
Primer ejercicio

La planeación familiar ha sido un tema de interés recurrente en salud pública, especialmente en las zonas rurales de Colombia. Según recientes reportes¹, pese a que el 80 % de mujeres en estas regiones informan que desea retrasar o prevenir un embarazo, sólo el 30 % de ellas usan algún tipo de método anticonceptivos. Por esta razón, y para comprender la persistencia de la brecha entre los deseos y el control de fecundidad, el Instituto de Salud Sexual y Reproductiva (ISSR) realizó un experimento aleatorio con el fin de entender el posible efecto que tiene dar información acerca de planificación familiar en el comportamiento de fertilidad de las parejas.

En esta ocasión, la muestra de estudio incluye 50,000 hogares, seleccionados al azar, y provenientes de zonas rurales de 300 municipios. Dentro de los 300 municipios encuestados, se seleccionaron 200 municipios de manera aleatoria para ser candidatos a tratamiento. No obstante, solo una fracción de los hogares de estos lugares fueron seleccionados para ser tratados mediante una lotería justa. Por su parte, todos los hogares de los otros 100 municipios fueron asignados al grupo de control. En general, el tratamiento consistió en una discusión acerca de los beneficios para la salud materno-infantil del espaciamiento de los nacimientos, junto con consejos acerca del proceso de adquisición de anticonceptivos. Sin embargo, solo un porcentaje de los hogares que fueron asignados a tratamiento participaron en más de tres sesiones durante el transcurso de la intervención. Hay que tener en cuenta que tres sesiones es el **número mínimo** necesario para recibir la suficiente información con respecto a los métodos de planificación para considerarse **efectivamente tratado**.

Ahora bien, para los investigadores uno de los factores claves para entender la dinámica en los hogares es el rol de los esposos o parejas en la planificación. Según un informe previo del ISSR, entre las mujeres que indicaron que no querían tener más hijos, resalta el hecho de que aquellas cuyo esposo sí quería tener hijos tenían casi el doble de probabilidad de dar a luz que aquellas cuya pareja no deseaba prole. Así, para explorar el efecto que tienen los maridos en las decisiones de fertilidad de sus esposas, se dividió en dos grupos de forma aleatoria a los hogares tratados. En el grupo 1 de tratamiento, los consultores le dieron la información sólo a la mujer cabeza del hogar (grupo de tratamiento individual), mientras que en el grupo 2, la información fue impartida a esposos y esposas juntos (grupo de tratamiento de parejas).

Con el fin de estudiar el efecto de interés, Profamilia los ha contratado a ustedes para efectuar una evaluación de impacto de la capacitación. Para ello, les provee los siguientes datos post y pre intervención para todos los hogares entrevistados en los 300 municipios estudiados en la base de datos titulada “planificación DID”. Estos datos provienen de la encuesta de hogares que se implementó en agosto de 2012, antes de que comenzara el programa de planificación familiar, y nuevamente a partir de julio de 2014, doce meses después de finalizar el programa. Esta encuesta incluyen módulos sobre estatus socioeconómico, salud y planificación familiar, relaciones conyugales y agricultura.². Algunas variables de interés son:

- $Y_{i,t}$: Dummy de si utiliza o no algún tipo de método de planificación en el año t .
- P_pooled_i : indicadora de si el hogar i fue seleccionado para la capacitación y es parte del grupo 1 o 2.
- P_sola_i : indicadora de si el hogar i fue seleccionado para la capacitación y es parte del grupo 1.
- P_pareja_i : indicadora de si el hogar i fue seleccionado para la capacitación y es parte del grupo 2.
- D_pooled_i : indicadora de si el hogar i fue a más de 3 capacitaciones, es decir, recibió la información necesaria para ser capacitado y es parte del grupo 1 o 2.

¹El experimento descrito y los datos reportados a este son ficticios y tienen un único fin pedagógico.

²Los resultados obtenidos en este taller no proveen información alguna sobre el efecto de las capacitaciones sobre el uso de anticonceptivos.

- D_{sola_i} : indicadora de si el hogar i fue a más de 3 capacitaciones, es decir, recibió la información necesaria para ser capacitado y es parte del grupo 1.
- D_{pareja_i} : indicadora de si el hogar i fue a más de 3 capacitaciones, es decir, recibió la información necesaria para ser capacitado y es parte del grupo 2.
- $Sexo_i$: Sexo del capacitador del hogar i .
- $X_{i,t}$: vector de variables de control individuales y del hogar.
 - Estrato del hogar, distancia al centro médico más cercano y población rural del municipio.
- $Post_t$: variable indicadora para el período de seguimiento (2014).

a) Dada la estructura de panel “2x2”, propongan un modelo de Diferencias en Diferencias (DID) que les permita estimar el efecto de interés de ser tratados en el grupo 1, en el grupo 2 y el efecto combinado. En este caso el efecto que se busca capturar es el de la asistencia efectiva a más de tres sesiones de capacitaciones sobre la probabilidad de usar métodos anticonceptivos para cada grupo de interés.

- i. Planteen el modelo de regresión a estimar.
- ii. ¿Cuáles son sus supuestos de identificación? ¿Son estos plausibles?
- iii. Realice un balance muestral entre los grupos tratados (grupo 1 y grupo 2) antes y después de la intervención para las variables de control relevantes. Interprete.
- iv. Para cada modelo ¿Qué coeficiente captura el efecto de interés bajo dichos supuestos?
- v. Estime el efecto para cada uno de los tres grupos de interés (grupo 1, el grupo 2 y efecto combinado) usando la base de datos.

Dado que tenemos un panel “2x2”, no es posible evaluar de forma empírica el cumplimiento del supuesto de tendencias paralelas. Conociendo la aleatorización de la asignación, un asistente de investigación le propone un modelo alternativo de Diferencias en Diferencias (DID) usando la variable de selección a las capacitaciones para medir el impacto sobre el uso de anticonceptivos (ITT).

b) Dado el anterior contexto responda las siguientes preguntas:

- i. Si la idea del asistente es correcta, ¿por qué el coeficiente planteado en el inciso anterior ya no captura el efecto de interés?
- ii. Si la idea del asistente es correcta, ¿sería posible estimar un efecto consistentemente que cumpla los supuestos necesarios? Si su respuesta es afirmativa, proponga la ecuación que captura dicho efecto.
- iii. Estime el efecto para cada uno de los tres grupos de interés (grupo 1, el grupo 2 y efecto combinado) usando la base de datos.

Sin embargo y dado que se busca capturar es el efecto del tratamiento efectivo, es decir, haber asistido a más de tres sesiones. Usted se aventura a una metodología que involucra utilizar un diferencias en diferencias instrumentado (DID-IV), esta especificación aprovecha explícitamente la aleatoriedad exógena en una tendencia de exposición para estimar el efecto del tratamiento promedio. Para conocer mas sobre su implementación le recomendaron leer el artículo de [Hudson, Hull & Liebersohn \(2017\)](#).

c) Dado el anterior contexto responda las siguientes preguntas, para ello utilicé el artículo citado anteriormente.

- i. Planteen el modelo de regresión a estimar.
- ii. ¿Cuáles son sus supuestos de identificación? ¿Son estos plausibles?
- iii. ¿Qué coeficiente representa su efecto de interés bajo dichos supuestos? Interprete.
- iv. Implemente la metodología anterior planteada en Stata. Para ello utilice el paquete *xtivreg*, que le permite realizar regresión de variables instrumentales para datos panel. En una tabla para tipo de grupo (grupo 1, grupo 2 y combinado) presente los resultados de la primera etapa, segunda etapa y la forma reducida. Interprete.
- v. Compare con los resultados obtenidos con aquellos en el inciso a) y b).

Una de las participantes del programa le hace conocer que factiblemente el efecto de las capacitaciones puede cambiar dependiendo del sexo del capacitador (i.e. quien imparte la información). Esto debido a que existen sesgos de género marcados y sistemáticos que puedan ser decisivos para determinar el efectos del tratamiento en un contexto en donde la sexualidad y planificación femenina siguen siendo un tema tabú. Es así como usted se plantea explorar la heterogeneidad en los efectos del tratamiento según el sexo del capacitador. Suponga que se cumple tendencias paralelas entre el grupo de control y tratamiento en los tres casos de tratamiento (pooled, grupo 1 y grupo 2).

- d) Para medir lo anterior usted plantea una interacción entre el sexo del capacitador con el estado y el tiempo del tratamiento, es decir, una estimación de triple diferencia. Para ello:
- Planteen el modelo de regresión a estimar.
 - ¿Cuáles son sus supuestos de identificación? ¿Son estos plausibles?
 - ¿Qué coeficiente representa su efecto de interés bajo dichos supuestos? Interprete.
 - Estime el efecto para cada uno de los tres grupos de interés (grupo 1, el grupo 2 y efecto combinado) usando la base de datos. *planificación_DID.dta* e interprete. Inserte el código utilizado.

Segundo ejercicio

El método de diferencias en diferencias es quizá el método más utilizado en la última década por los economistas en las investigaciones académicas empíricas. Las numerosas aplicaciones condujeron a un interés de desarrollar extensiones a esta metodología. Quizá el desarrollo más importante en esta índole llegó con el método de “estudio de eventos”, el cual permitió generalizar diferencias en diferencias a escenarios donde las unidades recibían el tratamiento en distintos momentos del tiempo.

Las primeras aproximaciones de los economistas con los estudios de eventos se basaron en la intuición que se tenía acerca de la metodología de diferencias en diferencias, sin mucho cuidado acerca de si esta generalización era, en principio, siquiera posible. Consecuentemente, los econométricos teóricos empezaron a encontrar varios problemas asociados a la manera en que se utilizaba comúnmente esta metodología, concluyendo que existían principalmente dos problemas: (1) contrario a otras metodologías, no era muy claro el efecto causal que identifica los estudios de eventos; y (2), la estimación por mínimos cuadrados ordinarios de un modelo de panel lineal con efectos fijos, conocida como “two-way fixed effect estimation (TWFE)”, es inadecuada, en tanto no captura un efecto causal salvo para casos muy restrictivos para el análisis empírico.

Quizá el primer trabajo en poner en discusión las falencias del TWFE en estudios de evento es el trabajo de [Goodman-Bacon \(2021\)](#). En breve, este trabajo tiene dos grandes contribuciones. La primera, descompone el estimador de mínimos cuadrados ordinario que resulta de estimar un TWFE de manera que se puede entender de manera intuitiva lo que se está estimando. Lo segundo, muestra que incluso bajo el cumplimiento de tendencias paralelas, el estimador no recupera un efecto causal en presencia de heterogeneidad en los efectos de tratamiento.

El objetivo de este punto es entender mejor el planteamiento que se hace en Goodman-Bacon (2021) y de paso, repasar algunos temas que vimos en clase acerca de los estudios de eventos.

- a) Corran el Do File *Sim1_Do* que simula una pequeña “base de datos” con la cual nos aproximaremos al problema.³
- Describan en palabras qué hace el código.
 - Ilustren en una *gráfica* cómo varía la variable dependiente (con nombre Y en la base de datos) en función del tiempo para cada unidad. Ilustren con una línea punteada el momento donde cada unidad es tratada. ¿Cuántas cohortes hay? Describa cada una brevemente.
 - Describan matemáticamente $Y_{i,t}(g)$, donde g hace referencia a la cohorte. ¿Cuál es el ATT para cada cohorte tratada?

³Para los que usen R, corran la base en Stata y luego exporten a R.

En el modelo de diferencias en diferencias (DiD) tradicional, se distinguen dos dimensiones importantes. Primero se divide la muestra en dos grupos: aquellos tratados y aquellos no tratados, donde estos últimos sirven como controles. Además, distinguimos dos periodos temporales, pre y post tratamiento. En ese caso, el estimador de DiD es

$$\hat{\beta} = (\bar{y}_{\text{Tratados}}^{\text{Post}} - \bar{y}_{\text{Controles}}^{\text{Post}}) - (\bar{y}_{\text{Pre}}^{\text{Post}} - \bar{y}_{\text{Pre}}^{\text{Pre}})$$

donde \bar{y}_k^τ representa el promedio de la variable y observado en el grupo k en el periodo de tiempo τ .

Una de las observaciones que hizo [Goodman-Bacon \(2021\)](#) es que el estimador de TWFE en los estudios de eventos ($\hat{\beta}^{\text{DD}}$) guarda similitudes con el estimador clásico de diferencias en diferencias. En particular, el autor logró demostrar que $\hat{\beta}^{\text{DD}}$ es una función de todos los estimadores de diferencias en diferencias que se pueden formular entre las distintas cohortes.

Formalmente, el resultado es el siguiente:

Theorem 1. Difference-in-Differences Decomposition Theorem

Assume that the data contain $k = 1, \dots, K$ groups of units ordered by the time when they receive a binary treatment, $t_k^* \in (1, T]$. There may be one group, U , that never receives treatment. The OLS estimate, $\hat{\beta}^{\text{DD}}$, in a two-way fixed-effects regression (2) is a weighted average of all possible two-by-two DD estimators.

$$\hat{\beta}^{\text{DD}} = \sum_{k \neq U} s_{kU} \hat{\beta}_{kU}^{2 \times 2} + \sum_{k \neq U} \sum_{\ell > k} [s_{k\ell}^k \hat{\beta}_{k\ell}^{2 \times 2, k} + s_{k\ell}^\ell \hat{\beta}_{k\ell}^{2 \times 2, \ell}]. \quad (10a)$$

Where the 2×2 DD estimators are:

$$\hat{\beta}_{kU}^{2 \times 2} \equiv (\bar{y}_k^{\text{POST}(k)} - \bar{y}_k^{\text{PRE}(k)}) - (\bar{y}_U^{\text{POST}(j)} - \bar{y}_U^{\text{PRE}(j)}), \quad (10b)$$

$$\hat{\beta}_{k\ell}^{2 \times 2, k} \equiv (\bar{y}_k^{\text{MID}(k, \ell)} - \bar{y}_k^{\text{PRE}(k)}) - (\bar{y}_\ell^{\text{MID}(k, \ell)} - \bar{y}_\ell^{\text{PRE}(k)}), \quad (10c)$$

$$\hat{\beta}_{k\ell}^{2 \times 2, \ell} \equiv (\bar{y}_\ell^{\text{POST}(\ell)} - \bar{y}_\ell^{\text{MID}(k, \ell)}) - (\bar{y}_k^{\text{POST}(\ell)} - \bar{y}_k^{\text{MID}(k, \ell)}). \quad (10d)$$

The weights are:

$$s_{kU} = \frac{\overbrace{(n_k + n_U)^2 n_{kU} (1 - n_{kU}) \bar{D}_k (1 - \bar{D}_k)}^{\hat{V}_{kU}^D}}{\hat{V}^D}, \quad (10e)$$

$$s_{k\ell}^k = \frac{\overbrace{((n_k + n_\ell)(1 - \bar{D}_\ell))^2 n_{k\ell} (1 - n_{k\ell}) \frac{\bar{D}_k - \bar{D}_\ell}{1 - \bar{D}_\ell} \frac{1 - \bar{D}_k}{1 - \bar{D}_\ell}}^{\hat{V}_{k\ell}^{D, k}}}{\hat{V}^D}, \quad (10f)$$

$$s_{k\ell}^\ell = \frac{\overbrace{((n_k + n_\ell) \bar{D}_k)^2 n_{k\ell} (1 - n_{k\ell}) \frac{\bar{D}_\ell}{\bar{D}_k} \frac{\bar{D}_k - \bar{D}_\ell}{\bar{D}_k}}^{\hat{V}_{k\ell}^{D, \ell}}}{\hat{V}^D}. \quad (10g)$$

and $\sum_{k \neq U} s_{kU} + \sum_{k \neq U} \sum_{\ell > k} [s_{k\ell}^k + s_{k\ell}^\ell] = 1$.

Figura 1: Teorema 1, [Goodman-Bacon \(2021\)](#)

Donde para cada pareja (i, j) de cohortes

- $\text{Pre}(i) :=$ el periodo donde i no es tratado.
- $\text{Mid}(i, j) :=$ es el periodo donde i ya fue tratado pero j todavía no lo es.
- $\text{Post}(j) :=$ es el periodo posterior al tratamiento de j .

Notemos con la letra U la cohorte que nunca es tratada, con k la cohorte que es tratada más temprano y con l aquella que es tratada de últimas.

b) Usando la muestra simulada, calcularemos todos los posibles estimadores tipo 2×2 :

- i) Calcule $\hat{\beta}_{kU}^{2 \times 2}$, $\hat{\beta}_{lU}^{2 \times 2}$, $\hat{\beta}_{kl}^{2 \times 2, k}$, $\hat{\beta}_{kl}^{2 \times 2, l}$. Para cada estimador, resalte en la gráfica que realizaron en a. ii) los periodos y cohortes utilizados para la estimación de cada pareja.
- ii) Pensando en un modelo de diferencias en diferencias estándar, ¿todas estas comparaciones parecen deseables? Justifiquen su respuesta.

Nota: Hay un typo en el teorema del paper: en el 10b), la j es una k .

Ahora exploramos un poco los pesos (ver Teorema 1). Note que \bar{D}_j es el promedio de la variable D_{it} , que indica si una unidad está tratada o no en el tiempo t , para el grupo j . Así las cosas, \bar{D}_j es la fracción de tiempo que una cohorte pasa tratada. Similarmente n_j es el tamaño de la cohorte j (como proporción del total), y sea $n_{ij} = n_i / (n_i + n_j)$ la proporción relativa de la cohorte i y la cohorte j (como proporción del total). Finalmente \hat{V}^D es la varianza muestral de D_{it} .

c) Usando la muestra simulada, calcularemos todos los posibles pesos y verificaremos la identidad de Goodman-Bacon:

- i) Calcule s_{kU} , s_{lU} , s_{kl}^k y s_{kl}^l .
- ii) Usando la fórmula 10e) explique las siguientes afirmaciones que realiza el autor:
 - ◊ “El peso que recibe el grupo k es proporcional a la varianza de D_{it} que se observa en el grupo.”
 - ◊ “La varianza de tratamiento se maximiza en la mitad del panel.”
- iii) ¿Es la propiedad de que aquellas cohortes que están en la mitad del panel reciban el mayor peso deseable?
- iv) Calcule $\hat{\beta}^{DD}$ estimando el modelo de TWFE. Verifique que la descomposición es cierta.

Entendida la descomposición del estimador, procederemos a ver por qué el estimador de TWFE puede tener problemas identificando un efecto causal:

En términos poblacionales tenemos que

$$\hat{\beta}_{kU}^{2 \times 2} \xrightarrow{P} \text{ATT}_k(\text{Post}) + \Delta Y_k^0(\text{Post}(k), \text{Pre}(k)) - \Delta Y_U^0(\text{Post}(k), \text{Pre}(k))$$

donde $\text{ATT}_k(\text{Post})$ es el efecto causal promedio del tratamiento del grupo k después del tratamiento y

$$\Delta Y_k^0(\text{Post}(k), \text{Pre}(k)) - \Delta Y_U^0(\text{Post}(k), \text{Pre}(k))$$

es el sesgo de tendencias paralelas entre la cohorte k y la cohorte U .

Similarmente

$$\hat{\beta}_{kl}^{2 \times 2, k} \xrightarrow{P} \text{ATT}_k(\text{Mid}) + \Delta Y_k^0(\text{Mid}(k, l), \text{Pre}(k)) - \Delta Y_l^0(\text{Mid}(k, l), \text{Pre}(k))$$

donde $\text{ATT}_k(\text{Mid})$ es el efecto causal promedio del tratamiento del grupo k después del tratamiento pero antes del tratamiento de l y

$$\Delta Y_k^0(\text{Mid}(k, l), \text{Pre}(k)) - \Delta Y_l^0(\text{Mid}(k, l), \text{Pre}(k))$$

el sesgo de tendencias paralelas entre la cohorte k y l , antes de que l sea tratado.

Como se puede apreciar, si existen tendencias paralelas entre cohortes, cada uno de los estimadores 2×2 captura un ATT . Ello es porque siempre se compara a un grupo tratado con otro que no ha sido tratado o que nunca será tratado.

No obstante, cuando analizamos el último término, donde usamos a la cohorte tratada al principio del panel como control, tenemos que

$$\hat{\beta}_{kl}^{2 \times 2, l} \xrightarrow{P} \text{ATT}_l(\text{Post}) + \Delta Y_l^0(\text{Post}(l), \text{Mid}(k, l)) - \Delta Y_k^0(\text{Post}(l), \text{Mid}(k, l)) - (\text{ATT}_k(\text{Post}) - \text{ATT}_k(\text{Mid}))$$

Note que en este caso tenemos lo mismo que en anteriores expresiones, pero con un factor adicional:

$$\text{ATT}_k(\text{Post}) - \text{ATT}_k(\text{Mid})$$

el cual constituye un sesgo por la posible heterogeneidad en los efectos de tratamiento que no desaparece bajo el supuesto de tendencias paralelas. El origen de este sesgo es que se está usando como control unidades que ya han sido tratadas, por lo que si el valor observado para estas unidades varía en el tiempo por consecuencia del tratamiento, entonces no estaremos encontrando el verdadero efecto causal. La intuición es que no podemos discernir la evolución normal de la variable dependiente (la que tiene una cohorte no tratada) de aquella generada por el tratamiento, lo cual hace imposible identificar el efecto.

Utilizando las anteriores expresiones junto con la descomposición de Goodman-Bacon tenemos entonces que

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_n^{DD} = \text{VWAT} + \text{VWPT} - \Delta \text{ATT}$$

donde

$$\text{VWAT} = \sum_{k \neq U} \sigma_{kU} \text{ATT}_k(\text{Post}(k)) + \sum_{k \neq U} \sum_{l > k} \sigma_{kl}^k \text{ATT}_k(\text{MID}(k, l)) + \sigma_{kl}^l \text{ATT}_l(\text{Post}(l))$$

es el “Variance-weighted ATT”, esto es, es un promedio ponderado de todos los ATT entre cohortes (ponderados por las varianzas de tratamiento dado que $\hat{s}_{ij}^l \rightarrow \sigma_{ij}^l$). Esto es una cantidad que quisiéramos estimar en tanto es el promedio ponderado de efectos causales.

Por su parte, VWPT es un promedio ponderado de los sesgos de tendencias paralelas, mientras que

$$\Delta \text{ATT} = \sum_{k \neq U} \sum_{l > k} (1 - \nu_{lk}) [\text{ATT}_k(\text{Post}(l)) - \text{ATT}_k(\text{Mid}(k, l))]$$

es el sesgo por heterogeneidad que se origina por las malas comparaciones, donde ν_{lk} son algunos pesos proporcionales a σ_{kl}^l . Así las cosas, **incluso bajo el supuesto de tendencias paralelas (VWPT= 0)**, el estimador no recupera un efecto causal:

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_n^{DD} = \text{VWAT} - \Delta \text{ATT}$$

El gran problema de dicho sesgo es que, dado que ΔATT puede tomar cualquier signo, existe la probabilidad de que $\hat{\beta}_n^{DD}$ tenga el signo contrario al efecto real. Más aún el sesgo existirá salvo que *i*) haya efectos constantes de tratamiento para todas las cohortes en todos los tiempos; o *ii*), no hay efectos dinámicos, esto es, el tratamiento no genera efectos diferenciales en el tiempo. Lastimosamente, ambos supuestos son casi imposibles que se satisfagan en la práctica.

No obstante, dado que sabemos que el sesgo lo originan precisamente las comparaciones que usan a cohortes ya tratadas como grupo de control, sabemos que el sesgo será menos fuerte en tanto el peso que tengan estas comparaciones en el cálculo de $\hat{\beta}_n^{DD}$ sea menor. La idea de Goodman-Bacon es entonces evaluar el tamaño de los pesos de estas comparaciones como una noción de cuán sesgado puede estar el estimador de TWFE. Vamos a poner a prueba esta idea con datos más elaborados.

Considere ahora la simulación provista en “Sim.Do2”.

d) Para la nueva muestra simulada

- i) Grafique en una *única gráfica* la variable dependiente de cada unidad en el tiempo.

De acuerdo con la gráfica y el código que las genera: ¿Cuántas cohortes hay? ¿Hay tendencias paralelas? ¿Cuál parece ser el signo del efecto de tratamiento? ¿Varían los efectos con el tiempo?

- ii) Estime el modelo por TWFE. ¿Es coherente el signo encontrado con lo que se espera teóricamente? De acuerdo a lo visto anteriormente, ¿qué puede estar causando este problema?

Para evaluar qué tanto peso tiene cada comparación, podemos estimar los pesos de la descomposición. Este proceso lo realiza automáticamente el comando *bacondecomp*.

- e) Para la nueva muestra simulada, presente de manera gráfica la descomposición de Goodman-Bacon. Interprete lo que encontró en el inciso *d.ii*) a la luz de esta nueva evidencia.

En los últimos años se han desarrollado procedimientos estadísticos alternativos para estimar estudios de eventos que no tienen las dificultades del estimador TWFE. Uno de los más usados de esta nueva familia es el estimador de Callaway & Sant'Anna que vimos en clase.

- f) Usando el estimador de Callaway & Sant'Anna, estime los efectos causales dinámicos del tratamiento en la muestra simulada (i.e, el efecto promedio del tratamiento τ periodos antes o después de que una cohorte es tratada).

¿La estimación parece ser consistente con efectos causales de la simulación? ¿La estimación parece ser consistente con la información que tenemos acerca del cumplimiento de tendencias paralelas?

Tercer ejercicio

El objetivo de este ejercicio es indagar sobre las distintas estructuras de errores estándar que comúnmente se emplean al estimar un modelo de regresión lineal. En particular, queremos evaluar cómo distintas metodologías para la estimación de los errores estándar se comportan en diversos contextos que pueden hacer que uno u otro enfoque resulte más apropiado. El análisis se centrará en estimar y efectuar inferencia estadística sobre el siguiente modelo:

$$Y_i = \beta * X_i + \varepsilon_i, \quad \beta = 10 \quad (1)$$

Donde inicialmente los X_i son independientes e idénticamente distribuidos con distribución normal estándar, mientras que en los siguientes puntos variaremos la estructura de la distribución de los términos de error ε_i .

1. Inicialmente, considere un escenario homocedástico en el que $\varepsilon_i \sim \mathcal{N}(0,1)$ iid. En este contexto, estudiaremos los efectos de suponer erróneamente un modelo heterocedástico (pero sin autocorrelación).

- a) En primer lugar, consideraremos dos posibles estimadores de la varianza del vector de $\hat{\beta}$, que viene dada por

$$\text{var}(\hat{\beta}) = (X'X)^{-1}X'\Sigma X(X'X)^{-1}$$

Donde Σ es la matriz de varianzas y covarianzas del vector de ε_i 's y X es la matriz de variables explicativas.

Bajo errores homocedásticos, se supone $\Sigma = \sigma^2 I$, de manera que solo es necesario estimar σ^2 , lo cual se hace por medio de la fórmula

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n \hat{\varepsilon}_i^2}{n - k}$$

donde k es el número de parámetros y $\hat{\varepsilon}_i$ son los residuales de la regresión.

En cambio, bajo errores heterocedásticos, se hace necesario estimar directamente la matriz de varianzas y covarianzas Σ . Una alternativa usual son los errores tipo Huber-White o *hc1* (*heteroskedasticity consistent 1*, los cuales son la opción predeterminada de Stata al agregar la opción “*robust*”). En particular este estimador toma la forma:

$$\widehat{\text{var}}(\hat{\beta})_{hc1} = (X'X)^{-1}X'\text{diag}(\hat{\epsilon}_1^2, \dots, \hat{\epsilon}_n^2)X(X'X)^{-1}$$

donde el estimador de la matriz Σ es la matriz diagonal compuesta por los residuales del modelo al cuadrado ($\hat{\Sigma}_{hc1} = \text{diag}(\hat{\epsilon}_1^2, \dots, \hat{\epsilon}_n^2)$). En otras palabras, el estimador puntual del elemento Σ_{ii} es $\hat{\epsilon}_i^2$.

- a) Simulen 1000 muestras $\{X_i, Y_i\}_{n=1}^{1000}$ acorde al modelo (1). Para cada una, estimen el modelo (1) y calculen el intervalo de confianza del 95 % para $\hat{\beta}$ empleando errores estándar usuales y *hc1*. Para cada caso, calcule el porcentaje de las veces en que en dicho intervalo se encuentra el parámetro verdadero. Presenten e interpreten sus resultados.
- b) Los errores *hc1* resultan problemáticos cuando se cuenta con muestras pequeñas por dos razones principales. Por una parte, aunque el estimador obtenido de Σ es consistente, su convergencia es relativamente lenta. Por otra parte, se ha probado que es especialmente sensible a la presencia de datos atípicos. Para visualizar estos problemas, repitan el procedimiento del inciso anterior, esta vez con muestras de tamaño 10. ¿Cómo se comparan sus resultados?
- c) Una solución al problema de muestra pequeña al que se enfrentan los errores *hc1* fue propuesta por **Davidson y MacKinnon (1993)**, quienes sugieren usar un estimador alternativo que contiene un factor de corrección para muestras finitas⁴ $\hat{\Sigma}_{hc3} = \text{diag}(\hat{\epsilon}_1^2/(1 - h_{11})^2, \dots, \hat{\epsilon}_n^2/(1 - h_{nn})^2)$, en donde h_{ii} se refiere al i -ésimo elemento de la diagonal de la matriz $X(X'X)^{-1}X$. La intuición detrás de esta corrección es que cuantifica, en cierta medida, qué tan influyente es una observación a la hora de calcular $\hat{\beta}$. De esta manera, este estimador se hace más resistente a datos atípicos⁵. Repita el procedimiento del inciso b, pero esta vez emplee errores convencionales y *hc3*. Reporte sus resultados y compare.

2. Suponga ahora que $\varepsilon_i \sim \mathcal{N}(0, X_i^2)$ independientes.

- a) Simulen 1000 muestras $\{X_i, Y_i\}_{n=1}^{1000}$ acorde al modelo (1). Para cada una, estimen el modelo (1) y calculen el intervalo de confianza $\hat{\beta}$ empleando errores estándar usuales, *hc1* y *hc3*. Para cada caso, calcule el porcentaje de las veces en que en dicho intervalo se encuentra el parámetro verdadero. Presenten e interpreten sus resultados. Comparen con lo obtenido en el punto 1.
- b) Repitan el anterior procedimiento con muestras de tamaño 10. Presenten e interpreten sus resultados. Comparen con lo obtenido en el punto 1.

3. Otro problema común es la autocorrelación o correlación serial, la cual tiene que ver con la existencia de grupos de unidades u observaciones dentro de la muestra cuyos términos de error se encuentran correlacionados. Este panorama rompe con el supuesto clásico de independencia entre los términos de error y plantea nuevos problemas que exploraremos a continuación.

- a) Simulen 1000 muestras $\{X_i, Y_i\}_{n=1}^{1000}$ acorde a los siguientes pasos:
 - Simule las primeras 100 observaciones de la muestra siguiendo el modelo 1 tomando $\varepsilon_i \sim \mathcal{N}(0, 1)$ iid.
 - Por medio del comando **expand**, produzca 10 copias de los datos (de esta forma construya una base con 1000 observaciones en total)

Explique por qué este procedimiento induce un problema de autocorrelación. Para cada muestra, estimen el modelo (1) y calculen el intervalo de confianza $\hat{\beta}$ empleando errores estándar usuales, *hc1*, *hc3* y cluster⁶. Los errores cluster permiten que las covarianzas de errores de observaciones pertenecientes a un mismo cluster sean estimadas individualmente (i.e. da libertad sobre el valor de estas covarianzas), mientras que las demás covarianzas son restringidas a 0. En cada caso, calcule el porcentaje de las veces en que en dicho intervalo se encuentra el parámetro verdadero. Presenten e interpreten sus resultados.

- b) ¿Es el modelo heterocedástico sin autocorrelación un caso específico de errores con estructura cluster?

⁴Puede revisar este [link](#) o [este](#) para más información.

⁵Este procedimiento es, de hecho, el default en R al usar errores robustos.

⁶Pista: piense cuáles son las observaciones que se correlacionan para escoger el nivel de cluster

c) Un problema con los errores clusters es que el número de entradas de la matriz varcov de los errores que hay que estimar puede ser muy alto si el número de clusters es relativamente bajo. Como consecuencia, con pocos clusters se hace necesario estimar un mayor número de parámetros simultáneamente, lo que puede afectar la calidad de dicha estimación. Para ver esto, simulen 1000 muestras $\{X_i, Y_i\}_{n=1}^{1000}$ para $j = 10, 20, 50, 100$ según el siguiente procedimiento:

- Simulen las primeras j observaciones.
- Dupliquen las base de datos tantas veces sea necesario para producir 1000 observaciones en total.
- Para cada muestra, estimen el modelo (1) y calculen el intervalo de confianza $\hat{\beta}$ empleando errores estándar cluster.

Para cada j , calcule el porcentaje de las veces en que en dicho intervalo se encuentra el parámetro verdadero. Gráfique el j empleado contra el porcentaje de aciertos respectivo e interprete.

4. Una alternativa al problema de seleccionar la estructura de los errores para estimar su matriz de varianza y covarianza es a través de métodos no paramétricos. En particular, la metodología de bootstrap nos permite recuperar una estimación del error estándar incluso en contextos en los que no es claro cuál de todos los estimadores previamente vistos es más adecuado.

- Simulen 1000 muestras $\{X_i, Y_i\}_{n=1}^{1000}$ acorde al modelo (1) cuando $\varepsilon_i \sim \mathcal{N}(0, 1)$ iid. Para cada una, estimen el modelo (1) y calculen el intervalo de confianza $\hat{\beta}$ del 95 % empleando errores estándar bootstrap con 100 repeticiones. Calculen el porcentaje de las veces en que en dicho intervalo se encuentra el parámetro verdadero.
- Simulen 1000 muestras $\{X_i, Y_i\}_{n=1}^{1000}$ acorde al modelo (1) cuando $\varepsilon_i \sim \mathcal{N}(0, X_i^2)$ independientes. Para cada una, estimen el modelo (1) y calculen el intervalo de confianza $\hat{\beta}$ del 95 % empleando errores estándar bootstrap con 100 repeticiones. Calculen el porcentaje de las veces en que en dicho intervalo se encuentra el parámetro verdadero.
- Simulen 1000 muestras $\{X_i, Y_i\}_{n=1}^{1000}$ acorde a los siguientes pasos:

- Simule las primeras 100 observaciones de la muestra siguiendo el modelo 1 tomando $\varepsilon_i \sim \mathcal{N}(0, 1)$ iid.
- Por medio del comando **expand**, produzca 10 copias de los datos (de esta forma construya una base con 1000 observaciones en total)

Para cada muestra, estimen el modelo (1) y calculen el intervalo de confianza $\hat{\beta}$ del 95 % empleando errores estándar bootstrap con 100 repeticiones. Calculen el porcentaje de las veces en que en dicho intervalo se encuentra el parámetro verdadero.

d) Comparen sus resultados con lo obtenido en los anteriores incisos.