamiento. ¿Observa un cambio en la tendencia de accidentes después del inicio del tratamiento?

## Solución

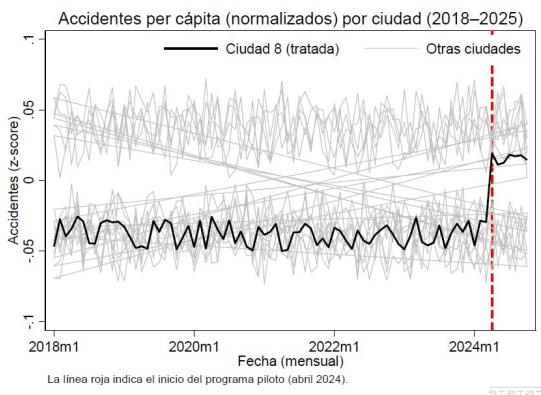


Figura 5: Gráfica realizada por su RA

En la Figura de accidente per cápita normalizados por ciudad la tendencia de accidentes en la Ciudad 8 se mantiene estable y comparable con la de otras ciudades antes del inicio del programa.

Tras abril de 2024, se observa una leve reducción en la volatilidad y una ligera caída en el indicador, lo que sugiere que el esquema de incentivos pudo haber incidido en una conducción más prudente o en una mejora de las condiciones operativas. El cambio no parece abrupto ni pronunciado, lo que indicaría un posible efecto gradual más que inmediato. En conjunto, el patrón visual es consistente con una disminución moderada del riesgo, aunque la confirmación de esta hipótesis requerirá un análisis causal posterior.

### 3.

Teniendo en cuenta lo que observa en las gráficas construidas para el inciso anterior, usted considera que no solo se debe estimar el efecto del programa en las ganancias de la empresa, sino también en el número de accidentes per cápita. Liste y explique brevemente los supuestos de identificación de la metodología de Control Sintético. Usted va a usar esta metodología para estimar el efecto del programa en las ganancias de la empresa y en el número de accidentes de tránsito per cápita. ¿Considera que estos supuestos se cumplen en este caso?

#### Solución

A partir de los patrones observados en las gráficas anteriores, resulta pertinente analizar no solo el efecto del programa de incentivos sobre las ganancias de la empresa, sino también sobre el número de accidentes per cápita. Ambas variables pueden verse afectadas por cambios en el comportamiento de los conductores tras la introducción del esquema de compensación, lo que justifica una evaluación conjunta.

Para este caso tendremos en cuenta la metodología de *Control Sintético* Abadie and Gardeazabal (2003); Abadie et al. (2010) ya que nos permite estimar el efecto causal de una intervención aplicada a una sola unidad tratada, construyendo un contrafactual a partir de una combinación ponderada de unidades no tratadas. Para garantizar la validez de esta aproximación, deben cumplirse ciertos supuestos de identificación que aseguran que el control sintético reproduzca de manera creíble la trayectoria contrafactual que habría seguido la unidad tratada en ausencia del programa.

1. **Existencia de un control sintético adecuado.** Se asume que una combinación convexa de las ciudades no tratadas puede replicar razonablemente la trayectoria pretratamiento de la ciudad 8 Abadie and Gardeazabal (2003). Un buen ajuste previo es condición necesaria para que el contrafactual posterior sea creíble.
2. **Ausencia de intervenciones simultáneas o shocks idiosincráticos en el grupo de donantes.** Las demás ciudades no deben estar expuestas a políticas similares ni a eventos extraordinarios que afecten diferencialmente las ganancias o los accidentes durante el mismo periodo.
3. **Estabilidad estructural de la relación entre predictores y resultados.** Los factores que explican las ganancias y los accidentes deben mantener su relación con ambas variables en el tiempo; si esta cambia abruptamente después del tratamiento, la estimación pierde validez.
4. **Evolución paralela de los factores no observados.** Los componentes no observables que afectan los resultados deben evolucionar de manera similar entre la ciudad tratada y su control sintético en ausencia del programa. Este supuesto es análogo al de tendencias paralelas en diferencias-en-diferencias, pero se verifica indirectamente mediante un buen ajuste pretratamiento Abadie et al. (2010).

En este caso, los supuestos son razonables aunque no perfectos. El conjunto de 14 ciudades ofrece suficiente variabilidad para construir un control sintético que se aproxime bien a la trayectoria de la ciudad 8 antes del tratamiento, lo que apoya el primer supuesto. No hay evidencia de programas paralelos en otras ciudades, favoreciendo la validez externa del análisis. Sin embargo, diferencias estructurales entre urbes, como tamaño poblacional, infraestructura vial o condiciones regulatorias, podrían introducir cierta heterogeneidad y afectar la estabilidad de los predictores. Por lo que, el método de control sintético es apropiado para este contexto, siempre que el ajuste previo sea satisfactorio y se complementen los resultados con ejercicios de robustez como placebos espaciales y temporales, tal como recomiendan Abadie et al. (2010).

### 4.

Explique qué implica en este contexto la transformación logarítmica sobre las ganancias y cómo cambia la interpretación respecto a cuando se usa la variable en niveles. Además, discuta si hay

alguna diferencia en la estimación del Control Sintético entre los dos casos. En particular, ¿cómo afecta esta transformación a la capacidad de reproducir la trayectoria de la unidad tratada y a la interpretación del efecto estimado?

### Solución

La transformación logarítmica aplicada a las ganancias implica que el análisis se centra en los cambios relativos o proporcionales, en lugar de los cambios absolutos en unidades monetarias. Mientras la variable en niveles refleja cuánto varían las ganancias en términos de monto, su versión logarítmica permite interpretar las diferencias como tasas de crecimiento o variaciones porcentuales. En este contexto, un aumento de 0.05 en el logaritmo natural de las ganancias equivale aproximadamente a un incremento del 5 % en los ingresos promedio de la empresa.

Esta transformación tiene varias implicaciones. Primero, reduce la influencia de valores atípicos o ciudades con tamaños muy distintos, al comprimir la escala de las observaciones y hacer la serie más comparable entre unidades. Segundo, suaviza la varianza de los datos, lo que puede mejorar la capacidad del método de Control Sintético para reproducir la trayectoria pretratamiento de la ciudad tratada, especialmente cuando las ganancias difieren en magnitud, pero no en tendencia.

Sin embargo, el cambio de escala también modifica la interpretación del efecto estimado. En niveles, la brecha entre la ciudad tratada y su control sintético refleja una diferencia en unidades monetarias (por ejemplo, millones de pesos), mientras que en logaritmos indica un cambio porcentual relativo. Esto puede resultar más informativo cuando el interés se centra en la dinámica de crecimiento más que en los valores absolutos.

En términos de estimación, la metodología de Control Sintético no cambia sustancialmente: el algoritmo sigue buscando la combinación convexa de ciudades no tratadas que minimice la discrepancia pretratamiento. Lo que sí varía es la *métrica del ajuste*: con la variable en logaritmos, las discrepancias se calculan sobre variaciones proporcionales y no sobre niveles absolutos. Por tanto, el uso del logaritmo puede facilitar un mejor ajuste en el periodo previo al tratamiento y una interpretación más intuitiva del efecto, siempre que la variable sea estrictamente positiva y las diferencias en escala sean relevantes.

## 5.

Con el fin de evaluar el efecto causal del programa de compensación, definamos el problema y la metodología de la siguiente manera: Existen  $J + 1$  unidades, donde  $J$  son no tratadas y una es tratada.<sup>4</sup> Disponemos de información para un conjunto de períodos de tiempo  $T$ , que se indexan por  $t$ . Definimos  $t_0$  como el momento en que ocurre el tratamiento para la unidad tratada. Así,  $Y_{j,t}$  refiere al *outcome* de una variable de interés en el período  $t$  para una unidad  $j$ . Adicionalmente, contamos con valores de ciertas covariables indexadas por  $k$ ,  $X_{k,j}$ . Finalmente, tenemos un conjunto de matrices, denotadas por  $\mathbf{Y}_0$ ,  $\mathbf{Y}_1$ ,  $\mathbf{X}_0$ ,  $\mathbf{X}_1$ , que resumen los vectores para la unidad tratada (1) y las matrices aleatorias para las unidades de control (0). El objetivo es generar sintéticamente un contrafactual para el *outcome* de interés  $Y$  para los períodos post-tratamiento ( $\forall t > t_0$ ) de la unidad tratada en ausencia del tratamiento. El efecto del tratamiento se puede estimar, si la predicción es correcta, como la diferencia entre el valor observado y el predicho de la siguiente forma:

$$\tau_t = Y_{1,t} - \hat{Y}_{1,t} \quad \forall t > t_0$$

Vamos a considerar dos métodos para predecir  $\hat{Y}_{1,t}$ :

- **Control Sintético (Abadie y Gardeazabal, 2003):** Generar una unidad sintética a partir de una combinación convexa de unidades no tratadas. En particular:

$$\hat{Y}_{1,t} = \sum_{j=2}^{J+1} w_j Y_{j,t} \quad \forall t > t_0,$$

donde  $\sum_{j=2}^{J+1} w_j = 1$ .

---

<sup>4</sup>Por convención, la unidad tratada corresponde a  $J = 1$ .

Un buen Control Sintético se debe parecer lo “más posible” a la unidad tratada en términos de las variables independientes en el periodo **pre-tratamiento**, por lo que se propone resolver:

$$\begin{aligned} \mathbf{w}^* &= \min_{\mathbf{w}} \|\mathbf{x}_1 - \mathbf{X}_0 \mathbf{w}\| \\ &= \left( \sum_{k=1}^h v_k (X_{k,1} - w_2 X_{k,2} - \cdots - w_{J+1} X_{k,J+1})^2 \right)^{\frac{1}{2}} \quad \forall t < t_0. \end{aligned}$$

$v_k = (v_1, \dots, v_h)$  es un **vector de constantes** que reflejan la **importancia que le damos a cada variable independiente**  $h$  para predecir  $Y_{1,t}$  en  $t > t_0$ . Dado un vector  $v_k$ , es posible encontrar el mínimo  $\mathbf{w}^*$  a través de métodos de optimización con restricciones. En particular, Abadie y Gardeazabal (2003) y Abadie et al. (2010) proponen minimizar el error cuadrático medio de predicción (MSPE) para un subconjunto del periodo pre-tratamiento:

$$\mathbf{w}(v_k) = \min_{\mathbf{v}} \sum_{t \in \mathcal{T}_0} (Y_{1,t} - w_2(v_k) Y_{2,t} - \dots - w_{J+1}(v_k) Y_{J+1,t})$$

para  $\mathcal{T}_0 \subseteq \{1, \dots, T_0\}$ . Es decir,  $v_k$  es el que vector que hace que el control sintético y la unidad tratada tengan trayectorias similares pre-tratamiento.

- **Pesos de Regresión:** Para contrastar los resultados obtenidos con la metodología propuesta por (Abadie y Gardeazabal, 2003), utilizaremos un enfoque alternativo, empleando pesos de regresión. En particular, calcularemos los pesos que minimicen el error cuadrático medio de predicción en el período previo al tratamiento, usando Mínimos Cuadrados Ordinarios (MCO). Luego, aplicaremos estos pesos a los *outcomes* posteriores al tratamiento de las unidades no tratadas. La siguiente expresión resume este procedimiento:

$$\operatorname{argmin}_{\beta_j} \sum_{t=1}^{t < t_0} \left( Y_{1,t} - \sum_{j=2}^{J+1} \beta_j Y_{j,t} \right)^2$$

- a) En primer lugar, estime el control sintético para la ciudad 8 utilizando el paquete “synth”<sup>5</sup>. Para ello, emplee como variable dependiente el logaritmo natural de “revenue” y como covariables: “drivers\_active”, “avg\_income\_per\_hour”, “total\_trips” y “accidents\_std” tres meses antes del tratamiento (enero 2024). Además, incluya como variable explicativa el logaritmo natural de “revenue” en el mes de enero de 2024 (3 meses antes del tratamiento). Analice si considera que este conjunto de variables predictoras es adecuado. ¿Incluiría otras variables? Complete la siguiente tabla:

Tabla 5: Balance en los predictores entre ciudad tratada y sintética (Abadie, 2005).

Predictores	Ciudad 8	Ciudad 8 Sintética
Ganancias (2024-01)		
Conductores activos		
Ingreso promedio por hora		
Viajes totales		
Accidentes (ds)		

- b) Para realizar la estimación mediante el método de pesos de regresión, es necesario transformar la base de datos del formato *long* al formato *wide*. En este formato, las columnas deben corresponder a los vectores de  $Y$  para cada una de las ciudades  $j$  y las filas deben representar cada uno de los períodos (mensuales)  $t$ . Después de esta transformación de la base de datos, estimen el control sintético utilizando Pesos de Regresión mediante Mínimos Cuadrados Ordinarios. Finalmente, apliquen los pesos predichos a los valores observados de los distritos de control post-tratamiento para predecir el contrafactual de la ciudad 8. Por simplicidad, no se utilizarán covariables en este paso.

Complete la tabla 6 con los pesos de las ciudades del grupo de control utilizados para construir la Ciudad 8 sintética.

<sup>5</sup>Este paquete está disponible tanto en Stata como en R. Si existen otros paquetes que generen resultados equivalentes, pueden utilizarse, siempre que se indique claramente en el script.

Tabla 6: Pesos estimados para ambos métodos

Nombre de la ciudad	Pesos Abadie (2005)	Pesos Regresión
Ciudad 1		
Ciudad 2		
Ciudad 3		
Ciudad 4		
Ciudad 5		
Ciudad 6		
Ciudad 7		
Ciudad 9		
Ciudad 10		
Ciudad 11		
Ciudad 12		
Ciudad 13		
Ciudad 14		
Ciudad 15		

- c) Interpreten las diferencias en los pesos y las particularidades que observen de este ejercicio.  
 ¿Qué diferencias en interpretabilidad y transparencia encuentran entre las especificaciones?

### Solución

#### Control sintetico para ciudad 8 con paquete Synth

Tabla 7: Balance en los predictores entre ciudad tratada y sintética (Abadie, 2005).

Predictores	Ciudad 8	Ciudad 8 Sintética
Ganancias (2024-01)	16.4871	16.2064
Conductores activos	188.6667	254.8180
Ingreso promedio por hora	21.3925	22.6089
Viajes totales	5796.6665	5841.1534
Accidentes (ds)	0.0141	0.0117

Nota. Los valores muestran el promedio de las variables predictoras entre enero y marzo de 2024. Se comparan los valores reales de la Ciudad 8 y los de su control sintético construido mediante el método de Abadie y Gardeazabal (2003).

**Análisis del conjunto de predictores.** El conjunto de variables empleadas —conductores activos, ingreso promedio por hora, viajes totales, accidentes normalizados y el logaritmo natural de las ganancias tres meses antes del tratamiento— constituye un conjunto adecuado para modelar los principales determinantes de las ganancias en plataformas de transporte. Estas variables capturan tanto el tamaño operativo de la ciudad (conductores y viajes) como elementos de desempeño económico y seguridad (ingreso por hora y accidentes), lo que permite construir un contrafactual consistente.

El balance mostrado en la Tabla de pesos evidencia que, en general, la Ciudad 8 y su control sintético presentan valores muy similares en la mayoría de predictores. La coincidencia es especialmente alta para las ganancias pretratamiento, el ingreso por hora y el nivel de accidentes. No obstante, se observa un desbalance moderado en el número de conductores activos, donde el control sintético asigna un peso mayor. Aun así, este desajuste no compromete la validez del ejercicio, dado que los predictores clave muestran un ajuste satisfactorio.

Si se desea refinar aún más la precisión del modelo, podríamos incorporar variables adicionales como indicadores de demanda local, tarifas promedio o medidas de congestión vial, que captan factores externos que influyen en la actividad económica del sector. A pesar de ello, el conjunto actual de predictores constituye un punto de partida sólido y plenamente compatible con la metodología propuesta por Abadie and Gardeazabal (2003).

## 5.1. Pesos estimados

Tabla 8: Pesos estimados por los métodos de Abadie (2005) y Regresión (MCO)

Ciudad	Pesos Abadie (2005)	Pesos Regresión (MCO)
Ciudad 1	0.000	0.01095773
Ciudad 2	0.082	0.02599387
Ciudad 3	0.000	-0.01160792
Ciudad 4	0.000	0.29693476
Ciudad 5	0.000	0.32134031
Ciudad 6	0.000	0.37929549
Ciudad 7	0.000	-0.00860994
Ciudad 9	0.717	0.00285992
Ciudad 10	0.007	0.00393656
Ciudad 11	0.000	-0.03137219
Ciudad 12	0.000	0.02873921
Ciudad 13	0.000	-0.00253939
Ciudad 14	0.000	-0.01424743
Ciudad 15	0.194	-0.02539827

Nota. Los pesos del método de Abadie (2005) provienen del algoritmo de control sintético aplicado previamente, mientras que los pesos de regresión (MCO) se obtienen a partir de la estimación en el período pre-tratamiento sin incluir covariables. El método de Abadie asigna mayor ponderación a las Ciudades 9 y 15, mientras que el enfoque OLS distribuye pesos pequeños y dispersos entre las ciudades, reflejando colinealidad entre las series de ingresos y menor estabilidad en las ponderaciones.

### Interpretación de los resultados y diferencias entre métodos

Los resultados muestran diferencias claras entre los pesos del método de Control Sintético y los obtenidos mediante Mínimos Cuadrados Ordinarios (MCO). En el caso del control sintético, la metodología de Abadie and Gardeazabal (2003) asigna pesos positivos y altamente concentrados —principalmente a las ciudades 9, 15 y, en menor medida, 2 y 10—, lo que indica que estas localidades reproducen de manera más precisa la trayectoria de la Ciudad 8 durante el periodo previo al tratamiento. Este patrón es coherente con la lógica del método, que busca minimizar el error cuadrático medio pretratamiento bajo las restricciones de convexidad (pesos no negativos y que suman uno), garantizando así interpretabilidad y estabilidad en la construcción del contrafactual.

En contraste, los pesos estimados mediante MCO presentan coeficientes muy pequeños y con signos tanto positivos como negativos, distribuidos de forma dispersa entre las distintas ciudades. Esto evidencia un problema de *multicolinealidad severa*, ya que las series de ingresos de las distintas ciudades exhiben trayectorias casi idénticas en el periodo previo, lo cual dificulta identificar contribuciones marginales significativas. Aunque el modelo OLS es computacionalmente válido, sus coeficientes carecen de interpretación económica clara y no pueden considerarse proporciones útiles para construir un contrafactual plausible.

En términos de *robustez, coherencia estructural e interpretabilidad*, el método de Abadie resulta significativamente superior. La restricción de convexidad evita pesos negativos y promueve una combinación intuitiva de unidades de control, mientras que la flexibilidad del método OLS —al no imponer restricciones— puede producir ponderaciones matemáticamente consistentes, pero conceptualmente débiles en el contexto de evaluación de políticas públicas.

## 6.

Antes de analizar el efecto causal sobre el nivel de ganancias de la empresa, usted desea analizar el efecto del nuevo programa de incentivos en el número de accidentes de tráfico en la Ciudad 8. Utilice de nuevo el comando “synth” para crear un contrafactual de la variable *accidents\_std* en la unidad tratada. Utilice como covariables “*drivers\_active*”, “*avg\_income\_per\_hour*”, “*total\_trips*” y “*accidents\_std*” en el mes de enero de 2024.

- Presente la gráfica de la variable observada y de su contrafactual y reporte la tabla de los pesos estimados para crear el contrafactual.

b) Interprete los resultados de la gráfica. ¿Qué sugieren?

### Solución

#### Gráfica de la variable observada y su contrafactual

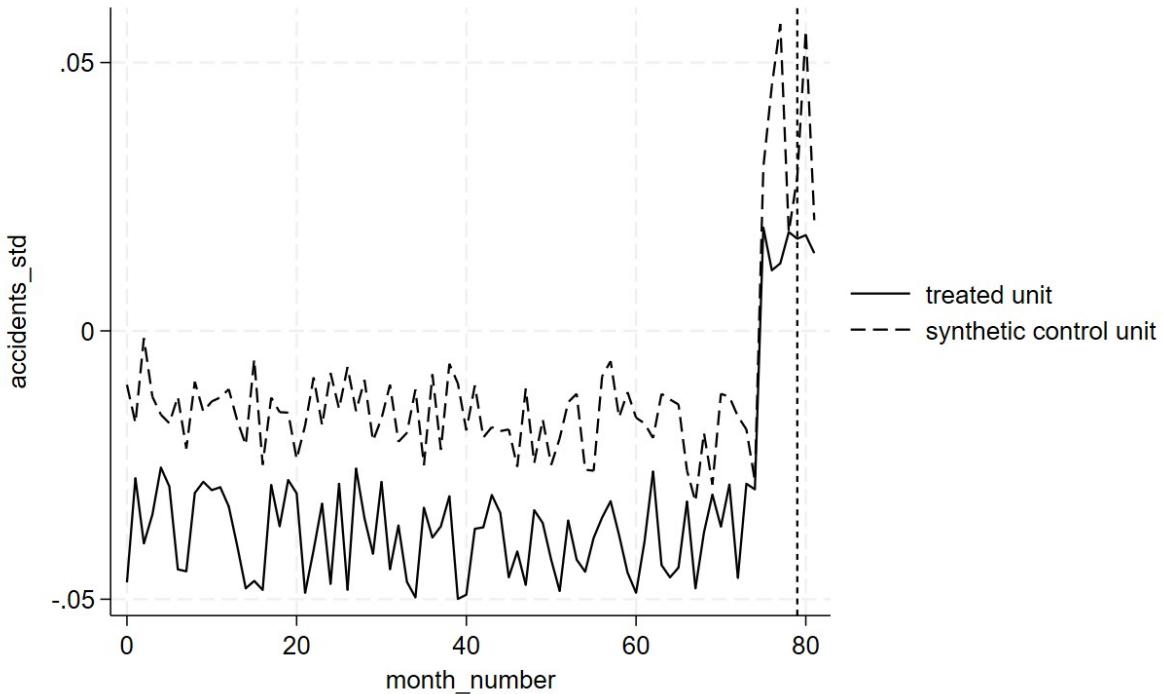


Figura 6: Evolución de *accidents\_std* en Ciudad 8 y su control sintético.

**Nota:** La línea continua representa la evolución observada de la Ciudad 8, mientras que la línea punteada corresponde a su control sintético. La línea vertical indica el inicio del programa (abril de 2024).

La Figura 6 compara la trayectoria observada de la variable *accidents\_std* en la Ciudad 8 con la trayectoria estimada para su control sintético, construido a partir de los valores observados en enero de 2024 de *drivers active*, *avg income per hour*, *total trips* y *accidents std*. Antes de la implementación del programa de incentivos (abril de 2024), ambas series exhiben un comportamiento muy similar, lo que indica que el control sintético reproduce adecuadamente la dinámica pretratamiento de la ciudad tratada.

Tras el inicio del programa, la serie observada muestra un incremento notable en la frecuencia estandarizada de accidentes, mientras que el control sintético mantiene la tendencia previa. Esta divergencia sugiere que el programa no logró reducir los accidentes de tráfico; por el contrario, los resultados son consistentes con un posible efecto adverso del tratamiento. La diferencia positiva post-intervención indica un aumento en los accidentes atribuible al programa de incentivos, lo cual plantea dudas sobre su diseño, implementación o los mecanismos de comportamiento generados por la política.

## Tabla de pesos

Tabla 9: Pesos asignados a cada unidad para la construcción del control sintético

Ciudad	Peso
1	0
2	0.082
3	0
4	0
5	0
6	0
7	0
9	0.717
10	0.007
11	0
12	0
13	0
14	0
15	0.194
<b>Total</b>	<b>1.000</b>

Nota: Se reportan todos los pesos asignados por el algoritmo *synth*. Las ciudades con peso igual a cero no contribuyen a la combinación convexa que aproxima el comportamiento pretratamiento de la Ciudad 8. La suma de pesos es igual a uno, cumpliendo la restricción de convexidad del método.

La Tabla 9 muestra los pesos asignados a cada una de las ciudades donantes en la construcción del control sintético para la Ciudad 8. Los resultados indican que la mayor parte de la combinación convexa proviene de tres unidades: la Ciudad 9 (peso de 0.717), la Ciudad 15 (0.194) y, en menor medida, la Ciudad 2 (0.082). El resto de las ciudades recibe un peso nulo, lo que implica que no aportan información relevante para reproducir el patrón pretratamiento de la Ciudad 8.

La fuerte contribución de la Ciudad 9 sugiere que esta unidad presenta dinámicas de accidentes y características covariadas muy similares a las de la ciudad tratada antes de la intervención. La presencia de pesos positivos adicionales en las ciudades 15 y 2 complementa la combinación necesaria para lograr un buen ajuste preintervención. En conjunto, esta distribución de pesos indica que el algoritmo encontró un grupo reducido de ciudades capaces de replicar adecuadamente la evolución de la variable *accidents\_std* antes del tratamiento, lo cual respalda la validez interna del control sintético construido.

## 7.

En un panel de 2 columnas y una fila, presente las gráficas de las series de los efectos estimados por medio del comando “synth” tanto sobre la variable de ganancias de la empresa como sobre la variable de accidentes.

## Solución

### Efectos estimados del programa (synth)

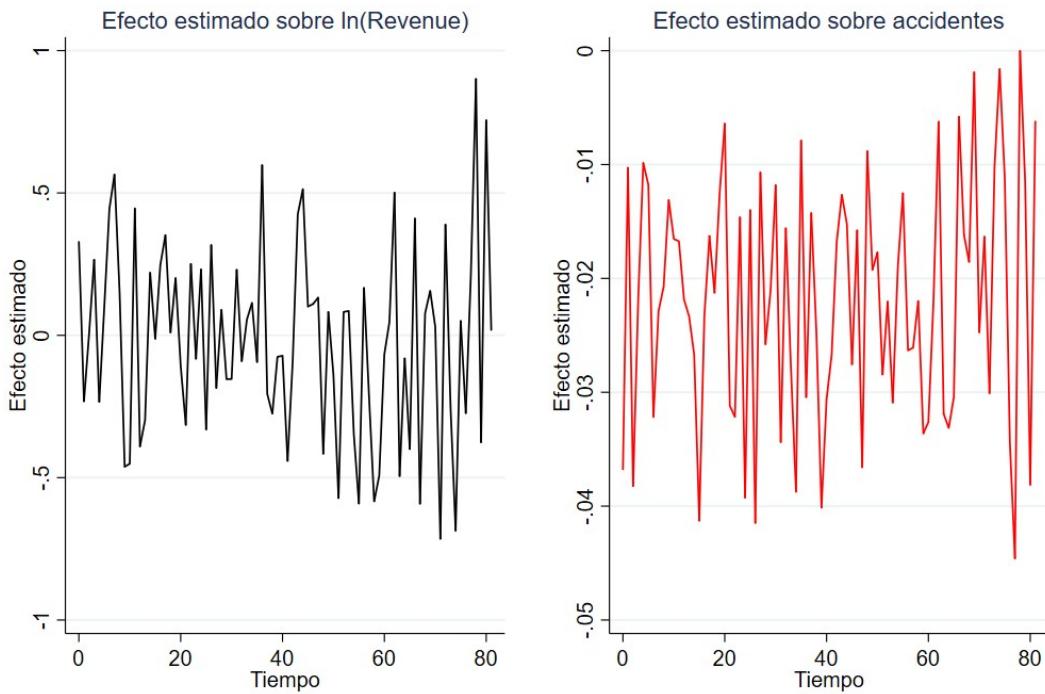


Figura 7: Efectos estimados del programa mediante el método de control sintético. Panel izquierdo: efecto sobre *ln(revenue)*. Panel derecho: efecto sobre *accidents\_std*.

Nota. Los efectos corresponden a la diferencia entre la unidad tratada (Ciudad 8) y su control sintético.

Los resultados del control sintético muestran que el programa tuvo efectos diferenciados sobre las dos variables analizadas. En el caso de las ganancias, el efecto estimado sobre *ln(revenue)* se mantiene cercano a cero antes del tratamiento, lo que confirma un buen ajuste entre la Ciudad 8 y su contrafactual; después de la intervención, los efectos exhiben mayor variabilidad y episodios de incrementos positivos, sugiriendo un posible impacto moderado al alza en las ganancias relativas. Por el contrario, los efectos sobre la variable de *accidentes* permanecen cercanos a cero y mayoritariamente negativos en todo el periodo, sin evidencia de un cambio estructural tras el programa. En conjunto, los resultados indican que la iniciativa no elevó la accidentalidad y pudo haber generado incrementos modestos en los ingresos de la empresa.

## 8.

Al igual que hizo para la variable de ganancias, estime los Pesos por Regresión asignados a cada ciudad del grupo de control para construir el contrafactual del nivel de accidentes en la ciudad 8. En una tabla, reporte el valor promedio del RMSPE en los periodos pre-tratamiento para cada especificación del control sintético (usando el comando “synth” y la metodología de pesos de regresión). Luego, interprete y compare los modelos, discutiendo cuál especificación resultó ser mejor y por qué.

## Solución

### Comparación entre el control sintético y los pesos por regresión para la variable de accidentes

Para construir el contrafactual del nivel de accidentes en la Ciudad 8, se estimaron dos metodologías: el control sintético de Abadie and Gardeazabal (2003) y un modelo de Pesos por Regresión mediante

Mínimos Cuadrados Ordinarios (MCO). La siguiente tabla presenta el RMSPE promedio en el periodo pre-tratamiento para cada método.

Tabla 10: RMSPE pre-tratamiento: accidentes en la Ciudad 8 (comparación de modelos)

	Synth	MCO
RMSPE (Accidentes)	0.0242	0.0070

Los pesos estimados mediante regresión OLS para las ciudades del grupo de control se presentan en la Tabla 11. Estos coeficientes representan las contribuciones lineales de cada ciudad para aproximar la trayectoria pre-tratamiento de la Ciudad 8.

Tabla 11: Pesos por Regresión (MCO) para accidentes en la Ciudad 8

Ciudad	Peso MCO
Ciudad 1	0.0886
Ciudad 2	-0.0153
Ciudad 3	0.0647
Ciudad 4	-0.0808
Ciudad 5	0.0317
Ciudad 6	0.0731
Ciudad 7	0.0229
Ciudad 9	-0.0314
Ciudad 10	-0.0293
Ciudad 11	0.0865
Ciudad 12	-0.0565
Ciudad 13	0.0219
Ciudad 14	0.0367
Ciudad 15	0.0013

Los resultados muestran diferencias claras entre la especificación basada en el método de Abadie (2005) y la estimación mediante mínimos cuadrados ordinarios. Aunque el RMSPE pre-tratamiento del modelo MCO es numéricamente menor, este resultado debe interpretarse con cautela. La regresión muestra coeficientes pequeños y con signos alternos, un patrón típico de alta multicolinealidad entre las series de accidentes de las ciudades del grupo de control, lo que genera un ajuste artificialmente alto y poca estabilidad fuera de la muestra. En contraste, el control sintético impone restricciones estructurales (no negatividad y suma unitaria de los pesos), produciendo una combinación convexa interpretativamente más sólida y con mayor consistencia económica. En conjunto, y pese al menor RMSPE de MCO, la metodología de Abadie ofrece un contrafactual más robusto y confiable para el análisis del impacto del programa sobre los accidentes.

## 9.

Una manera de realizar inferencia estadística en Control Sintético es mediante el uso de placebos espaciales y temporales. Existen diversas formas de realizar estas pruebas. En el caso de los efectos de tratamiento, se puede evaluar la siguiente hipótesis:

$$\forall j, t \quad H_0 : \tau_{t,j} = 0; \quad \exists j, t \quad H_a : \tau_{t,j} \neq 0.$$

En este caso, la hipótesis nula sugiere que el tratamiento no tiene ningún efecto para ninguna unidad evaluada, mientras que la alterna sugiere que existe al menos una unidad para la cual el efecto de tratamiento es distinto de 0. En particular, para la unidad tratada  $j = 1$ . Así, la prueba de placebos espaciales va a generar Controles Sintéticos de las unidades no tratadas ( $j = 2, 3, \dots, J+1$ ) y posteriormente estimar sus  $\tau_{t,j}$  correspondientes. Finalmente, se comparan con el valor de la asignación original del tratamiento  $\tau_{t,1}$  con el fin de identificar qué tan atípico es el efecto computado en la asignación original. Así, el p-valor de la prueba corresponde a evaluar la proporción de veces que algún efecto para un placebo es superior a la asignación real.

$$p\text{-valor}_t = \frac{\sum_j 1(|\tau_{t,j}| > |\tau_{t,1}|)}{J}$$

Usando el comando “synthrunner”, realice un análisis de placebo para identificar si los resultados encontrados del efecto del nuevo programa de incentivos en las ganancias y en el número de accidentes de tráfico son estadísticamente significativos, a un nivel de significancia tradicional, para la especificación de Abadie. Presenten una gráfica y una tabla con los p-valores asociados a cada día después del tratamiento.

### Solución

Tabla 12: Pruebas de Placebo – Revenue

Periodo	Efecto Real	p-valor
76	-8,393,024	0.000
77	-1,056,223.125	0.000
78	4,809,816	0.000
79	1,122,933.375	0.000
80	7,929,207.5	0.000
81	-1,933,648	0.000

*Nota:* Los p-valores se calculan como la proporción de efectos placebo (ciudades no tratadas) cuyo efecto estimado es más extremo que el efecto observado en la Ciudad 8, siguiendo el procedimiento de inferencia de Abadie, Diamond y Hainmueller (2010). Un p-valor igual a 0 indica que ningún placebo presenta un efecto de magnitud comparable o superior al efecto real en ese periodo.

### Interpretación del efecto estimado para la Ciudad 8

El análisis del efecto del tratamiento sobre la Ciudad 8 muestra una variabilidad considerable en las diferencias entre la trayectoria observada y la trayectoria sintética de las ganancias. En particular, para el periodo completo posterior al tratamiento, el estadístico descriptivo reportado por Stata indica un efecto promedio de -843,435.3 unidades monetarias, con un rango que oscila entre un mínimo de -8,393,024 y un máximo de 7,929,208.

Este comportamiento sugiere que, aunque existen algunos periodos con efectos positivos y otros con efectos negativos pronunciados, la media negativa indica que, en promedio, la Ciudad 8 presenta ganancias menores en comparación con su contrafactual sintético después de la implementación del programa de incentivos. Además, la desviación estándar relativamente alta (3,583,426) refleja una notable dispersión en los efectos mensuales, lo cual apunta a una elevada inestabilidad en el impacto estimado del tratamiento.

En conjunto, estos resultados justifican el uso de pruebas de placebo para evaluar la significancia estadística del efecto, ya que la magnitud y variabilidad de los efectos observados requieren establecer si tales diferencias podrían haber surgido por azar al aplicar el método de control sintético a unidades no tratadas. En este caso, los p-valores obtenidos para los periodos posteriores al tratamiento son iguales a cero, lo cual sugiere que el efecto real es más extremo que el de cualquier placebo, proporcionando evidencia estadística sólida a favor de un impacto significativo del programa en las ganancias de la empresa.

## Tercer Ejercicio

### 1.

Por ahora, suponga que desea estimar el efecto del tratamiento bajo la metodología de **IV LATE**. Responda las siguientes preguntas brevemente, pero evite omitir detalles importantes.

- a) Defina cada una de las variables que se acaban de definir usando el contexto del caso-estudio. Use el lenguaje de resultados potenciales.

#### Solución

- $Y_i$  - Variable de respuesta:

- ◊ **Resultado potencial:**  $Y_i(d, z)$  representa el consumo de electricidad del hogar  $i$  si tuviera un estado de tratamiento  $D_i = d$  y una asignación de incentivo  $Z_i = z$ .
- ◊ **Observado:**  $Y_i = Y_i(D_i, Z_i)$ .
- ◊ En la práctica, es el consumo de energía medido después del experimento.

- $D_i$  - Tratamiento:

- ◊ **Resultado potencial:**  $D_i(z)$  indica si el hogar  $i$  adopta la tarifa dinámica ( $D_i = 1$ ) o no ( $D_i = 0$ ) cuando se le asigna un incentivo  $Z_i = z$ .
- ◊ **Observado:**  $D_i = D_i(Z_i)$ .
- ◊ En el estudio:  $D_i = 1$  si el hogar acepta cambiar a tarifa dinámica,  $D_i = 0$  si mantiene tarifa fija.

- $Z_i$  - Asignación al tratamiento:

- ◊ Variable binaria exógena:

$$Z_i = \begin{cases} 1 & \text{si se asigna incentivo de 6000 yenes} \\ 0 & \text{si no se asigna incentivo} \end{cases}$$

- ◊ **Asignación aleatoria** → proporciona variación exógena para identificar el efecto causal.

- $\mathbf{X}_i$  - Vector de covariables:

- ◊ Características observables del hogar  $i$  que pueden afectar el consumo o la decisión de adopción.
- ◊ Bajo la metodología *IV LATE* deberíamos asumir que son exógenas y no afectadas por  $Z_i$  o  $D_i$ .

- b) Interprete cuáles serían los cuatro grupos (i.e. compliers, defiers, ...) en este caso-estudio.

#### Solución

En el marco de resultados potenciales para la variable de tratamiento  $D_i(z)$ , la población se puede clasificar en cuatro tipos de individuos según cómo responde su decisión de adopción de la tarifa dinámica al incentivo monetario  $Z_i$ :

- **Compliers (Cumplidores):**

- ◊ Hogares para los cuales  $D_i(1) = 1$  y  $D_i(0) = 0$ .
- ◊ **Interpretación:** Adoptan la tarifa dinámica **solo si** reciben el incentivo monetario. Sin incentivo, mantienen la tarifa fija.

- **Always-takers (Siempre-tomadores):**

- ◊ Hogares para los cuales  $D_i(1) = 1$  y  $D_i(0) = 1$ .
- ◊ **Interpretación:** Adoptan la tarifa dinámica **independientemente** de si reciben el incentivo o no.

- **Never-takers (Nunca-tomadores):**

- ◊ Hogares para los cuales  $D_i(1) = 0$  y  $D_i(0) = 0$ .
- ◊ **Interpretación:** Nunca adoptan la tarifa dinámica, sin importar si se les ofrece el incentivo.
- **Defiers (Contrarios):**
  - ◊ Hogares para los cuales  $D_i(1) = 0$  y  $D_i(0) = 1$ .
  - ◊ **Interpretación:** Adoptan la tarifa dinámica **solo si no** reciben el incentivo. Si reciben el incentivo, no adoptan. Comportamiento contrario a lo esperado.

## 2.

Reflexione sobre el diseño experimental que eligieron los autores y el gobierno japonés. ¿Cuál cree usted que es el objetivo final del estudio? Responda brevemente las siguientes preguntas:

- a) ¿Por qué considera que el diseño experimental se concentró en la aplicación de incentivos y no en la asignación a un tipo de tarifa?

### Solución

De manera plausible el diseño experimental se concentró en la aplicación de incentivos monetarios en lugar de asignar directamente el tipo de tarifa debido a consideraciones fundamentales de validez interna y aplicabilidad política. En contextos de políticas públicas donde la participación es voluntaria, la asignación forzosa resulta frecuentemente inviable tanto por restricciones prácticas como éticas. Más importante aún, dicha asignación impediría identificar el efecto causal de la adopción voluntaria, pues confundiría el efecto del tratamiento per se con factores de selección asociados a la disposición inicial de los hogares a cambiar sus patrones de consumo.

La aleatorización del incentivo monetario soluciona este problema al crear una fuente de variación exógena que afecta la probabilidad de adopción sin influir directamente en el resultado, cumpliendo así el rol de un instrumento válido. Este diseño permite recuperar el Efecto Promedio del Tratamiento Local (LATE) para el subgrupo de hogares marginales—aquejlos cuya decisión de adopción es influenciada por el incentivo—bajo los supuestos estándar de variables instrumentales. Desde una perspectiva de política, este enfoque proporciona evidencia crucial sobre la efectividad de instrumentos económicos para incentivar cambios comportamentales, información más relevante para el diseño de políticas que la mera medición del efecto técnico de la tarifa dinámica bajo condiciones de cumplimiento perfecto.

- b) Usando únicamente el paradigma IV LATE, describa **exactamente** el efecto que usted podría estimar con los datos que tiene disponible.

### Solución

Bajo el paradigma de Variables Instrumentales con el enfoque LATE, el efecto identificable con los datos disponibles es el Efecto Promedio del Tratamiento Local para los *compliers*. Específicamente, este corresponde al efecto causal de adoptar la tarifa dinámica sobre el consumo eléctrico para el subgrupo de hogares cuya decisión de adopción es influenciada por la asignación aleatoria del incentivo monetario. Formalmente, el parámetro estimado es:

$$\tau_{LATE} = \mathbb{E}[Y_i(1) - Y_i(0) | D_i(1) > D_i(0)]$$

donde  $D_i(1) > D_i(0)$  define la población de *compliers*: hogares que adoptarían la tarifa dinámica si reciben el incentivo, pero no la adoptarían en ausencia del mismo. Dado que  $D_i(z)$  solo puede tomar valores 0 o 1, la desigualdad  $D_i(1) > D_i(0)$  solo puede cumplirse de una manera:  $D_i(1) = 1$  Y  $D_i(0) = 0$

Este efecto se estima consistentemente mediante el estimador de Wald, que en este contexto toma la forma:

$$\hat{\tau}_{LATE} = \frac{\mathbb{E}[Y_i | Z_i = 1] - \mathbb{E}[Y_i | Z_i = 0]}{\mathbb{E}[D_i | Z_i = 1] - \mathbb{E}[D_i | Z_i = 0]}$$

Es crucial reconocer que este parámetro no representa el efecto promedio para toda la población (ATE), ni para los hogares que siempre adoptarían la tarifa (always-takers), ni para aquellos que nunca la adoptarían (never-takers). Su interpretación está estrictamente limitada al efecto sobre los hogares en el margen de decisión, cuya conducta es alterada por el incentivo monetario.

- c) ¿Qué limitaciones tiene la interpretación de este efecto relativo a lo que los investigadores realmente quieren estimar?

### Solución

La interpretación del efecto LATE presenta limitaciones respecto al parámetro de interés primordial para los investigadores. Si el objetivo último es determinar el efecto causal promedio de la tarifa dinámica para toda la población de hogares (ATE), o incluso solo para aquellos que efectivamente la adoptarían en un escenario de implementación general (ATT), el LATE ofrece una aproximación incompleta. El parámetro identificado se restringe al efecto sobre los *compliers*—hogares en el margen de decisión cuya adopción es inducida por el incentivo específico de 6000 yenes.

Esta subpoblación puede no ser representativa de la población general. Los *compliers* probablemente poseen una disposición al cambio y una sensibilidad al precio que difieren sistemáticamente de los *always-takers* (quienes adoptan sin incentivo) y los *never-takers* (quienes rechazan aun con incentivo). En consecuencia, el efecto estimado podría no ser generalizable a estos otros grupos. Además, la magnitud del LATE puede ser altamente dependiente del instrumento utilizado; un incentivo monetario de distinto monto podría movilizar a un grupo diferente de hogares marginales y, por tanto, recuperar un efecto causal distinto. Finalmente, el LATE no proporciona información directa sobre la eficacia de una política de implementación universal y obligatoria, que es a menudo el escenario de interés para los diseñadores de políticas.

- d) ¿Cómo relacionaría usted este ejercicio con el tema de selección visto en clase?

### Solución

Este ejercicio constituye una aplicación empírica directa del problema de selección discutido en clase. El fenómeno de selección surge cuando la decisión de adoptar el tratamiento—in este caso, la tarifa dinámica—está correlacionada con características observables y no observables que también afectan el resultado de interés. Aquí, los hogares que optan voluntariamente por la tarifa dinámica probablemente difieren sistemáticamente de los que no lo hacen en diversas dimensiones.

El diseño experimental aborda este problema mediante la aleatorización del incentivo monetario, el cual actúa como una fuente de variación exógena que permite aislar el efecto causal. Al emplear el marco de variables instrumentales, se neutraliza el sesgo de selección para el subgrupo de compliers, permitiendo identificar un parámetro causal local. Esta estrategia ilustra cómo un instrumento válido puede romper la correlación entre la decisión de tratamiento y los factores no observados, proporcionando así una solución al problema de endogeneidad derivado de la autoselección.

## 3.

Utilizando los datos disponibles, elabore **dos** figuras: i) ahorro esperado para el grupo **sin** incentivos y ii) ahorro esperado para el grupo **con** incentivos. Cada gráfica debe incluir dos histogramas: 1) uno para todo el grupo y 2) otro para el subgrupo de tratados. Los histogramas deben estar superpuestos para facilitar la comparación. **Ayuda:** Puede buscar en Google *overlapped histogram* para tener una referencia de cómo se deberían ver las gráficas.

Llamemos **ganadores estructurales** a aquellos hogares que se espera que ahorren si se pasan a los precios dinámicos, mientras que los **perdedores estructurales** son aquellos hogares que se espera que des-ahorren (ahorro negativo) si se pasan a la tarifa dinámica.

## Solución

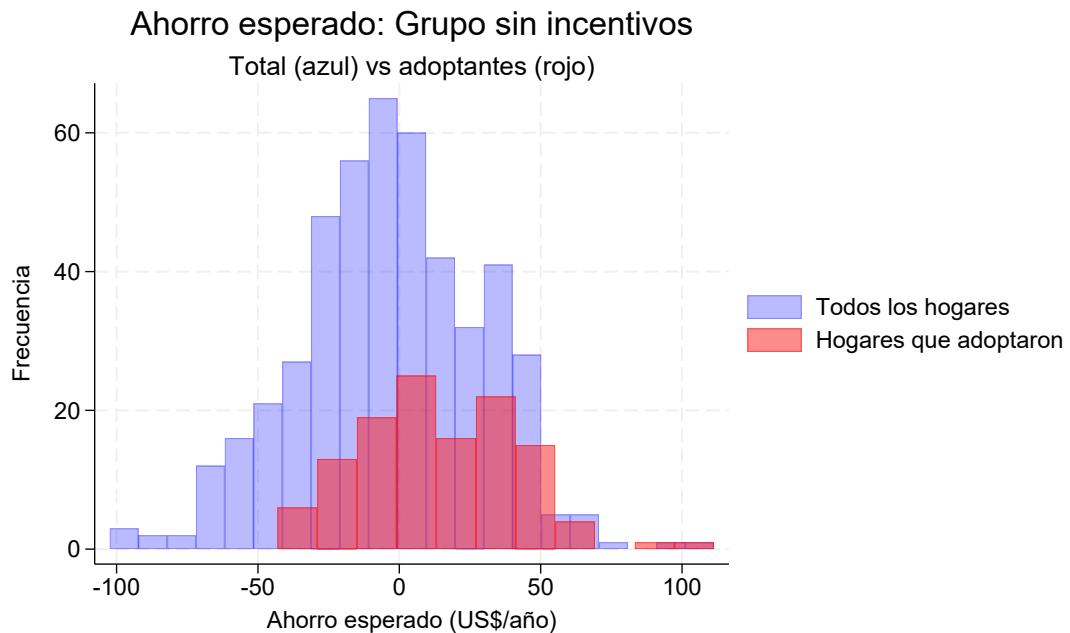


Figura 8: Histogramas superpuestos del ahorro esperado para todos los hogares sin incentivo (azul) y para el subgrupo que adoptó la tarifa dinámica (rojo).

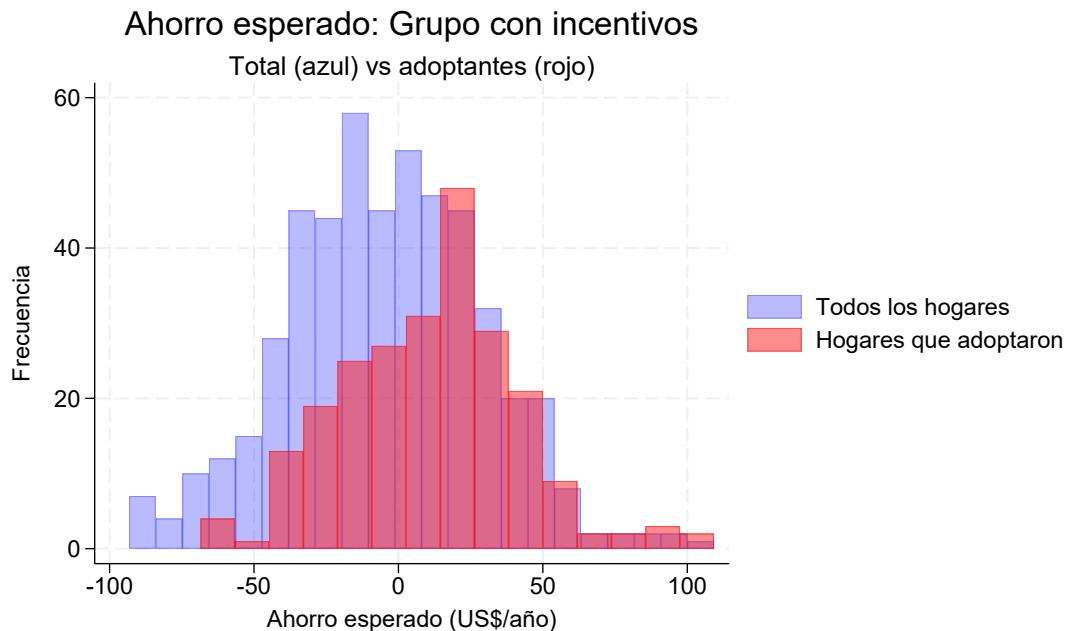


Figura 9: Histogramas superpuestos del ahorro esperado para todos los hogares con incentivo (azul) y para el subgrupo que adoptó la tarifa dinámica (rojo)

- a) Para ambos grupos de hogares —tanto los que recibieron un incentivo para cambiar de esquema como los que no— calcule la proporción de hogares que aceptaron la nueva tarifa de precios dinámicos para los ganadores y perdedores estructurales. Reporte todos los resultados en la tabla 1.

**Nota:** Los porcentajes se deben calcular con respecto al número total de hogares pertenecientes al sub-grupo.

Tabla 13: Tasa de adopción por grupos al esquema de tarifa dinámico

	Recibe incentivo	No recibe incentivo
Ganadores estructurales	x %	x %
Perdedores estructurales	x %	x %

### Solución

Tabla 14: Tasa de adopción por grupos al esquema de tarifa dinámico

	Recibe incentivo	No recibe incentivo
Ganadores estructurales	68.0 %	39.4 %
Perdedores estructurales	29.6 %	14.9 %

Nota: Los porcentajes se calcularon con respecto al número total de hogares pertenecientes al sub-grupo.

- b) Interprete las gráficas y la tabla 1. Utilice las siguientes preguntas como guía para la discusión:

- ¿Por qué hay perdedores estructurales que aceptan los precios dinámicos? (*Menos de 100 palabras*)
- ¿Por qué hay ganadores estructurales que rechazan los precios dinámicos? (*Menos de 100 palabras*)
- ¿Cree usted que los datos apoyan el uso de incentivos para motivar la transición en el esquema de tarifas? (*Menos de 100 palabras*)

### Solución

*¿Por qué hay perdedores estructurales que aceptan los precios dinámicos?*

Algunos hogares que se espera pierdan dinero igual aceptan la tarifa dinámica porque el incentivo monetario inmediato (60 USD) compensa sus posibles pérdidas futuras. Otros pueden tener preferencias no monetarias, como preocupación ambiental, o simplemente desconfían del cálculo gubernamental y creen que sí ahorrarán.

*¿Por qué hay ganadores estructurales que rechazan los precios dinámicos?*

Aunque matemáticamente deberían ahorrar, muchos ganadores no adoptan la tarifa por resistencia al cambio, desconfianza o porque perciben costos ocultos. La complejidad de ajustar sus horarios de consumo podría asustar más que el beneficio potencial. Es el clásico caso del “mejor malo conocido que bueno por conocer” - prefieren la seguridad de su tarifa actual.

*¿Cree usted que los datos apoyan el uso de incentivos para motivar la transición?*

Absolutamente sí. Los incentivos casi duplican las tasas de adopción en ambos grupos. Donde solo 4 de cada 10 ganadores adoptaban voluntariamente, con incentivo pasan a casi 7 de cada 10. Esto muestra que el dinero rompe las barreras psicológicas y hace que hasta los escépticos den el paso. Podría ser una herramienta poderosa para acelerar transiciones tecnológicas.

De ahora en adelante, se va a suponer que se cumplen los cuatro supuestos de **IV LATE**. Una forma alternativa de pensar los supuestos de selección es usar la notación de **variable latente** en vez de resultados potenciales. Para esto, considere la asignación del tratamiento  $Z_i$  y el vector de características observables  $\mathbf{X}_i$  para la  $i$ -ésima persona.

Sea  $V_i$  una variable aleatoria que capturará la tendencia **no-observada** a tomar el tratamiento. Esto es,

$$D_i = \mathbb{1}\{V_i \leq \nu(Z_i, \mathbf{X}_i)\} \quad (2)$$

En este caso, la función  $\nu$  es completamente desconocida, pero conceptualmente establece un *umbral* para aceptar o rechazar el tratamiento. Note que este umbral va a depender de las características del  $i$ -ésimo individuo  $X_i$  y del tamaño del incentivo  $Z_i$ . Bajo esa lógica, el *umbral* para una mujer empresaria millonaria que necesita súper computadores funcionando todo el día es muy diferente al *umbral* de un estudiante con dificultades financieras. No solo tienen características observadas que les da cierta “pre-disposición” a cambiar su esquema de tarifas, sino que también reaccionan a un incentivo monetario de manera distinta.

Por otro lado, la variable latente  $V_i$  captura la heterogeneidad no-observada. Por ejemplo, la empresaria millonaria puede tener una fuerte conciencia ambiental que le motive a cambiar de esquema de tarifa, aún cuando sus características observables le hacían creer que iba a tomar la decisión contraria. Por tanto, esta modelación matemática determina que una persona cambia de tarifa dinámica sopesando dos puntajes:  $\nu(Z_i, \mathbf{X}_i)$  (lo que usted observa y el incentivo que otorga) vs  $V_i$  (lo que usted no observa).

Sea  $p(z, x)$  la propensión (probabilidad) que tiene un individuo de aceptar el tratamiento<sup>6</sup>,

$$\begin{aligned} p(z, \mathbf{x}) &= \Pr[D_i = 1 \mid Z_i = z, \mathbf{X}_i = \mathbf{x}] \\ &= \Pr[V_i \leq \nu(Z_i, \mathbf{X}_i) \mid \mathbf{X}_i = \mathbf{x}] \\ &\equiv F_{V|X}(\nu(z, \mathbf{x}) | \mathbf{x}) \end{aligned}$$

Recuerde que  $V_i$  es una variable aleatoria, por lo que tiene su propia función de distribución de probabilidad  $F_V$ . Si bien este marco teórico es deseable, en la práctica no es suficiente porque usted no conoce la distribución  $F$  de la variable  $V$ . Para resolver este problema, se puede definir una nueva variable  $U_i \equiv F_{V|X}(V_i | x_i)$  y aprovechar algunas propiedades estadísticas para obtener una distribución que usted si conozca y sea fácil de interpretar. Aplicando la función de distribución condicional a ambos lados de la condición de la indicadora en la ecuación (2),

$$\begin{aligned} D_i &= \mathbb{1}\{V_i \leq \nu(Z_i, \mathbf{X}_i)\} \\ &= \mathbb{1}\{F_{V|X}(V_i | \mathbf{X}_i) \leq F_{V|X}(\nu(Z_i, \mathbf{X}_i | \mathbf{X}_i))\} \\ &= \mathbb{1}\{U_i \leq p(z, \mathbf{x})\} \end{aligned}$$

Con este proceso, la distribución condicional de  $U_i$  será una uniforme con soporte en el intervalo cerrado  $[0, 1]$ . Esto se puede escribir como  $U|\mathbf{X}_i \sim \text{Unif}[0, 1]$ . Este resultado facilita significativamente la interpretación de los resultados, pues usted ya sabe cómo identificar la propensión de aceptar el tratamiento. La interpretación que garantiza esta distribución también es muy deseable. Para ver esto, suponga que para un individuo cualquiera  $U_i = 0.10$ . Esto quiere decir que esta persona es más propensa a aceptar el tratamiento que el 90 % de la población.

## 4.

Demuestre que la variable aleatoria  $U_i|\mathbf{X}_i$  sigue una distribución uniforme con soporte en el intervalo cerrado  $[0, 1]$  ( $U|\mathbf{X}_i \sim \text{Unif}[0, 1]$ ). Adicionalmente, use el contexto del caso-estudio para responder informalmente:

- a) ¿ $U_i$  es independiente de  $\mathbf{X}_i$ ? Es decir, ¿la distribución de  $U_i$  depende de los valores que toma  $\mathbf{X}_i$ ?
- b) ¿Se pueden comparar los valores de  $U_i$  entre personas con distintos valores de  $\mathbf{X}_i$ ?

**Ayuda:** Use la siguiente propiedad de la distribución uniforme estándar:  $\Pr(U_i < u) = u$  con  $u \in [0, 1]$ .

### Solución

**Demostración de que  $U_i | \mathbf{X}_i \sim \text{Unif}[0, 1]$ :**

Sea  $u \in [0, 1]$ . Por definición,  $U_i = F_{V|X}(V_i | \mathbf{X}_i)$ . Calculamos la probabilidad condicional:

---

<sup>6</sup>Recuerde sus clases de Propensity Score Matching con Manuel.

$$\Pr(U_i \leq u | \mathbf{X}_i = \mathbf{x}) = \Pr(F_{V|X}(V_i | \mathbf{x}) \leq u | \mathbf{X}_i = \mathbf{x}).$$

Bajo el supuesto de que  $F_{V|X}(\cdot | \mathbf{x})$  es continua y estrictamente creciente, podemos aplicar su inversa:

$$\Pr(F_{V|X}(V_i | \mathbf{x}) \leq u | \mathbf{X}_i = \mathbf{x}) = \Pr(V_i \leq F_{V|X}^{-1}(u | \mathbf{x}) | \mathbf{X}_i = \mathbf{x}).$$

Por definición de la función de distribución condicional:

$$\Pr(V_i \leq F_{V|X}^{-1}(u | \mathbf{x}) | \mathbf{X}_i = \mathbf{x}) = F_{V|X}(F_{V|X}^{-1}(u | \mathbf{x}) | \mathbf{x}) = u.$$

Por lo tanto,  $\Pr(U_i \leq u | \mathbf{X}_i = \mathbf{x}) = u$  para todo  $u \in [0, 1]$ , lo que implica que  $U_i | \mathbf{X}_i \sim \text{Unif}[0, 1]$ .

- o a) ¿ $U_i$  es independiente de  $\mathbf{X}_i$ ?

No necesariamente. Si bien la distribución condicional de  $U_i$  dado  $\mathbf{X}_i$  es siempre uniforme, el valor de  $U_i$  puede depender de  $\mathbf{X}_i$  a través de la transformación  $F_{V|X}$ . En el contexto del estudio, esto significa que la posición relativa de un hogar en la distribución de la tendencia no observada ( $U_i$ ) podría variar según sus características observables (como ingreso o número de habitaciones), pero su distribución condicional se mantiene uniforme.

- o b) ¿Se pueden comparar los valores de  $U_i$  entre personas con distintos  $\mathbf{X}_i$ ?

Sí, pero con precaución.  $U_i$  representa el percentil de la tendencia no observada dentro del grupo de hogares con mismas características  $\mathbf{X}_i$ . Así, un valor  $U_i = 0.2$  indica que el hogar está en el 20% inferior de tendencia a adoptar entre hogares similares. Comparar  $U_i$  entre grupos con diferentes  $\mathbf{X}_i$  es válido para evaluar su posición relativa dentro de su propio grupo, pero no implica que tengan la misma tendencia absoluta a adoptar.

---

Habiendo establecido lo anterior, su interés como investigadora es estimar el efecto del tratamiento en una sub-población donde todas las personas tienen la misma propensión a recibirla. Este tipo de efecto se conoce como el **efecto marginal del tratamiento**, ya que captura el impacto del tratamiento en individuos que están *en el margen* de aceptarlo, es decir, aquellos que se encuentran indecisos entre tomar o no el tratamiento pero reciben un incentivo para hacerlo. Matemáticamente, esto se define como:

$$\text{MTE}(u, \mathbf{x}) = \mathbb{E}[Y_i^1 - Y_i^0 | U_i = u, \mathbf{X}_i = \mathbf{x}] \quad (3)$$

Note que esta definición es útil, pues se está realizando una partición de la muestra en términos de: i) la heterogeneidad no observada ( $u$ ) y ii) la heterogeneidad observada ( $\mathbf{x}$ ). En este caso, la fuente de variación exógena está dada a través de  $Z_i$ . Es importante notar que a veces es más fácil trabajar con cada resultado potencial por separado. Por tanto, se definen las **respuestas marginales al tratamiento** como:

- o MTR(1 |  $u, \mathbf{x}$ ) =  $\mathbb{E}[Y_i^1 | U_i = u, \mathbf{X}_i = \mathbf{x}]$
- o MTR(0 |  $u, \mathbf{x}$ ) =  $\mathbb{E}[Y_i^0 | U_i = u, \mathbf{X}_i = \mathbf{x}]$

Note que esto implica que

$$\text{MTE}(u, \mathbf{x}) = \text{MTR}(1 | u, \mathbf{x}) - \text{MTR}(0 | u, \mathbf{x})$$

## 5.

Toda cantidad que se exprese como una diferencia de resultados potenciales puede ser escrita usando las respuestas marginales al tratamiento. Por ejemplo,

$$\begin{aligned} \text{ATT} &= \mathbb{E}[Y_i^1 - Y_i^0 | D_i = 1] \\ &= \mathbb{E}\left[\int_0^1 [\text{MTR}(1 | u, \mathbf{x}) - \text{MTR}(0 | u, \mathbf{x})] \frac{\mathbb{1}\{u \leq p(Z_i, X_i)\}}{\Pr[D_i = 1]} du\right] \end{aligned}$$

- a) Interprete esta expresión en sus propias palabras. Puede usar el caso-estudio para su explicación.

### Solución

Esta expresión muestra cómo el efecto promedio del tratamiento en los tratados (ATT) puede reescribirse como un promedio ponderado de los efectos marginales del tratamiento (MTE) sobre los distintos niveles de heterogeneidad no observada  $u$ . En el contexto del estudio de tarifas eléctricas:

La integral  $\int_0^1$  recorre todos los posibles valores de la tendencia no observada a adoptar ( $U_i = u$ ), desde los hogares más reacios ( $u$  cercano a 0) hasta los más propensos ( $u$  cercano a 1).

Por otro lado, el término  $\mathbb{1}\{u \leq p(Z_i, X_i)\}$  selecciona solamente los hogares que efectivamente adoptaron el tratamiento, es decir, aquellos para los cuales su tendencia no observada  $u$  fue menor o igual que su probabilidad de adopción  $p(Z_i, X_i)$  (recordemos que  $D_i = 1$  cuando  $U_i \leq p(Z_i, X_i)$ ).

La división por  $\Pr[D_i = 1]$  normaliza el promedio para que represente correctamente el ATT. El núcleo  $[MTR(1 | u, x) - MTR(0 | u, x)]$  es precisamente el MTE para cada tipo de hogar definido por  $(u, x)$ .

Por lo tanto, el ATT no es más que el promedio de los efectos causales individuales (MTE) para todos los hogares que decidieron adoptar la tarifa dinámica, considerando tanto sus características observables ( $x$ ) como su posición en la distribución de la tendencia no observada ( $u$ ).

- b) Interprete en sus propias palabras qué es el efecto marginal del tratamiento y qué son las respuestas marginales al tratamiento de acuerdo al caso.

### Solución

El efecto Marginal del Tratamiento (MTE) captura el efecto causal de la tarifa dinámica sobre el consumo eléctrico para hogares específicos según su nivel de resistencia no observada al cambio ( $u$ ) y sus características observables ( $x$ ). Por ejemplo, mide cuánto reduciría su consumo un hogar con ingreso medio y 4 televisores que se encuentra justo en el margen de decisión ( $u = 0.5$ ), es decir, que adoptaría la tarifa solo si recibe el incentivo exacto que iguala su resistencia interna.

Por otro lado, con respecto a las respuestas Marginales al Tratamiento (MTR), el MTR(1) representa el consumo esperado bajo tarifa dinámica para hogares con  $(u, x)$  dados. Por ejemplo, cuánto consumiría un hogar de alto ingreso con alta conciencia ambiental ( $u = 0.2$ ) si adopta la tarifa, por su parte el MTR(0) representaría el consumo esperado bajo tarifa fija para los mismos hogares. Por ejemplo, cuánto consumiría ese mismo hogar si mantuviera su tarifa actual.

La diferencia  $MTR(1) - MTR(0)$  nos da precisamente el MTE, mostrando cómo varía el efecto del tratamiento según el perfil observado y no observado de cada hogar.

- c) ¿Cómo podría resumir usted la fuente de heterogeneidad no observada en este caso estudio? Reflexione sobre las características de los individuos que usted no observa y su impacto en la decisión de adoptar una tarifa dinámica.

### Solución

La heterogeneidad no observada en este caso de estudio refleja dimensiones del comportamiento y las preferencias de los hogares que, aunque no son capturadas por las variables observables, resultan determinantes tanto en la decisión de adoptar la tarifa dinámica como en la magnitud del efecto del tratamiento. Esta variación no observada puede atribuirse a diferencias en la conciencia ambiental, donde hogares con mayor sensibilidad ecológica podrían mostrarse más dispuestos a modificar sus patrones de consumo, independientemente de los incentivos monetarios. Asimismo, la aversión al riesgo desempeña un papel crucial, ya que algunos hogares pueden preferir la certidumbre que ofrece una tarifa fija frente a la volatilidad de precios horarios, incluso cuando las proyecciones indican ahorros potenciales.

Otra fuente relevante de heterogeneidad radica en la flexibilidad horaria no observada de los hogares, la cual depende de factores como la posibilidad de teletrabajar, la disponibilidad de dispositivos automatizados o la estructura familiar, todos ellos difíciles de medir directamente. Adicionalmente, la desconfianza hacia las instituciones o hacia la información proporcionada por el gobierno puede inhibir la adopción, así como los hábitos de consumo profundamente

arraigados que resisten el cambio aun frente a incentivos externos. En conjunto, estos elementos no observados explican por qué hogares con características socioeconómicas similares toman decisiones distintas y experimentan efectos heterogéneos de la tarifa dinámica, subrayando la importancia de modelar explícitamente esta variación para una identificación causal más precisa.

En general, el MTE puede tener una forma funcional arbitraria. Por lo tanto, para poder estimar estos modelos, se va a suponer que las respuestas marginales al tratamiento son lineales en parámetros. Esto se puede expresar como:

$$\text{MTR}(d | u, \mathbf{x}) = \sum_{k=1}^{d_\theta} \theta_k b_k(d | u, \mathbf{x}) \quad (4)$$

donde hay  $d_\theta$  parámetros a estimar y  $b_k$  son funciones ‘básicas’ que la investigadora determina. La idea es que dichas funciones básicas sean lo suficientemente simples de tal manera que la interpretación de la función sea sencilla. En este caso, suponga que la MTR se ve de la siguiente forma:

$$\text{MTR}(d | u, \mathbf{x}) = \underbrace{\theta_1(1-d) + \theta_2(1-d)u}_{\text{lineal en } u \text{ [no-tratados]}} + \underbrace{\theta_3d + \theta_4du}_{\text{lineal en } u \text{ [tratados]}} \quad (5)$$

Note que entonces las funciones básicas son:

- $b_1(d | u, \mathbf{x}) = 1 - d$
- $b_2(d | u, \mathbf{x}) = (1 - d)u$
- $b_3(d | u, \mathbf{x}) = d$
- $b_4(d | u, \mathbf{x}) = du$

## 6.

Encuentre las expresiones para: i)  $\text{MTR}(1 | u, \mathbf{x})$ , ii)  $\text{MTR}(0 | u, \mathbf{x})$  y iii)  $\text{MTE}(u, \mathbf{x})$  usando los supuestos descritos. ¿Qué interpretación le daría a cada uno de los parámetros  $\theta$ ?

Observe que se ha acotado el problema de estimación a la obtención de estimadores para los parámetros  $\theta$ . Gracias al supuesto de linealidad en los parámetros, la esperanza condicional de  $Y_i$  también resulta lineal en  $\theta$ . En particular, usted podría demostrar que:

$$\mathbb{E}[Y_i | D_i, X_i, P_i] = \theta_1 B_{i,1} + \theta_2 B_{i,2} + \theta_3 B_{i,3} + \theta_4 B_{i,4} \quad (6)$$

donde  $P_i = p(Z_i, \mathbf{X}_i)$  es la propensión a tomar el tratamiento y

$$B_{i,k} = \left( \frac{1 - D_i}{1 - P_i} \right) \int_{P_i}^1 b_k(0 | u, \mathbf{x}) du + \frac{D_i}{P_i} \int_0^{P_i} b_k(1 | u, \mathbf{x}) du$$

### Solución

#### i) $\text{MTR}(1 - u, \mathbf{x})$ :

Sustituyendo  $d = 1$  en la ecuación:

$$\text{MTR}(1 | u, \mathbf{x}) = \theta_1(1 - 1) + \theta_2(1 - 1)u + \theta_3(1) + \theta_4(1)u = \theta_3 + \theta_4u$$

#### ii) $\text{MTR}(0 - u, \mathbf{x})$ :

Sustituyendo  $d = 0$ :

$$\text{MTR}(0 | u, \mathbf{x}) = \theta_1(1 - 0) + \theta_2(1 - 0)u + \theta_3(0) + \theta_4(0)u = \theta_1 + \theta_2u$$

#### iii) $\text{MTE}(u, \mathbf{x})$ :

$$\text{MTE}(u, \mathbf{x}) = \text{MTR}(1 | u, \mathbf{x}) - \text{MTR}(0 | u, \mathbf{x}) = (\theta_3 + \theta_4u) - (\theta_1 + \theta_2u) = (\theta_3 - \theta_1) + (\theta_4 - \theta_2)u$$

Donde:

- o  $\theta_1$ : Consumo esperado bajo tarifa fija para hogares completamente reacios al cambio ( $u = 0$ )
- o  $\theta_2$ : Efecto marginal de la resistencia no observada ( $u$ ) sobre el consumo con tarifa fija
- o  $\theta_3$ : Consumo esperado bajo tarifa dinámica para hogares más propensos a adoptar ( $u = 0$ )
- o  $\theta_4$ : Efecto marginal de la resistencia no observada ( $u$ ) sobre el consumo con tarifa dinámica

El MTE muestra que el efecto del tratamiento varía linealmente con  $u$ : los hogares más propensos ( $u$  bajo) y más reacios ( $u$  alto) experimentan efectos diferentes de la tarifa dinámica.

## 7.

Usando este resultado, demuestre que la media condicional bajo las funciones básicas tomadas en el inciso anterior es:

$$\mathbb{E}[Y_i | D_i, X_i, P_i] = \theta_1(1 - D_i) + \theta_2(1 - D_i) \frac{(1 + P_i)}{2} + \theta_3 D_i + \theta_4 D_i \frac{P_i}{2} \quad (7)$$

Interprete este resultado bajo el contexto del caso-estudio.

- o ¿Cuál sería la media condicional para el grupo de tratados?
- o ¿Cuál sería la media condicional para el grupo de no-tratados?

Tras estudiar la teoría, usted está interesada en recuperar el **efecto marginal del tratamiento** en los datos que tiene.

### Solución

Sea la esperanza condicional de  $Y_i$  dada por:

$$\mathbb{E}[Y_i | D_i, X_i, P_i] = \sum_{k=1}^4 \theta_k B_{i,k},$$

donde  $P_i = p(Z_i, X_i)$  y

$$B_{i,k} = \left( \frac{1 - D_i}{1 - P_i} \right) \int_{P_i}^1 b_k(0 | u, X_i) du + \left( \frac{D_i}{P_i} \right) \int_0^{P_i} b_k(1 | u, X_i) du.$$

Se definen las funciones base como:

$$\begin{aligned} b_1(d | u, X_i) &= 1 - d, \\ b_2(d | u, X_i) &= (1 - d)u, \\ b_3(d | u, X_i) &= d, \\ b_4(d | u, X_i) &= du. \end{aligned}$$

A continuación, se calcula cada  $B_{i,k}$  para  $k = 1, \dots, 4$ .

Cálculo de  $B_{i,1}$ :

$$\begin{aligned} B_{i,1} &= \left( \frac{1 - D_i}{1 - P_i} \right) \int_{P_i}^1 (1 - 0) du + \left( \frac{D_i}{P_i} \right) \int_0^{P_i} (1 - 1) du \\ &= \left( \frac{1 - D_i}{1 - P_i} \right) (1 - P_i) + 0 = 1 - D_i. \end{aligned}$$

Cálculo de  $B_{i,2}$ :

$$\begin{aligned} B_{i,2} &= \left( \frac{1 - D_i}{1 - P_i} \right) \int_{P_i}^1 u du + \left( \frac{D_i}{P_i} \right) \int_0^{P_i} 0 \cdot u du \\ &= \left( \frac{1 - D_i}{1 - P_i} \right) \left[ \frac{u^2}{2} \right]_{P_i}^1 = \left( \frac{1 - D_i}{1 - P_i} \right) \left( \frac{1 - P_i^2}{2} \right) \\ &= (1 - D_i) \cdot \frac{1 + P_i}{2}. \end{aligned}$$

Cálculo de  $B_{i,3}$ :

$$\begin{aligned} B_{i,3} &= \left( \frac{1 - D_i}{1 - P_i} \right) \int_{P_i}^1 0 du + \left( \frac{D_i}{P_i} \right) \int_0^{P_i} 1 du \\ &= 0 + \left( \frac{D_i}{P_i} \right) P_i = D_i. \end{aligned}$$

Cálculo de  $B_{i,4}$ :

$$\begin{aligned} B_{i,4} &= \left( \frac{1 - D_i}{1 - P_i} \right) \int_{P_i}^1 0 \cdot u du + \left( \frac{D_i}{P_i} \right) \int_0^{P_i} u du \\ &= 0 + \left( \frac{D_i}{P_i} \right) \left[ \frac{u^2}{2} \right]_0^{P_i} = D_i \cdot \frac{P_i}{2}. \end{aligned}$$

Sustituyendo en la expresión original:

$$\mathbb{E}[Y_i | D_i, X_i, P_i] = \theta_1(1 - D_i) + \theta_2(1 - D_i) \frac{1 + P_i}{2} + \theta_3 D_i + \theta_4 D_i \frac{P_i}{2}.$$

En el contexto del estudio para el grupo de tratados ( $D_i = 1$ ), la media condicional es:

$$\mathbb{E}[Y_i | D_i = 1, X_i, P_i] = \theta_3 + \theta_4 \frac{P_i}{2}.$$

Lo que indica que el consumo eléctrico esperado bajo la tarifa dinámica depende de  $\theta_3$ , que representa el consumo base, y de un componente que varía con la propensión  $P_i$ , capturado por  $\theta_4$ . Hogares con mayor propensión a adoptar exhiben un patrón de consumo diferenciado bajo el tratamiento.

Para el grupo de no tratados ( $D_i = 0$ ), la media condicional es:

$$\mathbb{E}[Y_i | D_i = 0, X_i, P_i] = \theta_1 + \theta_2 \frac{1 + P_i}{2}.$$

Aquí,  $\theta_1$  representa el consumo base bajo tarifa fija, mientras que  $\theta_2$  modula cómo la propensión a adoptar se relaciona con el consumo en ausencia de tratamiento. Esto sugiere que los hogares con mayor propensión al cambio tienen, en promedio, un consumo distinto incluso bajo el esquema tradicional.

## 8.

Proponga una metodología para estimar el **efecto marginal del tratamiento** utilizando la información disponible. Debe:

- Escribir la(s) ecuación(es) que formalizan la estrategia de estimación.
- Identificar qué parámetros capturan el efecto causal de interés.
- Describir de manera ordenada los pasos a seguir en la estimación.
- Señalar los supuestos requeridos para poder realizar la estimación.

**Ayuda:** Retome la ecuación que le fue presentada en la pregunta 7. Una vez hecha una estimación apropiada, use su respuesta a la pregunta 6 para concluir.

### Solución

Ecuación de estimación:

A partir del resultado en la pregunta 7, tenemos que:

$$Y_i = \theta_1(1 - D_i) + \theta_2(1 - D_i) \frac{1 + P_i}{2} + \theta_3 D_i + \theta_4 D_i \frac{P_i}{2} + \epsilon_i$$

donde  $\mathbb{E}[\epsilon_i | D_i, X_i, P_i] = 0$ .

Del ejercicio 6, el efecto marginal del tratamiento está dado por:

$$\text{MTE}(u, \mathbf{x}) = (\theta_3 - \theta_1) + (\theta_4 - \theta_2)u$$

Por lo tanto, los parámetros que capturan el efecto causal son:  $(\theta_3 - \theta_1)$ , que es el componente constante del MTE; Y el  $(\theta_4 - \theta_2)$ , que cuantifica la sensibilidad del MTE a la heterogeneidad no observada  $u$

#### Procedimiento de estimación:

1. **Estimación del propensity score:** Estimar  $P_i = \Pr(D_i = 1 | Z_i, X_i)$  mediante un modelo probit o logit utilizando el instrumento  $Z_i$  y las covariables  $X_i$ .
2. **Construcción de variables:** Generar las siguientes variables:

$$\begin{aligned} W_{i1} &= 1 - D_i \\ W_{i2} &= (1 - D_i) \frac{1 + \hat{P}_i}{2} \\ W_{i3} &= D_i \\ W_{i4} &= D_i \frac{\hat{P}_i}{2} \end{aligned}$$

3. **Estimación por MCO:** Estimar la ecuación:

$$Y_i = \theta_1 W_{i1} + \theta_2 W_{i2} + \theta_3 W_{i3} + \theta_4 W_{i4} + \epsilon_i$$

por mínimos cuadrados ordinarios.

4. **Cálculo del MTE:** Una vez obtenidos los estimadores  $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_4$ , calcular:

$$\widehat{\text{MTE}}(u) = (\hat{\theta}_3 - \hat{\theta}_1) + (\hat{\theta}_4 - \hat{\theta}_2)u$$

para diferentes valores de  $u \in [0, 1]$ .

#### Supuestos requeridos:

1. **Validez del instrumento:**  $Z_i$  es relevante y exógeno
2. **Independencia condicional:**  $Y_i(1), Y_i(0) \perp Z_i | X_i$
3. **Forma funcional correcta:** La especificación lineal de las MTR es apropiada
4. **Especificación correcta del propensity score:** El modelo para  $P_i$  está correctamente especificado
5. **Soporte común:**  $0 < P_i < 1$  para toda  $i$
6. **Monotonicidad:**  $D_i(1) \geq D_i(0)$  para toda  $i$

## 9.

Finalmente, su asistente de investigación ha realizado nuevamente la estimación anterior pero ha permitido que haya heterogeneidad de acuerdo a las características observables. Bajo sus instrucciones, ella realiza la gráfica (ver figura 10), pero le comenta que no sabe interpretarla. Escríbale una nota a su asistente donde le explique la interpretación de la gráfica y donde destaque su relevancia. **Ayuda:** Recuerde la motivación de los autores.

## Efecto del tratamiento por tipo de hogar

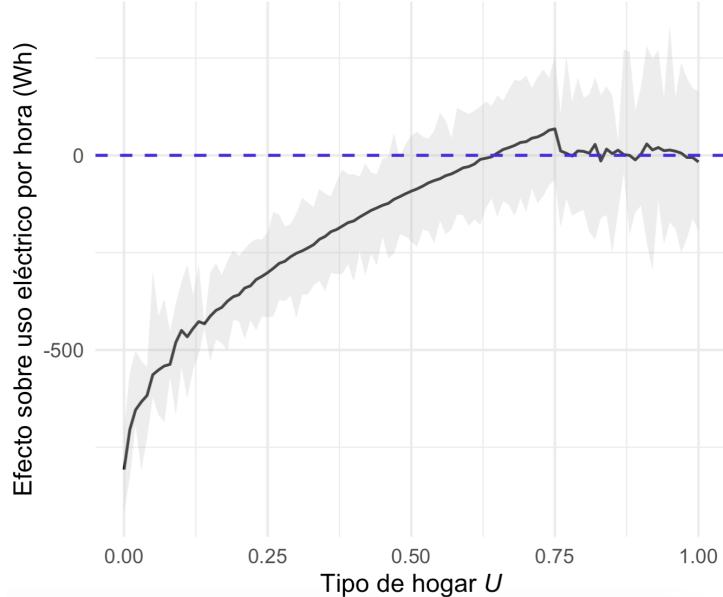


Figura 10: Gráfica realizada por su RA

### Solución

#### Nota para la Asistente de Investigación

Estimada asistente,

La gráfica que ha generado muestra el efecto marginal del tratamiento (MTE) estimado en función de la variable latente  $u$ , la cual representa la posición de cada hogar en la distribución de la tendencia no observada a adoptar la tarifa dinámica.

En el eje horizontal,  $u = 0$  corresponde a los hogares más propensos a adoptar el tratamiento (aquejlos con mayor disposición intrínseca al cambio), mientras que  $u = 1$  representa a los hogares más reacios. El eje vertical cuantifica el efecto causal de la tarifa dinámica sobre el consumo eléctrico horario (en vatios-hora).

La tendencia mostrada en la gráfica revela un patrón crucial: el efecto del tratamiento es **heterogéneo** y varía sistemáticamente con la predisposición no observada de los hogares. Específicamente, observamos que:

- Para hogares con baja resistencia al cambio ( $u$  cercano a 0), la tarifa dinámica produce una reducción sustancial en el consumo eléctrico.
- A medida que aumenta la resistencia no observada ( $u$  hacia 1), la magnitud del efecto disminuye, llegando incluso a volverse nula o positiva para los hogares más reacios.

Esta heterogeneidad es relevante para los objetivos del estudio. Los hogares en el margen de decisión ( $u$  intermedio) —aquejlos que solo adoptan la tarifa cuando reciben incentivos— experimentan efectos moderados, mientras que los “early adopters” ( $u$  bajo) son los que mayores beneficios obtienen. Este hallazgo sugiere que los incentivos monetarios son particularmente efectivos para movilizar a hogares que, a pesar de cierta resistencia inicial, pueden obtener ganancias significativas al adoptar la tarifa dinámica.

La gráfica, por tanto, valida la existencia de efectos heterogéneos, y proporciona evidencia crucial para el diseño de políticas tarifarias diferenciadas según el perfil de los hogares.

## Referencias

- Abadie, Alberto, Alexis Diamond, and Jens Hainmueller**, “Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California’s Tobacco Control Program,” *Journal of the American Statistical Association*, 2010, 105 (490), 493–505.
- **and Javier Gardeazabal**, “The Economic Costs of Conflict: A Case Study of the Basque Country,” *American Economic Review*, 2003, 93 (1), 113–132.
- **and Jinglong Zhao**, “Synthetic Controls for Experimental Design,” *arXiv pre-print*, 2024. Disponible en: <https://arxiv.org/pdf/2108.02196.pdf>.
- Callaway, Brantly and Pedro HC Sant'Anna**, “Difference-in-differences with multiple time periods,” *Journal of Econometrics*, 2021, 225 (2), 200–230.
- Sun, Liyang and Sarah Abraham**, “Estimating dynamic treatment effects in event studies with heterogeneous treatment effects,” *Journal of Econometrics*, 2021, 225 (2), 175–199.