Taller 3

Econometría Avanzada (2023-1)

Profesor: Manuel Fernández,

Integrantes:

Jorge Iván Mejía - 201922665

Lucas Rodríguez - 202021985

Primer ejercicio

Supongan que ustedes cuentan con una base de datos en la que observan a N individuos en dos periodos, $T_t \in \{0, 1\}$. Cada individuo i puede hacer parte del grupo de tratamiento (Di =1) o del grupo de control (Di = 0). En T_0 ningún individuo ha recibido el tratamiento, mientras que en T_1 solo aquellos pertenecientes al grupo de tratamiento han sido tratados. A ustedes les interesa conocer el efecto del tratamiento sobre una variable de resultado, $Y_{i,t}$. En particular, el efecto causal de interés es

$$\boldsymbol{ATT = E[Y_{i,1}^1 - Y_{i,1}^0 | D_i = 1]}$$

donde $Y_{i,t}^j$denota el resultado potencial del individuo i en el periodo t de tener estado de tratamiento j (= 1 si es tratado, = 0 de lo contrario). Para estimar este efecto, ustedes plantean el siguiente modelo:

$$\boldsymbol{Y_{i,t} = \beta + \alpha T_t + \delta D_i + \tau (T_t \times D_i) + \epsilon_{i,t}} \quad \textbf{(1)}$$

1) **En este contexto, para identificar el efecto de interés, los investigadores suelen suponer que existen tendencias paralelas entre tratados y controles en ausencia del tratamiento. En el marco del modelo (1), expresen formalmente este supuesto usando el lenguaje de resultados potenciales, expliquen intuitivamente su significado, y demuestren que bajo tendencias paralelas τ = ATT.**

$$ATT = \mathbb{E}(Y_{i1}^1 | D_i = 1) - \mathbb{E}(Y_{i1}^0 | D_i = 1)$$
$$\tau = \left(\mathbb{E}(Y_{i1}^1 | D_i = 1) - \mathbb{E}(Y_{i0}^0 | D_i = 1) \right) - \left(\mathbb{E}(Y_{i1}^0 | D_i = 0) - \mathbb{E}(Y_{i0}^0 | D_i = 0) \right)$$

Supuesto tendencias paralelas: $\mathbb{E}(Y_{i1}^0 | D_i = 1) - \mathbb{E}(Y_{i0}^0 | D_i = 1) = \mathbb{E}(Y_{i1}^0 | D_i = 0) - \mathbb{E}(Y_{i0}^0 | D_i = 0)$

Esto significa que, en ausencia de tratamiento, las unidades de tratamiento y control hubieran evolucionado de la misma manera

Reemplazando el supuesto de tendencias paralelas en la ecuación de ATT

$$ATT = \mathbb{E}(Y_{i1}^1 | D_i = 1) - \left(\mathbb{E}(Y_{i0}^0 | D_i = 1) + \mathbb{E}(Y_{i1}^0 | D_i = 0) - \mathbb{E}(Y_{i0}^0 | D_i = 0) \right)$$

Reorganizando

$$ATT = \left(\mathbb{E}(Y_{i1}^{1}\middle|D_i = 1) - \mathbb{E}(Y_{i0}^{0}\middle|D_i = 1)\right) - \left(\mathbb{E}(Y_{i1}^{0}\middle|D_i = 0) - \mathbb{E}(Y_{i0}^{0}\middle|D_i = 0)\right) = \tau$$

Q.E.D

2) No obstante, hay situaciones en las cuales el supuesto de tendencias paralelas es plausible solo tras condicionar en ciertas covariables. Esto es, se supone que tratados y controles con iguales valores de las covariables hubieran evolucionado paralelamente en ausencia de tratamiento. Por simplicidad, supongan que solo es necesario considerar una única covariable, Xi (noten que implícitamente estamos suponiendo que Xi no varía en el tiempo). Para capturar este hecho, se plantea el modelo alternativo

$$\boldsymbol{Y_{i,t} = \beta + \alpha T_t + \delta D_i + \tau (T_t \times D_i) + \gamma X_i + \epsilon'_{i,t}} \quad \textbf{(2)}$$

a) ¿Cuál es el equivalente del supuesto de tendencias paralelas para el modelo 2? Exprésenlo formalmente.

$$\mathbb{E}(Y_{i1}^{0} - Y_{i0}^{0}\middle|X_i, D_i = 1) = \mathbb{E}(Y_{i1}^{0} - Y_{i0}^{0}\middle|X_i, D_i = 0)$$

b) Un problema con el modelo (2) es que impone de manera implícita restricciones adicionales sobre el proceso generador de datos más allá de tendencias paralelas condicionales. En particular, demuestren que imponer la forma funcional (2) implica que

I) Los efectos de tratamiento son homogéneos en X:

$$\boldsymbol{E\left[Y_{i,1}^{1} - Y_{i,1}^{0}\middle|D_i = 1, X\right] = \tau}$$

Del literal (a) sabemos que $\mathbb{E}(Y_{i1}^{0}|X_i, D_i = 1) = \mathbb{E}(Y_{i0}^{0}|X_i, D_i = 1) + \mathbb{E}(Y_{i1}^{0}|X_i, D_i = 0) - \mathbb{E}(Y_{i0}^{0}|X_i, D_i = 0)$

En términos del modelo (2)

$$\mathbb{E}(Y_{i1}^{0}|X_i, D_i = 1) = (\beta + \delta + \gamma X_i) + (\beta + \alpha + \gamma X_i) - (\beta + \gamma X_i) = \beta + \delta + \alpha + \gamma X_i$$

$$\mathbb{E}(Y_{i1}^{1}|X_i, D_i = 1) = \beta + \alpha + \delta + \tau + \gamma X_i$$

Restando los dos

$$\mathbb{E}(Y_{i1}^{1}|X_i, D_i = 1) - \mathbb{E}(Y_{i1}^{0}|X_i, D_i = 1) = (\beta + \alpha + \delta + \tau + \gamma X_i) - (\beta + \delta + \alpha + \gamma X_i) = \tau$$

Esto significa que el tratamiento afecta a todas las unidades por igual, independiente de X

Q.E.D

II) **Descarta la posibilidad de tendencias específicas[1] según los valores de X en los grupos de tratamiento y de control.**

$$\boldsymbol{E[Y_{i,1} - Y_{i,0}|Di = d, X] = E[Y_{i,1} - Y_{i,0}|Di = d];} \quad \textbf{para } \boldsymbol{d = 0, 1}$$

Explique intuitivamente qué significan cada una de estos resultados. ¿Qué consecuencias tiene en la estimación del ATT por MCO que las implicaciones I) y II) no se cumplan?

Condicionando en X cuando D=1

$$(Y_{i1}|X_i\,, D_i = 1) - \mathbb{E}(Y_{i0}|X_i\,, D_i = 1) = (\beta + \alpha + \delta + \tau + \gamma X_i) - (\beta + \delta + \gamma X_i) = \alpha + \tau$$

Ahora sin condicionar en X

$$(Y_{i1}^1 \big| D_i = 1) - \mathbb{E}(Y_{i0}|D_i = 1)$$
$$= (\beta + \alpha + \delta + \tau + \gamma\mathbb{E}(X_i|D_i = 1)) - (\beta + \delta + \gamma\mathbb{E}(X_i|D_i = 1)) = \alpha + \tau$$

$$(Y_{i1} - Y_{i0}|X_i\,, D_i = 1) = (Y_{i1} - Y_{i0}|D_i = 1) = \alpha + \tau$$

Q.E.D

Esto significa que Y no es afectada por X a través del tiempo para el grupo de tratados

Condicionando en X cuando D=0

$$(Y_{i1}|X_i\,, D_i = 0) - \mathbb{E}(Y_{i0}|X_i\,, D_i = 0) = (\beta + \alpha + \gamma X_i) - (\beta + \gamma X_i) = \alpha$$

Ahora sin condicionar en X

$$(Y_{i1}^1 \big| D_i = 0) - \mathbb{E}(Y_{i0}|D_i = 0) = (\beta + \alpha + \gamma\mathbb{E}(X_i|D_i = 1)) - (\beta + \gamma\mathbb{E}(X_i|D_i = 1)) = \alpha$$

$$(Y_{i1} - Y_{i0}|X_i\,, D_i = 0) = (Y_{i1} - Y_{i0}|D_i = 0) = \alpha$$

Q.E.D

Esto significa que Y no es afectada por X a través del tiempo para el grupo de controles

En conclusión, la evolución de Y_i es independiente de X (tanto para el grupo de control como el de tratamiento).

Si (I) y/o (II) no se cumple entonces el $\tau \neq ATT$ ya que faltaría incluir la interacción entre el tratamiento y X (ϕ). Si no se incluye esta interacción entonces τ sería sesgado ya que se está incluyendo el efecto de X en Y dentro de este.

Pista: Para la segunda demostración se sugiere calcular la forma explícita de $\boldsymbol{E[Y_{i,1} - Y_{i,0}|D_i = d, X]}$ y mostrar que no depende de Xi. Recuerden que $\boldsymbol{Y_{i,t} = D_i Y_{i,t}^1 + (1 - D_i) Y_{i,t}^0}$. Se les sugiere empezar con $\boldsymbol{E[Y_{i,t}|D_i = d, X]}$

---

[1] Por tendencias específicas se quiere decir que unidades con distintos valores de las covariables evolucionan de manera diferenciada en el tiempo.

3) **Debido a estas limitaciones, ustedes quieren explorar especificaciones alternativas que les permitan flexibilizar los supuestos implícitos adicionales que encontraron en el punto anterior.**
   a) **Una compañera les recomienda estimar el siguiente modelo:**

$$\boldsymbol{Y_{i,t} = \beta + \alpha T_t + \delta D_i + \tau (T_t \times D_i) + \gamma X_i + \phi (D_i \times T_i) X_i + \epsilon''_{i,t}} \ \textbf{(3)}$$

¿Cuáles de los supuestos adicionales estudiados en el literal b) pueden relajarse con esta nueva especificación? Muéstrenlo formalmente.

Supuesto (I):

$$\mathbb{E}(Y_{i1}^0 | X_i, D_i = 1) = (\beta + \delta + \gamma X_i) + (\beta + \alpha + \gamma X_i) - (\beta + \gamma X_i) = \beta + \delta + \alpha + \gamma X_i$$

$$\mathbb{E}(Y_{i1}^1 | X_i, D_i = 1) = \beta + \alpha + \delta + \tau + \gamma X_i + \phi X_i$$

$$\mathbb{E}(Y_{i1}^1 - Y_{i1}^0 | X_i, D_i = 1) = (\beta + \alpha + \delta + \tau + \gamma X_i + \phi X_i) - (\beta + \delta + \alpha + \gamma X_i)$$
$$= \tau + \phi X_i$$

Supuesto (II):

Condicionando en X cuando D=1

$$(Y_{i1} | X_i, D_i = 1) - \mathbb{E}(Y_{i0} | X_i, D_i = 1) = (\beta + \alpha + \delta + \tau + \gamma X_i + \phi X_i) - (\beta + \delta + \gamma X_i)$$
$$= \alpha + \tau + \phi X_i$$

Ahora sin condicionar en X

$$(Y_{i1}^1 | D_i = 1) - \mathbb{E}(Y_{i0} | D_i = 1)$$
$$= \big(\beta + \alpha + \delta + \tau + \gamma \mathbb{E}(X_i | D_i = 1) + \phi \mathbb{E}(X_i | D_i = 1)\big)$$
$$- \big(\beta + \delta + \gamma \mathbb{E}(X_i | D_i = 1)\big) = \alpha + \tau + \phi \mathbb{E}(X_i | D_i = 1)$$

$$(Y_{i1} - Y_{i0} | X_i, D_i = 1) - (Y_{i1} - Y_{i0} | D_i = 1) = \phi X_i - \phi \mathbb{E}(X_i | D_i = 1)$$

Como la diferencia puede ser diferente de 0, esta especificación permite que haya tendencias específicas heterogéneas debido a que podemos identificar ϕ y tenerlo en cuenta

Condicionando en X cuando D=0

$$(Y_{i1} | X_i, D_i = 0) - \mathbb{E}(Y_{i0} | X_i, D_i = 0) = (\beta + \alpha + \gamma X_i) - (\beta + \gamma X_i) = \alpha$$

Ahora sin condicionar en X

$$(Y_{i1}^1 | D_i = 0) - \mathbb{E}(Y_{i0} | D_i = 0) = (\beta + \alpha + \gamma \mathbb{E}(X_i | D_i = 1)) - (\beta + \gamma \mathbb{E}(X_i | D_i = 1)) = \alpha$$

$$(Y_{i1} - Y_{i0} | X_i, D_i = 0) = (Y_{i1} - Y_{i0} | D_i = 0) = \alpha$$

Con esta nueva especificación se estaría relajando el supuesto (I) y (II), ya que se está añadiendo al modelo la interacción entre el tratamiento y X_i (ϕ). Como ϕ se puede estimar entonces ya no es necesario mantener estos supuestos.

b) **¿Qué aprendieron de este ejercicio? ¿Cómo podrían flexibilizar aún más los modelos anteriores para permitir que el efecto heterogéneo y la tendencia temporal no sean necesariamente lineales en Xi?**

Con este ejercicio se ejemplifica como añadir interacciones entre variables (ya sea entre control y tratamiento u otras) permite desagregar el efecto total en sus componentes (como el efecto causal del tratamiento, efectos heterogéneos del tratamiento y/o efectos de X en Y). Como consecuencia, ciertos supuestos se pueden relajar al poder identificar el efecto causal de interés sin que este sea "contaminado" por otros. Para flexibilizar aún más el modelo se propone incluir la interacción entre el tratamiento y polinomios de mayor grado de X. De esta manera, se permite que las interacciones entre el tratamiento y las variables de control no sean necesariamente lineales (por ejemplo pueden ser cuadráticas).

Segundo ejercicio

Los juzgados de descongestión son tribunales especializados creados en Colombia a partir del año 2015 para abordar el atraso acumulado de casos en el sistema judicial del país. Estos juzgados se crearon como parte de una reforma judicial que buscaba mejorar la eficiencia del sistema de justicia y resolver casos que habían estado pendientes por un período prolongado. Si bien esta medida se diseñó para ser temporal, los juzgados continuaron operando de forma permanente en el territorio.

Algunos expertos han manifestado que la implementación de juzgados de descongestión no ha sido efectiva para descongestionar el sistema judicial. Ante las críticas, el Consejo Superior de la Judicatura los contrata para examinar la efectividad de esta política. Para esto les proporcionan una base de datos con las siguientes variables:

- ***id_municipio*: una variable identificadora para cada municipio en el país.**
- ***new_juzgados*: variable indicadora igual a uno si el municipio recibió juzgados de descongestión entre 2015 y 2020, y cero de lo contrario.**
- ***year*: el año en el que se establecieron juzgados de descongestión en el municipio, con cero si en el municipio no se implementaron este tipo de juzgados.**
- ***IEP 2010 a IEP 2020*: el Índice de Evacuación Parcial Efectivo (IEP). Este índice corresponde a la división entre egresos efectivos (procesos judiciales finalizados) e ingresos efectivos (procesos judiciales iniciados) en cada municipio y en cada año. Si el índice es superior al 100 %, indica la desacumulación de procesos; un valor inferior al 100 %, representa la acumulación.**

Ustedes además saben que los juzgados de descongestión se asignaron a los municipios en los que más casos habían estado pendientes, junto con otras consideraciones sobre el tipo de casos más prevalentes en el municipio (de familia, laborales, penales, etc.), y el distrito judicial al que pertenecen. Sin embargo, desconocen las especificidades de la regla de asignación. Por último, les informan que el anuncio de dónde se iban a implementar los juzgados cada año se daba en enero y que ese mismo mes los juzgados comenzaban a operar en el municipio.

Respondan las siguientes preguntas con base en el contexto proporcionado.

a) **Según la información disponible y el contexto, discutan si las siguientes metodologías son o no apropiadas para estimar de manera consistente el efecto de interés. Justifiquen su respuesta en cada caso. Para aquella(s) que consideren pertinente(s), asegúrense de escribir la ecuación o ecuaciones que describen la estrategia de identificación, y mencionar qué parámetro o función de parámetros lo captura.**

   i) **Diferencia de medias entre todos los municipios que recibieron juzgados de descongestión en cualquiera de los años entre 2015 y 2020 y los que no.**
   Esta metodología no es adecuada ya que los municipios tratados y no tratados no son comparables. Como se dice en el enunciado, los municipios no fueron asignados aleatoriamente; los municipios tratados son los que tienen mayor número de casos pendientes. Si se estimara el efecto causal por esta metodología, este sería subestimado (o incluso cambiando de signo) debido a las diferencias estructurales entre tratados y controles (controles tienen una menor congestión de casos entonces ya habría una diferencia con los tratados).

   ii) **Regresión discontinua**
   Esta metodología no es adecuada ya que no se sabe con exactitud la regla de asignación. Sabemos que los municipios con más casos pendientes son asignados pero no sabemos a partir de cuántos casos (v.g $D_i = 1$ si hay más de mil casos pendientes y $D_i = 0$ de lo contrario). Además, el número de casos pendientes no es la única regla de asignación, también se consideran los tipos de casos más prevalentes en el y el distrito judicial al que pertenecen. Esto hace que no podamos identificar un "salto" en la probabilidad de ser tratado o no, por lo que RD no es el método adecuado.

   iii) **Variables instrumentales**
   Esta metodología no es adecuada ya que no hay una variable en la base de datos que sirva como instrumento para la variable de tratamiento.

   iv) **Diferencias en diferencias en contexto dinámico.**
   Esta metodología es pertinente debido a tenemos un buen grupo de control que sirva como contrafactual de los tratados (hay municipios que son tratados y otros que no). Específicamente, dif en dif permite que existan diferencias estructurales entre el grupo de tratados y de control (mientras que de diferencia de medias no); sólo nos pide que tengan tendencias paralelas (que ambas evolucionen de la misma manera a través del tiempo). Además, como hay distintos periodos de tratamiento y *perfect-*

complience, hacer dif en dif dinámico es la mejor opción para capturar el efecto causal.

b) Listen los supuestos que se deben cumplir para capturar el efecto causal de interés si se implementa un estudio de evento usando una especificación TWFE. Expliquen si estos supuestos se cumplen en el contexto del ejercicio. Escriban y expliquen las ecuaciones que describen estos supuestos.

Tendencias paralelas generalizadas: Este supuesto nos dice que el comportamiento del grupo de tratamiento en ausencia de tratamiento hubiera sido el mismo que el grupo de control.

La ecuación que caracteriza este supuesto es $\mathbb{E}(Y_{i1}^0 - Y_{i0}^0 \big| D_i = 1) = \mathbb{E}(Y_{i1}^0 - Y_{i0}^0 \big| D_i = 0)$

No existe razón para pensar que la actividad judicial de los municipios (tratados o control) evolucionen de manera diferente.

No hay efectos anticipatorios: Este supuesto nos dice que las causas son anteriores a las consecuencias. No pueden haber efectos del tratamiento antes que este sea implementado.

La ecuación que caracteriza este supuesto es: $Y_{it}(g) = Y_{it}(\infty)\ \forall t < g$

Es razonable pensar que no existen efectos anticipatorios ya que no tiene mucho sentido que las personas empiecen a interponer más demandas sólo porque se anuncia que van a abrir más juzgados en ese municipio.

Efectos constantes/homogéneos: Los efectos son constantes en el tiempo y no varían entre cohortes

La ecuación que caracteriza este supuesto es: $\mathbb{E}(Y_{ist}^1 - Y_{ist}^0 | s, t) = \tau\ \forall s, t$

No es tan plausible que se cumpla este supuesto debido a que probablemente los municipios tengan efectos diferentes gracias a las características idiosincráticas de cada municipio (no hay 2 municipios exactamente iguales por lo que es probable que su reacción al tratamiento tampoco sea igual). Por otra parte, también es probable que hayan efectos dinámicos debido a que las personas ajustan su comportamiento al ser tratados (intuición similar a efecto Hawthorne). No es irrazonable pensar que las personas aumenten su actividad judicial porque ahora saben que hay más jueces que pueden solucionar su caso.

c) Transformen la información proporcionada en la base de datos para generar una base de datos panel, con la cual puedan estimar las ecuaciones que propongan en los siguientes incisos. Asegúrense de definir las variables del tiempo relativo al tratamiento, los leads y los lags. Presenten el código que utilizaron como respuesta al enunciado.[2]

Código:

[2] Se debe presentar el código como parte de la respuesta aún si se entrega el código completo como anexo al taller.

```
use datos_CSJ.dta, replace
reshape long IEP_, i(id_municipio) j(año)
xtset id_municipio año, y
gen t = año - first_year if first_year != 0
replace t = 0 if t==.
tab t, gen(evt)
forvalues x = 1/9 {
local j= 10-`x'
ren evt`x' evt_l`j'
cap label var evt_l`j' "-`j'"
}
forvalues x= 0/4{
local j= 10 +`x'
ren evt`j' evt_f`x'
cap label var evt_f`x' "`x'"
}
replace evt_l1=0
```

d) Estimen el efecto de la creación de juzgados de descongestión por medio de un TWFE dinámico, con adopción escalonada del tratamiento. Presenten sus resultados de manera gráfica. ¿Existe un efecto de la implementación de juzgados de descongestión sobre el IEP? ¿Qué les sugiere este ejercicio acerca del cumplimiento de los supuestos de identificación de la metodología? Presenten el código utilizado.

Artículos recientes han mostrado que, en contextos en los que hay más de dos periodos y las unidades son tratadas en diferentes momentos, los coeficientes estimados de los modelos TWFE pueden no capturar un promedio ponderado de los efectos del tratamiento. Los problemas se presentan cuando hay efectos dinámicos (es decir, que no son constantes en el tiempo) y/o efectos heterogéneos (e.g., efectos diferentes según el valor de una covariable X). El problema es que las regresiones TWFE hacen comparaciones no válidas entre unidades que ya han sido tratadas y aquellas que reciben el tratamiento (Sun y Abraham, 2021). Cuando los efectos del tratamiento son heterogéneos, estas comparaciones pueden mostrar coeficientes que

incluso presenten el signo opuesto a todos los efectos de tratamiento a nivel individual debido a problemas de ponderación negativa (Goodman-Bacon, 2021).[3]

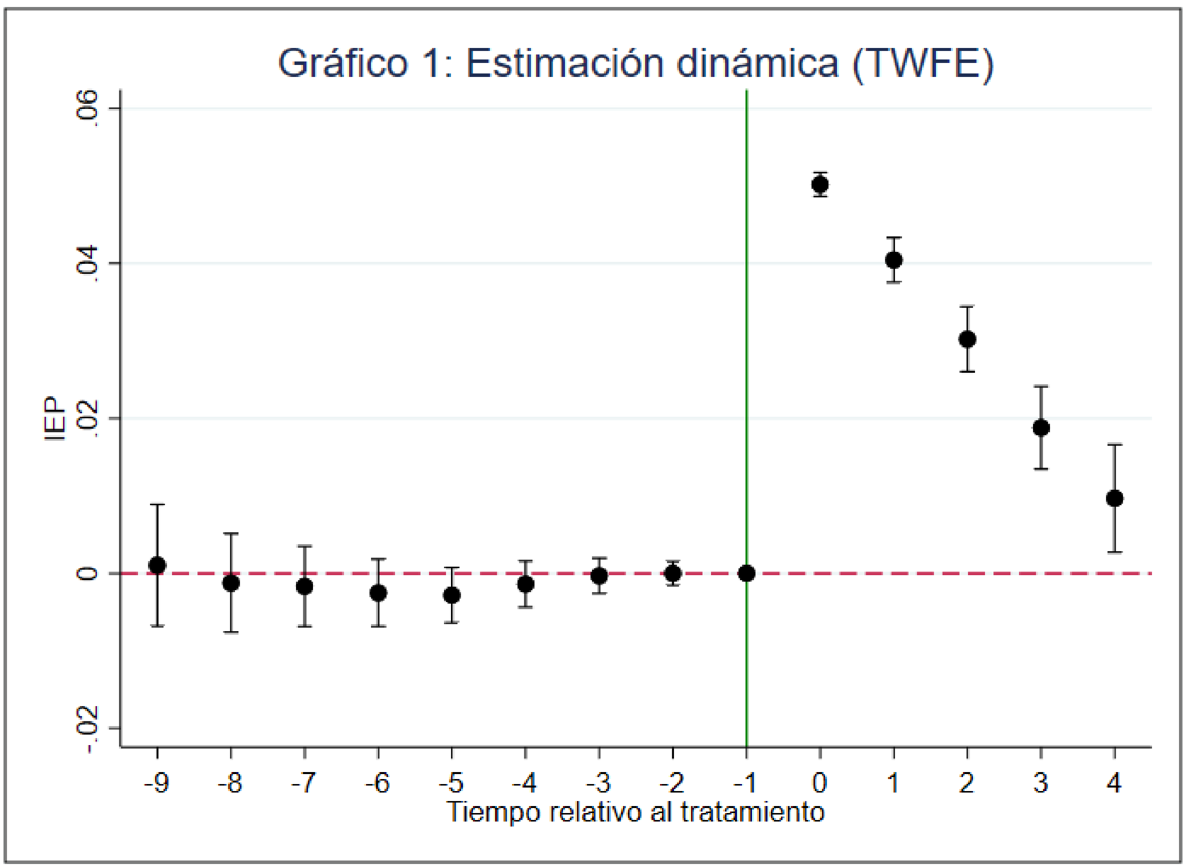
Código:

```
reghdfe IEP_ evt_l* evt_f*, nocon absorb(id_municipio año) cluster(id_municipio)

estimates store coefs_i

coefplot coefs_i, omitted vertical label drop(_cons) ytitle("IEP") xtitle("Tiempo relativo al tratamiento") yline(0, lpattern(dash)) xline(10, lcolor(green)) graphregion(lcolor(black) fcolor(white)) ciopts(lcol(black) recast(rcap) lwidth(*0.8)) mlcolor(black) mfcolor(black) levels(95)

graph export coefs_IEP.png, replace
```



Como se puede ver en la gráfica, después del tratamiento, los municipios tratados muestran una descongestión de casos. Específicamente, el primer año en el que fue implementado se evidencia un aumento de 0.05 puntos porcentuales en el IEP. Sin embargo, estos efectos al cabo de unos años van disminuyendo por lo que se puede concluir que existen efectos dinámicos o heterogéneos (se incumple supuesto de efectos constantes/homogéneos).

[3] Para una discusión general sobre los problemas que se han identificado ver Roth et al. (2022)

Además, en los periodos anteriores al tratamiento, se puede ver que los estimadores no son significativamente diferente de 0. Esto nos da indicios de que sí existen tendencias paralelas y no hay efectos anticipatorios (aunque tendencias paralelas no se puede comprobar al 100%). Si los estimadores fueran diferentes de 0 entonces eso significaría que habría un cambio en las diferencias entre tratados y controles antes del tratamiento (lo que significa que no hay tendencias paralelas o hay efectos anticipatorios).

e) Investiguen y expliquen en sus palabras cómo los estimadores de Sun y Abraham (2021) y los de Callaway y Sant'Anna (2021) lidian con los problemas mencionados.

Sun&Abraham (2021) y Callaway&Sant'Anna (2021) lidian con el problema de efectos dinámicos de manera similar (S&A es un caso especial de C&S). Para solucionar esto, los estimadores evitan hacer las comparaciones "prohibidas" entre los ya tratados (usados como control) y los recientemente tratados. Específicamente, lo que hacen es estimar los *leads* y *lags* de cada cohorte usando los no tratados como controles ponderado por el tamaño de su cohorte. De esta manera se identifica los efectos específicos de cada cohorte por año (vg. El efecto de los tratados en 2019, 2 años después de tratamiento). La diferencia entre C&S y S&A es que S&A toma como control el grupo de nunca-tratados (de haberlo) o el último grupo en ser tratado (excluyendo el periodo en el que este fue tratado) mientras que C&S toma como control todos los grupos que aún no han sido tratados. Por último, C&S también utiliza Bootstrap para corregir la posible autocorrelación entre los errores del mismo cohorte pero en diferentes periodos.

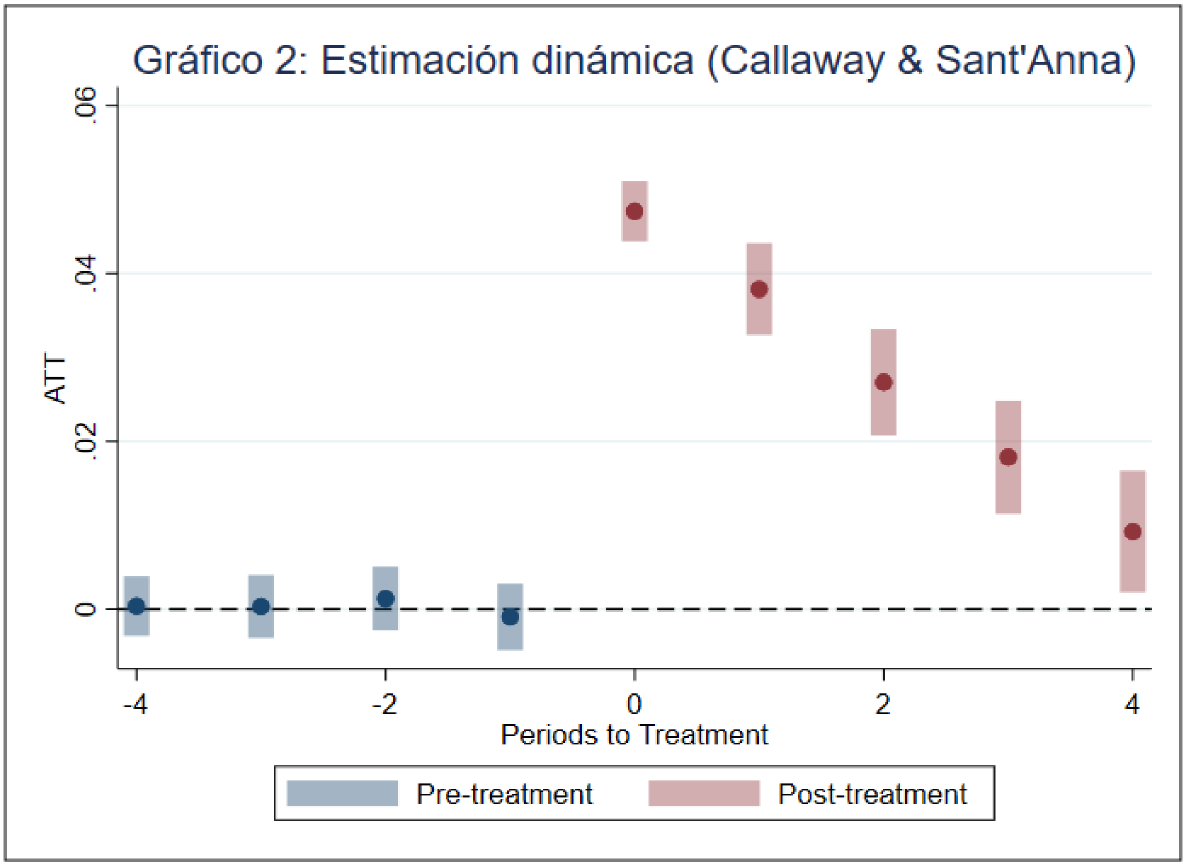
f) **Estimen el efecto del tratamiento utilizando el estimador de Callaway y Sant'Anna (2021) y presenten gráficamente los efectos dinámicos del tratamiento para quienes fueron tratados en 2015. ¿Cómo cambian sus conclusiones con respecto a lo encontrado en el inciso (d)? Presenten el código utilizado.**

Código:

```
qui csdid IEP_, ivar(id_municipio) time(año) gvar(first_year)

csdid_plot, group(2015) [title("Gráfico 2: Estimación dinámica (Callaway & Sant'Anna)") xline(9, lcolor(green)) graphregion(lcolor(black) fcolor(white))]

graph export graph_c&s.png, replace
```



Usando el estimador de C&S se puede ver que no hay diferencias grandes con la gráfica del literal (d). Ambos muestran que, antes del tratamiento, no hay cambios en las diferencias entre tratados y control lo que sugiere tendencias paralelas y no efectos anticipatorios. Después de tratamiento, se evidencia un aumento en el IEP de 0.05 puntos porcentuales que luego va disminuyendo con el tiempo. La diferencia entre (d) y (f) es que este último sólo tiene en cuenta el cohorte de 2015 (el efecto estimado de cada año es específico a esta cohorte). El hecho de que los efectos estimados, dado un año relativo, sean muy similares a los del literal (d) nos indica que no hay efectos heterogéneos entre cohortes sino más bien efectos dinámicos. En conclusión, aumentar los juzgados sí ayuda a descongestionar el sistema judicial, empero, este efecto va perdiendo peso con el paso de los años. Como consecuencia, al cabo de unos años se va a volver al mismo nivel de congestión que se tenía antes, a pesar del aumento de juzgados.

Tercer ejercicio

El objetivo básico del análisis estadístico es "extraer toda la información de los datos" (Rao, 1989)[4] para deducir propiedades de la población de la cual provienen. Para realizar inferencia estadística generalmente nos basamos en estadísticos, que son funciones de los datos y se seleccionan de acuerdo con alguna propiedad deseada (Shao & Tu, 1995)[5].

Antes de la recopilación de datos, un estadístico es una variable aleatoria que tiene una distribución de probabilidad, denominada distribución muestral del estadístico. La mayoría de los procedimientos estadísticos requieren cierto conocimiento de la distribución muestral del estadístico que se utiliza para el análisis. El tipo de conocimiento necesario depende del tipo de análisis. La construcción de intervalos de confianza y la prueba de hipótesis, por ejemplo, requieren el conocimiento de la distribución de muestreo en sí o de los percentiles de la distribución de muestreo. Por otro lado, en un problema de estimación, es fundamental tener una indicación de la precisión del estimador, ya que cualquier estimador tiene un error de estimación. Por lo tanto, en este caso se requiere el conocimiento de medidas de precisión como la varianza, el sesgo y el error cuadrático medio del estimador. Estas medidas de precisión son características de la distribución muestral del estimador (Shao & Tu, 1995).

La distribución muestral de un estadístico y sus características suelen depender de la población subyacente. Cuando desconocemos esta distribución, métodos como Jackknife y Bootstrap nos permiten aproximarla (Shao & Tu, 1995). En términos generales, el objetivo de Bootstrap es hacer inferencia sobre una estimación (como la media muestral) para un parámetro poblacional (como la media poblacional) usando los datos de la muestra (Yen, 2019). En este método no se hacen supuestos sobre la distribución de los datos y se puede evaluar la variabilidad de prácticamente cualquier estadístico (Yen, 2019; Shao & Tu, 1995), lo que lo hace poderoso y flexible. Considerando las ventajas de esta metodología, en este ejercicio vamos a centrarnos en entender Bootstrap y lo emplearemos para efectuar inferencia estadística. Para esto, vamos a partir de un caso donde usemos la metodología de MC2E.

Suponga que usted está interesada en la siguiente relación poblacional:

$$\boldsymbol{Y_i = \beta_0 + \tau D_i + \gamma M_i + u_i} \qquad \textbf{(1)}$$

---

[4] Rao, C. R. (1989). Statistics and Truth. Putting Chance to Work, International Cooperative Publishing House, Burtonsville, Md.

[5] Shao, J., & Tu, D., (1995). The Jackknife and Bootstrap. Chapter 4. Bootstrap Confidence Sets and Hypothesis Tests. Springer: New York.

donde $D_i \in \{0,1\}$ es una variable dicótoma cuyo valor depende de si el individuo fue tratado o no[6] y M_i tiene una distribución normal estándar. De la misma forma, u_i se distribuye como una normal estándar. Adicionalmente, usted sabe que $\beta_0 = 0.5$, $\tau = 2$, $\gamma = 1$.

Ahora bien, en la realidad M_i no es observado. Por tanto, el modelo que pueden estimar es el siguiente:

$$Y_i = \beta_0 + \tau D_i + \varepsilon_i, (2)$$

donde $\varepsilon_i = u_i + \gamma M_i$. Ustedes tienen (observan) una variable alterna, A_i, que denota la asignación al tratamiento. Saben que esta variable es independiente y se distribuye como una Bernoulli cuya probabilidad de éxito es 0.5. Del mismo modo, conocen la relación entre D_i y las demás variables:

$$D_i = 1\{0.3 * A_i + 0.15 * M_i + v_i \geq 0.5\} \quad (3)$$

donde $v_i \sim Normal(0.2, 0.05)$.

1. **Para comenzar, simularán el proceso generador de datos expuesto previamente y usarán la información suministrada para demostrar la existencia de endogeneidad y corregirla.**
   a. **Primero, simulen una base de datos acorde al proceso expuesto previamente con un tamaño de muestra de 10.000 individuos. Incluyan un identificador único para cada una de las observaciones. Guarden la base de datos resultante. Presenten el código utilizado en el ejercicio.[7]**

   Código:
   ```
   set obs 10000
   set seed 201922665
   gen id = _n
   gen Mi = rnormal(0,1)
   gen Ui = rnormal(0,1)
   gen Ai = rbinomial(1, 0.5)
   gen Vi = rnormal(0.2, 0.05)
   gen Di = 1 if (0.3*Ai + 0.15*Mi + Vi) >= 0.5
   replace Di = 0 if Di==.
   gen Yi = 0.5 + 2*Di + 1*Mi + Ui
   save “datos_simulados.dta”, replace
   ```

---

[6] Puede prescindir de la información de una política o programa puntual. La interpretación puede ser en unidades.

[7] Se debe presentar el código como parte de la respuesta aún si se entrega el código completo como anexo al taller.

b. **Estimen el modelo (2) por MCO. Interpreten los resultados obtenidos y expliquen si existe (o no) sesgo y cuál es su causa. Expliquen si la metodología utilizada es adecuada o, en su defecto, propongan una metodología que les permita capturar el efecto causal de D_i sobre la variable Y. Listen los supuestos de identificación que se deben cumplir para estimar un efecto causal y si es factible que se cumplan.[8] Sean detallados en la explicación de la metodología y de los supuestos[9].**

Estimación por MCO

	(1)
VARIABLES	Yi
Di	3.145***
	(0.0292)
Constant	0.182***
	(0.0153)
Observations	10,000
R-squared	0.537

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Dado que conocemos los valores que deberían de tomar los estimadores, podemos observar que existe un sesgo que sobreestima el valor de τ (debería ser igual a 2). Este sesgo se da por la existencia de una variable omitida no observable (M_i) la cual sabemos que afecta tanto a la variable de resultado como a la variable de tratamiento (dada la estructura de los datos).

Dado el problema de endogeneidad, estimar la ecuación (2) a través de mínimos cuadrados ordinarios no es adecuado puesto que no se cumple el supuesto de exogeneidad estricta. Dicho esto, para corregir el problema de endogeneidad es posible plantear un modelo con una variable instrumental (A_i) ya que esta está relacionada a la variable de tratamiento y afecta el resultado únicamente a través de la misma. Los supuestos de identificación de esta metodología son:

i. Relevancia: El instrumento debe ser relevante para la estimación de la variable endógena. Esto es factible ya que, conociendo el proceso generador de los datos, el instrumento hace parte del proceso que determina a la variable de tratamiento.
ii. Independencia: La variable instrumental no debe estar correlacionada con el termino de error. Lo anterior es factible ya que el termino de error no hace parte del proceso generador de la variable instrumental.
iii. Exclusión: La variable instrumental solo afecta a la variable de resultado a través de la variable endógena (D_i). Anteriormente se discutió por qué esto es factible.

[8] Consideren solo las variables expuestas en el ejercicio.
[9] Recuerden, M_i es una variable no observada y deseamos estimar de forma consistente el efecto de D_i sobre Y_i

iv. Monotonicidad: La variable instrumental afecta a todas las unidades de observación de la misma manera. En este caso también se cumple dado que el instrumento A_i solo aumenta la probabilidad de ser tratado para todas las unidades.

c. **Una compañera les sugiere usar MC2E. Planteen las ecuaciones para la primera y la segunda etapa. Estimen cada una de ellas por separado[10] y presenten los resultados. Reporten el valor del estadístico F en una prueba de relevancia. Concluyan sobre: i) si el instrumento propuesto cumple con los supuestos necesarios para realizar inferencia causal; ii) si el instrumento es lo suficientemente fuerte y, iii) si se corrige el problema de endogeneidad. ¿Consideran que esta recomendación es adecuada?**

Adicionalmente, expliquen si se puede usar el error estándar de $\hat{\tau}(\hat{\sigma}_\tau)$de la segunda etapa para realizar inferencia estadística. De no ser así, expliquen qué metodología podría corregir este problema.

Las ecuaciones necesarias para estimar el efecto causal del tratamiento mediante la metodología de variables instrumentales son las siguientes:
Primera etapa: $\widehat{D_i} = \gamma_0 + \gamma_1 A_i + \eta_i$
Segunda etapa: $Y_i = \beta_0 + \tau\widehat{D_i} + \epsilon_i$
Al estimar estas ecuaciones obtenemos la siguiente tabla de regresión:

Estimación por MC2E

VARIABLES	(1) primera etapa	(2) segunda etapa
Ai	0.480***	
	(0.00753)	
Di_hat		2.083***
		(0.0771)
Constante	0.0305***	0.474***
	(0.00537)	(0.0281)
Observaciones	10,000	10,000
R-cuadrado	0.289	0.068
F calculado	4065	
F tabulado	6.637	

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

[10] Esto es, estimen la primera etapa, y luego usen sus resultados para estimar la segunda etapa.

Los supuestos que discutimos en el inciso anterior se cumplen dadas las mismas razones expuestas anteriormente. La significancia estadística y el estadístico F de la primera etapa demuestran el supuesto de relevancia de la variable instrumental, esto significa que es un instrumento lo suficientemente fuerte para corregir el problema de endogeneidad y así poder realizar inferencia causal del coeficiente τ. Ahora bien, el error estándar del estimador no es el apropiado dado que desconocemos la distribución muestral del estadístico. El estimador, por muy cercano que es al valor real, no es completamente preciso, de manera que debemos calcular su error estándar a través de la metodología de bootstrapping. Dicha metodología consiste en realizar la estimación del parámetro a través de una cantidad B de submuestras de la muestra original para así aprovechar la ley de los grandes números y aproximarse a la media y varianza del parámetro en cuestión de forma asintótica. Visto en ecuaciones sería lo siguiente:

$$\bar{\tau} = \frac{1}{B}\sum_{j=1}^{B} \tau_j \rightarrow E(\tau)\,, en\ cuanto\ B \rightarrow \infty$$

$$s^2 = \frac{1}{B}\sum_{j=1}^{B} \left(\tau_j\right)^2 - \left(\frac{1}{B}\sum_{j=1}^{B} \tau_j\right)^2 \rightarrow Var(\tau), en\ cuanto\ B \rightarrow\ \infty$$

2. **Una vez realizada la estimación usando MC2E manualmente, sabemos que no se puede realizar inferencia estadística sobre los estimadores con los errores estándar reportados en la segunda etapa. De esta forma, vamos a crear un programa para corregir este problema.**
   a. **Creen un programa en donde se realice el proceso de MC2E y bootstrap sobre el coeficiente asociado a $\widehat{D}_i(\hat{\tau})$. Denotaremos la distribución obtenida al realizar bootstrap como $\hat{\tau}_n^*$. Para ello, utilicen el comando *program define*.[11] Presenten el código utilizado.**

Código:

```
clear all
use datos_simulados.dta, replace

program define bstrp, rclass
preserve
bsample
quiet reg Di Ai
predict Di_hat
quiet reg Yi Di_hat
```

[11] Recuerden que, en palabras, el algoritmo detrás de bootstrap consiste en hacer un remuestreo aleatorio con repetición del mayor número de unidades posibles, estimar el modelo de MC2E (para este problema en particular) y guardar las estimaciones para el coeficiente que les interesa

```
return scalar tau_hat = _b[Di_hat]
return scalar se_bstrp = _se[Di_hat]
restore
end
```

b. Usen el comando simulate para ejecutar el programa. Realicen 5000 iteraciones y recuerden fijar una semilla para que sus resultados sean replicables. Guarden la información obtenida en una base de datos. Presenten el código utilizado e incluyan una tabla de estadísticas descriptivas (que tenga la siguiente forma) donde comparen los resultados de la segunda etapa y los resultados luego de aplicar bootstrap. En particular, la información para $\hat{\tau}$ y $\hat{\tau}_n^*$ con sus respectivos errores estándar. Interpreten los resultados.

Tabla 1: MC2E y *bootstrap*

Método	Estimador
MC2E	$(\hat{\tau})$
	$(\hat{\sigma})$
bootstrap	$(\hat{\tau}_n^*)$
	$(\hat{\sigma}_n^*)$

Tabla 1: MC2E y bootstrap

Método	Estimador
MC2E	2.08251
	0.07709
bootstrap	2.08155
	0.05835

El parámetro estimado mediante mínimos cuadrados en dos etapas es insesgados y consistente, de manera que al calcularlo mediante el método de bootstrap cambia muy poco su valor. En cuanto al error estándar que se obtiene mediante mínimos cuadrados en dos etapas, este no es el apropiado para el ejercicio dado que el error ϵ_i está compuesto por $u_i + \gamma M_i$ de manera que el error que se estima en la segunda etapa no corresponde al error poblacional. Para corregir esto, al utilizar el método de bootstrap, aprovechamos la ley de los grandes números para así encontrar el error que corresponde a la población total. En este caso encontramos que el error estimado mediante bootstrap es menor al estimado mediante MC2E.

Código:

```
simulate tau_estimado = r(tau_hat) error_s = r(se_bstrp), reps(5000) seed(201922665): bstrp
save "coefs_bootstrap.dta", replace
use coefs_bootstrap.dta, replace
mat tabla = J(4,1,.)
estpost sum tau_estimado
```

```
mat tabla[3,1] = e(mean)
mat tabla[4,1] = e(sd)

use datos_simulados.dta, replace
reg Di Ai
predict Di_hat
reg Yi Di_hat
mat se = r(table)
mat tabla[1,1] = _b[Di_hat]
mat tabla[2,1] = se[2,1]
frmttable using "tabla_MC2E&bootstrap.doc", replace sdec(5) stat(tabla) title("Tabla 1: MC2E y bootstrap") ctitle("Metodo","Estimador") rtitles("MC2E"\""\"bootstrap"\"")
```

3. **Ahora bien, una vez estimada la distribución del estimador, queda la pregunta respecto a cómo realizar inferencia estadística. En especial, se desea saber cómo construir los intervalos de confianza del estimador dado que no existe certeza sobre su distribución. Así, hay diversas formas de aproximarse a los valores críticos de una distribución a partir de bootstrap. En este ejercicio vamos a explorar 3 formas de computar los intervalos de confianza: bootstrap percentile, normal bootstrap y bootstrap-t.**
   a. **Primero, vamos a iniciar con el método más simple: bootstrap percentile. Con bootstrap percentile se construyen los intervalos de confianza a partir de la distribución aproximada del estimador luego de aplicar bootstrap tomando los percentiles de dicha distribución. Así, la forma del intervalo de confianza está dado por:**

$$\boldsymbol{IC_{1-\alpha} = [\tau^*_{n,\alpha/2}, \tau^*_{n,1-\alpha/2}]}$$

Donde $\boldsymbol{\tau^*_{n,p}}$ denota el percentil p de la distribución $\boldsymbol{\tau^*_n}$. Presenten el código para estimar los intervalos de confianza, con 95 % de confianza, usando este método.

Tabla 2: Intervalo de confianza bootstrap percentile (95%)

	$\tau^*_{n,\alpha/2}$	$\tau^*_{n,1-\alpha/2}$
IC	1.96637	2.19442

Código:

```
use coefs_bootstrap.dta, replace
mat IC_p = J(1,2,.)
_pctile tau_estimado, p(2.5, 97.5)
return list
mat IC_p[1,1] = `r(r1)'
mat IC_p[1,2] = `r(r2)'
```

```
frmttable using "IC_p.doc", replace stat(IC_p) sdec(5) title("Tabla 2: Intervalo de confianza bootstrap percentile (95%)") ctitle("", "", "") rtitles("IC")
```

b. Otra posible metodología es normal bootstrap. En este caso, suponemos que la distribución poblacional es una normal. De esta forma, los puntos críticos están dados los percentiles de una normal, a la cual se le imputa el error estándar estimado con bootstrap como desviación estándar. De esta forma, los intervalos de confianza toman esta forma:

$$\boldsymbol{IC_{1-\alpha} = [\hat{\tau} - z_{1-\frac{\alpha}{2}} * \hat{\sigma}^*;\ \hat{\tau} + z_{1-\frac{\alpha}{2}} * \hat{\sigma}^*]}$$

Presenten el código para estimar los intervalos de confianza del 95 % usando este método.

Tabla 3: Intervalo de confianza normal bootstrap (95%)

	$\hat{\tau} - z_{1-\frac{\alpha}{2}} * \hat{\sigma}^*$	$\hat{\tau} + z_{1-\frac{\alpha}{2}} * \hat{\sigma}^*$
IC	1.96814	2.19688

Código:

```
use datos_simulados.dta, replace
reg Di Ai
predict Di_hat
reg Yi Di_hat
scalar t_hat = _b[Di_hat]

use coefs_bootstrap.dta, replace
estpost sum tau_estimado
scalar sigma = e(sd)[1,1]
display sigma

*Los valores de z vienen de la tabla de distribución normal. Entonces Z_1-alpha/2 = 1.96
mat IC_n = J(1,2,.)
mat IC_n[1,1] = t_hat - (1.96*sigma)
mat IC_n[1,2] = t_hat + (1.96*sigma)
frmttable using "IC_n.doc", replace stat(IC_n) sdec(5) title("Tabla 3: Intervalo de confianza normal bootstrap (95%)") ctitle("", "", "") rtitles("IC")
```

c. Finalmente, examinaremos el método bootstrap-t. Con este método se centra la distribución estimada con bootstrap ($\boldsymbol{\hat{\tau}_n^*}$) en torno al parámetro estimado ($\boldsymbol{\hat{\tau}}$). De esta forma, se construye una nueva distribución

$$\boldsymbol{t}_n^*(x) = \hat{\tau}_n^*(\frac{x-\hat{\tau}}{\hat{\sigma}_n^*})$$

Donde $\widehat{\boldsymbol{\sigma}}_{\boldsymbol{n}}$ es la desviación estándar de la distribución $\hat{\boldsymbol{\tau}}_{\boldsymbol{n}}^*$. A partir de esta distribución se computar los valores críticos de forma que los intervalos estarían dados por:

$$\boldsymbol{IC}_{\boldsymbol{1-\alpha}\,\%} = [\hat{\boldsymbol{\tau}}_{\boldsymbol{n}} - \boldsymbol{t}_{\boldsymbol{n},\boldsymbol{1-\alpha/2}}^* * \widehat{\boldsymbol{\sigma}}_{\boldsymbol{n}}^* \,;\, \hat{\boldsymbol{\tau}}_{\boldsymbol{n}} + \boldsymbol{t}_{\boldsymbol{n},\boldsymbol{1-\alpha/2}}^* * \widehat{\boldsymbol{\sigma}}_{\boldsymbol{n}}^*]$$

Presente el código para estimar los intervalos de confianza, con 95 % de confianza, usando este método.

Tabla 4: Intervalo de confianza bootstrap t (95%)

	$\hat{\tau}_n - t_{n,1-\frac{\alpha}{2}}^* * \hat{\sigma}_n^*$	$\hat{\tau}_n + t_{n,1-\frac{\alpha}{2}}^* * \hat{\sigma}_n^*$
IC	1.97060	2.19442

```
use datos_simulados.dta, replace
reg Di Ai
predict Di_hat
reg Yi Di_hat
scalar t_iv = _b[Di_hat]

use coefs_bootstrap.dta, replace
estpost sum tau_estimado
scalar sigma = e(sd)[1,1]
gen t_n = (tau_estimado-t_iv)/sigma
_pctile t_n, p(97.5)
mat IC_t = J(1,2,.)
mat IC_t[1,1] = t_iv - (`r(r1)'*sigma)
mat IC_t[1,2] = t_iv + (`r(r1)'*sigma)
frmttable using "IC_t.doc", replace stat(IC_t) sdec(5) title("Tabla 4: Intervalo de confianza bootstrap t (95%)") ctitle("", "", "") rtitles("IC")
```

d. En una sola tabla presenten los intervalos de confianza estimados con cada una de las metodologías previas. Comparen el valor del estimador, el error estándar y los intervalos de confianza con los resultados al estimar IV de forma directa (usando el comando ivreg2). Interpreten los resultados.

Tabla 5: Intervalos de confianza (95%)

Metodo	IC (Minimo)	IC (Maximo)
Bootstrap percentile	1.96637	2.19442
Normal Bootstrap	1.96814	2.19688

Bootstrap t		1.97060	2.19442

Tabla 6: resultados del comando ivreg2

Método	Coeficiente	Error Estándar	IC (Mínimo)	IC (Máximo)
MC2E	2.08251	0.05782	1.96918	2.19584

La estimación de los coeficientes, los errores estándar y los intervalos de confianza mediante bootstrap y MC2E utilizando el comando ivreg2 es equivalente. Nótese que los resultados de las tablas 1 y 6 exponen la similitud entre los coeficientes y errores estándar de bootstrap y ivreg2; y las tablas 5 y 6 muestran lo mismo para los intervalos de confianza. Sin embargo, el error estimado por el comando de ivreg2 no es igual al calculado mediante MC2E manual. Lo anterior se debe a que este comando toma estadísticos de muestras grandes (z y chi-cuadrado) y calcula sus errores de forma asintótica. Como explicamos anteriormente, bootstrap también es una metodología que realiza el cálculo de los parámetros y errores asintóticamente, es por esto que los resultados son tan similares entre ivreg2 y bootstrap, pero no lo son para MC2E manual.

4. **Ahora bien, hasta el momento hemos realizado las estimaciones suponiendo un muestreo aleatorio que es i.i.d, lo que implica errores homocedásticos y sin correlación serial. Vamos a levantar el segundo supuesto, es decir, vamos a suponer que existe correlación serial entre las unidades. Para ello, vamos a modificar la base de datos usada al principio. Sigan los siguientes pasos:**
   - **Abran la base de datos creada al principio del ejercicio.**
   - **Por medio del comando expand, produzcan 10 copias de los datos (ahora tendrán una base de datos con 100.000 observaciones).**

   a. **Expliquen por qué este procedimiento induce un problema de autocorrelación.**

Este método de expansión de la muestra genera duplicados de las observaciones ya existentes. Por definición, un par de observaciones duplicadas tendrá errores correlacionados entre sí, lo cual genera el problema de autocorrelación en la muestra expandida.

   b. **Vuelvan a correr el programa que implementaron en el inciso 2.a.) sobre la nueva base de datos. Presenten el error estándar estimado usando el método de bootstrap. Comparen el resultado del inciso 2.b.). ¿Qué pueden concluir?**

Tabla 7: coeficientes y errores estandar

Metodo	Estimador
MC2E	2.08251
	0.07709
bootstrap	2.08155
	0.05835

bootstrap expandido	2.08278 0.01860

Las filas de MC2E y bootstrap son aquellas encontradas en el inciso 2.b). La tabla 7 añade el coeficiente y error estándar del método bootstrap expandido. Como pueden notar, al aplicar bootstrap a la muestra expandida, el error estándar disminuye significativamente. Esto es ocasionado por el problema de autocorrelación que discutimos en el inciso anterior ya que las submuestras creadas pueden contener varias observaciones con el mismo id. Esto induce una menor variación de los parámetros estimados y por ende un error estándar menor.

c. **Estimen los intervalos de confianza expuestos en los incisos 3.a.), 3.b.) y 3.c.). Presenten el código utilizado.**

Código:

```
*IC percentile
use bootstrap_expanded.dta, replace
mat IC_p_expanded = J(1,2,.)
_pctile tau_estimado, p(2.5, 97.5)
return list
mat IC_p_expanded[1,1] = `r(r1)'
mat IC_p_expanded[1,2] = `r(r2)'

*IC normal
use datos_simulados.dta, replace
reg Di Ai
predict Di_hat
reg Yi Di_hat
scalar t_hat = _b[Di_hat]

use bootstrap_expanded.dta, replace
estpost sum tau_estimado2
scalar sigma_expanded = e(sd)[1,1]
display sigma_expanded

mat IC_n_expanded = J(1,2,.)
mat IC_n_expanded[1,1] = t_hat - (1.96*sigma_expanded)
mat IC_n_expanded[1,2] = t_hat + (1.96*sigma_expanded)

*IC t
use datos_simulados.dta, replace
reg Di Ai
predict Di_hat
reg Yi Di_hat
scalar t_iv = _b[Di_hat]
```

```
use bootstrap_expanded.dta, replace
estpost sum tau_estimado2
scalar sigma = e(sd)[1,1]
gen t_n = (tau_estimado-t_iv)/sigma
_pctile t_n, p(97.5)
mat IC_t_expanded = J(1,2,.)
mat IC_t_expanded[1,1] = t_iv - (`r(r1)'*sigma)
mat IC_t_expanded[1,2] = t_iv + (`r(r1)'*sigma)
```

d. **Realicen en una tabla los resultados obtenidos y compárenlos con los resultados al realizar IV de forma directa. Presenten una tabla similar a la presentada en el inciso 3.d.). Interpreten los resultados. En especial, ¿cómo se ve afectado el método de bootstrap, tanto para la corrección del error estándar como en la estimación de intervalos de confianza, cuando existe autocorrelación entre las unidades?**

Tabla 8: Intervalos de confianza (95%)

Metodo	IC (Minimo)	IC (Maximo)
Bootstrap percentile	2.04657	2.11862
Normal Bootstrap	2.04604	2.11897
Bootstrap t	2.04640	2.11862

Tabla 6: resultados del comando ivreg2 (copia)

Método	Coeficiente	Error Estándar	IC (Mínimo)	IC (Máximo)
MC2E	2.08251	0.05782	1.96918	2.19584

La menor variación de los estimadores inducida por el problema de autocorrelación originado en la expansión de la muestra se ve también reflejada en intervalos de confianza más pequeños; Nótese que el valor mínimo de los intervalos de confianza aumenta y a su vez disminuye el valor máximo. Por otro lado, el coeficiente estimado sigue siendo similar ya que el problema de autocorrelación solo induce ineficiencia del estimador, mas no afecta el insesgamiento del mismo.

5. **Dada la existencia de autocorrelación entre las unidades, una alternativa para corregir este problema es mediante el uso de errores cluster.**

a. **En este caso, creen un programa en el que corran la primera y segunda etapa usando errores estándar convencionales. Luego, utilicen el comando bootstrap. Realicen este proceso sobre los coeficientes de la segunda etapa con 5000**

repeticiones e incluyan clusters al momento de la estimación.[12] Comparen el error estándar corrigiendo por clusters con el error estándar del inciso inmediatamente anterior. Presenten esta información en una tabla. Incluyan en la solución del punto el código utilizado.

Tabla 9: Errores estandar (Estimación del error mediante bootstrap)

Método	Error Estandar
Sin corrección de clusters	0.01856
Corregido con clusters	0.05807

Al comparar el error estándar de bootstrap corregido por clusters con aquel obtenido a través del comando ivreg2 notamos que es el mismo. Esto se da pues al clusterizar los errores por unidad de observación estamos anulando el problema de autocorrelación. El comando bootstrap clusterizado tomará únicamente submuestras cuyas observaciones a clusters distintos, de manera que no se reduce la variación de los parámetros estimados y obtenemos el mismo error calculado asintóticamente por ivreg2.

Código:

```
use datos_simulados.dta, replace
expand(10)

program define bstrp_cluster, rclass
preserve
quiet reg Di Ai
cap drop Di_hat
predict Di_hat
quiet reg Yi Di_hat
end

bootstrap _b, rep(5000): bstrp_cluster
save "bootstrap_cluster_1.dta", replace
mat ee_bstrp_cluster = J(2,1,.)
mat ee_bstrp_cluster[1,1] = e(se)[1,1]

bootstrap _b, rep(5000) cluster(id): bstrp_cluster
save "bootstrap_cluster_2.dta", replace
mat ee_bstrp_cluster[2,1] = e(se)[1,1]
```

[12] **Pista:** Piensen cuáles son las observaciones que se correlacionan para escoger el nivel de cluster.

frmttable using "tabla_p5.doc", replace sdec(5) stat(ee_bstrp_cluster) title("Tabla 9: Errores estandar (Estimación del error mediante bootstrap)") ctitle("Metodo","Error Estandar") rtitles("Sin corrección de clusters"\"Corregido con clusters")

Bibliografía

Cunningham, S. (2022). *Illutrating the bias of TWFE in event studies.* Scott's substack. Recuperado de https://causalinf.substack.com/p/illustrating-the-bias-of-twfe-in

Cunningham, S. (2021).*Waiting for event studies: A play in three acts.* Scott's substack. Recuperado de https://causalinf.substack.com/p/waiting-for-event-studies-a-play

Anexos:

Código:

```
clear all

set more off

cd "C:\Users\Lenovo\OneDrive\Documents\8vo semestre\Econometría Avanzada\Taller 3"

cls


*************************************************************************
**

*EJERCICIO
2:************************************************************

*************************************************************************
**

*c)

use datos_CSJ.dta, replace

reshape long IEP_, i(id_municipio) j(año)

xtset id_municipio año, y


gen t = año - first_year if first_year != 0

replace t = 0 if t==.


tab t, gen(evt)
```

```
forvalues x = 1/9 {
    local j= 10-`x'
ren evt`x' evt_l`j'
cap label var evt_l`j' "-`j'"
}
forvalues x= 0/4{
local j= 10 +`x'
ren evt`j' evt_f`x'
cap label var evt_f`x' "`x'"
}
replace evt_l1=0


*d)
reghdfe IEP_ evt_l* evt_f*, nocon absorb(id_municipio año) cluster(id_municipio)


estimates store coefs_i


coefplot coefs_i, omitted vertical label drop(_cons) title("Gráfico 1: Estimación dinámica (TWFE)") ytitle("IEP") xtitle("Tiempo relativo al tratamiento") yline(0, lpattern(dash)) xline(9, lcolor(green)) graphregion(lcolor(black) fcolor(white)) ciopts(lcol(black) recast(rcap) lwidth(*0.8)) mlcolor(black) mfcolor(black) levels(95) xlabel(, labsize(small) nogextend labc(black)) ylabel(,nogrid nogextend labc(black) format(%9.2f)) xlabel(, labsize(small) nogextend labc(black))
graph export coefs_IEP.png, replace


*f)
qui csdid IEP_, ivar(id_municipio) time(año) gvar(first_year)
csdid_plot, group(2015) [title("Gráfico 2: Estimación dinámica (Callaway & Sant'Anna)") xline(9, lcolor(green)) graphregion(lcolor(black) fcolor(white))]
graph export graph_c&s.png, replace
```

```
*****************************************************************************
**
*EJERCICIO
3:*****************************************************************
*****************************************************************************
**
*1)
*a)simulacion de datos
clear all
set obs 10000
set seed 201922665
gen id = _n
gen Mi = rnormal(0,1)
gen Ui = rnormal(0,1)
gen Ai = rbinomial(1, 0.5)
gen Vi = rnormal(0.2, 0.05)
gen Di = 1 if (0.3*Ai + 0.15*Mi + Vi) >= 0.5
replace Di = 0 if Di==.
gen Yi = 0.5 + 2*Di + 1*Mi + Ui
save "datos_simulados.dta", replace

*b) estimacion por MCO
reg Yi Di
outreg2 using "MCO_1.doc",replace title(Estimacion por MCO) ctitle(Yi)

*c)
reg Di Ai
```

```
outreg2 using "MC2E_1.doc", replace title(Estimacion por MC2E) ctitle(primera etapa) addstat("F calculado", `e(F)', "F tabulado", 6.63743)

predict Di_hat


reg Yi Di_hat

outreg2 using "MC2E_1.doc", append ctitle(segunda etapa)



*2)--------------------------------------------------------------------

*2a)

clear all

use datos_simulados.dta, replace


program define bstrp, rclass

preserve

bsample

quiet reg Di Ai

predict Di_hat

quiet reg Yi Di_hat

return scalar tau_hat = _b[Di_hat]

return scalar se_bstrp = _se[Di_hat]

restore

end


simulate tau_estimado = r(tau_hat) error_s = r(se_bstrp), reps(5000) seed(201922665): bstrp

save "coefs_bootstrap.dta", replace


use coefs_bootstrap.dta, replace

mat tabla = J(4,1,.)
```

```
estpost sum tau_estimado
mat tabla[3,1] = e(mean)
mat tabla[4,1] = e(sd)

use datos_simulados.dta, replace
reg Di Ai
predict Di_hat
reg Yi Di_hat
mat se = r(table)
mat tabla[1,1] = _b[Di_hat]
mat tabla[2,1] = se[2,1]
frmttable using "tabla_MC2E&bootstrap.doc", replace sdec(5) stat(tabla) title("Tabla 1: MC2E y bootstrap") ctitle("Metodo","Estimador") rtitles("MC2E"\""\"bootstrap"\"")

*3)---------------------------------------------------------------------
*a) IC percentiles
use coefs_bootstrap.dta, replace
mat IC_p = J(1,2,.)
_pctile tau_estimado, p(2.5, 97.5)
return list
mat IC_p[1,1] = `r(r1)'
mat IC_p[1,2] = `r(r2)'
frmttable using "IC_p.doc", replace stat(IC_p) sdec(5) title("Tabla 2: Intervalo de confianza bootstrap percentile (95%)") ctitle("", "", "") rtitles("IC")

*b) IC normal
use datos_simulados.dta, replace
reg Di Ai
predict Di_hat
```

```
reg Yi Di_hat
scalar t_hat = _b[Di_hat]

use coefs_bootstrap.dta, replace
estpost sum tau_estimado
scalar sigma = e(sd)[1,1]
display sigma

*Los valores de z vienen de la tabla de distribución normal. Entonces Z_1-alpha/2 = 1.96
mat IC_n = J(1,2,.)
mat IC_n[1,1] = t_hat - (1.96*sigma)
mat IC_n[1,2] = t_hat + (1.96*sigma)
frmttable using "IC_n.doc", replace stat(IC_n) sdec(5) title("Tabla 3: Intervalo de confianza normal bootstrap (95%)") ctitle("", "", "") rtitles("IC")

*c)
use datos_simulados.dta, replace
reg Di Ai
predict Di_hat
reg Yi Di_hat
scalar t_iv = _b[Di_hat]

use coefs_bootstrap.dta, replace
estpost sum tau_estimado
scalar sigma = e(sd)[1,1]
gen t_n = (tau_estimado-t_iv)/sigma
_pctile t_n, p(97.5)
mat IC_t = J(1,2,.)
mat IC_t[1,1] = t_iv - (`r(r1)'*sigma)
```

```
mat IC_t[1,2] = t_iv + (`r(r1)'*sigma)

frmttable using "IC_t.doc", replace stat(IC_t) sdec(5) title("Tabla 4: Intervalo de confianza bootstrap t (95%)") ctitle("", "", "") rtitles("IC")


*d)

use datos_simulados.dta, replace

mat tabla_IC = (IC_p\IC_n\IC_t)

frmttable using "IC_todos.doc", replace stat(tabla_IC) sdec(5) title("Tabla 5: Intervalos de confianza (95%)") ctitle("Metodo", "IC (Minimo)", "IC (Maximo)") rtitles("Bootstrap percentile"\"Normal Bootstrap"\"Bootstrap t")


preserve

ivreg2 Yi (Di = Ai)

regsave Di, ci

mkmat coef stderr ci_lower ci_upper, matrix(tabla_ivreg2)

restore


frmttable using "tabla_ivreg2", replace stat(tabla_ivreg2) sdec(5) title("Tabla 6: resultados del comando ivreg2") ctitle("Metodo","Coeficiente", "Error Estandar", "IC (Minimo)","C (Maximo)") rtitles("MC2E")


*4)----------------------------------------------------------------------

*b)

use datos_simulados.dta, replace

expand 10


simulate tau_estimado2 = r(tau_hat) error_s2 = r(se_bstrp), reps(5000) seed(201922665): bstrp

save "bootstrap_expanded.dta", replace
```

```
use bootstrap_expanded.dta, replace
mat tabla2 = J(2,1,.)
estpost sum tau_estimado2
mat tabla2[1,1] = e(mean)
mat tabla2[2,1] = e(sd)

mat tabla3 = (tabla\tabla2)
mat list tabla
mat list tabla2
mat list tabla3

frmttable using "tabla_p4.doc", replace sdec(5) stat(tabla3) title("Tabla 7: coeficientes y errores estandar") ctitle("Metodo","Estimador") rtitles("MC2E"\""\"bootstrap"\""\"bootstrap expandido"\"")

*c)
*IC percentile
use bootstrap_expanded.dta, replace
mat IC_p_expanded = J(1,2,.)
_pctile tau_estimado, p(2.5, 97.5)
return list
mat IC_p_expanded[1,1] = `r(r1)'
mat IC_p_expanded[1,2] = `r(r2)'

*IC normal
use datos_simulados.dta, replace
reg Di Ai
predict Di_hat
reg Yi Di_hat
```

```
scalar t_hat = _b[Di_hat]

use bootstrap_expanded.dta, replace
estpost sum tau_estimado2
scalar sigma_expanded = e(sd)[1,1]
display sigma_expanded

mat IC_n_expanded = J(1,2,.)
mat IC_n_expanded[1,1] = t_hat - (1.96*sigma_expanded)
mat IC_n_expanded[1,2] = t_hat + (1.96*sigma_expanded)

*IC t
use datos_simulados.dta, replace
reg Di Ai
predict Di_hat
reg Yi Di_hat
scalar t_iv = _b[Di_hat]

use bootstrap_expanded.dta, replace
estpost sum tau_estimado2
scalar sigma = e(sd)[1,1]
gen t_n = (tau_estimado-t_iv)/sigma
_pctile t_n, p(97.5)
mat IC_t_expanded = J(1,2,.)
mat IC_t_expanded[1,1] = t_iv - (`r(r1)'*sigma)
mat IC_t_expanded[1,2] = t_iv + (`r(r1)'*sigma)

mat list IC_t_expanded
```

```
mat list IC_p_expanded

mat list IC_n_expanded


mat bootstrap_expanded_IC = (IC_p_expanded\IC_n_expanded\IC_t_expanded)

frmttable using "IC_todos_expandido.doc", replace stat(bootstrap_expanded_IC) sdec(5) title("Tabla 8: Intervalos de confianza (95%)") ctitle("Metodo", "IC (Minimo)", "IC (Maximo)") rtitles("Bootstrap percentile"\"Normal Bootstrap"\"Bootstrap t")


*5)----------------------------------------------------------------------

*a)

use datos_simulados.dta, replace

expand(10)


program define bstrp_cluster, rclass

preserve

quiet reg Di Ai

cap drop Di_hat

predict Di_hat

quiet reg Yi Di_hat

end


bootstrap _b, rep(5000): bstrp_cluster

save "bootstrap_cluster_1.dta", replace

mat ee_bstrp_cluster = J(2,1,.)

mat ee_bstrp_cluster[1,1] = e(se)[1,1]


bootstrap _b, rep(5000) cluster(id): bstrp_cluster

save "bootstrap_cluster_2.dta", replace

mat ee_bstrp_cluster[2,1] = e(se)[1,1]
```

```
frmttable using "tabla_p5.doc", replace sdec(5) stat(ee_bstrp_cluster) title("Tabla 9: Errores estandar (Estimación del error mediante bootstrap)") ctitle("Metodo","Error Estandar") rtitles("Sin corrección de clusters"\"Corregido con clusters")
```