

## Primer ejercicio

La idea del siguiente ejercicio será familiarizarnos con los estimadores de triple diferencias (DDD) y su fundamento teórico. Para ello, los guiaremos a través de una serie de pasos que nos permitirán entender mejor en qué contextos dichos estimadores pueden ser útiles, y comprender bajo qué supuestos nos permiten recuperar un efecto causal de interés. Inicialmente, supongan que contamos con un panel balanceado de  $i = 1, \dots, n$  individuos que son observados en dos periodos de tiempo  $t = -1, 0$ . En su base de datos, ustedes cuentan con un identificador de individuo y de periodo. Además, disponen de información sobre la variable de resultado de interés,  $Y_{i,t}$ . Un subconjunto conocido de los individuos de su muestra es sujeto a un mismo tratamiento en algún instante de tiempo entre el periodo  $-1$  y el periodo  $0$ .

1. Consideren el siguiente modelo de diferencias en diferencias:

$$Y_{i,t} = \beta_0 + \beta_1 G_i + \beta_2 T_t + \beta_3 G_i \cdot T_t + \epsilon_{i,t}, \quad (1)$$

donde  $G_i$  codifica si el individuo  $i$  hace parte del grupo de tratamiento,  $T_t$  codifica si el periodo es  $t = 0$ , y  $\epsilon_{i,t}$  es un término de error idiosincrático.

- a) ¿Bajo qué supuestos es posible recuperar un estimador consistente del  $ATT$ ? Exprésenlos formalmente y denles una interpretación intuitiva.

**Solución:** Para recuperar un estimador consistente del  $ATT$  se debe cumplir el supuesto de tendencias paralelas, el cual implica que en ausencia del tratamiento, el cambio esperado en la variable dependiente del grupo de tratamiento es igual al cambio observado en esa variable para el grupo de control.

Esto se puede ver expresado de la siguiente manera:

$$\mathbb{E}[Y_{i,0}(0) | G_i = 1] - \mathbb{E}[Y_{i,-1}(0) | G_i = 1] = \mathbb{E}[Y_{i,0}(0) | G_i = 0] - \mathbb{E}[Y_{i,-1}(0) | G_i = 0]$$

En términos intuitivos, esto quiere decir que el supuesto de tendencias paralelas implica que la brecha entre el grupo de tratamiento y de control no se ve alterada en ausencia de un tratamiento. Previamente al tratamiento, la brecha entre tratamiento y control debería mantenerse constante.

Asimismo, se debe cumplir el supuesto de no anticipación, el cual implica que los valores de la variable independiente observados antes del momento de tratamiento no dependen de si voy a ser o no tratado en un futuro. Esto se puede ver expresado formalmente a continuación:

$$Y_{i,t}(1) = Y_{i,t}(0) \text{ para todos los } i \text{ tales que } D_i = 1$$

- b) Dada su respuesta al inciso anterior, ¿Qué parámetro del modelo (1) captura el efecto causal de interés? Demuéstrenlo formalmente.

**Solución:**  $\beta_3$  A partir del modelo (1) se puede recuperar el parámetro que captura el efecto causal de interés:

$$\tau_{DiD} = \mathbb{E}[Y_{i,t} | T_t = 1, G_i = 1] - \mathbb{E}[Y_{i,t} | T_t = 0, G_i = 1]$$

$$-(\mathbb{E}[Y_{i,t} | T_t = 1, G_i = 0] - \mathbb{E}[Y_{i,t} | T_t = 0, G_i = 0])$$

$$= \mathbb{E}[\beta_0 + \beta_1 G_i + \beta_2 + \beta_3 G_i + \epsilon_{i,0} | G_i = 1] - \mathbb{E}[\beta_0 + \beta_1 G_i + \epsilon_{i,-1} | G_i = 1] \\ - (\mathbb{E}[\beta_0 + \beta_1 G_i + \beta_2 + \beta_3 G_i + \epsilon_{i,0} | G_i = 0] - \mathbb{E}[\beta_0 + \beta_1 G_i + \epsilon_{i,-1} | G_i = 0])$$

Teniendo en cuenta el supuesto de exogeneidad de los errores y el de tendencias paralelas, se tiene entonces que :

$$= \beta_0 + \beta_1 + \beta_2 + \beta_3 - \beta_0 - \beta_1 \\ - (\beta_0 + \beta_2 + -\beta_0)$$

$$= \beta_2 + \beta_3 - \beta_2$$

$$= \beta_3$$

$$\tau_{DiD} = \beta_3$$

De esta forma, el parámetro  $\beta_3$  captura el efecto causal de interés.

2. Una forma más general del modelo (1) es conocida como especificación *Two Way Fixed Effects* (TWFE):

$$Y_{i,t} = \alpha_i + \gamma_t + \delta D_{i,t} + \epsilon'_{i,t}, \quad (2)$$

donde  $\alpha_i$  son efectos fijos de unidad (i.e., variables indicadoras de cada unidad),  $\gamma_t$  son efectos fijos de tiempo, y  $D_{i,t}$  es la indicadora de si el individuo  $i$  se encuentra tratado en el periodo  $t$ .

a) Demuestren que bajo los mismos supuestos propuestos en el inciso anterior que  $\delta$  identifica el efecto causal de interés.

**Solución:**

Se parte de la expresión utilizada en el inciso anterior para encontrar el efecto causal, pero esta vez adaptada al modelo (2).

$$\tau_{DiD} = \mathbb{E}[Y_{i,t} | t = 0, D_i = 1] - \mathbb{E}[Y_{i,t} | t = -1, D_i = 1] \\ - (\mathbb{E}[Y_{i,t} | t = 0, D_i = 0] - \mathbb{E}[Y_{i,t} | t = -1, D_i = 0])$$

$$\tau_{DiD} = \mathbb{E}[\alpha_i + \gamma_t + \delta D_{i,t} + \epsilon'_{i,t} | t = 0, D_i = 1] - \mathbb{E}[\alpha_i + \gamma_t + \epsilon'_{i,t} | t = -1, D_i = 1] \\ - (\mathbb{E}[\alpha_i + \gamma_t + \epsilon'_{i,t} | t = 0, D_i = 0] - \mathbb{E}[\alpha_i + \gamma_t + \epsilon'_{i,t} | t = -1, D_i = 0])$$

Bajo el supuesto de exogeneidad de los errores, se tiene entonces que:

$$\tau_{DiD} = \alpha_1 + \gamma_0 + \delta - \alpha_1 - \gamma_{-1} \\ - (\alpha_0 + \gamma_0 - \alpha_0 - \gamma_{-1})$$

$$\tau_{DiD} = \gamma_0 - \gamma_{-1} + \delta \\ - (\gamma_0 - \gamma_{-1})$$

$$\tau_{DiD} = \delta$$

De esta forma, el parámetro  $\delta$  captura el efecto causal de interés.

3. ¿Es posible en este contexto dar evidencia que soporte la validez de los supuestos de identificación? Argumenten.

**Solución:** El cumplimiento del supuesto de tendencias paralelas se puede validar parcialmente si se tienen datos para varios periodos pre-tratamiento. Si bien no se puede probar de forma definitiva, existen varias maneras de examinar el supuesto formalmente (por ejemplo, la Causalidad de Granger).

Ahora bien, debido a que en este caso solo tenemos un periodo pre-tratamiento, no sería posible validar este supuesto de identificación. Esto pues no habría forma de verificar que antes del periodo de tratamiento, el grupo de control y el grupo de tratamiento tuvieran un comportamiento semejante que permita sugerir el cumplimiento del supuesto de tendencias paralelas. En otras palabras, no es posible ver que ambos grupos tuvieran una evolución similar antes del tratamiento.

Supongan que ahora ustedes logran ampliar su panel, de manera que cuentan información para los periodos  $t = -k_1, \dots, 0, \dots, k_2$ , con  $k_1, k_2 > 0$ . En este escenario, se proponen estimar el siguiente modelo:

$$Y_{i,t} = \alpha_i + \gamma_t + \sum_{j=2}^{k_1} \delta_{-j} D_{i,t+j} + \sum_{j=0}^{k_2} \delta_j D_{i,t-j} + \epsilon''_{i,t}, \quad (3)$$

donde  $D_{i,t}$  es la indicadora de si el individuo  $i$  se encuentra tratada en el periodo  $t$ . Definan  $\forall \eta \geq 0$

$$ATT(\eta) = \mathbb{E}[Y_{i,\eta}(1) - Y_{i,\eta}(0) | G_i = 1]$$

Intuitivamente, este es el efecto del tratamiento sobre los tratados en el periodo  $\eta$ .

4. Bajo los supuestos de identificación previamente enunciados:

a) ¿Qué parámetro, si alguno, captura el  $ATT(\eta)$ ? Demuéstrenlo formalmente.

**Solución:**

Suponiendo el cumplimiento de los supuestos de identificación previamente anunciados, se tiene entonces que se cumple lo siguiente:

$$\mathbb{E}[Y_{i,\eta}(0) | G_i = 1] - \mathbb{E}[Y_{i,-1}(0) | G_i = 1] = (\mathbb{E}[Y_{i,\eta}(0) | G_i = 0] - \mathbb{E}[Y_{i,-1}(0) | G_i = 0])$$

De esta forma, en cualquier momento del tiempo, el grupo de tratamiento y de control van a tener la misma evolución en ausencia del tratamiento. Teniendo esto en cuenta, podemos despejar  $ATT(\eta)$

$$\begin{aligned} ATT(\eta) &= \mathbb{E}[Y_{i,\eta}(1) - Y_{i,\eta}(0) | G_i = 1] \\ &= \mathbb{E}[Y_{i,\eta}(1) - Y_{i,\eta}(0) | G_i = 1] \\ &\quad + \underbrace{\mathbb{E}[Y_{i,\eta}(0) | G_i = 1] - \mathbb{E}[Y_{i,-1}(0) | G_i = 1] - (\mathbb{E}[Y_{i,\eta}(0) | G_i = 0] - \mathbb{E}[Y_{i,-1}(0) | G_i = 0])}_{\text{Esto es igual a 0 por el supuesto de tendencias paralelas}} \\ &= \mathbb{E}[Y_{i,\eta}(1) | G_i = 1, t = \eta] - \mathbb{E}[Y_{i,\eta}(0) | G_i = 1, t = \eta] + \mathbb{E}[Y_{i,\eta}(0) | G_i = 1, t = \eta] \\ &\quad - \mathbb{E}[Y_{i,-1}(0) | G_i = 1, t = -1] - (\mathbb{E}[Y_{i,\eta}(0) | G_i = 0, t = \eta] - \mathbb{E}[Y_{i,-1}(0) | G_i = 0, t = -1]) \\ &= \mathbb{E}[Y_{i,\eta}(1) | G_i = 1, t = \eta] - \mathbb{E}[Y_{i,-1}(0) | G_i = 1, t = -1] \\ &\quad - (\mathbb{E}[Y_{i,\eta}(0) | G_i = 0, t = \eta] - \mathbb{E}[Y_{i,-1}(0) | G_i = 0, t = -1]) \end{aligned}$$

Ahora, vamos a encontrar cada uno de los casos teniendo en cuenta el supuesto de exogeneidad de los errores:

CASO 1: Tratados en  $\eta$

$$\mathbb{E}[Y_{i,t} | G_i = 1, t = \eta] = \mathbb{E}[\alpha_i | G_i = 1] + \gamma_\eta + \delta_\eta = \underbrace{\mathbb{E}[Y_{i,\eta}(1) | G_i = 1, t = \eta]}_{\text{Dado que } t=\eta}$$

## CASO 2: Tratados antes del tratamiento

$$\mathbb{E}[Y_{i,t}|G_i = 1, t = -1] = \mathbb{E}[\alpha_i|G_i = 1] + \gamma_{-1} = \underbrace{\mathbb{E}[Y_{i,-1}(0)|G_i = 1, t = -1]}_{\text{Dado que } t=-1}$$

## CASO 3: No tratados en $\eta$

$$\mathbb{E}[Y_{i,t}|G_i = 0, t = \eta] = \mathbb{E}[\alpha_i|G_i = 0] + \gamma_\eta = \underbrace{\mathbb{E}[Y_{i,\eta}(0)|G_i = 0, t = \eta]}_{\text{Dado que } t=\eta}$$

## CASO 4: No tratados antes del tratamiento

$$\mathbb{E}[Y_{i,t}|G_i = 0, t = -1] = \mathbb{E}[\alpha_i|G_i = 0] + \gamma_{-1} = \underbrace{\mathbb{E}[Y_{i,-1}(0)|G_i = 0, t = -1]}_{\text{Dado que } t=-1}$$

Volvemos a la expresión del inicio y reemplazamos:

$$\begin{aligned} ATT(\eta) &= \mathbb{E}[Y_{i,\eta}(1)|G_i = 1, t = \eta] - \mathbb{E}[Y_{i,-1}(0)|G_i = 1, t = -1] \\ &\quad - (\mathbb{E}[Y_{i,\eta}(0)|G_i = 0, t = \eta] - \mathbb{E}[Y_{i,-1}(0)|G_i = 0, t = -1]) \\ &= \mathbb{E}[\alpha_i|G_i = 1] + \gamma_\eta + \delta_\eta - \mathbb{E}[\alpha_i|G_i = 1] - \gamma_{-1} \\ &\quad - \mathbb{E}[\alpha_i|G_i = 0] - \gamma_\eta + \mathbb{E}[\alpha_i|G_i = 0] + \gamma_{-1} \\ &= \delta_\eta \end{aligned}$$

De esta forma, el parámetro  $\delta_\eta$  captura el  $ATT(\eta)$ .

- b) ¿Cuál es el valor teórico de los  $\delta_{-j}$  si se cumplen los supuestos de identificación? Demuéstrenlo formalmente. ¿Cómo cambia esto su respuesta a la pregunta del inciso 3?

### Solución:

Se tiene que  $j \in \{2, \dots, k_1\}$ , con lo cual podemos partir siguiendo la misma lógica del inciso anterior utilizando el supuesto de tendencias paralelas con estas condiciones:

$$\mathbb{E}[Y_{i,\eta}(0)|G_i = 1] - \mathbb{E}[Y_{i,-1}(0)|G_i = 1] = (\mathbb{E}[Y_{i,\eta}(0)|G_i = 0] - \mathbb{E}[Y_{i,-1}(0)|G_i = 0])$$

$$\begin{aligned} \delta_{-j} &= \mathbb{E}[Y_{i,-j}(1) - Y_{i,-j}(0)|G_i = 1] \\ &= \mathbb{E}[Y_{i,-j}|G_i = 1, t = -j] - \mathbb{E}[Y_{i,-1}|G_i = 1, t = -1] \\ &\quad - (\mathbb{E}[Y_{i,-j}|G_i = 0, t = -j] - \mathbb{E}[Y_{i,-1}|G_i = 0, t = -1]) \\ &= \mathbb{E}[Y_{i,-j}(0)|G_i = 1, t = -j] - \mathbb{E}[Y_{i,-1}(0)|G_i = 1, t = -1] \\ &\quad - (\mathbb{E}[Y_{i,-j}(0)|G_i = 0, t = -j] - \mathbb{E}[Y_{i,-1}(0)|G_i = 0, t = -1]) \end{aligned}$$

Si se cumple el supuesto de no anticipación, debido a que nos encontramos en un periodo previo al tratamiento, se puede decir que los resultados potenciales de ambos grupos deben ser iguales. Esto debido a que estos valores no dependen de si se va a ser o no tratado en un futuro. De esta forma, se encuentra que:

$$\delta_{-j} = 0$$

Teniendo en cuenta esto, se puede decir que existe una forma para hallar evidencia que sugiera el cumplimiento del supuesto de tendencias paralelas. Esto debido a que al estimar el parámetro  $\delta_{-j}$ , se puede realizar una prueba de significancia estadística con el parámetro resultante de la prueba. Por ejemplo, si nuestra hipótesis nula es 0 y no se rechaza esta a partir de nuestro coeficiente estimado, se puede decir que hay evidencia que sugiera el cumplimiento del supuesto de tendencias paralelas. En el caso contrario, en el que sí se rechaza la hipótesis nula, no habría evidencia que sugiriera el cumplimiento del supuesto.

5. Triples diferencias (DDD) es una herramienta comúnmente empleada en contextos en los cuales los supuestos requeridos por diferencias en diferencias no se cumplen. En particular, si el ejercicio de validación propuesto en el inciso 4.b apunta a posibles violaciones de los supuestos de identificación, se hace necesario encontrar una nueva estrategia de identificación.

La intuición detrás de emplear DDD es la siguiente. Incluso si la variable de resultado para los dos grupos relevantes, tratados y controles, no evoluciona paralelamente, es posible que podamos estimar esa brecha, la diferencia en tendencias, ampliando nuestra muestra. Para aterrizar esta idea, consideren el siguiente ejemplo. Supongan que en la ciudad  $A$  se implementa una política educativa que beneficia a las mujeres y solo a ellas, y se desea conocer el efecto de dicha política sobre el salario. Se podría pensar que, aun en ausencia de tratamiento, el salario esperado de hombres y mujeres de la ciudad  $A$  hubiera evolucionado de manera diferencial. Ahora, supongan que obtenemos acceso a datos de hombres y mujeres de la ciudad  $B$ , en la cual no fue implementada ninguna política. Si pensamos que nuevamente hombres y mujeres de la ciudad  $B$  hubieran evolucionado de manera diferencial, pero que dicha diferencia de tendencias es la misma que aquella entre hombres y mujeres de la ciudad  $A$ , podemos usar esta nueva muestra (los individuos de  $B$ ) para “remover” la diferencia en tendencias y obtener estimadores consistentes de nuestros efectos causales de interés.

Supongan entonces que, adicional a su muestra original, ustedes cuentan con información de una nueva muestra. En esta nueva muestra, los individuos también pertenecen a uno de dos grupos, tratados o controles, que nuevamente vienen indicados por la variable  $G_i$ . Además,  $U_i$  es la indicadora de si el individuo  $i$  pertenece a la muestra original. Supongan que

i)  $\forall t$  :

$$\mathbb{E}[Y_{i,t}(0) - Y_{i,t-1}(0)|G_i = 1, U_i] - \mathbb{E}[Y_{i,t}(0) - Y_{i,t-1}(0)|G_i = 0, U_i] = DT(t) (\neq 0 \text{ Posiblemente}) \quad (4)$$

ii)

$$\mathbb{E}[Y_{i,\eta}(1) - Y_{i,\eta}(0)|G_i = 1, U_i = 0] = 0 \quad (5)$$

- a) Interpreten intuitivamente cada uno de estos supuestos.

**Solución:** Bajo el supuesto i), el primer término indica el valor esperado de la diferencia entre la variable de resultado de dos individuos consecutivos que hacen parte del grupo de tratados cuando aún no ha ocurrido el tratamiento. Por su parte, el segundo término indica el valor esperado de la diferencia entre la variable de resultado de dos individuos consecutivos que hacen parte del grupo de control cuando aún no ha ocurrido el tratamiento. De esta forma, la diferencia entre estos dos términos nos da un valor denotado por  $DT(t)$ , el cual puede ser diferente de cero. Esto implicaría que las diferencias en las tendencias de las muestras no sería igual a cero, por lo cual la variable de resultado para los distintos grupos no estaría evolucionando paralelamente. Por lo cual, se estaría incumpliendo el supuesto de tendencias paralelas.

El supuesto ii) indica que los individuos que hacen parte de la nueva muestra, aunque hagan parte del grupo de tratados, no van a percibir ningún efecto sobre su variable de resultado una vez el tratamiento haya sido efectivo. Es decir, aunque en esta segunda muestra existan los respectivos grupos de tratados y no tratados, no pareciera haber evidencia de que fue implementada alguna política significativa para la variable de resultado de estos individuos.

- b) ¿Qué implicaciones tienen estos supuestos sobre la estimación del modelo (3) a partir de únicamente la muestra original?

**Solución:**

El primer supuesto, en particular, implicaría que se puede no cumplir el supuesto de tendencias paralelas, situación bajo la cual la estimación del modelo por Mínimos Cuadrados Ordinarios no arrojaría estimadores consistentes. De esta forma, no se estaría capturando un efecto causal a través de este modelo, ya que no se tendría claridad sobre la forma en la que evolucionaban los

grupos que fueron tratados antes del tratamiento. Por su parte, el segundo supuesto implica que la segunda muestra no estaría percibiendo ningún efecto del tratamiento o no estaría expuesta al tratamiento.

6. Consideren el siguiente modelo:

$$Y_{i,t} = \alpha_i + \gamma_t + \gamma'_t U_i + \sum_{j=2}^{k_1} \left( \delta_{-j} D_{i,t+j} + \delta'_{-j} D_{i,t+j} U_i \right) + \sum_{j=0}^{k_2} \left( \delta_j D_{i,t-j} + \delta'_j D_{i,t-j} U_i \right) + \epsilon'''_{i,t} \quad (6)$$

a) ¿Qué parámetro, si alguno, captura el  $ATT(\eta)$ ? Demuéstrenlo formalmente.

**Solución:**

El supuesto i) nos indica que

$$\begin{aligned} \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 1, U_i = 1] - \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 0, U_i = 1] &= DT(\eta) \\ \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 1, U_i = 0] - \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 0, U_i = 0] &= DT(\eta) \\ \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 1, U_i = 1] - \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 0, U_i = 1] &= \\ \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 1, U_i = 0] - \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 0, U_i = 0] &= \end{aligned}$$

El supuesto ii) nos indica que

$$\begin{aligned} \mathbb{E}[Y_{i,\eta}(1) - Y_{i,\eta}(0) | G_i = 1, U_i = 0] &= 0 \\ \mathbb{E}[Y_{i,\eta}(1) | G_i = 1, U_i = 0] &= \mathbb{E}[Y_{i,\eta}(0) | G_i = 1, U_i = 0] \end{aligned}$$

Reemplazamos la expresión anterior en el primer supuesto,

$$\begin{aligned} \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 1, U_i = 1] - \mathbb{E}[Y_{i,t}(0) - Y_{i,\eta-1}(0) | G_i = 0, U_i = 1] &= \\ \mathbb{E}[Y_{i,\eta}(1) - Y_{i,\eta-1}(0) | G_i = 1, U_i = 0] - \mathbb{E}[Y_{i,t}(0) - Y_{i,\eta-1}(0) | G_i = 0, U_i = 0] &= \end{aligned}$$

De esta forma, podemos despejar  $\mathbb{E}[Y_{i,\eta}(0) | G_i = 1, U_i = 1]$  de la anterior expresión,

$$\begin{aligned} \mathbb{E}[Y_{i,\eta}(0) | G_i = 1, U_i = 1] &= \mathbb{E}[Y_{i,\eta-1}(0) | G_i = 1, U_i = 1] + \mathbb{E}[Y_{i,t}(0) - Y_{i,\eta-1}(0) | G_i = 0, U_i = 1] \\ &+ \mathbb{E}[Y_{i,\eta}(1) - Y_{i,\eta-1}(0) | G_i = 1, U_i = 0] - \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 0, U_i = 0] \end{aligned}$$

Ahora, reemplazamos esto último en la expresión de  $ATT(\eta)$ ,

$$\begin{aligned} ATT(\eta) &= \mathbb{E}[Y_{i,\eta}(1) - Y_{i,\eta}(0) | G_i = 1, U_i = 1] \\ ATT(\eta) &= \mathbb{E}[Y_{i,\eta}(1) - Y_{i,\eta-1}(0) | G_i = 1, U_i = 1] - \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 0, U_i = 1] \\ &- (\mathbb{E}[Y_{i,\eta}(1) - Y_{i,\eta-1}(0) | G_i = 1, U_i = 0] - \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 0, U_i = 0]) \end{aligned}$$

Esto se puede reescribir de la siguiente forma,

$$\begin{aligned} ATT(\eta) &= \mathbb{E}[Y_{i,\eta}(1) - Y_{i,\eta}(0) | G_i = 1, U_i = 1] \\ ATT(\eta) &= \Delta_t \mathbb{E}[Y_{i,t} | G_i = 1, U_i = 1] - \Delta_t \mathbb{E}[Y_{i,t} | G_i = 0, U_i = 1] \\ &- (\Delta_t \mathbb{E}[Y_{i,t} | G_i = 1, U_i = 0] - \Delta_t \mathbb{E}[Y_{i,t} | G_i = 0, U_i = 0]) \\ ATT(\eta) &= \tau_{DiD_{U_i=1}} - \tau_{DiD_{U_i=0}} \end{aligned}$$

Similar a lo que hicimos para el inciso 4, vamos a ver los casos y posteriormente vamos a reemplazar en la anterior expresión,

$$\Delta_t \mathbb{E}[Y_{i,t} | G_i = 1, U_i = 1] = (\gamma_\eta - \gamma_{\eta-1}) + (\gamma'_\eta - \gamma'_{\eta-1}) + \delta_0 + \delta'_0$$

$$\Delta_t \mathbb{E}[Y_{i,t} | G_i = 0, U_i = 1] = (\gamma_\eta - \gamma_{\eta-1}) + (\gamma'_\eta - \gamma'_{\eta-1})$$

$$\Delta_t \mathbb{E}[Y_{i,t} | G_i = 1, U_i = 0] = (\gamma'_\eta - \gamma'_{\eta-1}) + \delta_0$$

$$\Delta_t \mathbb{E}[Y_{i,t} | G_i = 0, U_i = 0] = (\gamma_\eta - \gamma_{\eta-1})$$

Reemplazando estas expresiones en nuestro  $ATT(\eta)$  y operando, encontramos que:

$$ATT(\eta) = \tau_{DiD_{U_i=1}} - \tau_{DiD_{U_i=0}} = \delta_0 + \delta'_0 - \delta_0 = \delta'_0$$

De esta forma, el  $ATT(\eta)$  es capturado por el parámetro  $\delta'_0$ .

- b) ¿Cuál es el valor teórico de los  $\delta'_{-j}$ ? Demuéstrenlo formalmente. ¿Cómo cambia esto su respuesta a la pregunta del inciso 3?

**Solución:** Esta demostración es semejante a la demostración realizada en el inciso 4b, en donde se espera encontrar entonces que  $\delta'_{-j} = 0$ .

Sabemos que:

$$\Delta_t \mathbb{E}[Y_{i,t} | G_i = 1, U_i = 1] = (\gamma_\eta - \gamma_{\eta-1}) + (\gamma'_\eta - \gamma'_{\eta-1}) + \delta_{-j} + \delta'_{-j}$$

$$\Delta_t \mathbb{E}[Y_{i,t} | G_i = 0, U_i = 1] = (\gamma_\eta - \gamma_{\eta-1}) + (\gamma'_\eta - \gamma'_{\eta-1})$$

$$\Delta_t \mathbb{E}[Y_{i,t} | G_i = 1, U_i = 0] = (\gamma'_\eta - \gamma'_{\eta-1}) + \delta_{-j}$$

$$\Delta_t \mathbb{E}[Y_{i,t} | G_i = 0, U_i = 0] = (\gamma_\eta - \gamma_{\eta-1})$$

De esta forma,

$$\Delta_t \mathbb{E}[Y_{i,t} | G_i = 1, U_i = 1] - \Delta_t \mathbb{E}[Y_{i,t} | G_i = 0, U_i = 1] = \delta_{-j} + \delta'_{-j}$$

$$\Delta_t \mathbb{E}[Y_{i,t} | G_i = 1, U_i = 0] - \Delta_t \mathbb{E}[Y_{i,t} | G_i = 0, U_i = 0] = \delta_{-j}$$

De acuerdo al supuesto de tendencias paralelas y de no anticipación, las anteriores dos expresiones se pueden igualar, ya que el supuesto i) indica que:

$$\begin{aligned} & \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 1, U_i = 1] - \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 0, U_i = 1] \\ &= \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 1, U_i = 0] - \mathbb{E}[Y_{i,\eta}(0) - Y_{i,\eta-1}(0) | G_i = 0, U_i = 0] \end{aligned}$$

Por lo tanto,

$$\delta_{-j} + \delta'_{-j} = \delta_{-j}$$

$$\delta'_{-j} = 0$$

Teniendo en cuenta esto, siguiendo la misma lógica del inciso 4b, se puede decir que ahora sí existe una forma para hallar evidencia que sugiera el cumplimiento del supuesto de tendencias paralelas. Esto debido a que al estimar el parámetro  $\delta'_{-j}$ , se puede realizar una prueba de significancia estadística con el parámetro resultante de la prueba. Por ejemplo, si nuestra hipótesis nula es 0 y no se rechaza esta a partir de nuestro coeficiente estimado, se puede decir que hay evidencia que sugiera el cumplimiento del supuesto de tendencias paralelas. En el caso contrario, en el que sí se rechaza la hipótesis nula, no habría evidencia que sugiriera el cumplimiento del supuesto.

## Segundo ejercicio

Actualmente, se están implementando diversas estrategias a nivel global para atenuar los impactos del cambio climático. La mayoría de países han ratificado acuerdos internacionales, como el Acuerdo de París de 2015, en los cuales se comprometen a minimizar su impacto ambiental. Del mismo modo, la inmensa mayoría de países han logrado reducir la cantidad de emisiones de  $CO_2$  y otros gases de efecto invernadero (OWiD, 2023).

No obstante, en algunas regiones, particularmente en Estados Unidos, las tendencias recientes de concentración de material particulado en el ambiente han mostrado signos de estancamiento o incluso de reversión. El material particulado (PM por sus siglas en inglés) hace referencia a la mezcla de partículas sólidas y gotas líquidas que se encuentran en el aire (por ejemplo: polvo, suciedad, hollín, etc.). Los tipos de material particulado son:

- $PM_{10}$ : Estas son partículas inhalables que tienen diámetros de 10 micrómetros o menos.
- $PM_{2.5}$ : Estas son partículas inhalables finas que tienen diámetro de 2,5 micrómetros o menos<sup>1</sup>

Este tipo de contaminación implica un riesgo para la salud ya que puede inhalarse y acumularse en el sistema respiratorio. En especial, el  $PM_{2.5}$  puede llegar a alojarse profundamente en los pulmones, causando infecciones respiratorias, enfermedades cardiovasculares, accidentes cerebrovasculares y cáncer de pulmón (OPS, 2016).

Según Burke et. al. (2023)<sup>2</sup>, para el caso de Estados Unidos, desde 2016 se ha encontrado que el humo de los incendios forestales ha influido en la tendencia de la concentración promedio de  $PM_{2.5}$ .

Ustedes empiezan a trabajar en el Banco Mundial y se les encomienda la tarea de estudiar esta relación. En particular, están interesados en estudiar los efectos a corto y mediano plazo de los incendios forestales sobre la emisión de material particulado de  $2.5 \mu g/m^3$ <sup>3</sup>. Para ello, ustedes cuentan con una base de datos para todos los municipios (*counties*) de Estados Unidos suministrada por el Programa de Análisis de Incendios (Fire Program Analysis (FPA)). En esta base tienen información sobre el municipio donde ocurrió un incendio forestal. Adicionalmente, cuentan con información sobre la concentración promedio de material particulado para cada día del año para el verano de 2019<sup>4</sup> “Air Quality Statistics Report”.

La base de datos con esta información se llama “*smoke\_day\_v3.dta*”. Donde tiene información de los incendios para cada municipio durante el verano de 2019. Algunas de las variables relevantes son:

- *county\_fip*: Código único del municipio
- *smoke\_day<sub>m,t</sub>*: Una variable indicadora que toma el valor de 1 si en el día  $t$  en el municipio  $m$  hubo un incendio forestal.
- *date*: Es una variable que indica la fecha.
- *pm25*: Esta variable indica la concentración promedio diaria para cada municipio.
- *day*: Esta variable indica el día de verano para cada condado.

Con estos datos deben estimar el efecto causal de los incendios sobre la emisión de  $PM_{2.5}$  en las ciudades y determinar si este efecto persiste o no en el mediano plazo.

- a. De acuerdo con la información disponible y el contexto proporcionado, expliquen cual(es) de la(s) siguiente(s) metodología(s) es(son) apropiada(s) para estimar de manera consistente el efecto de interés. Justifiquen de forma detallada sus respuestas. Para el/los método(s) que consideren pertinente(s), presenten las ecuaciones que describen la estrategia de identificación, explique cual sería el parámetro de interés y cuales son los supuestos de identificación.

<sup>1</sup>Como referencia, un cabello humano mide en promedio 70 micrómetros de diámetro

<sup>2</sup>Burke, M., Childs, M.L., de la Cuesta, B. et al. The contribution of wildfire to  $PM_{2.5}$  trends in the USA. Nature (2023). <https://doi.org/10.1038/s41586-023-06522-6>

<sup>3</sup>Micrómetros por metro cúbico. Es decir, cantidad de material particulado por cada metro cúbico de aire.

<sup>4</sup>Considere que el verano en el hemisferio norte se suele considerar que empieza el 20 de junio y acaba el 22 de septiembre



- i. Diferencia de medias entre los municipios que tuvieron un incendio forestal en algún momento entre 2007 y 2019 respecto a los municipios donde no hubo un incendio.
- ii. Diferencia de medias entre la emisión de material particulado previo a la ocurrencia de incendios forestales y posterior a la ocurrencia de un incendio forestal para cada uno de los municipios.
- iii. Regresión discontinua
- iv. Variables instrumentales
- v. Diferencias en diferencias  $2 \times 2$
- vi. Diferencias en diferencias con adopción escalonada del tratamiento

**Solución:**

En este caso, estamos interesados en estudiar los efectos a corto y mediano plazo de los incendios forestales sobre la emisión de PM2.5 en las ciudades. Para esto, en nuestra base de datos contamos con dos variables que son particularmente de nuestro interés. En primer lugar, una variable indicadora que toma el valor de 1 si en el día  $t$  en el municipio  $m$  hubo un incendio forestal y, adicionalmente, una variable que indica la concentración promedio diaria para cada municipio de PM2.5.

Teniendo en cuenta lo anterior, nuestro tratamiento sería la ocurrencia de un incendio en un determinado municipio. Por lo cual, un municipio podría considerarse no tratado hasta el momento en el que sufra de un incendio, caso en el cual entraría al grupo de tratamiento.

Ahora bien, para aterrizar mejor estas ideas vamos a analizar algunas de las alternativas metodológicas que propone el enunciado para ver cuál sería la más apropiada:

- i. Diferencia de medias entre los municipios que tuvieron un incendio forestal en algún momento entre 2007 y 2019 respecto a los municipios donde no hubo un incendio.

Desde una primera instancia, esta metodología no sería apropiada debido a que no contamos con datos desde el 2007 para estudiar. Asimismo, se estaría desperdiciando el potencial que tienen los datos diarios para estudiar los efectos a corto y mediano plazo que pide el enunciado.

- ii. Diferencia de medias entre la emisión de material particulado previo a la ocurrencia de incendios forestales y posterior a la ocurrencia de un incendio forestal para cada uno de los municipios.

Para encontrar un efecto causal a través de esta metodología se necesitaría que la asignación al tratamiento (i.e. la ocurrencia de un incendio) fuera completamente aleatoria. Si bien en algunos municipios puedes ser el caso (i.e. accidentes), seguramente en varios municipios existan factores que no estamos considerando que puedan aumentar la incidencia de estos incendios. Asimismo, para comparar cada uno de los municipios se necesitaría saber más características para poder entender si son efectivamente comparables o no. En nuestra base de datos no contamos con suficiente información al respecto.

- iii. Regresión discontinua

Para lograr utilizar esta metodología se necesitaría establecer un umbral a través del cual un municipio sería tratado (ocurre un incendio) o no (no ocurre un incendio). Sin embargo, teniendo en cuenta la naturaleza del tratamiento, no es posible establecer un umbral que permita explotar esta metodología, ya que el incendio ocurrió o no ocurrió. Es decir, no es claro si este tratamiento es realmente una función de un score determinado aleatoriamente. Por lo cual, tampoco sería posible crear un bando de ancha que permita estudiar los municipios que están alrededor del corte”, incrementando aún más el problema de utilizar esta metodología.

- iv. Variables instrumentales

En este caso, teniendo en cuenta la información disponible en la base de datos suministrada, no se encuentra suficientes datos para proponer una variable instrumental. Si bien resulta interesante estudiar los problemas de endogeneidad que pueden existir en este ejercicio (factores que no estamos teniendo en cuenta que pueden incidir sobre la contaminación), desafortunadamente no podemos proponer una solución por esta metodología ya que no contamos con una variable que logre instrumentar nuestra variable explicativa (la ocurrencia de un incendio), ya que solo contamos con variables descriptivas aparte de las de nuestro interés.

v. Diferencias en diferencias  $2 \times 2$

En este caso, no sería conveniente utilizar esta metodología ya que los municipios sufren de incendios en diferentes periodos, no necesariamente en el mismo momento, por lo cual no habría un único tratamiento. Asimismo, teniendo en cuenta que necesitamos ver los efectos tanto a corto como largo plazo, el hecho de que solo hayan dos periodos de estudio hace que se pierda este componente dinámico del problema.

vi. Diferencias en diferencias con adopción escalonada del tratamiento

Ahora bien, esta metodología permite capturar el componente dinámico que decíamos que se perdía en la del anterior inciso. Esto debido a que ya permite que las unidades a estudiar reciban el tratamiento en diferentes momentos del tiempo, lo cual es precisamente nuestro caso, ya que las ciudades se incendian en diferentes momentos. Con lo cual se pueden analizar los efectos a corto y largo plazo, al estudiar la evolución de los municipios a través de distintos periodos posteriores al incendio.

De esta forma, la metodología que consideramos más apropiada para estimar de manera consistente el efecto de interés sería la metodología de diferencias en diferencias con adopción escalonada del tratamiento. A continuación, presentamos las ecuaciones que describen la estrategia de identificación, explicando cuál sería el parámetro de interés y cuáles son los supuestos de identificación:

$$Y_{m,t} = \alpha_m + \gamma_t + \sum_{j=-q}^{-k_1} \tau_j D_{jm,t} + \sum_{j=0}^{k_2} \tau_j D_{jm,t} + \epsilon_{m,t},$$

El parámetro de interés en este caso sería  $\tau_j$ , el cual recupera un efecto causal si se cumplen los supuestos de identificación de tendencias paralelas y no anticipación. El primero implica que, en ausencia del tratamiento, el cambio esperado en la variable dependiente del grupo de tratamiento es igual al cambio observado en esa variable para el grupo de control. Por su parte, el segundo supuesto implica que los valores de la variable independiente observados antes del momento de tratamiento no dependen de si voy a ser o no tratado en el futuro.

Luego de discutir con su jefa, llegan a la conclusión de que la mejor metodología para estimar el efecto causal sería utilizar una aproximación de Diferencias en Diferencias. Sin embargo, dado que cada uno de los municipios puede ser tratado en momentos diferentes del tiempo, es decir, tienen un tratamiento escalonado, la especificación que ella les propone es una de Estudio de Eventos.

Considerando que tiene información diaria para cada municipio, ella les solicita ajustar la base de datos con el fin de estimar este modelo.

- b. Transformen la información que se encuentra en la base del datos para generar una base de datos panel, con la cual pueda estimar los modelos expuestos más adelante. Recuerde definir las variables relativas a la primera vez tratado<sup>5</sup>, una variable dicótoma que indique si el municipio  $i$  es tratada en el tiempo  $t^6$ ,

<sup>5</sup>que indique la primera vez que el municipio tiene un incendio

<sup>6</sup>vamos a suponer que se trata de un estado absorbente

variables del tiempo relativo al tratamiento, los *leads* y los *lags*. Presenten el código que utilizaron a continuación

```

1  *****
2  *****PUNTO  B*****
3  *****
4
5  *) Ordenamos la base de acuerdo a la variable panel (countyfip) y a la variable de tiempo (
   date)
6  order countyfip date
7
8  *) Declaramos la base para ser panel
9  xtset countyfip date, d
10
11 *) Generamos una variable que nos indique la fecha en la que hubo cada incendio
12 gen first_fire_date = day if smoke_day==1
13
14 *) A partir de lo anterior, generamos una variable que nos indique cuando fue el primer
   incendio
15 egen first_fire = min(first_fire_date), by(countyfip)
16 replace first_fire = 0 if first_fire==.
17
18 *) Generamos variable de tratamiento
19 gen Dit = (day >= first_fire & first_fire!=0)
20
21 *) Creamos una variable que nos indique el tiempo relativo
22 bys countyfip: gen rel_time=day-first_fire
23 replace rel_time=. if first_fire==0
24
25
26 *) Generamos dummies para cada categoria posible de tiempos relativos
27 tab rel_time, gen(evt)
28
29 *) Cambiamos los labels y los nombres para no confundirnos
30
31 *) Generamos los Leads
32
33 forvalues x = 1/80 {
34
35     local j= 81-'x'
36     ren evt 'x' evt_l 'j'
37     cap label var evt_l 'j' "- 'j'"
38 }
39
40
41 *) Generamos los Lags
42
43 forvalues x = 0/82 {
44
45     local j= 81+'x'
46     ren evt 'j' evt_f 'x'
47     cap label var evt_f 'x' "'x'"
48 }
49
50 *) Periodo base (t=-1) - Omitir por multicolinealidad
51 replace evt_l1=0

```

- c. Expliquen los supuestos para que el estimador por MCO del modelo TWFE provea estimadores consistentes del efecto causal de interés en el caso de un estudio de eventos. Justifiquen si consideran que son factibles para el contexto planteado. Adicionalmente, escriban las ecuaciones que describen estos supuestos.

### Solución:

Para que el estimador por MCO del modelo TWFE provea estimadores consistentes del efecto causal de interés en el caso de un estudio de eventos, se deben cumplir los siguientes supuestos:

- **Tendencias paralelas generalizadas:** El supuesto de tendencias paralelas implica que, en ausencia del tratamiento, el cambio en la variable de resultado es el mismo para el grupo de tratamiento y de control. En este caso, se cumple el supuesto de tendencias paralelas si, en ausencia de incendios forestales, el cambio en la concentración promedio de PM2.5 es igual para el grupo de tratamiento y de control. Esto implica que para cada t:

$$\mathbb{E}[Y_{m,t}(0) - Y_{m,t-1}(1)|D_m = 1] = \mathbb{E}[Y_{m,t}(0) - Y_{m,t-1}(1)|D_m = 0]$$

En cuanto la factibilidad, es posible que haya una evolución distinta de PM2.5 entre

**municipios si se comparan ciudades, aglomeraciones, municipios rurales y rural disperso.**

- **No anticipación:** El supuesto de no efectos anticipatorios antes del tratamiento implica que los valores de la variable independiente observados antes del momento de tratamiento no dependen de si los individuos van a ser o no tratados en un futuro. Así, se asume que, antes de que ocurran los incendios forestales, la concentración de PM2.5 no se ve afectada por la expectativa o anticipación del evento. En otras palabras, cualquier cambio en la concentración de PM2.5 antes del incendio no debe estar relacionado con la previsión o conocimiento previo del incendio. La ecuación que describe este supuesto es la siguiente:

$$Y_{m,t}(0) = Y_{m,t}(1) \forall t : D_m = 1$$

Este supuesto es más plausible si no se conoce la existencia del tratamiento o la trayectoria de este. Es decir, si los individuos tienen conocimiento privado sobre el futuro del tratamiento, pueden cambiar su comportamiento en anticipación y, por lo tanto, el resultado potencial antes del tratamiento puede no representar los resultados iniciales.

En este caso, los individuos no tienen cómo saber en qué municipios habrá un incendio forestal, pues estos eventos son impredecibles. Por esta razón, no tiene sentido pensar que los individuos tendrán anticipación alguna.

- **Hay efectos constantes y no hay efectos dinámicos:** Este supuesto implica que los efectos varían entre cohortes, pero no en el tiempo. Adicionalmente, el cambio en los efectos del tratamiento entre municipios debe ser nulo. Particularmente esto se ve a través de la ecuación:

$$\Delta ATT = 0$$

También entendido como:

$$\tau_l = \tau'_l$$

- d. Estimen por MCO el efecto de la ocurrencia de un incendio forestal sobre la concentración de PM2.5 con un modelo de TWFE dinámico con adopción escalonada del tratamiento.

Se les solicita estimar un modelo donde identifiquen el efecto en cada periodo 10 días antes y 10 días después del la ocurrencia del incendio para los municipios. Para la estimación del modelo, vamos a suponer que los incendios son un estado absorbente, es decir, luego del primer incendio se puede considerar al municipio como tratado. Adicionalmente, incluyan efectos fijos por día, municipio y día de la semana. Utilicen errores estándar clusterizados a nivel municipio.

Presenten la ecuación del modelo que se les solicita, una tabla con los resultados del modelo estático, una gráfica donde presenten los coeficientes y sus respectivos intervalos de confianza.

Interpreten los resultados obtenidos. Primero, concluyan respecto al cumplimiento de los supuestos de identificación. Con ello, expliquen si el efecto estimado es plausiblemente causal o no. Expliquen cual es la importancia e intuición de los efectos fijos incluidos. Luego, indiquen si efectivamente parece que existen efectos de corto y mediano plazo sobre la emisión de material particulado luego de incendios forestales.

#### **Solución:**

A continuación, se presenta la ecuación del modelo que se solicita:

$$Y_{mt} = \alpha_m + \gamma_t + \theta_d + \sum_{\substack{j \neq -1 \\ -10 \leq j \leq 10}} \beta_j D_{mt}^j + \epsilon_{mt}$$

Donde  $D_{mt}^j$  es una variable dicótoma que es igual a 1 si el día  $t$  está a  $j$  periodos desde el momento del tratamiento de  $m$ ,  $t_m$ . Es decir,  $D_{mt}^j = \mathbb{1}\{t - t_m = j\}$ . Por su parte,  $\beta_j$ : para  $j \geq 0$ , se piensa

como el efecto del tratamiento  $j$  periodos después de ser tratado. Para  $j < 1$ , se piensa como el efecto anticipatorio del tratamiento  $|j|$  periodos antes. Finalmente,  $\alpha_m$  es el efecto fijo por municipio,  $\gamma_t$  es el efecto fijo por día y  $\theta_d$  es el efecto fijo por día de la semana.

Ahora, se presentan los resultados de la estimación del anterior modelo a través de MCO:

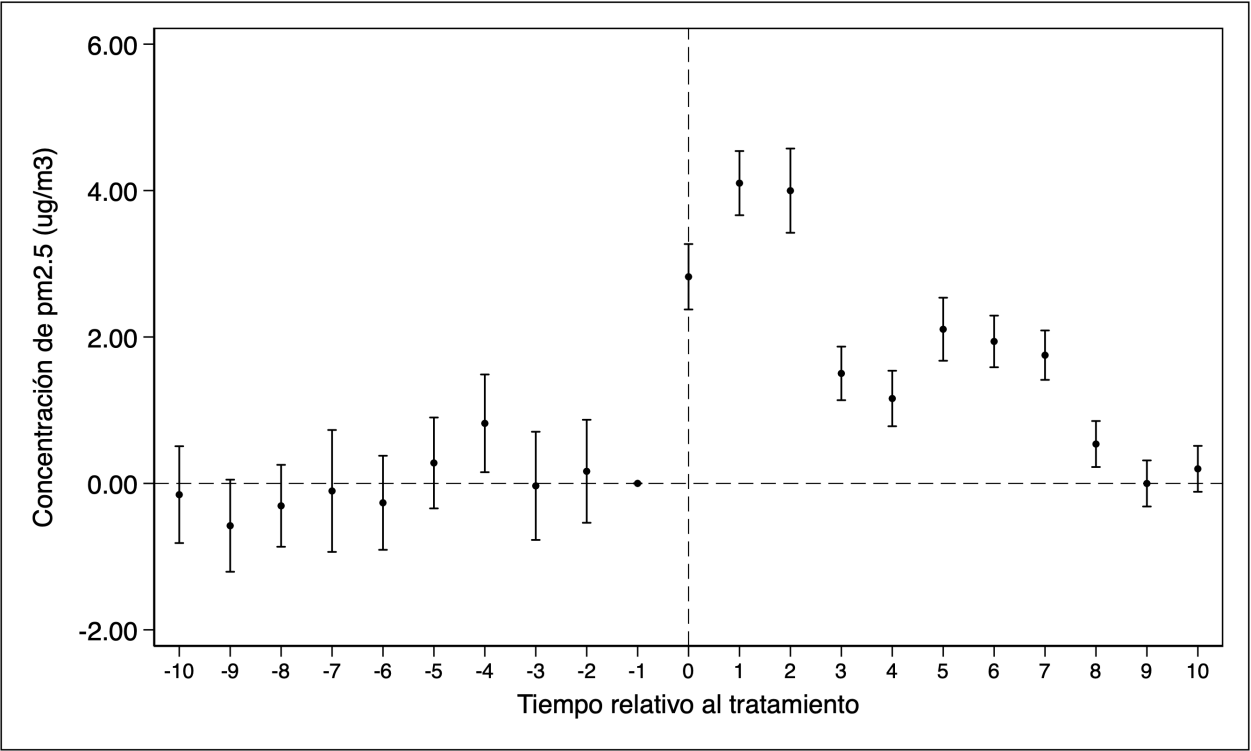
Tabla 1: Estimación por MCO del modelo de TWFE

	PM2.5 concentración (ug/m3)
$D_{mt}^j$	2.694*** (0.211)
Observaciones	36,069
R-cuadrado	0.236
Efectos fijos por día	Sí
Efectos fijos por municipio	Sí
Efectos fijos por día de la semana	Sí

\*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Errores estándares clusterizados a nivel municipio.

Efectos fijos: Día (day), Municipio (countyfip), Día de la semana (weekday).

Figura 1: Coeficientes del modelo de TWFE dinámico con adopción escalonada del tratamiento



Los resultados de la estimación se encuentran en la Tabla 1. Efectivamente, los incendios forestales incrementan la emisión de material particulado de  $2.5 \mu g/m^3$ . Aquellos municipios que sufren un incendio forestal, es decir hacen parte del tratamiento, tienen una mayor concentración diaria de PM2.5; específicamente  $2.7 \text{ ug/m}^3$  más, que los municipios que nunca han sufrido de un incendio forestal, ceteris paribus. El efecto es significativo, con un nivel de significancia del 1 % sabemos que un incendio forestal aumenta la concentración diaria de PM2.5.

A través de la Figura 1, observamos la evidencia que sugiere que se cumplen los supuestos de identificación, tanto el supuesto de tendencias paralelas como el de no anticipación. Primero, los estimadores del periodo pre tratamiento son cercanos a cero. Segundo, 9 de los 10 intervalos de confianza de los coeficientes incluyen cero. Tercero, para 8 de los 10 estimadores no es posible rechazar la hipótesis nula la cual indica que el estimador es efectivamente 0. Adicionalmente, no se

observa una desviación de la tendencia para los periodos inmediatamente anteriores al tratamiento. En este sentido, es posible afirmar que la evidencia sugiere que se cumplen los supuestos de identificación del modelo y el efecto estimado es plausiblemente causal.

La incorporación de efectos fijos por día, municipio y día de la semana fortalece la validez interna de las estimaciones al controlar factores temporales y geográficos. Sin embargo, es importante destacar que, a pesar de la significancia estadística, el modelo explica aproximadamente el 23.7% de la variabilidad en la concentración de PM2.5, lo que sugiere la existencia de otros determinantes que podrían influir en la presencia del PM2.5 en el aire.

Por último, en la Figura 1 se observa nuevamente que los niveles de PM2.5 en los días antes del incendio es menor que los días posterior al incendio. Es posible notar que si bien los días posterior al incendio tienen mayor concentración de PM2.5, la tendencia es decreciente a medida que pasan los días. Incluso, en el periodo  $t=9$  y  $t=10$ , la concentración de PM2.5 disminuye casi al mismo nivel que los días antes del incendio forestal y las diferencias ya no son significativas. Hay evidencia de la existencia de efectos dinámicos positivos del tratamiento sobre la variable de resultado. Definitivamente, existen efectos de corto plazo sobre la emisión de material particulado luego de incendios forestales. En cuanto a efectos a mediano plazo, también los hay, pero en menor medida.

Recientemente varios artículos han mostrado que, en contextos de tratamiento escalonado, es decir, hay diversas cohortes de tratamiento, los estimadores por MCO del modelo de TWFE pueden ser inconsistentes (Goodman-Bacon, 2021).

- e. Expliquen intuitivamente por qué aparece este problema y cuáles son las situaciones donde los estimadores por MCO del modelo TWFE no proveen estimadores consistentes<sup>7</sup>. ¿Consideran que para esta estimación hay alguno de estos problemas?

**Solución:**

Los estimadores de los modelos de TWFE no capturan el efecto de forma consistente si sucede alguna de las siguientes dos cosas: hay efectos dinámicos del tratamiento o hay efectos heterogéneos entre cohortes X periodos después del tratamiento. Intuitivamente, cuando se presenta alguna de las dos situaciones, o las dos, este tipo de modelos hace comparaciones no válidas entre unidades que ya han sido tratadas y aquellas que reciben el tratamiento.

Adicionalmente, los pesos que se otorgan en la metodología son una función del tamaño de cada grupo y del tiempo de exposición al tratamiento. Si hay efectos heterogéneos, la distribución temporal del tratamiento afecta el resultado. Si hay efectos dinámicos, algunos pesos pueden ser negativos. Así, es posible que el ATT asociado a todas las diferencias en diferencias de  $2 \times 2$  sea positivo, pero  $\tau$  es negativo.

Como respuesta a los problemas de las estimaciones de un tratamiento escalonado usando un modelo de TWFE se han desarrollado una serie de nuevos estimadores que resuelven los problemas enunciados. El más popular ha sido el propuesto por Callaway y Sant'anna (2021). Su método se basa en separar el análisis de Diferencias en Diferencias en 3 pasos: i) la estimación de los efectos causales desagregados por cohorte usando grupos de control “adecuados”; ii) la agregación de estos efectos al nivel de interés, y iii) procedimientos para realizar inferencia estadística sobre los valores estimados.<sup>8</sup>

- f. Estimen el efecto del tratamiento utilizando el estimador de Callaway y Sant'anna (2021)<sup>9</sup> y presenten gráficamente los efectos dinámicos del tratamiento. Interpreten sus resultados. ¿Cómo cambia el efecto respecto a las conclusiones obtenidas en el inciso d.?

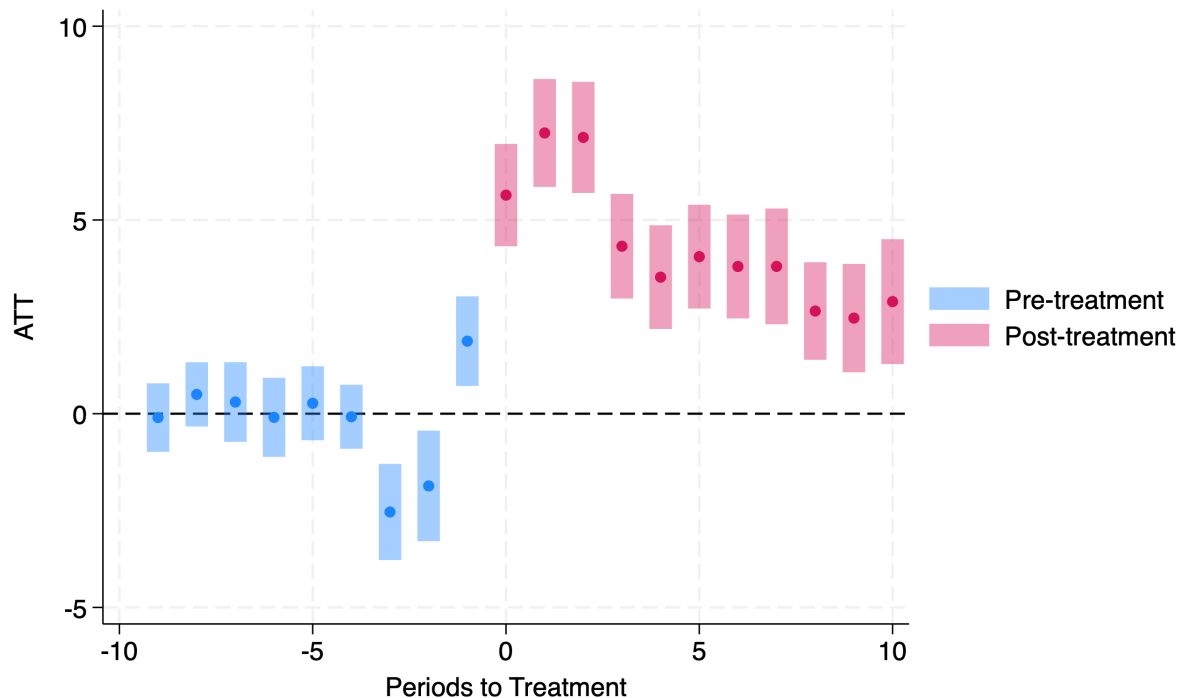
<sup>7</sup>Otro autor que realiza una discusión general sobre los problemas de usar modelos de TWFE es Roth et. al. (2022)

<sup>8</sup>Este resumen de Scott Cunningham (2021) puede ser de ayuda

<sup>9</sup>Recuerden que solo están interesados en ver el efecto 10 días antes y 10 días después, no el efecto en todos los días para todas las cohortes

### Solución:

Figura 2: Efectos dinámicos del tratamiento



A primera vista, la gráfica sugiere que no se cumple ningún supuesto de identificación. Primero, hay tres periodos de tiempos para los cuales los coeficientes son significativos, indicando que no se cumple el supuesto de tendencias paralelas. Segundo, se observa una caída en el cambio en los periodos justo antes del tratamiento. Así, existe evidencia que sugiere que no se cumple el supuesto de identificación de no anticipación. Esto afecta los resultados porque invalidaría las comparaciones que se hagan y los efectos no se podrían pensar como limpios y causales. Por último, se observa que el efecto a mediano plazo en los periodos post tratamiento es contundente, pues todos los periodos post-tratamiento tienen coeficientes significativos cuyos intervalos de confianza no incluyen cero.

- g. Investiguen y expliquen brevemente y en sus palabras cómo los estimadores de [Sun y Abraham \(2021\)](#) y [Borusyak et. al. \(2022\)](#) lidian con los problemas de TWFE. Adicionalmente, estimen el efecto del tratamiento con estos estimadores y, en una sola gráfica, muestren los resultados comparando TWFE, [Callaway y Sant'anna \(2021\)](#), [Sun y Abraham \(2021\)](#) y [Borusyak et. al. \(2022\)](#). Interpreten sus resultados.<sup>10</sup>

### Solución:

Los autores Sun y Abraham (2021) hacen una descomposición del coeficiente estimado de TWFE bajo los supuestos de tendencias paralelas y no anticipación: i) Uno que es deseable, el promedio de los efectos dinámicos de todas las observaciones que están en una misma caja; ii) Uno que es indeseable, contaminación de otras cajas; iii) Otro que es indeseable, contaminación por periodos omitidos. De esta forma, el estimador puede no preservar el signo de los efectos causales, lo cual invalida la validación de tendencias paralelas e interpretación de efectos dinámicos.

Ellos proponen estimadores paramétricos de  $ATT(g,t)$  usando modelos de regresión saturados que permiten recuperar efectos causales aún en presencia de heterogeneidad entre cohortes y efectos dinámicos. Se pueden estimar tanto efectos causales dinámicos como en otros niveles de agregación. Para esto, lo que hacen prácticamente es no usar el grupo de los 'ya tratados', que son los que generan problemas, como controles en las ecuaciones de DiD. Esto lo hacen a través de una estimación de los leads y los lags utilizando controles no tratados ponderados por una

<sup>10</sup>La gráfica debe tener un formato diferente al mostrado en la clase complementaria

proporción estimada de unidades de una cohorte. De esta forma, como las ponderaciones ahora son consistentes, y los leads y los lags también son estimados consistentemente, la estimación de la interacción ponderada también lo es.

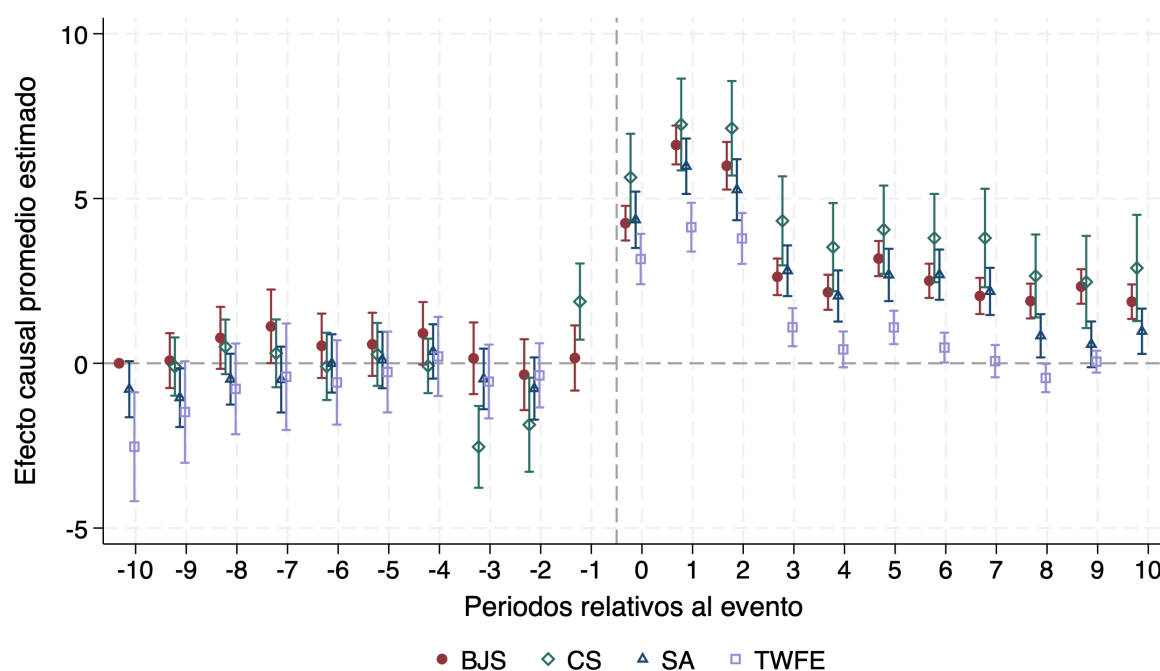
En otras palabras, lo que hacen es estimar los tres casos que vimos en el primer párrafo usando un modelo TWFE saturado con interacciones entre todas las variables. De esta forma, cuando hacen la estimación, se le puede asignar los pesos que se requieran a estos para asegurarse que los pesos sean mayores a cero y que logren sumar 1. Asimismo, este procedimiento elimina las contaminaciones no deseadas de otras cajas.

Por su parte, Borusyak et. al. (2022) proponen (1) un marco más claro y explícito para definir los objetivos de estimación y (2) un estimador eficiente y robusto para abordar la heterogeneidad en los efectos del tratamiento en estudios de eventos con adopción escalonada. El marco permite probar de mejor manera el cumplimiento de los supuestos de tendencias paralelas y no efectos anticipados. El estimador se basa en un enfoque de imputación que busca contrarrestar los desafíos asociados con la estimación de tratamientos dinámicos en presencia de heterogeneidad. Primero, los efectos fijos de unidad y período se ajustan con regresiones que solo utilizan observaciones no tratadas. Segundo, utilizan los efectos fijos para imputar los resultados potenciales no tratados y obtener un efecto de tratamiento para cada observación tratada. Finalmente, se hace una suma ponderada de estas estimaciones de efectos de tratamiento, con pesos que corresponden al objetivo de estimación planteado desde un principio.

Los autores también garantizan que las inferencias realizadas a partir del estimador propuesto sean válidas. Para esto, derivan condiciones bajo las cuales el estimador es consistente y asintóticamente normal, y proponen estimaciones de errores estándar que atribuyen alguna variación en los efectos de tratamiento estimados al error.

El enfoque evita la contaminación de las pruebas por la heterogeneidad en los efectos del tratamiento mostrada por Sun y Abraham (2021). Además, la estrategia evita una pérdida de poder en las pruebas y evita los problemas de prepruebas bajo homocedasticidad.

Figura 3: Efectos dinámicos del tratamiento para diferentes estimadores de estudios de eventos



La figura 3 muestra los resultados de los efectos dinámicos del tratamiento para los estimadores de TWFE, Callaway y Sant'anna (CA), Sun y Abraham (SA) y Borusyak et. al. (BJS). En general,



los estimadores son similares entre sí. Los estimadores de BJS, SA y TWFE parecen cumplir ambos supuestos de identificación: tendencias paralelas y no anticipación. Por su parte, y como mencionado anteriormente, CA no cumple con estos supuestos porque algunos coeficientes pre tratamiento son estadísticamente significativos y existe una caída en el cambio justo antes de ser tratados.

En los periodos post tratamiento, los estimadores de BJS, CS, y SA todos tienen efectos significativos a corto y mediano plazo pues los coeficientes son significativos y los intervalos de confianza no contienen cero (excepto CS  $t=9$ ). En cambio, los estimadores de TWFE no parecen tener efectos en el mediano plazo, solo en el corto plazo. Estos, al igual que los estimadores de SA Y BJS, tienen intervalos de confianza más acotados que los de CS. Esto concuerda con lo anteriormente discutido sobre los estimadores menos eficientes de CA.

h. ¿Por qué es relevante realizar un análisis comparativo entre los resultados de distintos tipos de estimadores?

**Solución:**

Realizar un análisis comparativo entre distintos estimadores es crucial para evaluar la robustez y validez de los resultados empíricos. Este análisis no solo ayuda a identificar posibles sesgos o errores sistemáticos, sino también a entender la precisión y consistencia de los estimadores utilizados. Al verificar que los resultados son consistentes a través de diferentes métodos, se refuerza la robustez de los estudios y se fortalece la confianza en las conclusiones obtenidas.

### Tercer ejercicio

Minimos Cuadrados Ordinarios (MCO) es un método de estimación que utiliza como función objetivo (función de pérdida) el error cuadrático medio (MSE por sus siglas en inglés) para obtener estimadores muestrales para un modelo de regresión lineal. Concretamente, bajo una variable aleatoria independiente  $x_i$ , un error de predicción  $u_i$ , un parámetro poblacional de un proceso generador de datos  $\beta$  y una variable dependiente  $y_i$  soluciona el siguiente proceso de minimización:

$$\min \quad MSE = \frac{1}{n} \sum_{i=1}^n (u_i)^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \beta x_i)^2 \quad (7)$$

donde  $MSE = (u_i)^2$  establece la función de pérdida del error de predicción al cuadrado promedio, que puede ser visualizada en la figura 4.

Figura 4: Función de pérdida en MCO

Talleres/2023-2/Figures/x2.png

Así, noten que la función cuadrática que penaliza el error es convexa, por lo que puntos muy distantes a la recta ajustada generan errores muy grandes. Por esto, para encontrar un mínimo, la suma de los residuales positivos (por encima de la recta) tiene que ser igual en magnitud a la suma de los residuales negativos (por debajo de la recta). Si no fuese así, por la naturaleza convexa de la función existiría una recta ajustada distinta que reduzca marginalmente el error de predicción en alguno de los dos sentidos, afectando mínimamente el otro.

1. Para ahondar más en esto ustedes cuentan con una base de datos simulada llamada “sim.dta” con 54 observaciones con información de dos variables  $X \sim \text{Gamma}(2, 0.5)$  e  $Y \sim N(0, X/10) + X$ . Ajusten por MCO un modelo de regresión lineal entre ambas variables usando  $Y$  como dependiente.
- a) Reporten en una tabla la suma de los errores positivos, la suma de los errores negativos, el número de observaciones con errores positivos y el número de observaciones con errores negativos.

**Solución:**

Tabla 2: Errores positivos y negativos

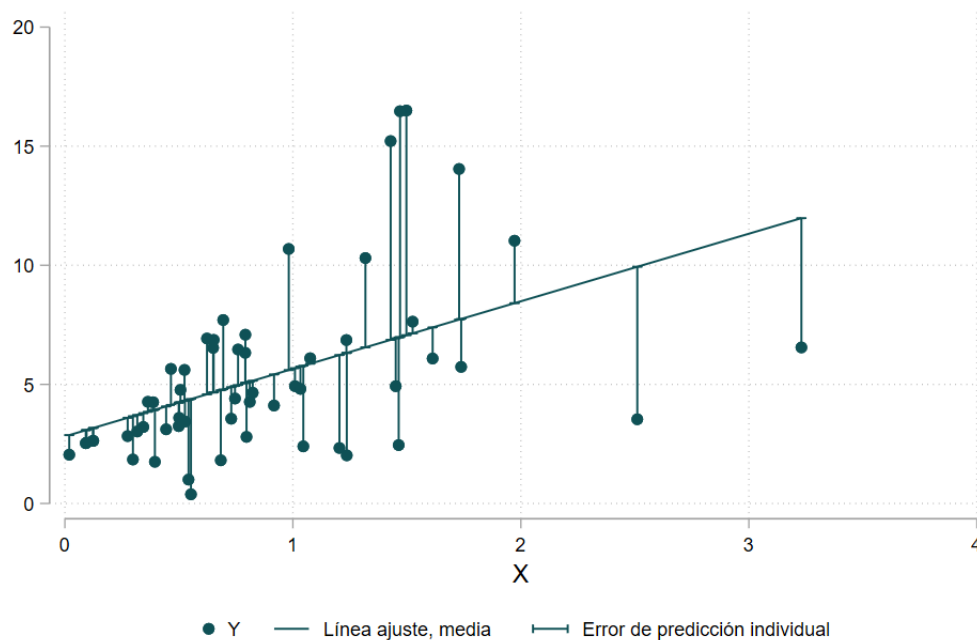
	No. de obs	Suma
Positivos	22	64.52114
Negativos	34	-64.52114
Total	56	0.00000

En primer lugar, observamos de un total de 54 observaciones, la regresión ha producido errores de predicción positivos en 22 observaciones y errores de predicción negativos en 34 observaciones. En 22 ocasiones, las predicciones del modelo están por encima de los valores reales de la variable dependiente mientras que en 34 ocasiones, las predicciones del modelo están por debajo de los valores reales de la variable dependiente. La suma de los errores positivos y negativos es de 64.52114 y -64.52114, respectivamente. Dado que los errores por encima de la recta son de igual magnitud a aquellos por debajo de la recta, se ha encontrado un mínimo.

- b) Presenten una gráfica de dispersión entre la variable Y e X, incluyan dentro de esta la línea de regresión ajustada por MCO y adicioneen al gráfico líneas verticales que conecten cada valor de Y observado con el valor de Y en la línea de ajuste (visualmente esta última línea corresponde al error de predicción individual).

**Solución:**

Figura 5: Gráfica de dispersión



La figura 2 es una representación gráfica de los resultados de la tabla del inciso anterior. Se observa que los círculos redondos, indicando las predicciones del modelo, están concentrados por debajo de la línea de ajuste. Esto concuerda con los resultados de la Tabla 1, pues la mayoría de errores son negativos. Adicionalmente, sabemos que la magnitud de la suma de los errores negativos y positivos respectivamente es la misma. Se podría decir que si bien los círculos por debajo de la línea de ajuste son mayores, la distancia entre el valor predicho y el valor real de Y es mayor en aquellos errores positivos que negativos.

Los estimadores asociados a MCO presentan una interpretación sobre la media, hecho que puede ser problemático bajo distribuciones asimétricas de las variables aleatorias o en presencia de efectos heterogéneos. Por ejemplo,

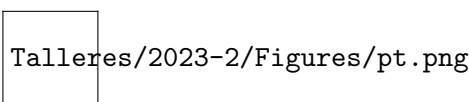
piensen que el efecto promedio de alguna intervención ofrece una medida de tendencia central, pero no indica sobre el efecto diferencial entre sus partes. Así mismo, si la distribución es asimétrica, la media de la distribución puede estar desplazada hacia alguna parte de la cola distribucional y, por tanto, el efecto estimado se refleja sobre esta cola donde no se encuentra la mayoría de la muestra. Finalmente, aunque conocemos que asintóticamente MCO no se ve alterado por heteroscedasticidad, en muestras pequeñas la heteroscedasticidad puede generar que los coeficientes de MCO no se comporten adecuadamente, hecho que comprobarán más adelante en el desarrollo de este punto.

Pensando en esto, el artículo seminal de [Koenker & Bassett \(1978\)](#) estudia una metodología denominada “Quantile Regression” la cual se basa en ajustar funciones condicionales para cada cuantil presente en los datos. Un cuantil es un valor que divide un conjunto de datos ordenados en partes iguales, de manera que cada parte contenga una fracción específica de los mismos. Por ejemplo, un percentil divide la distribución de los datos en 100 partes iguales de forma que  $\forall \tau \in \{1, 2, \dots, 99\}$  el percentil  $\tau$  es el valor por debajo del cual se encuentra el  $\tau\%$  de los datos. En este sentido, “Quantile Regression” estima para un o múltiples  $\tau$ , estimadores muestrales para el cuantil condicional del modelo de regresión lineal. En esta línea, el problema es similar al que resuelve MCO, pero se realiza un cambio en la función de pérdida que ahora se define como  $p_\tau()$  en el siguiente problema de minimización:

$$\min \frac{1}{n} \sum_{i=1}^n p_\tau(u_i) = \frac{1}{n} \sum_{i=1}^n p_\tau(y_i - \hat{y}_i) = \frac{1}{n} \sum_{i=1}^n p_\tau(y_i - \beta x_i) \quad (8)$$

donde  $p_\tau()$  es una función definida como en la figura 6:

Figura 6: Función de pérdida en QR



Como se puede observar  $p_\tau()$  es una función valor absoluto que mapea el error de predicción  $(y_i - \hat{y}_i)$  de forma ponderada según si este error es positivo o negativo para un  $\hat{\beta}$  escogido, es decir, alguna línea o hiperplano que ajuste los datos. En particular, dependiendo del cuantil  $\tau$  la función  $p_\tau()$  pondera los errores negativos  $(y_i - \hat{y}_i < 0)$  de la forma  $\tau - 1$  y los errores positivos  $(y_i - \hat{y}_i > 0)$  de la forma  $1 - \tau$ , hecho que se vuelve evidente al analizar la pendiente de la figura 6.

Como ejemplo, consideren la mediana, es decir,  $\tau = 0.5$ . Noten que para este valor  $\tau - 1 = -1/2$  y  $1 - \tau = 1/2$ , por lo que la pendiente por el dominio negativo y positivo son iguales y la función  $p_\tau()$  resulta siendo simétrica. Intuitivamente, la función pondera equitativamente los errores de predicción tanto positivos como negativos. Adicionalmente, específicamente para la mediana, la función  $p_\tau()$  es equivalente a la función de valor absoluto convencional  $(|u|)$ , por lo cual, el problema de minimización se resume en:

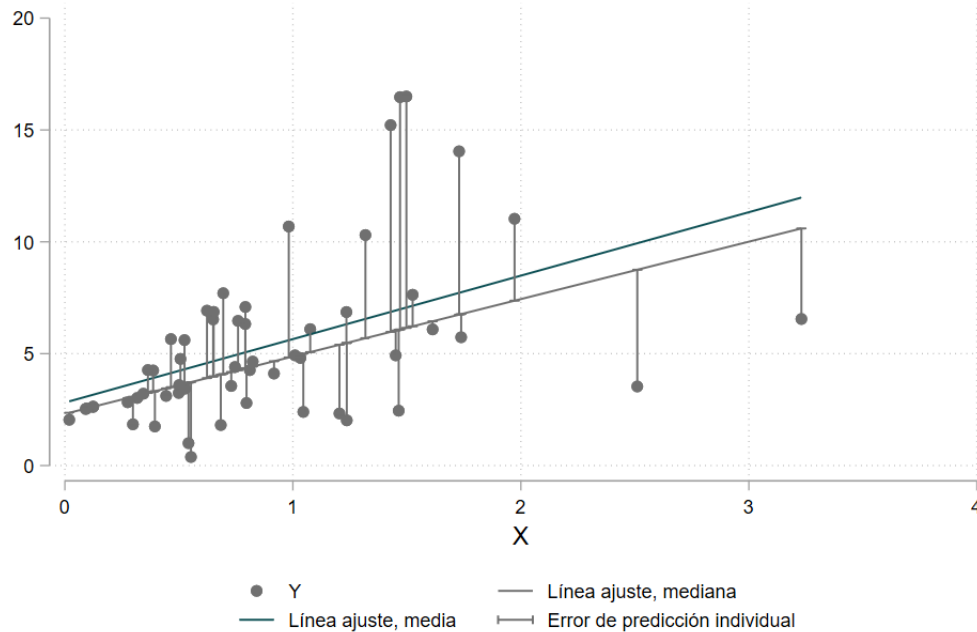
$$\min \frac{1}{n} \sum_{i=1}^n |y_i - \beta x_i| \quad (9)$$

A diferencia de la minimización con una función cuadrática, debido a las propiedades del valor absoluto -y en particular su simetría- implica que la minimización de la suma de residuos absolutos debe igualar el número de residuos positivos y negativos, asegurando así que haya el mismo número de observaciones por encima y por debajo de la recta ajustada, en particular, configurando la mediana.

2. En primera instancia, para entender un poco mejor el problema, estimen una regresión lineal entre Y y X ajustada en la mediana, para eso, revisen el comando *qreg* en Stata y respondan lo siguiente:
  - a) Realicen un gráfico de dispersión entre las variables X e Y, adiciónen al gráfico la línea de ajuste sobre la media estimada por MCO en el literal anterior y la línea de ajuste sobre la mediana. Incluyan en el gráfico líneas verticales que conecten cada valor de Y observado con el valor de Y en la línea de ajuste **de la mediana** (visualmente esta última línea corresponde al error de predicción individual).

**Solución:**

Figura 7: Gráfica de dispersión



- b) Interpreten los aspectos más relevantes de la gráfica ¿Cómo son los errores de los outliers frente a la recta ajustada en la mediana vs en el promedio? ¿Esto como se relaciona con la diferencia entre la función cuadrática y la función valor absoluto?

**Solución:** Se observa que la línea de ajuste de la mediana está por debajo de la línea de ajuste sobre la media estimada por MCO. También, los errores de los outliers se encuentran a mayor distancia de la línea de ajuste de la mediana que de la media estimada por MCO. En este sentido, los outliers afectan más a la recta de la media estimada por MCO.

La gráfica expone la diferencia entre función cuadrática y la función valor absoluto. Cuando se utiliza la función de valor absoluto convencional ( $\tau = 0.5$ ), la asimetría y otras propiedades del valor absoluto implican que haya el mismo número de observaciones por encima y por debajo de la recta ajustada, en particular, configurando la mediana.

Adicionalmente, es más fácil reconocer los outliers en la estimación que utiliza la mediana. Dado que todos los errores de predicción son ponderados de igual manera, los outliers no tienen tanto peso en la estimación (a diferencia de si fuera la estimación por MCO) y están más lejos de la recta. En la estimación por MCO, la función de pérdida utiliza la función cuadrática la cual es convexa. En este sentido, los puntos muy distantes a la recta ajustada generan errores muy grandes.

- c) ¿Cómo afectan los valores outliers la recta estimada por MCO vs por QR sobre la mediana? Si en una muestra pequeña con alta variabilidad se registrara una nueva observación ¿Cuál de las dos rectas sería más afectada?

**Solución:** La mayoría de los outliers están por encima de las dos líneas y, como mencionado anteriormente, la línea de ajuste de la mediana está por debajo de la línea de ajuste sobre la media estimada por MCO. En este sentido, se podría pensar que los datos atípicos halan la línea de ajuste de la media estimada por MCO hacia arriba. Por esta razón, se entiende que la estimación por MCO es más sensible a los outliers, incluso a pequeñas cantidades como los outliers de esta muestra. Esto resulta particularmente conflictivo en situaciones con muestras pequeñas con datos de alta variabilidad o en situaciones en las que la distribución de los datos no es normal.

Si se pensara en que en una muestra pequeña con alta variabilidad se registrara una nueva observación, la recta de la media estimada por MCO sería más afectada que la recta de la mediana. La estimación ajustada a la mediana es más robusta que la estimación por MCO, por tanto, sería la recta menos afectada.

En general, para cualquier cuantil  $\tau$  Koenker & Bassett (1978) resumen el problema en la siguiente especificación:

$$\min \frac{1}{n} \left[ \sum_{i \in (i: y_i > \beta x_i)} \tau |y_i - \beta x_i| + \sum_{i \in (i: y_i < \beta x_i)} (1 - \tau) |y_i - \beta x_i| \right] \quad (10)$$

En esta especificación se puede apreciar de una forma más intuitiva que la metodología es una minimización del error absoluto promedio (MAE), pero que le da un peso mayor a alguno de los errores, ya sean positivos o negativos, según el cuantil utilizado. Siguiendo la lógica anterior, debido a las propiedades del valor absoluto, la minimización de la ecuación 10 ocurre tomando un número de residuos positivos y negativos ponderados asimétricamente por el valor de  $\tau$ , con la particularidad de la simetría en el número de observaciones tomadas para el cálculo de la mediana.

Finalmente, como pueden observar en la ecuación 10, esta no es diferenciable en 0. Por ello, se utilizan métodos computacionales para estimar en una muestra dada los coeficientes asociados a cada cuantil ( $\beta^\tau$ ). Konker (2017) <sup>11</sup> expone la historia del proceso de computación abarcando técnicas como puntos interiores, exteriores, métodos de aproximación, entre otros (p.55). Por otra parte, para derivar la varianza y desviación estándar de los estimadores con el fin de realizar inferencia estadística se calculan a través de la metodología de bootstrap a través del remuestreo y estimación del coeficiente asociado al cuantil de interés para estas muestras.

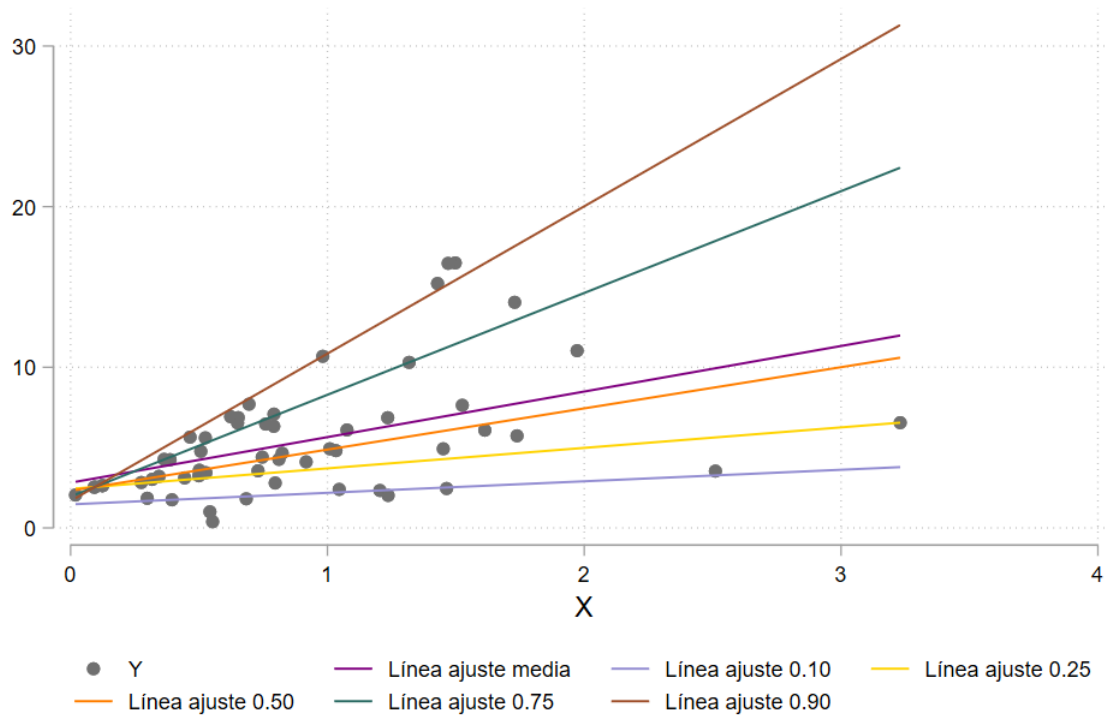
A continuación vamos a explorar las propiedades de la función valor absoluto, comprobar como este MAE ajustado estima un cuantil e identificar el número de residuales positivos y negativos. Para eso:

4. Estimen una regresión por cuantiles entre la variable Y y X para los cuantiles 0.1, 0.25, 0.5, 0.75 y 0.9 de la distribución.
- a) Realicen un gráfico de dispersión entre las variables X e Y, adicionen la línea de ajuste sobre la media estimada por MCO en el literal anterior y la línea de ajuste sobre cada uno de los cuantiles mencionados.

**Solución:**

<sup>11</sup>Handbook of Quantile Regression, doi: <https://doi.org/10.1201/9781315120256>

Figura 8: Gráfica de dispersión



b) Interpreten los aspectos más relevantes de la gráfica

**Solución:** La gráfica muestra las rectas correspondientes a cada estimación por MCO y por los cuantiles 0.1, 0.25, 0.50, 0.75 y 0.90. Se observa que a medida que aumenta el cuantil, la recta se vuelve más empinada y se mueve hacia arriba. Esto corresponde a la metodología explicada, que corresponde a una minimización del error absoluto promedio, pero que le da un peso mayor a algunos errores (positivos o negativos) dependiendo del cuantil utilizado. En los cuantiles más altos, el peso que se le otorga a los outliers que tienen mayor valor de  $Y$  predicho es mayor. Esto desplaza la recta hacia arriba, dejando a la mayoría de las observaciones por debajo de la recta. En este sentido, hay más observaciones con un error negativo.

Por otro lado, en los cuantiles más bajos, el peso que se le otorga a los outliers con menor valor de  $Y$  predicho es mayor. Por esta razón, la recta se desplaza hacia abajo y la mayoría de las observaciones quedan por encima de la recta. Esto indica que el número de observaciones con errores positivos es mayor.

c) Reporten en una tabla para cada cuantil la suma absoluta de los errores positivos, la suma absoluta de los errores negativos, el número de observaciones con errores positivos y el número de observaciones con errores negativos. Adicionen también a esta tabla la realizada en el punto 1. a) para la estimación por MCO.

**Solución:**

Tabla 3: Errores positivos y negativos

Estimación	# obs con $\epsilon > 0$	Suma $\epsilon > 0$	# obs con $\epsilon < 0$	Suma $\epsilon < 0$
MCO	22	64.521	34	-64.521
Cuantil 0.10	49	183.632	7	-2.874
Cuantil 0.25	41	117.615	15	-18.513
Cuantil 0.50	27	81.908	29	-40.147
Cuantil 0.75	14	22.947	42	-148.279
Cuantil 0.90	5	3.025	51	-253.833

- d) Interpreten la tabla y los resultados para los cuantiles, adicionalmente, comparenlo con lo encontrado por MCO.

**Solución:** La tabla permite observar que a medida que nos aproximamos a los cuantiles más altos, el número de observaciones con errores negativos aumenta significativamente, mientras que el número de observaciones con errores positivos disminuye. Es decir, la estimación del cuantil 0.10 tiene 49 observaciones con errores positivos y 7 observaciones con errores negativos. Mientras, la estimación del cuantil 0.90 tiene 5 observaciones con errores positivos y 51 observaciones con errores negativos. Esto indica que en los cuantiles superiores, hay más observaciones con errores negativos, lo que sugiere que en estos puntos de la distribución de ingresos, la estimación tiene una tendencia a subestimar los valores reales.

De la misma manera, a medida que el cuantil aumenta, la suma de errores positivos disminuye y la suma de errores negativos crece. En la estimación del cuantil 0.10, la suma de errores positivos es de 183.632 y la suma de errores negativos es de -2.874. En la estimación del cuantil 0.90, la suma de errores positivos es de 3.025 y la suma de errores negativos es de -253.833.

También se observa que, como esperado, en el cuantil 0.50, el número de observaciones negativas y positivas es casi igual, 29 y 27 respectivamente. La suma de los errores positivos y negativos es de 81.908 y -40.147, respectivamente. Si bien es una diferencia notoria, es la menor diferencia entre todos los cuantiles.

Finalmente, tras entender la teoría detrás de “Quantile Regression” (QR) es momento de aplicarlo en un caso práctico. En particular, ustedes están interesados en identificar la relación que hay entre la educación y el salario. Más aún, están interesados en explorar también si para los trabajadores con mayores ingresos un año de educación adicional influye más en el salario que para los de bajos ingresos. Para esto, Deirdre McCloskey les dice que la metodología de QR se ajusta perfectamente a lo que quieren comprobar y deciden aplicarla. Concretamente, tienen los datos del artículo de Card (1992) titulado “Earning, Schooling and Ability Revisited” en una base de datos llamada “Card.dta”. En esta cuentan con información del salario por hora de una muestra de trabajadores en 1976 y los años de educación para 1976, además de otras variables sociodemográficas que no son directamente relevantes para el estudio. Luego de analizar el problema se encuentran interesados en la siguiente ecuación:

$$\ln(Wage)_i = \beta_0 + \beta_1 * educ_i + e_i \quad (11)$$

donde  $\ln(Wage)_i$  es el logaritmo natural del salario en horas de 1976 para el individuo  $i$ ,  $educ_i$  es el número de años de educación del mismo y  $e_i$  es un error estocástico a nivel individual.

5. Respondan:

- a). Enuncien dos ventajas que podrían evidenciar de utilizar la metodología de QR en este contexto.

**Solución:** Utilizar la metodología de QR en este contexto tiene dos ventajas: mayor robustez ante posibles datos atípicos y posibilidad de captura de efectos heterogéneos. Primero, y como se ha discutido en los incisos anteriores, la estimación por cuantiles es menos sensible a los outliers en una muestra, en comparación con la regresión por MCO. Es posible que los salarios, a nivel poblacional, presenten una gran variación, pues esta variable puede variar dependiendo de factores como edad, carrera estudiada, horas trabajadas, etc. Además, si los salarios son autoreportados, puede que exista algún tipo de sesgo. La estimación por cuantiles permite generar resultados robustos y poco sensibles, en los cuales los puntos muy distantes a la estimación ajustada generan errores no tan grandes.

Por otro lado, la QR permite analizar la variación en los retornos para trabajadores con el mismo nivel educativo. Se podría demostrar que la educación no tiene un efecto ‘único’ en los salarios, sino que un efecto de un año adicional de educación es diferente para trabajadores con salarios bajos en comparación con trabajadores con salarios altos. A su vez, este tipo de análisis conlleva a otros: por ejemplo, si existe heterogeneidad significativa en los retornos a lo largo de la distribución de ingresos, con retornos más altos para aquellos con niveles de ingresos más altos (que

se asume indican individuos con alta capacidad), la inversión en educación podría producir más desigualdad. Capturar la heterogeneidad en la relación entre educación y salarios es una ventaja importante en comparación con modelos de regresión lineal tradicionales, que no permite este tipo de análisis.

b). ¿Cómo se interpreta el parámetro  $\beta_1$  en la ecuación 11?

**Solución:** El parámetro  $\beta_1$  en la ecuación 11 es la relación entre el número de años de educación y el salario en términos logarítmicos. Es la semi-elasticidad del salario con respecto a la educación. Específicamente, se debe multiplicar por 100 el resultado del  $\beta_1$ . Entonces, por ejemplo, si la regresión se hizo para el cuantil 0.3 y el resultado del  $\beta_1$  es igual a 0.1, esto significa que un año adicional de educación se asocia con un aumento del 10 % en el salario para aquellos individuos que tienen un salario por encima del percentil 30 del resto de la población, manteniendo todo lo demás constante.

c). ¿Qué supuesto(s) consideran necesario(s) para que exista una identificación limpia del efecto causal de interés a través de la metodología de QR? ¿En este contexto creen que se cumplirían?

**Solución:** Para que este estimador sea consistente y exista una identificación limpia del efecto causal de interés a través de la metodología de QR, se deben cumplir tres condiciones, las mismas que en una estimación por MCO:

- 1 **Linealidad en los parámetros:** El modelo de regresión es lineal en los parámetros, aunque puede ser o no lineal en las variables.  $y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{ik}\beta_k + \epsilon_i$

Es probable que este supuesto se cumpla dado que los parámetros de la regresión son lineales.  $\beta_0$  y  $\beta_1$  están multiplicando las variables de manera lineal. La relación entre las variables independientes y dependientes es logarítmica, lo que implica que el cambio en el salario no es constante, sino proporcional al nivel de educación. De igual manera, se cumple el supuesto.

- 2 **Exogeneidad de las variables independientes:** Este supuesto de identificación indica que el valor esperado condicional del error no es una función de las variables independientes de ninguna observación. Este supuesto permite establecer causalidad en el efecto estimado.

Es posible que la educación esté correlacionada con el error de la regresión, pues existen algunas variables que podrían estar contenidas en el error y no se están teniendo en cuenta en la regresión, por ejemplo: la experiencia laboral previa, las habilidades innatas, las decisiones de carrera, la salud de cada individuo, etc.

- 3 **X es estocástica y de rango completo:** Este supuesto de identificación implica que el conjunto de variables utilizadas son aleatorias ordenadas. Además, que la matriz de estas variables sea invertible.

Si hay una variabilidad sustancial en las variables relacionadas con la educación y el salario entre las personas, es plausible que se cumpla el supuesto de estocasticidad y rango completo. Se podría asumir que la educación, la experiencia laboral o habilidades específicas y otras variables relevantes varían entre las personas. También, se podría pensar que la educación formal, la experiencia laboral, las habilidades técnicas e interpersonales entre otras podrían ser variables independientes, que no están completamente correlacionadas entre sí.

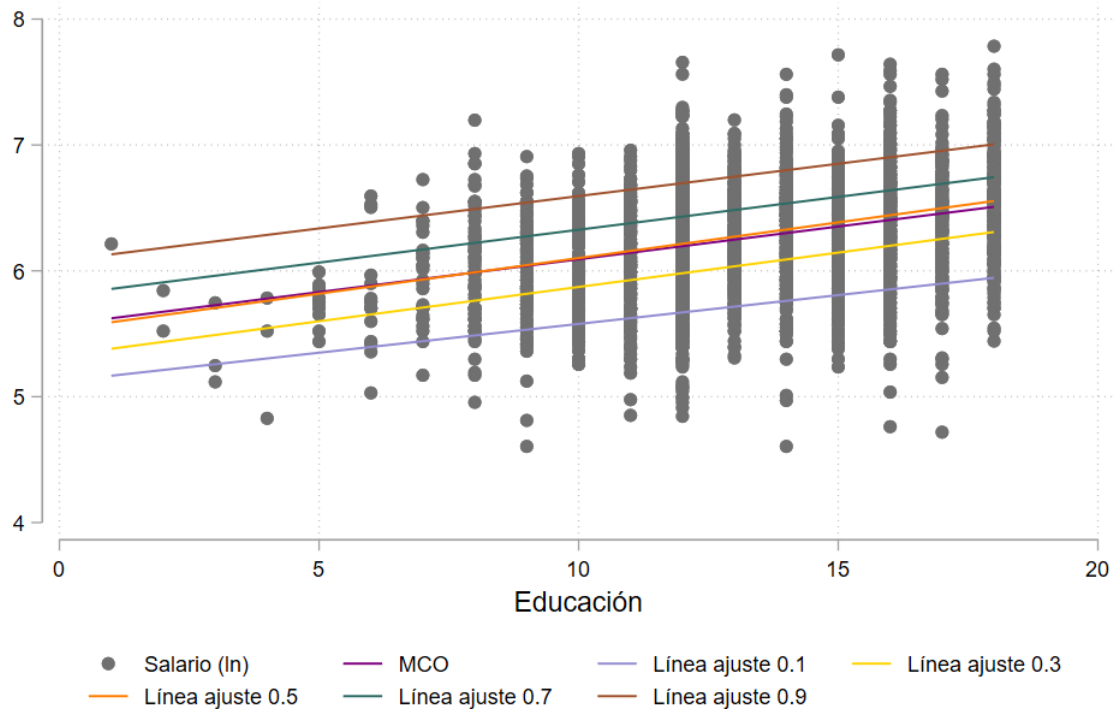
6. Suponga que los supuestos mencionados anteriormente se cumplen. Estimen una regresión lineal entre  $\ln(Wage)_i$  y  $educ_i$  por MCO y por QR para los deciles (cuantiles) del 0.1 al 0.9.

a) Presenten un gráfico que reúna un scatter entre el logaritmo natural del salario y los años de educación. Añadan las líneas de las regresiones estimadas por MCO, y para los cuantiles 0.1 0.3, 0.5, 0.7 y 0.9.



**Solución:**

Figura 9: Gráfica de dispersión



- b) Presenten en una tabla la estimación para la especificación por MCO y para los cuantiles 0.1, 0.3, 0.5, 0.7 y 0.9.

**Solución:**

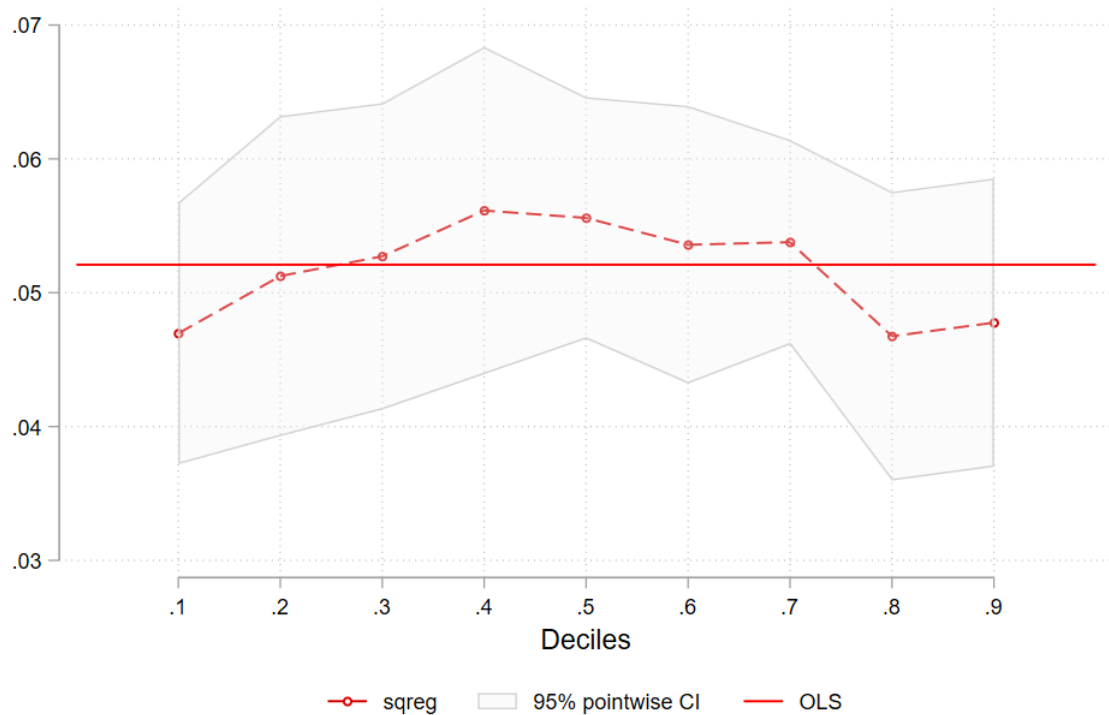
Tabla 4: Estimación por MCO y QR

	Salario (logaritmo natural)					
	(1) MCO	(2) Decil 0.1	(3) Decil 0.3	(4) Decil 0.5	(5) Decil 0.7	(6) Decil 0.9
Educación	0.052*** [0.003]	0.047*** [0.006]	0.053*** [0.004]	0.056*** [0.004]	0.054*** [0.003]	0.048*** [0.005]
Constante	5.571*** [0.039]	5.086*** [0.079]	5.355*** [0.055]	5.548*** [0.052]	5.784*** [0.045]	6.132*** [0.064]
N	3010	3010	3010	3010	3010	3010

- c) Presenten una gráfica donde en el eje X esté cada uno de los deciles desde el 0.1 al 0.9 y en el eje Y esté el coeficiente estimado para el respectivo decil. Grafiquen también los intervalos de confianza del 95 % para cada coeficiente, añadan también una comparación con MCO

**Solución:**

Figura 10: Coeficientes e intervalos de confianza de deciles



- d) Interpreten los resultados de las gráficas y tabla anterior, comparen la estimación por cuantiles con MCO y los resultados al interior del cuantil. Asegúrense de incluir la interpretación de los coeficientes, su tendencia, la significancia económica y estadística.

**Nota:** Revisen con cuidado la interpretación de los coeficientes de los cuantiles.

#### Solución:

Los resultados de este inciso confirman una y otra vez que los años de educación efectivamente tienen un efecto positivo sobre los salarios de los individuos. En la Figura 9, observamos que las pendientes de todas las estimaciones tienen una tendencia similar y son bastante cercanas. Esto se refleja en la Tabla 4, pues vemos que todos los coeficientes estimados se encuentran entre 0.047 y 0.056. Como es de esperar, la línea de ajuste de la mediana (decil 0.5) se aproxima a la línea de ajuste de la media (MCO). Finalmente, cabe resaltar de la figura que ocurre lo mismo que en los incisos anteriores. A medida que aumenta el decil, aumenta el número de observaciones con errores negativos y disminuye el número de observaciones con errores positivos.

Los resultados de la Tabla 4 se resumen de la siguiente manera:

- Mediante la estimación por MCO, se tiene que el efecto de un año adicional de educación aumenta el salario del individuo promedio en aproximadamente 5.2 %, *ceteris paribus*.
- Mediante la estimación por la metodología QR, se tiene que el efecto de un año adicional de educación aumenta el salario del 10 % más bajo de la distribución salarial en aproximadamente 4.7 %, *ceteris paribus*.
- Mediante la estimación por la metodología QR, se tiene que el efecto de un año adicional de educación aumenta el salario del 30 % más bajo de la distribución salarial en aproximadamente 5.3 %, *ceteris paribus*.
- Mediante la estimación por la metodología QR, se tiene que el efecto de un año adicional de educación aumenta el salario del 50 % más bajo de la distribución salarial en aproximadamente 5.6 %, *ceteris paribus*.
- Mediante la estimación por la metodología QR, se tiene que el efecto de un año adicional de

educación aumenta el salario del 70 % más bajo de la distribución salarial en aproximadamente 5.4 %, *ceteris paribus*.

- Mediante la estimación por la metodología QR, se tiene que el efecto de un año adicional de educación aumenta el salario del 10 % más alto de la distribución salarial en aproximadamente 4.8 %, *ceteris paribus*.

Sabemos que estos resultados son todos significativos al 1 %. Confirmamos esto a través de la Figura 10, pues observamos que ningún intervalo de confianza incluye 0. También, vemos reflejada la tendencia de los coeficientes, la cual se asemeja a una función concava. Esto se puede entender como que una persona con un salario bajo que decide invertir en su educación, no verá un aumento considerable en su salario. De manera de ejemplo, una persona con título de bachiller que trabaja en el sector informal decide hacer un técnico. Esta inversión no se verá reflejada en su salario.

Por otro lado, una persona que ya tiene altos ingresos no verá un aumento significativo en su salario si invierte en un año adicional de educación. Por ejemplo, una persona con doctorado decide hacer un post doctorado. Esta inversión no se verá reflejada en su salario.

7. Finalmente, les interesa evaluar si el efecto de un año adicional para los que tienen mayores ingresos es mayor que para los que menores ingresos tienen. Para esto, desarrollen un procedimiento de prueba de hipótesis y aplíquelo para obtener una respuesta. Para facilitar su procedimiento supongan que no hay autocorrelación serial.

- a) Planteen un estadístico de prueba, un nivel de significancia y una región de rechazo.

**Solución:** La hipótesis nula establece que no hay diferencias en el efecto de un año adicional para los que tienen mayores ingresos que para los que tienen menores ingresos. La hipótesis alternativa establece que el efecto de un año adicional de educación es mayor para los que tienen mayores ingresos que para los que tienen menores ingresos.

$$H_0 : \beta_1^{(0.9)} - \beta_1^{(0.1)} = 0$$

$$H_0 : \beta_1^{(0.9)} - \beta_1^{(0.1)} > 0$$

El estadístico de prueba sería el siguiente:

$$\tau = \frac{\beta_1^{(0.9)} - \beta_1^{(0.1)}}{\sqrt{\text{var}(\beta_1^{(0.9)} - \beta_1^{(0.1)})}}$$

Finalmente, comparamos el p-valor de este estadístico de prueba con los niveles de significancia habituales: al 10 %, al 5 % o al 1 %. Luego, rechazamos o no rechazamos la hipótesis nula y concluimos.

- b) Revisen la documentación de `qreg`, `sqreg` e `iqreg` en Stata y apliquen la prueba de hipótesis que sugirieron en el inciso anterior.

**Solución:**

Tabla 5: Aplicación de prueba de hipótesis

	Valor F	P-valor
Prueba de hipotesis	0.012	0.912

- c) Interpreten sus resultados ¿Hay desigualdad en este aspecto?

### Solución:

Con un valor F de 0.012 y un p-valor de 0.912 no tenemos suficiente evidencia para rechazar la hipótesis nula. En este sentido, no existen diferencias en el efecto de un año adicional para los que tienen mayores ingresos que para los que tienen menores ingresos. En este sentido, no hay desigualdad en el retorno que tiene la educación en los salarios de los individuos.

Como mencionado anteriormente, esto se puede entender como que una persona con un salario bajo que decide invertir en su educación, no verá un aumento considerable en su salario. (una persona con título de bachiller que trabaja en el sector informal decide hacer un técnico). También, se entiende como que una persona que ya tiene altos ingresos no verá un aumento significativo en su salario si invierte en un año adicional de educación. (una persona con doctorado decide hacer un post doctorado). ninguna de estas inversiones se verán reflejadas en sus salarios.

## Anexo

```
1
2      *Taller 3
3
4  clear all
5  cap log close
6  set more off
7  cls
8
9
10 /*=====
11 1: Main Paths
12 =====*/
13
14 **# Mariana Bonet -----
15 if "`c(username)'"=="maria" {
16     global git      "C:\Users\maria\OneDrive - Universidad de los andes\MECA\Econometria Avanzada\
17     econavanzadatalles"
18 }
19
20 **# Andres Molano -----
21 if "`c(username)'"=="andresmolano" {
22     global git      "/Users/andresmolano/Desktop/econavanzadatalles"
23 }
24
25 /*=====
26 2: Secondary Paths
27 =====*/
28 * Output
29 global taller      "${git}/3"
30 global do           "${taller}/do"
31 global output      "${taller}/output"
32 global data         "${taller}/data"
33
34 /*=====
35 3: Desarrollo Punto 2
36 =====*/
37
38 *Cargar datos
39
40 use "${data}/P3/smoke_day_v3.dta", clear
41
42 *****
43 *****PUNTO B*****
44 *****
45
46 *) Ordenamos la base de acuerdo a la variable panel (countyfip) y a la variable de tiempo (date)
47 order countyfip date
48
49 *) Declaramos la base para ser panel
50 xtset countyfip date, d
51
52 *) Generamos una variable que nos indique la fecha en la que hubo cada incendio
53 gen first_fire_date = day if smoke_day==1
54
55 *) A partir de lo anterior, generamos una variable que nos indique cuando fue el primer incendio
56 egen first_fire = min(first_fire_date), by(countyfip)
```

```

57 replace first_fire = 0 if first_fire==.
58
59 *) Generamos variable de tratamiento
60 gen Dit = (day >= first_fire & first_fire!=0)
61
62 *) Creamos una variable que nos indique el tiempo relativo
63 by countyfip: gen rel_time=day-first_fire
64 replace rel_time=. if first_fire==0
65
66
67 *) Generamos dummies para cada categoria posible de tiempos relativos
68 tab rel_time, gen(evt)
69
70 *) Generamos los Leads
71
72 forvalues x = 1/80 {
73     local j= 81-'x'
74     ren evt'x' evt_l'j'
75     cap label var evt_l'j' "-'j'"
76 }
77
78
79
80 *) Generamos los Lags
81
82 forvalues x = 0/82 {
83     local j= 81+'x'
84     ren evt'j' evt_f'x'
85     cap label var evt_f'x' "'x'"
86 }
87
88
89 *) Periodo base (t=-1) - Omitir por multicolinealidad
90 replace evt_l1=0
91
92 *****
93 *****PUNTO D*****
94 *****
95
96 *) Vamos a generar la variable que indique el dia de la semana
97
98 gen weekday = dow(date)
99 label variable weekday "0=Sunday,...,6=Saturday"
100
101 *) Estimamos por MCO el modelo de TWFE con todos los leads y lags
102 reghdfe pm25 Dit, nocon absorb(countyfip day) cluster(countyfip)
103
104 *) Realizamos la tabla
105 eststo clear
106 eststo: reghdfe pm25 Dit, nocon absorb(countyfip day) cluster(countyfip)
107 estadd local FE_day = "S ", replace
108 estadd local FE_city = "S ", replace
109 estadd local FE_weekday = "S ", replace
110
111 esttab using tabla2d.tex, replace ///
112 se(%9.3f) b(%9.3f) ///
113 star(* 0.10 ** 0.05 *** 0.01) ///
114 stats(N r2 FE_day FE_city FE_weekday, labels("Observaciones" "R-cuadrado" "Efectos fijos por d a"
115     "Efectos fijos por municipio" "Efectos fijos por d a de la semana") fmt(%9.0fc %9.3fc %9.0fc
116     %9.0f %9.0f)) wrap ///
117 nonumbers nogaps label ///
118 nonotes ///
119 addnotes("\sym{*} \(<0.1\), \sym{**} \(<0.05\), \sym{***} \(<0.01\). Errores est ndares
120     clusterizados a nivel municipio." "Efectos fijos: D a (day), Municipio (countyfip), D a de
121     la semana (weekday).")
122
123 *) Estimamos por MCO el modelo de TWFE para ver el efecto 10 dias antes y 10 dias despues
124 reghdfe pm25 evt_l10 evt_l9 evt_l8 evt_l7 evt_l6 evt_l5 evt_l4 evt_l3 evt_l2 evt_l1 evt_f1
125     evt_f2 evt_f3 evt_f4 evt_f5 evt_f6 evt_f7 evt_f8 evt_f9 evt_f10, nocon absorb(day countyfip
126     weekday) cluster(countyfip)
127
128 estimates store coefs_i
129
130 *) Graficamos
131 coefplot coefs_i, omitted ///
132     vertical ///
133     label drop(_cons) ///
134     yline(0, lpattern(dash) lwidth(*0.5)) ///
135     ytitle("Concentraci n de pm2.5 (ug/m3)") ///
136     xtitle("Tiempo relativo al tratamiento", size(medsmall)) ///
137     xlabel(, labsize(small) nogextend labc(black)) ///

```

```

132 ylabel(,nogrid nogextend labc(black) format(%9.2f)) ///
133 msymbol(0) ///
134 mlcolor(black) ///
135 mfcolor(black) ///
136 msize(vsmall) ///
137 levels(95) ///
138 xline(11, lpattern(dash) lwidth(*0.5)) ///
139 ciopts(lcol(black) recast(rcap) lwidth(*0.8)) ///
140 plotregion(lcolor(black) fcolor(white)) ///
141 graphregion(lcolor(black) fcolor(white)) ///
142 yscale(lc(black)) ///
143 xscale(lc(black)) ///
144 name(Punto_2d, replace)
145 graph export Punto_2d.png, replace
146
147 *****
148 *****PUNTO F*****
149 *****
150
151 *) Para tener solo 10 dias antes y 10 dias despues, eliminamos las observaciones que estan en los
    demas periodos
152 gen rel_time_1 = cond(rel_time >= -10 & rel_time <= 10, rel_time, .)
153 drop if rel_time_1==. & first_fire!=0
154
155 *) Tambi n eliminamos las variables que no ser n necesarias de estos periodos
156 forvalues x=11/80{
157     drop evt_l`x'
158 }
159 forvalues x=11/82{
160     drop evt_f`x'
161 }
162
163 *) Estimacion a partir de Callaway & Sant'Anna (2020)
164 *net install csdid, from ("https://raw.githubusercontent.com/friosavila/csdid_drdid/main/code/")
    replace
165
166 csdid pm25, ivar(countyfip) time(day) gvar(first_fire)
167
168 *) Graficamos los efectos dinamicos del tratamiento
169 estat event
170 csdid_plot
171 graph export Punto_2f.png, replace
172
173 *****
174 *****PUNTO G*****
175 *****
176
177 *) Renombramos el a o de referencia
178 rename evt_l1 ref
179
180 *) Generamos una variable para el grupo de los nunca tratados
181 gen never=(first_fire==0)
182
183 *) Creamos los globals importantes para hacer el grafico segun la complementaria
184 global post 10 /* N mero de periodos post sin contar el 0*/
185 global pre 10 /* N mero de periodos pre*/
186 global ep event_plot
187 global g0 "default_look"
188 global g1 xla(-$pre (1) $post)
189 global g2 xt("Periodos relativos al evento")
190 global g3 yt("Efecto causal")
191 global g $g1 $g2 $g3
192 global t "together"
193
194
195 *) Estimaci n por TWFE
196 reghdfe pm25 evt_l10 evt_l9 evt_l8 evt_l7 evt_l6 evt_l5 evt_l4 evt_l3 evt_l2 ref evt_f0 evt_f1
    evt_f2 evt_f3 evt_f4 evt_f5 evt_f6 evt_f7 evt_f8 evt_f9 evt_f10, nocon absorb(day countyfip
    weekday) cluster(countyfip)
197 estimates store ols // guardando los estimadores para graficar
198 $ep ols, stub_lag(evt_f#) stub_lead(evt_l#) $t $g0 graph_opt($g ti("OLS") name(gOLS, replace))
199
200 *) Estimaci n con csdid de Callaway & Sant'Anna (2020)
201 csdid pm25, ivar(countyfip) time(day) gvar(first_fire)
202 estat event, estore(cs)
203 $ep cs, stub_lag(Tp#) stub_lead(Tm#) $t $g0 graph_opt($g ti("CS 20") name(gCS, replace))
204
205
206 *) Estimaci n con eventstudyinteract de Sun & Abraham (2020)
207 eventstudyinteract pm25 evt_l* evt_f*, ///
208 absorb(countyfip) cohort(first_fire) ///

```

```

209     control_cohort(never) vce(cluster i.countyfip)
210 $ep e(b_iw)#e(V_iw), stub_lag(evt_f#) stub_lead(evt_l#) $t $g0 graph_opt($g ti("SA 20") name(gSA,
    replace))
211 matrix sa_b = e(b_iw) // guardando los estimadores para graficar
212 matrix sa_v = e(V_iw)
213
214
215 *) Estimaci n con did_imputation de Borusyak et al. (2021)
216 gen first_fire_miss = cond(first_fire==0, ., first_fire)
217
218 did_imputation pm25 countyfip day first_fire_miss, horizons(0/$post) autosample pretrend($pre)
    minn(0) delta(1)
219 estimates store bjs // guardando los estimadores para graficar
220 $ep bjs, $t $g0 graph_opt($g ti("BJS 21") name(gBJS, replace))
221
222
223 *) Combinamos los plots almacenados
224 event_plot ///
225 bjs cs sa_b#sa_v ols, ///
226     stub_lag(tau# Tp# evt_f# evt_f#) ///
227     stub_lead(pre# Tm# evt_l# evt_l#) ///
228     plottype(scatter) ciplottype(rcap) ///
229     together perturb(-0.325(0.1)0.325) trimlead(5) noautolegend ///
230     graph_opt(xtitle("Periodos relativos al evento", size(med)) ///
231     ytitle("Efecto causal promedio estimado", size(med)) xlabel(-$pre(1)$post) ///
232     legend(order(1 "BJS" 3 "CS" 5 "SA" 7 "TWFE") rows(1) position(6) region(style(none))) ///
233     xline(-0.5, lcolor(gs10) lpattern(dash)) yline(0, lcolor(gs10)) ///
234     graphregion(color(white)) bgcolor(white) ylabel(, angle(horizontal)) ///
235     ) ///
236     lag_opt1(msymbol(0) color(maroon)) lag_ci_opt1(color(maroon)) ///
237     lag_opt2(msymbol(Dh) color(emerald)) lag_ci_opt2(color(emerald)) ///
238     lag_opt3(msymbol(Th) color(navy)) lag_ci_opt3(color(navy)) ///
239     lag_opt4(msymbol(Sh) color(lavender)) lag_ci_opt4(color(lavender))
240 graph export Punto_2g.png, replace
241
242 /*****
243 5: Desarrollo Punto 3
244 *****/
245 clear all
246 use "${data}/P2/sim.dta", clear
247
248 ***1a. Tabla de suma de errores
249
250 reg Y X
251 predict residual, res
252 predict Y_hat, xb
253
254 gen var_errors= cond(residual>0, 1, 0)
255
256 egen sumapos = sum(residual) if residual>0
257 egen sumaneg = sum(residual) if residual<0
258 sum sumapos sumaneg
259
260 eststo total: estpost summ sumapos sumaneg
261
262 ***1b. Grafica dispersion
263 twoway (scatter Y X, msymbol(0) mlcolor(stc15) mfcolor(stc15) xlabel(0(1)4) xscale(range(0 4))) (lfit
    Y X, lcolor(stc15) legend(label (2 "L nea ajuste, media") pos(6) col(3))) (rcap Y Y_hat
    X, lcolor(stc15) legend(label (3 "Error de predicci n individual") pos(6)))
264 gr export "${output}/grafica1b.png", replace as(png)
265
266 ***2 Estimacion
267 qreg Y X
268 predict Y_hatb, xb
269 predict residualm, res
270
271 ***2a. Grafica dispersion
272 twoway (scatter Y X, msymbol(0) mlcolor(gs7) mfcolor(gs7) xlabel(0(1)4) xscale(range(0 4))) (lfit
    Y_hatb X, lcolor(gs7) legend(label (2 "L nea ajuste, mediana") pos(6) col(2))) (lfit Y X,
    lcolor(stc15) legend(label (3 "L nea ajuste, media") pos(6))) (rcap Y Y_hatb X, lcolor(gs7)
    legend(label (4 "Error de predicci n individual") pos(6)))
273 gr export "${output}/grafica2a.png", replace as(png)
274
275 ***4 Estimacion 0.1, 0.25, 0.5, 0.75 y 0.9
276 qreg Y X, quantile(0.1)
277 predict residual10, res
278 predict Y_hat_10, xb
279
280 qreg Y X, quantile(0.25)
281 predict residual25, res
282 predict Y_hat_25, xb

```

```

283 qreg Y X, quantile(0.5)
284 predict residual50, res
285 predict Y_hat_50, xb
286
287
288 qreg Y X, quantile(0.75)
289 predict residual75, res
290 predict Y_hat_75, xb
291
292 qreg Y X, quantile(0.9)
293 predict residual90, res
294 predict Y_hat_90, xb
295
296 ***4a Grafico con estimaciones anteriores 0.1, 0.25, 0.5, 0.75 y 0.9
297 twoway (scatter Y X, msymbol(0) mlcolor(gs7) mfcolor(gs7) xlabel(0(1)4) xscale(range(0 4))) (lfit
    Y X, lcolor(purple) legend(label(2 "L nea ajuste media") pos(6) col(4))) (lfit Y_hat_10 X,
    lcolor(lavender) legend(label(3 "L nea ajuste 0.10"))) (lfit Y_hat_25 X, lcolor(styellow)
    legend(label(4 "L nea ajuste 0.25"))) (lfit Y_hat_50 X, lcolor(orange) legend(label(5 "L nea
    ajuste 0.50"))) (lfit Y_hat_75 X, lcolor(emerald) legend(label(6 "L nea ajuste 0.75"))) (
    lfit Y_hat_90 X, lcolor(sienna) legend(label(7 "L nea ajuste 0.90")))
298 gr export "${output}/grafica4a.png", replace as(png)
299
300 ***4c Tabla de errores
301
302 preserve
303
304 keep residual*
305
306 global res residual residualm residual10 residual25 residual50 residual75 residual90
307
308 foreach var in $res {
309
310     gen errors`var' = cond(`var'>0, 1, 0)
311     sum errors`var'
312     dis "'`var' count positive errors:"
313     count if errors`var'==1
314     dis "'`var' count negative errors: "
315     count if errors`var'==0
316
317     egen sumapos`var' = sum(`var') if `var'>0
318     sum sumapos`var'
319     local pos = r(mean)
320     dis "'`var' sum of positive errors: 'pos'"
321
322     egen sumaneg`var' = sum(`var') if `var'<0
323     sum sumaneg`var'
324     local neg = r(mean)
325     dis "'`var' sum of negative errors: 'neg'"
326 }
327
328 clear
329 input Estimaci n ObsPos SumaPos ObsNeg SumaNeg
330 Estimaci n ObsPos SumaPos ObsNeg SumaNeg
331 1 22 64.521 34 -64.521
332 2 27 81.908 29 -40.147
333 10 49 183.632 7 -2.874
334 25 41 117.615 15 -18.513
335 50 27 81.908 29 -40.147
336 75 14 22.947 42 -148.279
337 90 5 3.025 51 -253.833
338     end
339
340 mkmat Estimaci n ObsPos SumaPos ObsNeg SumaNeg, matrix(Tablaerrores)
341
342 esttab matrix(Tablaerrores) using "${output}/Tablaerrores4.tex", replace
343
344 ***6 Estimacion reg lineal lnwage y educ por MCO y QR [0.1, 0.9]
345 clear all
346 use "${data}/P2/Card.dta", clear
347
348 reg lwage i.educ
349 predict residual, res
350 predict Y_hat, xb
351
352 qreg lwage i.educ, quantile(0.1)
353 predict residual1, res
354 predict Y_hat_1, xb
355
356 qreg lwage i.educ, quantile(0.2)
357 predict residual2, res
358 predict Y_hat_2, xb

```



```

359
360 qreg lwage i.educ, quantile(0.3)
361 predict residual3, res
362 predict Y_hat_3, xb
363
364 qreg lwage i.educ, quantile(0.4)
365 predict residual4, res
366 predict Y_hat_4, xb
367
368 qreg lwage i.educ, quantile(0.5)
369 predict residual5, res
370 predict Y_hat_5, xb
371
372 qreg lwage i.educ, quantile(0.6)
373 predict residual6, res
374 predict Y_hat_6, xb
375
376 qreg lwage i.educ, quantile(0.7)
377 predict residual7, res
378 predict Y_hat_7, xb
379
380 qreg lwage i.educ, quantile(0.8)
381 predict residual8, res
382 predict Y_hat_8, xb
383
384 qreg lwage i.educ, quantile(0.9)
385 predict residual9, res
386 predict Y_hat_9, xb
387
388 ***6a Grafico regresiones estimadas por MCO, y para los cuantiles 0.1 0.3, 0.5, 0.7 y 0.9.
389 twoway (scatter lwage educ, msymbol(0) mlcolor(gs7) mfcolor(gs7) xtitle("Educaci n") xlabel(0(5)
      20) xscale(range(0 20)) legend(label(1 "Salario (ln)") (lfit lwage educ, lcolor(purple)
      legend(label(2 "MCO") pos(6) col(4))) (lfit Y_hat_1 educ, lcolor(lavender) legend(label(3 "
      L nea ajuste 0.1"))) (lfit Y_hat_3 educ, lcolor(styellow) legend(label(4 "L nea ajuste 0.3")
      )) (lfit Y_hat_5 educ, lcolor(orange) legend(label(5 "L nea ajuste 0.5"))) (lfit Y_hat_7 educ
      , lcolor(emerald) legend(label(6 "L nea ajuste 0.7"))) (lfit Y_hat_9 educ, lcolor(sienna)
      legend(label(7 "L nea ajuste 0.9"))))
390 gr export "${output}/grafica6a.png", replace as(png)
391
392
393 ***6b regresiones MCO y para los cuantiles 0.1, 0.3, 0.5, 0.7 y 0.9.
394 eststo MCO: reg lwage educ
395 eststo q1: qreg lwage educ, quantile(0.1)
396 eststo q2: qreg lwage educ, quantile(0.2)
397 eststo q3: qreg lwage educ, quantile(0.3)
398 eststo q4: qreg lwage educ, quantile(0.4)
399 eststo q5: qreg lwage educ, quantile(0.5)
400 eststo q6: qreg lwage educ, quantile(0.6)
401 eststo q7: qreg lwage educ, quantile(0.7)
402 eststo q8: qreg lwage educ, quantile(0.8)
403 eststo q9: qreg lwage educ, quantile(0.9)
404
405 #d;
406             esttab MCO q1 q3 q5 q7 q9
407                 using "${output}/results6b.tex", replace
408                 style(tex) title(" ")
409                 cells("b(fmt(3) star)" se(par([ ])fmt(3)))
410                 starlevels(* 0.1 ** 0.05 *** 0.01)
411             ;
412 #d cr
413
414 ***6c grafica deciles, coef estimado e intervalos de confianza
415
416 eststo educ: sqreg lwage educ, quantiles(0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9)
417 estat coefplot educ, title("") xlabel(0.1(0.1)0.9) xscale(range(0.1 0.9)) ytitle("") xtitle("
      Deciles") legend(pos(6) col(3) label(1 "Regresi n" 2 "Intervalo de confianza 95%"))
418
419 using "${output}/grafica6c.png", replace as(png)
420
421 ***7b prueba de hipotesis
422 set seed 1001
423 sqreg lwage educ, quantile(.1 .9)
424 eststo prueba: test [q10]educ = [q90]educ
425
426 mat MAT = J(1,2,..)
427 mat MAT[1,1]=r(F)
428 mat MAT[1,2]=r(p)
429 mat list MAT
430 #d;
431             esttab matrix(MAT,fmt(3)) using "${output}/tabla7.tex",
432             cell("c1 c2") noobs nonumber nomtitles

```

```
433 collabels("Valor F" "P-valor") varlabels(r1 "Prueba de hipotesis")
434 label replace;
435 #d cr
```