

Microeconometría II
Práctica 1
Modelo de Resultados Potenciales

1. Resultados potenciales, sesgo de selección y el médico perfecto

Considere el ejemplo hipotético simple del Cuadro 1. Este ejemplo involucra una población de once pacientes, cada uno de los cuales está infectado con COVID-19. Hay dos tratamientos: **ventiladores y reposo en cama**. El Cuadro 1 muestra los resultados potenciales de cada paciente en términos de años de supervivencia después del tratamiento con cada tratamiento. Los valores de resultado más grandes corresponden a mejores resultados de salud.

Cuadro 1: Resultados Potenciales

Paciente	$Y_T(u)$	$Y_C(u)$	TE	S	Y
1	1	10			
2	1	5			
3	1	4			
4	5	6			
5	5	1			
6	9	7			
7	6	8			
8	7	10			
9	8	2			
10	9	6			
11	10	7			

1. Calcule el efecto del tratamiento para cada paciente (columna TE)
- 2.Cuál es el efecto tratamiento promedio (ATE) para ventiladores (Y_T) comparado con reposo en cama (Y_C)? Qué tipo de intervención es más efectiva en promedio?
3. Supongamos que el “médico perfecto” conoce los resultados potenciales de cada paciente y, como resultado, elige el mejor tratamiento para cada paciente. Si asigna a cada paciente el tratamiento más beneficioso para ese paciente, qué pacientes recibirán ventiladores y cuáles recibirán reposo en cama (columna S)? Complete la última columna en función de lo que elija el médico perfecto.
- 4.Cuál es el efecto tratamiento promedio para ventiladores comparado con reposo en cama en el caso del médico perfecto (calculado con los datos observados)?
- 5.Cuál es la diferencia entre el ATE del médico perfecto y el ATE calculado en el punto b)? Explique con sus palabras por qué el resultado es diferente en ambos casos y justifique cual de los resultados sería el correcto.

2. Simulación de Monte Carlo

En este ejercicio se busca mostrar, utilizando simulaciones de Monte Carlo, que bajo el supuesto de independencia, la diferencia de medias para los individuos que recibieron el tratamiento y aquellos que no lo recibieron identifica el ATE. Considere los siguientes datos.

1. Calcule el ATE.
2. Genere una variable que, para cada observación, obtenga una realización de una normal estándar. Ordene las observaciones de menor a mayor en función del valor de esta normal estándar.

Cuadro 2: Resultados Potenciales

Paciente	$Y_T(u)$	$Y_C(u)$
1	7	1
2	5	6
3	5	1
4	7	8
5	4	2
6	10	1
7	1	10
8	5	6
9	3	7
10	9	8

3. Genere una variable d de otorgamiento del tratamiento que valga 1 para las primeras 5 observaciones ordenadas y que valga 0 para las restantes.
4. Compute la diferencia de medias en los promedios muestrales.
5. Repita el procedimiento anterior 10000 veces y reporte la media de las diferencias de medias de cada simulación.

3. Efectos de Tratamiento y Heterogeneidad

Sean Y_T, Y_C los resultados potenciales y D la variable de otorgamiento del tratamiento. Se define el efecto promedio de tratamiento como

$$ATT = E[Y_T - Y_C], \quad (1)$$

el efecto promedio de tratamiento sobre los tratados como

$$ATT = E[Y_T - Y_C | D = 1] \quad (2)$$

y el efecto promedio de tratamiento sobre los no tratados como

$$ATT = E[Y_T - Y_C | D = 0]. \quad (3)$$

En este ejercicio se propone una descomposición del ATE diferente que la analizada en clase. La misma sigue el capítulo 4 de *Causal Inference: The Mixtape* de Scott Cunningham.

1. Muestre que $ATT = \pi ATT + (1 - \pi)ATU$, y describa los ponderadores π .
2. En base al libro mencionado arriba, interprete que

$$\bar{y}_1 - \bar{y}_0 = ATE + \text{Sesgo de Selección} + \text{Sesgo por heterogeneidad en tratamiento.}$$

en términos de efectos de tratamiento.

3. Inicialice una muestra con 100 observaciones. Genere resultados potenciales de no recibir el tratamiento como

$$Y_0 \sim \mathcal{N}(100, 30)$$

.

4. Genere ahora un efecto de tratamiento constante e igual a 20 para todos, es decir, $TE_i = 20$ para todo $i = 1, \dots, n$. Genere una variable aleatoria normal estándar. Genere una variable de tratamiento D_i igual a 1 para aquellas observaciones que poseen un valor positivo en la variable aleatoria normal. Genere la variable Y observada como $Y = DY_1 + (1 - D)Y_0$. Compute la diferencia de medias y el test t . Luego, calcule ATE, ATT y ATU.
5. Repita el inciso anterior pero utilice que $TE \sim \mathcal{N}(20, 10)$.
6. Repita el inciso anterior pero genere una variable aleatoria normal estándar $rand$ y genere $W = 1\{rand > 0\}$. Genere el tratamiento como $TE \sim \mathcal{N}(20, 10)$ si $W = 1$ y $TE \sim \mathcal{N}(10, 10)$ si $W = 0$. Utilice W como la variable utilizada para asignar el tratamiento.
7. Repita el inciso anterior pero ahora luego de generar los efectos de tratamiento en función de W asigne el tratamiento aleatoriamente como en el primer inciso.
8. Comente las conclusiones obtenidas con respecto a la heterogeneidad del tratamiento.

Trabajo Práctico N° 1: **Modelo de Resultados Potenciales.**

Ejercicio 1.

Considerar el ejemplo hipotético simple del Cuadro 1. Este ejemplo involucra una población de once pacientes, cada uno de los cuales está infectado con COVID-19. Hay dos tratamientos: ventiladores y reposo en cama. El Cuadro 1 muestra los resultados potenciales de cada paciente en términos de años de supervivencia después del tratamiento con cada tratamiento. Los valores de resultado más grandes corresponden a mejores resultados de salud.

Cuadro 1. Resultados potenciales.

Paciente	$Y_T(u)$	$Y_C(u)$	TE	S	Y
1	1	10	-9	0	10
2	1	5	-4	0	5
3	1	4	-3	0	4
4	5	6	-1	0	6
5	5	1	4	1	5
6	9	7	2	1	9
7	6	8	-2	0	8
8	7	10	-3	0	10
9	8	2	6	1	8
10	9	6	3	1	9
11	10	7	3	1	10

(a) Calcular el efecto del tratamiento para cada paciente (columna TE).

Cuadro 1.

(b) ¿Cuál es el efecto tratamiento promedio (ATE) para ventiladores (Y_T) comparado con reposo en cama (Y_C)? ¿Qué tipo de intervención es más efectiva en promedio?

$$\begin{aligned} ATE &= \frac{\sum_{i=1}^{11} TE_i}{11} \\ ATE &= \frac{-4}{11} \\ ATE &= -0,36. \end{aligned}$$

Por lo tanto, el efecto tratamiento promedio (ATE) para ventiladores (Y_T) comparado con reposo en cama (Y_C) es $-0,36$. En promedio, la intervención reposo en cama es más efectiva.

(c) Se supone que el “médico perfecto” conoce los resultados potenciales de cada paciente y, como resultado, elige el mejor tratamiento para cada paciente. Si se asigna a cada paciente el tratamiento más beneficioso para ese paciente, ¿qué pacientes recibirán

ventiladores y cuáles recibirán reposo en cama (columna S)? Completar la última columna en función de lo que elija el médico perfecto.

Cuadro 1.

(d) ¿Cuál es el efecto tratamiento promedio para ventiladores comparado con reposo en cama en el caso del médico perfecto (calculado con los datos observados)?

$$\begin{aligned} \text{ATE} &= \frac{\sum_{i=1}^{11} Y_i}{11} \\ \text{ATE} &= \frac{84}{11} \\ \text{ATE} &= 7,63. \end{aligned}$$

Por lo tanto, el efecto tratamiento promedio para ventiladores comparado con reposo en cama en el caso del médico perfecto (calculado con los datos observados) es 7,63.

(e) ¿Cuál es la diferencia entre el ATE del médico perfecto y el ATE calculado en el inciso (b)? Explicar por qué el resultado es diferente en ambos casos y justificar cuál de los resultados será el correcto.

La diferencia entre el ATE del médico perfecto y el ATE calculado en el inciso (b) es que el primero aparece un sesgo de selección cuando se calculan los resultados potenciales. El resultado correcto es el del inciso (b), ya que es el verdadero ATE.

Ejercicio 2.

En este ejercicio, se busca mostrar, utilizando simulaciones de Monte Carlo, que bajo el supuesto de independencia, la diferencia de medias para los individuos que recibieron el tratamiento y aquellos que no lo recibieron identifica el ATE. Considerar los siguientes datos.

Cuadro 2. Resultados potenciales.

<i>Paciente</i>	$Y_T(u)$	$Y_C(u)$
1	7	1
2	5	6
3	5	1
4	7	8
5	4	2
6	10	1
7	1	10
8	5	6
9	3	7
10	9	8

(a) Calcular el ATE.

$$\text{ATE} = \frac{\sum_{i=1}^{10} TE_i}{10}$$

$$\text{ATE} = \frac{6}{10}$$

$$\text{ATE} = 0,6.$$

(b) Generar una variable que, para cada observación, obtenga una realización de una normal estándar. Ordenar las observaciones de menor a mayor en función del valor de esta normal estándar.

Stata.

(c) Generar una variable d de otorgamiento del tratamiento que valga 1 para las primeras 5 observaciones ordenadas y que valga 0 para las restantes.

Stata.

(d) Computar la diferencia de medias en los promedios muestrales.

Stata.

(e) Repetir el procedimiento anterior 10000 veces y reportar la media de las diferencias de medias de cada simulación.

Stata.

Ejercicio 3.

Sean Y_T , Y_C los resultados potenciales y D la variable de otorgamiento del tratamiento. Se define el efecto promedio de tratamiento como

$$ATE = E[Y_T - Y_C], \quad (1)$$

el efecto promedio de tratamiento sobre los tratados como

$$ATT = E[Y_T - Y_C \mid D = 1], \quad (2)$$

y el efecto promedio de tratamiento sobre los no tratados como

$$ATU = E[Y_T - Y_C \mid D = 0], \quad (3)$$

En este ejercicio, se propone una descomposición del ATE diferente que la analizada en clase. La misma sigue el capítulo 4 de *Causal Inference: The Mixtape* de Scott Cunningham.

(a) Mostrar que $ATE = \pi ATT + (1 - \pi) ATU$ y describir los ponderadores π .

$$ATE = E(Y_T - Y_C)$$

$$ATE = E(E(Y_T - Y_C) \mid D)$$

$$ATE = \pi E(Y_T - Y_C \mid D = 1) + (1 - \pi) E(Y_T - Y_C \mid D = 0)$$

$$ATE = \pi ATT + (1 - \pi) ATU.$$

(b) En base al libro mencionado arriba, interpretar que $\bar{y}_1 - \bar{y}_0 = ATE + \text{Sesgo de Selección} + \text{Sesgo por heterogeneidad en tratamiento}$, en términos de efectos de tratamiento.

$$\begin{aligned} \frac{1}{N_T} \sum_{i=1}^n (y_i \mid d_i = 1) - \frac{1}{N_C} \sum_{i=1}^n (y_i \mid d_i = 0) &= E(Y^1) - E(Y^0) + \\ &= E(Y^0 \mid D = 1) - E(Y^0 \mid D = 0) + \\ &= (1 + \pi) (ATT - ATU). \end{aligned}$$

- Sesgo de selección: Es una descripción de las diferencias entre los dos grupos si nunca se hubiera realizado un tratamiento.
- Sesgo por heterogeneidad en tratamiento: Es la diferencia en los retornos del tratamiento para los dos grupos multiplicada por la proporción de la población que está en el grupo de control.

(c) Inicializar una muestra con 100 observaciones. Generar resultados potenciales de no recibir el tratamiento como $Y_0 \sim \mathcal{N}(100, 30)$.

Stata.

(d) Generar, ahora, un efecto de tratamiento constante e igual a 20 para todos, es decir, $TE_i = 20$ para todo $i = 1, \dots, n$. Generar una variable aleatoria normal estándar. Generar una variable de tratamiento D_i igual a 1 para aquellas observaciones que poseen un valor positivo en la variable aleatoria normal. Generar la variable Y observada como $Y = D Y_1 + (1 - D) Y_0$. Computar la diferencia de medias y el test t . Luego, calcular ATE, ATT y ATU.

Stata.

(e) Repetir el inciso anterior, pero utilizar que $TE \sim \mathcal{N}(20, 10)$.

Stata.

(f) Repetir el inciso anterior, pero generar una variable aleatoria normal estándar $rand$ y generar $W = 1\{rand > 0\}$. Generar el tratamiento como $TE \sim \mathcal{N}(20, 10)$ si $W = 1$ y $TE \sim \mathcal{N}(10, 10)$ si $W = 0$. Utilizar W como la variable utilizada para asignar el tratamiento.

Stata.

(g) Repetir el inciso anterior, pero, ahora, luego de generar los efectos de tratamiento en función de W , asignar el tratamiento aleatoriamente como en el primer inciso.

Stata.

(h) Comentar las conclusiones obtenidas con respecto a la heterogeneidad del tratamiento.

Microeconometría II
Práctica 2
Matching

1. Exact Matching

(Basado en Cunningham) Este ejercicio propone utilizar el procedimiento de matching sobre una variable para pensar en términos simples de dónde provienen los estimadores de Matching. Hoy en día es conocimiento común que fumar aumenta la tasa de mortalidad. Sin embargo, esta afirmación no proviene de datos experimentales.

1. Considere los datos de mortalidad y condición de fumador de la Tabla 1. ¿Qué resultado da la diferencia de medias entre grupos de fumadores en términos de tasa de mortalidad? ¿Qué pide el supuesto de independencia? En particular, comente qué espera que pase con otras variables observables.
2. Considere la edad de las personas. ¿Qué puede decir sobre el supuesto de independencia?
3. Estime el efecto correcto.

Table 5.1: Death rates per 1,000 person-years (Cochran 1968)

Smoking group	Canada	UK	US
Non-smokers	20,2	11,3	13,5
Cigarettes	20,5	14,1	13,5
Cigars/pipes	35,5	20,7	17,4

Table 5.2: Mean ages, years (Cochran 1968).

Smoking group	Canada	UK	US
Non-smokers	54,9	49,1	57,0
Cigarettes	50,5	49,8	53,2
Cigars/pipes	65,9	55,7	59,7

Table 5.2: Mean ages, years (Cochran 1968).

Smoking group	Canada	UK	US
Non-smokers	54,9	49,1	57,0
Cigarettes	50,5	49,8	53,2
Cigars/pipes	65,9	55,7	59,7

Solution:

1. La diferencia de medias simple sugiere que fumar pipas y habanos está asociado a una mayor tasa de mortalidad. El supuesto de independencia sugiere que la media los resultados potenciales es iguales entre grupos,

$$\begin{aligned} E[Y^1 \mid \text{Cigarette}] &= E[Y^1 \mid \text{Pipe}] = E[Y^1 \mid \text{Cigar}] \\ E[Y^0 \mid \text{Cigarette}] &= E[Y^0 \mid \text{Pipe}] = E[Y^0 \mid \text{Cigar}]. \end{aligned} \quad (1)$$

Es decir, si se cumple, debería suceder que la mortalidad si se deciden a fumar no debería variar entre no fumadores, fumadores de cigarrillos y fumadores de habanos y pipas, tanto si fumaran como si no fumaran. También debería ser cierto que que las características observables entre fumadores no deberían ser, en promedio, diferentes. Es decir, los grupos deberían estar balanceados.

2. La diferencia en edades por grupo (variable observada) está relacionada al no cumplimiento del supuesto de independencia.

2. Teorema de Rosenbaum y Rubin

Demuestre el siguiente teorema. Sean $Y(0), Y(1)$ resultados potenciales, X un tratamiento binario, W un vector de características observables. Suponiendo que vale el supuesto de *unconfoundedness*, se define $e(W)$ como la probabilidad de recibir el tratamiento en función de variables observables. Luego $(Y(0), Y(1) \perp X | e(W))$. Como corolario,

$$ATE = E[E[Y^{obs}|X=1, e(W)] - E[Y^{obs}|X=0, e(W)]] . \quad (2)$$

Solution: Por la definición de independencia, es suficiente probar que

$$P(X=1|Y(0), Y(1), e(W)) = P(X=1|e(W)) \quad (3)$$

Como X es binaria,

$$\begin{aligned} P(X=1|Y(0), Y(1), e(W)) &= E(X|Y(0), Y(1), e(W)) \\ &= E(E(X|Y(0), Y(1), e(W), W)|Y(0), Y(1), e(W)) \quad \text{LIE} \\ &= E(E(X|Y(0), Y(1), W)|Y(0), Y(1), e(W)) \\ &= E(E(X|W)|Y(0), Y(1), e(W)) \quad \text{unconfoundedness} \\ &= E(e(W)|Y(0), Y(1), e(W)) \quad e: W \rightarrow [0, 1] \\ &= e(W) \end{aligned}$$

Por lo tanto,

$$P(X=1|Y(0), Y(1), e(W)) = P(X=1|e(W)) = P(X=1|W) = e(W) \quad (4)$$

El resultado implica que los resultados potenciales son independientes del tratamiento habiendo condicionado en la probabilidad de recibir el tratamiento, $e(W)$ (el llamado *propensity score*). Como tenemos el resultado demostrado al nivel de la probabilidad, esto implica el resultado deseado para las esperanzas, y podemos calcular el ATE habiendo condicionado en el puntaje de propensión en vez de en todas las características W .

$$ATE = E[E[Y^{obs}|X=1, e(W)] - E[Y^{obs}|X=0, e(W)]] . \quad (5)$$

Notar que

1. Si hay 2 unidades en tratamiento y control con igual probabilidad de recibir el tratamiento, en promedio tienen los mismos resultados potenciales. Es decir, si antes *unconfoundedness* pedía

$$E[Y(d)|X=1, W] = E[Y(d)|X=0, W], \quad d \in \{0, 1\}, \quad (6)$$

ahora podemos relajar la condición a

$$E[Y(d)|X=1, e(W)] = E[Y(d)|X=0, e(W)], \quad d \in \{0, 1\}. \quad (7)$$

2. El teorema de Rosenbaum y Rubin sugiere un procedimiento aplicado para estimar los efectos promedios de tratamiento: calcular la probabilidad $e(W)$ de recibir el tratamiento en función de las características W , y hacer un matching en esta variable que resume las anteriores. Esto, además, simplifica el problema de dimensionalidad de emparejar muchas variables con algunas potencialmente continuas.
3. El supuesto de *unconfoundedness* es necesario para el teorema de Rosenbaum y Rubin. No se puede reemplazar fallas en este supuesto y esperar que se solucione haciendo Propensity Score Matching.

3. Propensity Score Matching

Para solucionar el problema de la dimensionalidad utilizando matching sobre múltiples variables se puede reducir el conjunto de variables en una sola, la llamada *propensity score*, puntaje de propensión o probabilidad de recibir el tratamiento. En este ejercicio se propone realizar una evaluación del impacto de fumar durante el embarazo sobre el peso de los bebés en base a datos observacionales. Utilizando la base de datos `cattaneo2.dta` que utiliza Cattaneo (XX), responda las siguientes preguntas.

1. Compruebe si los grupos de control y de tratamiento están balanceados.

2. ¿Cuál es la diferencia de medias simple?
3. Estime la probabilidad de recibir el tratamiento con un modelo Logit. Utilice como variables explicativas `mmarried` `deadkids` `nprenatal` `months1b` `prenatal` `fbaby` `alcohol`. Defina una sección de soporte común.
4. Estime el efecto promedio de tratamiento sobre los tratados utilizando Propensity Score Matching implementando nearest-neighbor matching, radius matching, kernel matching y stratification matching.
5. Estime el efecto promedio de tratamiento utilizando nearest-neighbour matching.
6. Estime el efecto promedio de tratamiento utilizando ajuste por regresión.
7. Estime el efecto promedio de tratamiento utilizando *inverse probability weighting*.
8. Estime el efecto promedio de tratamiento utilizando el estimador doblemente robusto.

4. Estimador de matching sin características oservables

Suponga que vale el supuesto de *unconfoundedness*. En este ejercicio se propone computar la expresión analítica del estimador de PSM cuando la probabilidad de recibir el tratamiento no depende de características oservables. Utilice el corolario del ejercicio 2.

Solution: Si vale *unconfoundedness* sin condicionar en otras variables explicativas, entonces $X \perp Y$. El ATE es

$$ATE = E [E[Y^{obs}|X = 1] - E[Y^{obs}|X = 0]] . \quad (8)$$

Definimos

$$\mu_X = E(Y^{obs}|X = x) \quad (9)$$

Como estimador, tomamos el análogo muestral, porque converge en probabilidad al parámetro de interés. Entonces,

$$\hat{\mu}_X = \frac{n^{-1} \sum_{i=1}^n Y_i^{obs} \mathbb{I}\{X_i = x\}}{n^{-1} \sum_{i=1}^n \mathbb{I}\{X_i = x\}} \quad (10)$$

Luego, el estimador del ATE es la diferencia de medias.

$$\widehat{ATE} = \hat{\mu}_1 - \hat{\mu}_0.$$

Utilizando la expresión del estimador de las medias,

$$\begin{aligned} \widehat{ATE} &= \hat{\mu}_1 - \hat{\mu}_0 \\ &= \frac{n^{-1} \sum_{i=1}^n Y_i^{obs} x_i}{n^{-1} \sum_{i=1}^n x_i} - \frac{n^{-1} \sum_{i=1}^n Y_i^{obs} (1 - x_i)}{n^{-1} \sum_{i=1}^n (1 - x_i)} \\ &= \frac{n^{-1} \sum_{i=1}^n Y_i^{obs} x_i}{\bar{x}} - \frac{n^{-1} \sum_{i=1}^n Y_i^{obs} (1 - x_i)}{(1 - \bar{x})} \\ &= n^{-1} \sum_{i=1}^n Y_i^{obs} \left(\frac{x_i}{\bar{x}} - \frac{1 - x_i}{1 - \bar{x}} \right) \\ &= n^{-1} \sum_{i=1}^n Y_i^{obs} \frac{x_i - \bar{x}}{\bar{x}(1 - \bar{x})} \end{aligned}$$

Trabajo Práctico N° 2: **Matching.**

Ejercicio 1: Exact Matching.

(Basado en Cunningham) Este ejercicio propone utilizar el procedimiento de matching sobre una variable para pensar en términos simples de dónde provienen los estimadores de Matching. Hoy en día, es conocimiento común que fumar aumenta la tasa de mortalidad. Sin embargo, esta afirmación no proviene de datos experimentales.

Table 5.1: Death rates per 1,000 person-years (Cochran 1968)

Smoking group	Canada	UK	US
Non-smokers	20,2	11,3	13,5
Cigarettes	20,5	14,1	13,5
Cigars/pipes	35,5	20,7	17,4

Table 5.2: Mean ages, years (Cochran 1968).

Smoking group	Canada	UK	US
Non-smokers	54,9	49,1	57,0
Cigarettes	50,5	49,8	53,2
Cigars/pipes	65,9	55,7	59,7

(a) Considerar los datos de mortalidad y condición de fumador de la Tabla 1. ¿Qué resultado da la diferencia de medias entre grupos de fumadores en términos de tasa de mortalidad? ¿Qué pide el supuesto de independencia? En particular, comentar qué espera que pase con otras variables observables.

El resultado que da la diferencia de medias entre grupos de fumadores en términos de tasa de mortalidad es 15, 6,6 y 3,9 para Canadá, UK, US, respectivamente. El supuesto de independencia pide que la media de la tasa de mortalidad entre grupos sea igual. En particular, se espera que las variables observables entre fumadores no deberían ser, en promedio, diferentes, es decir, los grupos deberían ser balanceados.

(b) Considerar la edad de las personas. ¿Qué se puede decir sobre el supuesto de independencia?

Considerando la edad de las personas, lo que se puede decir sobre el supuesto de independencia es que no se cumpliría, ya que esta característica no es independiente del resultado potencial.

(c) Estimar el efecto correcto.

Ejercicio 2: Teorema de Rosenbaum y Rubin.

*Demostrar el siguiente teorema. Sean $Y(0)$, $Y(1)$ resultados potenciales, X un tratamiento binario, W un vector de características observables. Suponiendo que vale el supuesto de *unconfoundedness*, se define $e(W)$ como la probabilidad de recibir el tratamiento en función de variables observables. Luego $(Y(0), Y(1) \perp X \mid e(W))$. Como corolario:*

$$ATE = E[E[Y^{obs} \mid X=1, e(W)] - E[Y^{obs} \mid X=0, e(W)]]$$

$$P(X=1 \mid Y(0), Y(1), e(W)) = E(X \mid Y(0), Y(1), e(W))$$

$$P(X=1 \mid Y(0), Y(1), e(W)) = E(E(X \mid Y(0), Y(1), e(W), W) \mid Y(0), Y(1), e(W))$$

$$P(X=1 \mid Y(0), Y(1), e(W)) = E(E(X \mid W) \mid Y(0), Y(1), e(W))$$

$$P(X=1 \mid Y(0), Y(1), e(W)) = E(e(W) \mid Y(0), Y(1), e(W))$$

$$P(X=1 \mid Y(0), Y(1), e(W)) = e(W)$$

El resultado implica que los resultados potenciales son independientes del tratamiento habiendo condicionado en la probabilidad de recibir el tratamiento, $e(W)$ (el llamado *propensity score*). Como se tiene el resultado demostrado al nivel de la probabilidad, esto implica el resultado deseado para las esperanzas y se puede calcular el ATE habiendo condicionado en el *propensity score* en vez de en todas las características W .

El teorema de Rosenbaum y Rubin sugiere un procedimiento aplicado para estimar los efectos promedios de tratamiento: calcular la probabilidad $e(W)$ de recibir el tratamiento en función de las características W y hacer un *matching* en esta variable que resume las anteriores. Esto, además, simplifica el problema de dimensionalidad de emparejar muchas variables con algunas, potencialmente, continuas. El supuesto de *unconfoundedness* es necesario para el teorema de Rosenbaum y Rubin. No se puede reemplazar fallas en este supuesto y esperar que se solucione haciendo *Propensity Score Matching*.

Ejercicio 3: Propensity Score Matching.

Para solucionar el problema de la dimensionalidad utilizando matching sobre múltiples variables, se puede reducir el conjunto de variables en una sola, la llamada propensity score, puntaje de propensión o probabilidad de recibir el tratamiento. En este ejercicio, se propone realizar una evaluación del impacto de fumar durante el embarazo sobre el peso de los bebés en base a datos observacionales. Utilizando la base de datos "cattaneo2.dta" que utiliza Cattaneo (XX), responder las siguientes preguntas.

(a) Comprobar si los grupos de control y de tratamiento están balanceados.

Stata.

(b) ¿Cuál es la diferencia de medias simple?

Stata.

(c) Estimar la probabilidad de recibir el tratamiento con un modelo Logit. Utilizar como variables explicativas mmarried, deadkids, nprenatal, monthslb, prenatal, fbaby y alcohol. Definir una sección de soporte común.

Stata.

(d) Estimar el efecto promedio de tratamiento sobre los tratados utilizando Propensity Score Matching implementando nearest-neighbor matching, radius matching, kernel matching y stratification matching.

Nearest-Neighbor Matching:

ATT estimation with Nearest Neighbor Matching method
(random draw version)
Bootstrapped standard errors

n. treat.	n. contr.	ATT	Std. Err.	t

864	659	-235.279	37.700	-6.241

Note: the numbers of treated and controls refer to actual
nearest neighbour matches

Radius Matching:

ATT estimation with the Radius Matching method
Bootstrapped standard errors

n. treat.	n. contr.	ATT	Std. Err.	t
863	864	-246.371	34.574	-7.126

Note: the numbers of treated and controls refer to actual matches within radius

Kernel Matching:

ATT estimation with the Kernel Matching method
Bootstrapped standard errors

n. treat.	n. contr.	ATT	Std. Err.	t
864	864	-229.593	31.817	-7.216

Stratification Matching:

ATT estimation with the Stratification method
Bootstrapped standard errors

n. treat.	n. contr.	ATT	Std. Err.	t
864	864	-227.836	25.709	-8.862

(e) Estimar el efecto promedio de tratamiento utilizando nearest-neighbour matching.

Matching estimator: Average Treatment Effect

Number of obs = 1728
Number of matches (m) = 1

bweight	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
SATE	-209.5832	28.82159	-7.27	0.000	-266.0725	-153.0939

Matching variables: e_x

Bias-adj variables: mmarried alcohol deadkids nprenatal monthslb prenatal fbaby

(f) Estimar el efecto promedio de tratamiento utilizando ajuste por regresión.

Treatment-effects estimation
Estimator : regression adjustment
Outcome model : linear
Treatment model: none

Number of obs = 1,728

	bweight	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ATE							
	mbsmoke						
(smoker vs nonsmoker)		-203.0403	26.27955	-7.73	0.000	-254.5473	-151.5334
POmean							
	mbsmoke						
	nonsmoker	3343.881	19.0128	175.88	0.000	3306.616	3381.145

(g) Estimar el efecto promedio de tratamiento utilizando inverse probability weighting.

Treatment-effects estimation
Estimator : inverse-probability weights
Outcome model : weighted mean
Treatment model: logit

Number of obs = 1,728

	bweight	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ATE							
	mmarried						
(married vs notmarried)		103.099	28.54709	3.61	0.000	47.14773	159.0503
POmean							
	mmarried						
	notmarried	3191.532	20.32589	157.02	0.000	3151.694	3231.37

(h) Estimar el efecto promedio de tratamiento utilizando el estimador doblemente robusto.

		Coefficient	Robust std. err.	z	P> z	[95% conf. interval]	
ATE							
	mbsmoke						
	(smoker vs nonsmoker)	-203.1215	26.27575	-7.73	0.000	-254.621	-151.622
POmean							
	mbsmoke						
	nonsmoker	3343.847	19.01012	175.90	0.000	3306.588	3381.107

Ejercicio 4: Estimador de Matching sin Características Observables.

Suponer que vale el supuesto de unconfoundedness. En este ejercicio, se propone computar la expresión analítica del estimador de PSM cuando la probabilidad de recibir el tratamiento no depende de características observables. Utilizar el corolario del Ejercicio 2.

$$ATE = E [E [Y^{obs} | X = 1, e(W)] - E [Y^{obs} | X = 0, e(W)]].$$

$$\widehat{ATE} = E (\hat{\mu}_1 - \hat{\mu}_0)$$

$$\widehat{ATE} = E \left[\frac{n^{-1} \sum_{i=1}^n Y_i^{obs}(X_i=1)}{n^{-1} \sum_{i=1}^n I(X_i=1)} - \frac{n^{-1} \sum_{i=1}^n Y_i^{obs}(X_i=0)}{n^{-1} \sum_{i=1}^n I(X_i=0)} \right]$$

$$\widehat{ATE} = E \left[\frac{n^{-1} \sum_{i=1}^n Y_i^{obs} x_i}{n^{-1} \sum_{i=1}^n x_i} - \frac{n^{-1} \sum_{i=1}^n Y_i^{obs} (1-x_i)}{n^{-1} \sum_{i=1}^n (1-x_i)} \right]$$

$$\widehat{ATE} = E \left[\frac{n^{-1} \sum_{i=1}^n Y_i^{obs} x_i}{\bar{x}} - \frac{n^{-1} \sum_{i=1}^n Y_i^{obs} (1-x_i)}{1-\bar{x}} \right]$$

$$\widehat{ATE} = E \left[n^{-1} \sum_{i=1}^n Y_i^{obs} \left(\frac{x_i}{\bar{x}} - \frac{1-x_i}{1-\bar{x}} \right) \right]$$

$$\widehat{ATE} = E \left[n^{-1} \sum_{i=1}^n Y_i^{obs} \frac{x_i - \bar{x}}{\bar{x}(1-\bar{x})} \right]$$

$$\widehat{ATE} = E \left[\frac{\sum_{i=1}^n Y_i^{obs} \frac{x_i - \bar{x}}{\bar{x}(1-\bar{x})}}{n} \right]$$

$$\widehat{ATE} = E \left(\frac{\sum_{i=1}^n Y_i^{obs} \omega_i}{n} \right),$$

donde:

$$\omega_i = \frac{x_i - \bar{x}}{\bar{x}(1-\bar{x})}$$

$$\omega_i = \frac{x_i - p}{p(1-p)}.$$

Microeconometría II
Práctica 3
Diferencias-en-Diferencias

1. Evaluación de Impacto con DiD

Una práctica común en la evaluación de un programa cuando se tienen datos de panel para dos períodos es la siguiente: sea y_{it} el resultado observado para i en el período t . En $t = 1$ nadie está en el programa. En $t = 2$ algunos están en el grupo de control y otros en el grupo de tratamiento. Sea $prog_{it}$ una variable que vale 1 si el individuo i está en el grupo de tratamiento en el período t ; y cero en caso contrario. Note que $prog_{i1} = 0$ para todo i . Se puede plantear el siguiente modelo:

$$y_{it} = \theta_1 + \theta_2 d2_t + \delta_1 prog_{it} + c_i + u_{it}$$

Con $E(u_{it}/prog_{i2}, c_i) = 0$. En el que $d2_t$ es una variable dummy que vale 1 si $t = 2$, cero si $t = 1$; c_i es el efecto no observado. Usando el método de primeras diferencias, muestre que $\hat{\theta}_2 = \overline{\Delta y_c}$ y $\hat{\delta}_1 = \overline{\Delta y_t} - \overline{\Delta y_c}$, donde $\overline{\Delta y_c}$ es el cambio promedio en y a lo largo de los dos períodos para el grupo con $prog_{i2} = 0$, y $\overline{\Delta y_t}$ es el cambio promedio en y a lo largo de los dos períodos para el grupo con $prog_{i2} = 1$.

Solution:

Tenemos el siguiente modelo

$$y_{it} = \theta_1 + \theta_2 d2_t + \delta_1 prog_{it} + c_i + u_{it}$$

donde y_{it} es el resultado observado para i en el período t , $prog_{it}$ vale 1 si el individuo i está en el grupo de tratamiento en t y $d2_t$ es una variable dicotómica que vale 1 si $t = 2$. c_i es el efecto no observado por individuo, que asumimos constante en el tiempo.

El modelo en $t = 1$ es

$$y_{i1} = \theta_1 + \theta_2 * 0 + \delta_1 * 0 + c_i + u_{i1} = \theta_1 + c_i + u_{i1}$$

El modelo en $t = 2$ es

$$y_{i2} = \theta_1 + \theta_2 * 1 + \delta_1 * prog_{i2} + c_i + u_{i2} = \theta_1 + \theta_2 + \delta_1 * prog_{i2} + c_i + u_{i2}$$

Tomando primeras diferencias

$$y_{i2} - y_{i1} = (\theta_1 + \theta_2 + \delta_1 * prog_{i2} + c_i + u_{i2}) - (\theta_1 + c_i + u_{i1}) = \theta_2 + \delta_1 * prog_{i2} + u_{i2} - u_{i1}$$

Reordenando

$$\Delta y_i = \theta_2 + \delta_1 * prog_{i2} + \Delta u_i$$

Los estimadores OLS de θ_2 y δ_1 son

$$\delta_1^{OLS} = \frac{Cov(prog_{i2}, \Delta y_i)}{Var(prog_{i2})} \quad \theta_2^{OLS} = \overline{\Delta y} - \delta_1^{OLS} \overline{prog_{i2}}$$

Sea N_1 la cantidad de veces que $prog_{i2} = 1$. Computemos los distintos elementos. Utilizaremos algunos atajos de la Práctica 1. En particular, del Ejercicio 0, Ejercicio 7.

$$\begin{aligned}
Cov(prog_{i2}, \Delta y_i) &= \sum_{i=1}^N prog_{i2} \Delta y_i - N \overline{prog_{i2}} \overline{\Delta y} \\
&= \sum_{prog_{i2}=1} \Delta y_i - N \left(\frac{\sum_{i=1}^N prog_{i2}}{N} \right) \overline{\Delta y} \\
&= N_1 \overline{\Delta y_t} - N_1 \overline{\Delta y} \\
&= N_1 \left(\frac{\sum_{prog_{i2}=1} \Delta y_i}{N_1} - \frac{\sum_{i=1}^N \Delta y_i}{N} \right) \\
&= N_1 \left(\frac{N \sum_{prog_{i2}=1} \Delta y_i - N_1 \sum_{i=1}^N \Delta y_i}{N_1 N} \right) \\
&= \frac{(N - N_1) \sum_{prog_{i2}=1} \Delta y_i - N_1 \sum_{prog_{i2}=0} \Delta y_i}{N} \\
&= \frac{(N - N_1) N_1 \overline{\Delta y_t} - (N - N_1) N_1 \overline{\Delta y_c}}{N} \\
&= \left(1 - \frac{N_1}{N} \right) N_1 (\overline{\Delta y_t} - \overline{\Delta y_c})
\end{aligned}$$

$$\begin{aligned}
Var(prog_{i2}) &= \sum_{i=1}^N prog_{i2}^2 - N \overline{prog_{i2}}^2 \\
&= N_1 - N \left(\frac{N_1}{N} \right)^2 = N_1 - \frac{N_1^2}{N} \\
&= N_1 \left(1 - \frac{N_1}{N} \right)
\end{aligned}$$

Luego

$$\delta_1^{OLS} = \overline{\Delta y_t} - \overline{\Delta y_c}$$

Computamos θ_2^{OLS}

$$\begin{aligned}
\theta_2^{OLS} &= \overline{\Delta y} - \delta_1^{OLS} \overline{prog_{i2}} \\
&= \overline{\Delta y} - (\overline{\Delta y_t} - \overline{\Delta y_c}) \frac{N_1}{N} \\
&= \frac{\sum_{i=1}^N \Delta y_i}{N} - \frac{N_1}{N} \left(\frac{\sum_{prog_{i2}=1} \Delta y_i}{N_1} - \frac{\sum_{prog_{i2}=0} \Delta y_i}{N - N_1} \right) \\
&= \frac{\sum_{i=1}^N \Delta y_i}{N} + \frac{\sum_{prog_{i2}=1} \Delta y_i}{N} - \frac{N_1 \sum_{prog_{i2}=0} \Delta y_i}{N(N - N_1)} \\
&= \frac{(N - N_1) \sum_{i=1}^N \Delta y_i - (N - N_1) \sum_{prog_{i2}=1} \Delta y_i + N_1 \sum_{prog_{i2}=0} \Delta y_i}{N(N - N_1)} \\
&= \frac{N \sum_{prog_{i2}=0} \Delta y_i}{N(N - N_1)} \\
&= \frac{\sum_{prog_{i2}=0} \Delta y_i}{N(N - N_1)} \\
&= \overline{\Delta y_c}
\end{aligned}$$

2. DiD Simple en Stata

1. Abra la base `Panel101.dta` y genere las siguientes variables de resultado, de tiempo y de tratamiento. Los “países” tratados son los países 5 a 7 y el tratamiento se otorgó en 1994.

$$\begin{aligned}Y &= y/1000000000 \\time &= \mathbb{I}\{year \geq 1994\} \\treated &= \mathbb{I}\{country > 4\}\end{aligned}$$

2. Compute e interprete el estimador de diferencias en diferencias utilizando una regresión lineal.
3. Compute e interprete el estimador de diferencias en diferencias utilizando una especificación con efectos fijos.
4. Compute e interprete el estimador de diferencias en diferencias utilizando el paquete de Stata `diff`.

3. Card & Krueger (1994)

Este ejercicio se basa en el artículo *Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania*.

1. ¿Qué efecto intentan estimar los autores en el artículo?
2. ¿Cuál es la estrategia de identificación?
3. Utilice el archivo `CardKrueger1994.dta`. Utilizando `diff`, compute el estimador de diferencias-en-diferencias.
4. Repita el inciso (c) utilizando errores estándar de *bootstrap*.
5. Repita el inciso (c) utilizando la cadena de restaurantes como variables explicativas.

4. didregress

Hasta ahora estimamos utilizando la especificación

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 T_{it} + \beta_3 (D \times T)_{it} + \mathbf{z}_{it}\theta + u_{it} \quad (1)$$

O en el caso de datos longitudinales, ampliamos la ecuación con efectos fijos. En Stata, usamos los comandos `regress` y/o `xtreg`. Stata 17 incluyó nuevos comandos para estimar modelos con la siguiente forma

$$Y_{ist} = \gamma_s + \gamma_t + \mathbf{z}_{ist}\beta + \delta D_{st} + u_{ist}$$

con el comando `didregress` o incluyendo efectos fijos por individuo en el caso de datos longitudinales con el comando `didregress`

1. Explique en qué difieren estos comandos con respecto al *setup* usual de DiD.
2. ¿Puede replicar las regresiones de los ejercicios anteriores con estos comandos?
3. Un proveedor de salud está interesado en estudiar el efecto de un nuevo procedimiento de ingreso hospitalario en la satisfacción de los pacientes. El proveedor dispone de datos mensuales de pacientes de enero a julio. El nuevo procedimiento de admisiones fue implementado en abril por hospitales que estaban bajo nueva administración. De los 46 hospitales del estudio, 18 implementaron el nuevo procedimiento.

El proveedor de salud utilizará una regresión DID para analizar el efecto del nuevo procedimiento de admisiones en los hospitales que participaron en el programa. El resultado de interés es la satisfacción del paciente, `satis`, que se registra como un promedio de las respuestas a un conjunto de cuatro preguntas realizadas a los pacientes. `satis` puede tomar valores entre 0 y 10, donde 10 es el mayor nivel de satisfacción posible y 0 es la decepción total. La variable procedimiento marca las observaciones tratadas; es 1 si una persona encuestada ingresó al hospital utilizando el nuevo procedimiento después de marzo y 0 en caso contrario.

Los datos están en la base `hospdd.dta`. Evalúe el impacto del nuevo procedimiento sobre la satisfacción de los pacientes.

4. ¿Cómo interpreta el coeficiente obtenido? ¿Se cumple el supuesto de tendencias paralelas?
5. Comente sobre los errores estándar utilizados y estudie las distintas opciones que el comando tiene pre-configuradas para usar. ¿Hay diferencias en la inferencia?

Trabajo Práctico N° 3: **Diferencias en Diferencias.**

Ejercicio 1: Evaluación de Impacto con DiD.

Una práctica común en la evaluación de un programa cuando se tienen datos de panel para dos períodos es la siguiente: sea y_{it} el resultado observado para i en el período t . En $t=1$, nadie está en el programa. En $t=2$, algunos están en el grupo de control y otros en el grupo de tratamiento. Sea $prog_{it}$ una variable que vale 1 si el individuo i está en el grupo de tratamiento en el período t y cero en caso contrario. Notar que $prog_{i1}=0$ para todo i . Se puede plantear el siguiente modelo:

$$y_{it} = \theta_1 + \theta_2 d2_t + \delta_1 prog_{it} + c_i + u_{it},$$

con $E(u_{it} / prog_{i2}, c_i) = 0$. En el que $d2_t$ es una variable dummy que vale 1 si $t=2$, 0 si $t=1$; c_i es el efecto no observado. Usando el método de primeras diferencias, mostrar que $\hat{\theta}_2 = \overline{\Delta y_t}$ y $\hat{\delta}_1 = \overline{\Delta y_t} - \overline{\Delta y_c}$, donde $\overline{\Delta y_c}$ es el cambio promedio en y a lo largo de los dos períodos para el grupo con $prog_{i2}=0$, y $\overline{\Delta y_t}$ es el cambio promedio en y a lo largo de los dos períodos para el grupo con $prog_{i2}=1$.

El modelo en $t=1$ es:

$$y_{i1} = \theta_1 + \theta_2 * 0 + \delta_1 * 0 + c_i + u_{i1}$$

$$y_{i1} = \theta_1 + c_i + u_{i1}.$$

El modelo en $t=2$ es:

$$y_{i2} = \theta_1 + \theta_2 * 1 + \delta_1 prog_{i2} + c_i + u_{i2}$$

$$y_{i2} = \theta_1 + \theta_2 + \delta_1 prog_{i2} + c_i + u_{i2}.$$

Tomando primeras diferencias, se tiene:

$$\Delta y_i = y_{i2} - y_{i1}$$

$$\Delta y_i = (\theta_1 + \theta_2 + \delta_1 prog_{i2} + c_i + u_{i2}) - (\theta_1 + c_i + u_{i1})$$

$$\Delta y_i = \theta_1 + \theta_2 + \delta_1 prog_{i2} + c_i + u_{i2} - \theta_1 - c_i - u_{i1}$$

$$\Delta y_i = \theta_2 + \delta_1 prog_{i2} + \Delta u_i.$$

Los estimadores OLS de θ_2 y δ_1 son:

$$\hat{\delta}_1 = \frac{Cov(prog_{i2}, \Delta y_i)}{Var(prog_{i2})}, \quad \hat{\theta}_2 = \overline{\Delta y} - \hat{\delta}_1 \overline{prog_{i2}}.$$

Sea N_1 la cantidad de veces que $prog_{i2}=1$, se tiene:

$$Cov(prog_{i2}, \Delta y_i) = \sum_{i=1}^N prog_{i2} \Delta y_i - N \overline{prog_{i2}} \overline{\Delta y}$$

$$Cov(prog_{i2}, \Delta y_i) = \sum_{prog_{i2}=1} \Delta y_i - N \frac{\sum_{i=1}^N prog_{i2}}{N} \overline{\Delta y}$$

$$Cov(prog_{i2}, \Delta y_i) = N_1 \frac{\sum_{prog_{i2}=1} \Delta y_i}{N_1} - \sum_{i=1}^N prog_{i2} \overline{\Delta y}$$

$$\text{Cov}(\text{prog}_{i2}, \Delta y_i) = N_1 \overline{\Delta y_t} - N_1 \overline{\Delta y_c}$$

$$\text{Cov}(\text{prog}_{i2}, \Delta y_i) = N_1 (\overline{\Delta y_t} - \overline{\Delta y_c})$$

$$\text{Cov}(\text{prog}_{i2}, \Delta y_i) = N_1 \left(\frac{\sum_{\text{prog}_{i2}=1} \Delta y_i}{N_1} - \frac{\sum_{i=1}^N \Delta y_i}{N} \right)$$

$$\text{Cov}(\text{prog}_{i2}, \Delta y_i) = N_1 \frac{N \sum_{\text{prog}_{i2}=1} \Delta y_i - N_1 \sum_{i=1}^N \Delta y_i}{N_1 N}$$

$$\text{Cov}(\text{prog}_{i2}, \Delta y_i) = \frac{N \sum_{\text{prog}_{i2}=1} \Delta y_i - N_1 \sum_{i=1}^N \Delta y_i}{N}$$

$$\text{Cov}(\text{prog}_{i2}, \Delta y_i) = \frac{N \sum_{\text{prog}_{i2}=1} \Delta y_i - N_1 (\sum_{\text{prog}_{i2}=0} \Delta y_i + \sum_{\text{prog}_{i2}=1} \Delta y_i)}{N}$$

$$\text{Cov}(\text{prog}_{i2}, \Delta y_i) = \frac{N \sum_{\text{prog}_{i2}=1} \Delta y_i - N_1 \sum_{\text{prog}_{i2}=0} \Delta y_i - N_1 \sum_{\text{prog}_{i2}=1} \Delta y_i}{N}$$

$$\text{Cov}(\text{prog}_{i2}, \Delta y_i) = \frac{(N - N_1) \sum_{\text{prog}_{i2}=1} \Delta y_i - N_1 \sum_{\text{prog}_{i2}=0} \Delta y_i}{N}$$

$$\text{Cov}(\text{prog}_{i2}, \Delta y_i) = \frac{(N - N_1) N_1 \overline{\Delta y_t} - (N - N_1) N_1 \overline{\Delta y_c}}{N}$$

$$\text{Cov}(\text{prog}_{i2}, \Delta y_i) = \frac{(N - N_1) N_1 (\overline{\Delta y_t} - \overline{\Delta y_c})}{N}$$

$$\text{Cov}(\text{prog}_{i2}, \Delta y_i) = (1 - \frac{N_1}{N}) N_1 (\overline{\Delta y_t} - \overline{\Delta y_c}).$$

$$\text{Var}(\text{prog}_{i2}) = \sum_{i=1}^N \text{prog}_{i2}^2 - N (\overline{\text{prog}_{i2}})^2$$

$$\text{Var}(\text{prog}_{i2}) = N_1 - N \left(\frac{N_1}{N} \right)^2$$

$$\text{Var}(\text{prog}_{i2}) = N_1 - N \frac{N_1^2}{N^2}$$

$$\text{Var}(\text{prog}_{i2}) = N_1 - \frac{N_1^2}{N}$$

$$\text{Var}(\text{prog}_{i2}) = (1 - \frac{N_1}{N}) N_1.$$

Por lo tanto, se tiene:

$$\hat{\delta}_1 = \frac{(1 - \frac{N_1}{N}) N_1 (\overline{\Delta y_t} - \overline{\Delta y_c})}{(1 - \frac{N_1}{N}) N_1}$$

$$\hat{\delta}_1 = \overline{\Delta y_t} - \overline{\Delta y_c}.$$

$$\hat{\theta}_2 = \overline{\Delta y} - (\overline{\Delta y_t} - \overline{\Delta y_c}) \overline{\text{prog}_{i2}}$$

$$\hat{\theta}_2 = \frac{\sum_{i=1}^N \Delta y_i}{N} - \frac{N_1}{N} \left(\frac{\sum_{\text{prog}_{i2}=1} \Delta y_i}{N_1} - \frac{\sum_{\text{prog}_{i2}=0} \Delta y_i}{N - N_1} \right)$$

$$\hat{\theta}_2 = \frac{\sum_{i=1}^N \Delta y_i}{N} - \frac{\sum_{\text{prog}_{i2}=1} \Delta y_i}{N} + N_1 \frac{\sum_{\text{prog}_{i2}=0} \Delta y_i}{N(N - N_1)}$$

$$\hat{\theta}_2 = \frac{(N - N_1) \sum_{i=1}^N \Delta y_i - (N - N_1) \sum_{\text{prog}_{i2}=1} \Delta y_i + N_1 \sum_{\text{prog}_{i2}=0} \Delta y_i}{N(N - N_1)}$$

$$\hat{\theta}_2 = \frac{(N - N_1) (\sum_{\text{prog}_{i2}=0} \Delta y_i + \sum_{\text{prog}_{i2}=1} \Delta y_i) - (N - N_1) \sum_{\text{prog}_{i2}=1} \Delta y_i + N_1 \sum_{\text{prog}_{i2}=0} \Delta y_i}{N(N - N_1)}$$

$$\hat{\theta}_2 = \frac{(N - N_1) \sum_{\text{prog}_{i2}=0} \Delta y_i + (N - N_1) \sum_{\text{prog}_{i2}=1} \Delta y_i - (N - N_1) \sum_{\text{prog}_{i2}=1} \Delta y_i + N_1 \sum_{\text{prog}_{i2}=0} \Delta y_i}{N(N - N_1)}$$

$$\hat{\theta}_2 = \frac{(N - N_1) \sum_{\text{prog}_{i2}=0} \Delta y_i + N_1 \sum_{\text{prog}_{i2}=0} \Delta y_i}{N(N - N_1)}$$

$$\hat{\theta}_2 = \frac{N \sum_{\text{prog}_{i2}=0} \Delta y_i - N_1 \sum_{\text{prog}_{i2}=0} \Delta y_i + N_1 \sum_{\text{prog}_{i2}=0} \Delta y_i}{N(N - N_1)}$$

$$\hat{\theta}_2 = \frac{N \sum_{\text{prog}_{i2}=0} \Delta y_i}{N(N - N_1)}$$

$$\hat{\theta}_2 = \frac{\sum_{\text{prog}_{i2}=0} \Delta y_i}{N - N_1}$$

$$\theta_2^{OLS} = \overline{\Delta y_c}.$$

Ejercicio 2: DiD Simple en Stata.

(a) Abrir la base "Panel101.dta" y generar las siguientes variables de resultado, de tiempo y de tratamiento. Los "países" tratados son los países 5 a 7 y el tratamiento se otorgó en 1994.

Stata.

(b) Computar e interpretar el estimador de diferencias en diferencias utilizando una regresión lineal.

Source	SS	df	MS	Number of obs	=	70
Model	51.8976208	3	17.2992069	F(3, 66)	=	1.98
Residual	575.397385	66	8.71814219	Prob > F	=	0.1249
Total	627.295005	69	9.09123196	R-squared	=	0.0827
				Adj R-squared	=	0.0410
				Root MSE	=	2.9527

Y	Coefficient	Std. err.	t	P> t	[95% conf. interval]
time	2.289455	.9529637	2.40	0.019	.3868012 4.192108
treated	1.77597	1.127562	1.58	0.120	-.4752802 4.02722
did	-2.519512	1.455676	-1.73	0.088	-5.425863 .3868395
_cons	.358144	.7381625	0.49	0.629	-1.115645 1.831933

(c) Computar e interpretar el estimador de diferencias en diferencias utilizando una especificación con efectos fijos.

Fixed-effects (within) regression	Number of obs	=	70
Group variable: country	Number of groups	=	7
R-squared:	Obs per group:		
Within = 0.0968	min =		10
Between = 0.0116	avg =		10.0
Overall = 0.0341	max =		10
corr(u_i, Xb) = -0.3880	F(2, 61)	=	3.27
	Prob > F	=	0.0448

Y	Coefficient	Std. err.	t	P> t	[95% conf. interval]
time	2.289455	.8986787	2.55	0.013	.4924355 4.086474
treated	0 (omitted)				
did	-2.519512	1.372754	-1.84	0.071	-5.264504 .2254806
_cons	1.119274	.5262124	2.13	0.037	.0670471 2.1715
sigma_u	1.6151513				
sigma_e	2.7844542				
rho	.25176021	(fraction of variance due to u_i)			

F test that all u_i=0: F(6, 61) = 2.67 Prob > F = 0.0230

(d) *Computar e interpretar el estimador de diferencias en diferencias utilizando el paquete de Stata diff.*

DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS

Number of observations in the DIFF-IN-DIFF: 70

	Before	After		
Control:	16	24	40	
Treated:	12	18	30	
	28	42		

Outcome var.	Y	S. Err.	t	P> t

Before				
Control	0.358			
Treated	2.134			
Diff (T-C)	1.776	1.128	1.58	0.120
After				
Control	2.648			
Treated	1.904			
Diff (T-C)	-0.744	0.921	0.81	0.422
Diff-in-Diff	-2.520	1.456	1.73	0.088*

R-square: 0.08

* Means and Standard Errors are estimated by linear regression

Inference: * p<0.01; ** p<0.05; * p<0.1

Ejercicio 3: Card & Krueger (1994).

Este ejercicio se basa en el artículo “Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania”.

(a) ¿Qué efecto intentan estimar los autores en el artículo?

(b) ¿Cuál es la estrategia de identificación?

(c) Utilizar el archivo “CardKrueger1994.dta”. Utilizando diff, computar el estimador de diferencias en diferencias.

DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS

Number of observations in the DIFF-IN-DIFF: 801

	Before	After		
Control:	78	77	155	
Treated:	326	320	646	
	404	397		

Outcome var.	fte	S. Err.	t	P> t

Before				
Control	19.949			
Treated	17.065			
Diff (T-C)	-2.884	1.135	-2.54	0.011**
After				
Control	17.542			
Treated	17.573			
Diff (T-C)	0.030	1.143	0.03	0.979
Diff-in-Diff	2.914	1.611	1.81	0.071*

R-square: 0.01

* Means and Standard Errors are estimated by linear regression

Inference: * p<0.01; ** p<0.05; * p<0.1

(d) Repetir el inciso (c) utilizando errores estándar de bootstrap.

DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS

Number of observations in the DIFF-IN-DIFF: 801

	Before	After	
Control:	78	77	155
Treated:	326	320	646
	404	397	

Bootstrapped Standard Errors

Outcome var.	fte	S. Err.	t	P> t
Before				
Control	19.949			
Treated	17.065			
Diff (T-C)	-2.884	1.329	-2.17	0.030**
After				
Control	17.542			
Treated	17.573			
Diff (T-C)	0.030	1.080	0.03	0.977
Diff-in-Diff	2.914	1.620	1.80	0.072*

R-square: 0.01

* Means and Standard Errors are estimated by linear regression

Inference: * p<0.01; ** p<0.05; * p<0.1

(e) Repetir el inciso (c) utilizando la cadena de restaurantes como variables explicativas.

DIFFERENCE-IN-DIFFERENCES ESTIMATION RESULTS

Number of observations in the DIFF-IN-DIFF: 801

	Before	After	
Control:	78	77	155
Treated:	326	320	646
	404	397	

Report - Covariates and coefficients:

Variable(s)	Coeff.	Std. Err.	z	P> z
bk	0.917	1.060	0.865	0.387
kfc	-9.205	0.961	-9.575	0.000
roys	-0.897	1.136	-0.790	0.430

Bootstrapped Standard Errors

Outcome var.	fte	S. Err.	t	P> t
Before				
Control	21.161			
Treated	18.837			
Diff (T-C)	-2.324	1.276	-1.82	0.069*
After				
Control	18.758			
Treated	19.369			
Diff (T-C)	0.611	0.946	0.65	0.518
Diff-in-Diff	2.935	1.563	1.88	0.060*

R-square: 0.19

* Means and Standard Errors are estimated by linear regression

Inference: * p<0.01; ** p<0.05; * p<0.1

Ejercicio 4: didregress.

Hasta ahora, se estimó utilizando la especificación:

$$Y_{it} = \beta_0 + \beta_1 D_{it} + \beta_2 T_{it} + \beta_3 (D \times T)_{it} + z_{it} \theta + u_{it} \quad (1)$$

O, en el caso de datos longitudinales, se amplía la ecuación con efectos fijos. En Stata, se usa los comandos `regress` y/o `xtreg`. Stata 17 incluyó nuevos comandos para estimar modelos con la siguiente forma:

$$Y_{ist} = \gamma_s + \gamma_t + z_{ist} \beta + \delta D_{st} + u_{ist}$$

con el comando `didregress` o incluyendo efectos fijos por individuo en el caso de datos longitudinales con el comando `didregress`.

(a) Explicar en qué difieren estos comandos con respecto al setup usual de DiD.

(b) ¿Se puede replicar las regresiones de los ejercicios anteriores con estos comandos?

(c) Un proveedor de salud está interesado en estudiar el efecto de un nuevo procedimiento de ingreso hospitalario en la satisfacción de los pacientes. El proveedor dispone de datos mensuales de pacientes de enero a julio. El nuevo procedimiento de admisiones fue implementado en abril por hospitales que estaban bajo nueva administración. De los 46 hospitales del estudio, 18 implementaron el nuevo procedimiento. El proveedor de salud utilizará una regresión DID para analizar el efecto del nuevo procedimiento de admisiones en los hospitales que participaron en el programa. El resultado de interés es la satisfacción del paciente, `satis`, que se registra como un promedio de las respuestas a un conjunto de cuatro preguntas realizadas a los pacientes. `satis` puede tomar valores entre 0 y 10, donde 10 es el mayor nivel de satisfacción posible y 0 es la decepción total. La variable `procedimiento` marca las observaciones tratadas; es 1 si una persona encuestada ingresó al hospital utilizando el nuevo procedimiento después de marzo y 0 en caso contrario. Los datos están en la base "hospdd.dta". Evaluar el impacto del nuevo procedimiento sobre la satisfacción de los pacientes.

Number of groups and treatment time

Time variable: month

Control: procedure = 0

Treatment: procedure = 1

	Control	Treatment
Group		
hospital	28	18
Time		
Minimum	1	4
Maximum	1	4

Difference-in-differences regression

Number of obs = 7,368

Data type: Repeated cross-sectional

(Std. err. adjusted for 46 clusters in hospital)

	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]
satis					
ATET					
procedure					
(New vs Old)	.8479879	.0321121	26.41	0.000	.7833108 .912665

Note: ATET estimate adjusted for covariates, group effects, and time effects.

(d) ¿Cómo se interpreta el coeficiente obtenido? ¿Se cumple el supuesto de tendencias paralelas?

El coeficiente obtenido de la regresión DID (0,8479879) representa el efecto promedio del nuevo procedimiento de admisión hospitalaria sobre la satisfacción de los pacientes en los hospitales que implementaron el nuevo procedimiento, en comparación con aquellos que no lo hicieron. Específicamente, se estima que el nuevo procedimiento incrementa la satisfacción de los pacientes en, aproximadamente, 0,85 puntos en la escala de 0 a 10.

El supuesto de tendencias paralelas es crucial para la validez de la estimación del modelo DID. Este supuesto implica que, en ausencia del tratamiento (nuevo procedimiento), las tendencias de la satisfacción de los pacientes habrían sido similares en los hospitales tratados y no tratados. Sin embargo, la evaluación de este supuesto no se puede realizar directamente a partir de los coeficientes del modelo; se requiere un análisis gráfico o pruebas formales previas.

Para verificar si este supuesto se cumple, se podría graficar las tendencias de la satisfacción en ambos grupos (tratados y no tratados) antes de la implementación del nuevo procedimiento (es decir, de enero a marzo) y observar si las tendencias son similares. Si las tendencias son paralelas antes de la implementación, entonces, es razonable suponer que el supuesto se cumple.

Si los datos muestran que las tendencias de satisfacción entre los grupos tratados y no tratados divergen antes de abril, entonces, el supuesto de tendencias paralelas no se cumple y la interpretación causal del coeficiente podría ser sesgada. En tal caso, se deberían considerar ajustes adicionales o utilizar métodos alternativos para analizar el efecto del nuevo procedimiento.

(e) Comentar sobre los errores estándar utilizados y estudiar las distintas opciones que el comando tiene pre-configuradas para usar. ¿Hay diferencias en la inferencia?

Errores estándar robustos:

Number of groups and treatment time

Time variable: month

Control: procedure = 0

Treatment: procedure = 1

	Control	Treatment
Group		
hospital	28	18
Time		
Minimum	1	4
Maximum	1	4

Difference-in-differences regression
Data type: Repeated cross-sectional

Number of obs = 7,368

		Robust				
	satis	Coefficient	std. err.	z	P> z	[95% conf. interval]
ATET						
procedure						
(New vs Old)		.8479879	.0353158	24.01	0.000	.7787701 .9172056

Note: ATET estimate adjusted for covariates, group effects, and time effects.

Errores estándar hc2:

Number of groups and treatment time

Time variable: month

Control: procedure = 0

Treatment: procedure = 1

	Control	Treatment
Group		
hospital	28	18
Time		
Minimum	1	4
Maximum	1	4

Difference-in-differences regression
Data type: Repeated cross-sectional

Number of obs = 7,368
No. of clusters = 46

		Robust HC2				
	satis	Coefficient	std. err.	t	P> t	[95% conf. interval]
ATET						
procedure						
(New vs Old)		.8479879	.0325552	26.05	0.000	.7819941 .9139816

Note: ATET estimate adjusted for covariates, group effects, and time effects.

Errores estándar bootstrap:

Number of groups and treatment time

Time variable: month

Control: procedure = 0

Treatment: procedure = 1

	Control	Treatment
Group		
hospital	28	18
Time		
Minimum	1	4
Maximum	1	4

Difference-in-differences regression

Number of obs = 7,368

Replications = 50

Wald chi2(1) = 888.21

Prob > chi2 = 0.0000

Data type: Repeated cross-sectional

(Replications based on 46 clusters in hospital)

	Observed	Bootstrap			Normal-based
satis	coefficient	std. err.	z	P> z	[95% conf. interval]
ATET					
procedure					
(New vs Old)	.8479879	.0284533	29.80	0.000	.7922204 .9037553

Note: ATET estimate adjusted for covariates, group effects, and time effects.

Por lo tanto, se puede observar que no hay grandes diferencias en la inferencia, ya que el ATET es estadísticamente significativo al 1%.

Microeconometría II
Práctica 4
Extensiones de Diferencias-en-Diferencias

1. Static Two-Way Fixed Effects

Una forma de extender el *framework* de diferencias-en-diferencias cuando hay más períodos temporales disponibles es utilizar efectos fijos temporales adicionalmente a los efectos fijos por individuo. Es decir, si el resultado de interés es Y y el tratamiento está definido por D , entonces se puede estimar la ecuación

$$Y_{it} = \alpha_i + \phi_t + D_{it}\beta_{post} + \epsilon_{it}. \quad (1)$$

α_i son efectos fijos por individuo, ϕ_t son efectos fijos temporales y

1. ¿Cómo debe definirse D_{it} para que esto sea la generalización de la ecuación de DiD 2x2?
2. Escriba como queda la variable si $T = 3$ $N = 3$ con la unidad 1 sin tratar, la unidad 2 tratada en el período 1, la unidad 2 tratada en el período 2.
3. ¿Cuál es la expresión del estimador de β^{post} ?
4. ¿Cuándo podría este estimador ser bueno para recuperar el efecto promedio de tratamiento sobre los tratados? ¿Qué problemas podría tener esta especificación?
5. Utilizando la base de datos `organ_donations.dta` de Kessler & Roth (2014) estime el coeficiente β^{post} de esta especificación.

2. Dynamic Two-Way Fixed Effects

Con la especificación anterior se estima un único efecto de tratamiento. Sin embargo, si se posee información de varios períodos, ésta podría utilizarse para evaluar, por un lado, cómo cambian los efectos en el tiempo, y por otro, evaluar cómo se comportaba la variable de interés previo del otorgamiento del tratamiento. La especificación dinámica de TWFE es

$$Y_{it} = \alpha_i + \phi_t + \sum_{\substack{r \neq 0 \\ -T \leq r \leq T}} 1[R_{it} = r] \beta_r + \epsilon_{it}.$$

1. Escriba como quedan las variables si $T = 3$ $N = 3$ con la unidad 1 sin tratar, la unidad 2 tratada en el período 1, la unidad 2 tratada en el período 2.
2. Compute los efectos de tratamiento dinámicos β_r . Notar que se normaliza $\beta_0 = 0$. (¿Por qué?)
3. Grafique los efectos de tratamiento pre y post otorgamiento junto con sus intervalos de confianza.
4. Utilice los datos de los períodos anteriores al tratamiento para hacer un “*test de placebo*”:
 - Utilice solo los datos que llegaron antes de que el tratamiento entrara en vigor.
 - Elija un período de tratamiento falso.
 - Calcule el mismo modelo de diferencias en diferencias que planeaba usar (por ejemplo), pero cree la variable igual a 1 si está en el grupo tratado y después de la fecha de tratamiento falso que eligió.
 - Si encuentra un “efecto” para esa fecha de tratamiento donde realmente no debería haberlo, eso es evidencia de que hay algo mal con su diseño, lo que puede implicar una violación de tendencias paralelas.

3. Callaway & Sant'Anna y csdid

Una solución a los problemas de TWFE es la que proponen Callaway & Sant'Anna (2020). Ellos proponen computar todos los ATT válidos y ponderarlos adecuadamente. En Stata esto se puede hacer con el comando `csdid`. Utilizando la base de datos `mpdta.dta`, se busca estimar el impacto de una suba del salario mínimo en el empleo joven.

1. Estime todos los $ATT(g, t)$ sin variables explicativas.
2. Evalúe si es plausible el efecto de tendencias paralelas en base a las tendencias previas al otorgamiento del tratamiento. ¿Considera que puede haber habido factores que hayan afectado la evolución del empleo en todos los estados tratados que **no** se deba al otorgamiento del tratamiento? Reflexione acerca del rol de la forma funcional de las variables (por ejemplo, en niveles vs. en logaritmos).
3. Compute el efecto agregado simple, el efecto agregado por grupos, el efecto agregado por período y el efecto agregado por períodos tras el otorgamiento del tratamiento.
4. Repita los incisos anteriores condicionando en la variable de población.
5. Hasta ahora se utilizaron los nunca tratados como grupo de control Repita los incisos anteriores utilizando el grupo de los no tratados todavía. ¿Observa cambios?
6. Compare con los resultados que surgen de hacer TWFE estático y dinámico.

4. Enfoque de Wooldridge y jwddid

Frente a las críticas de la literatura a la presunta incapacidad de la especificación de TWFE para estimar los efectos promedio de tratamiento sobre los tratados, Wooldridge (2021) propone que el problema no es intrínseco a que la ecuación de estimación sea lineal con efectos fijos ni a que se estime con los métodos tradicionales de datos de panel.¹ En particular, con T períodos temporales y el tratamiento otorgado en cada período desde el período q hasta el último se propone estimar la siguiente ecuación:

$$Y_{it} = \alpha + \lambda_q d_{iq} + \dots + \lambda_T d_{iT} + \sum_{r=q}^T \sum_{s=r}^T \tau_{rs} (d_{ir} \cdot 1\{t = s\}) + \theta_t + \epsilon_{it}$$

donde τ_{rs} representa el ATT en el período s para el grupo r , α es una constante, d_{ir} es una dummy que vale 1 para el grupo tratado en r , con $r = q, \dots, T$. Replique el ejercicio anterior con el comando `jwddid`.²

¹En particular, el artículo plantea una equivalencia entre una especificación de efectos fijos y una especificación two-way Mundlak, dejando el caso de DiD como una aplicación.

²Debe agregar los archivos `.ado` a su directorio. No va a poder replicar el test de tendencias previas ni los gráficos, al menos no al momento en el que se escribe esto.

Trabajo Práctico N° 4: **Extensiones de Diferencias en Diferencias.**

Ejercicio 1: Static Two-Way Fixed Effects.

Una forma de extender el framework de diferencias en diferencias cuando hay más períodos temporales disponibles es utilizar efectos fijos temporales adicionalmente a los efectos fijos por individuo. Es decir, si el resultado de interés es Y y el tratamiento está definido por D , entonces, se puede estimar la ecuación:

$$Y_{it} = \alpha_i + \phi_t + D_{it}\beta_{post} + \epsilon_{it} \quad (1)$$

donde α_i son efectos fijos por individuo, ϕ_t son efectos fijos temporales.

(a) ¿Cómo debe definirse D_{it} para que esto sea la generalización de la ecuación de DiD 2x2?

$D_{it} = \mathbb{1}\{i \text{ tratado}\} \{t \geq \text{fecha de otorgamiento del tratamiento}\}.$

(b) Escribir cómo queda la variable si $T = 3$ y $N = 3$ con la unidad 1 sin tratar, la unidad 2 tratada en el período 1, la unidad 3 tratada en el período 2.

id	t	Cuándo	D_t	D_{it}
1	1	0	0	0
1	2	0	0	0
1	3	0	0	0
2	1	1	1	1
2	2	1	1	1
2	3	1	1	1
3	1	2	1	0
3	2	2	1	1
3	3	2	1	1

(c) ¿Cuál es la expresión del estimador de β_{post} ?

La expresión del estimador de β_{post} es:

$$\hat{\beta}_{post} = \frac{\sum_{i,t} (D_{it} - \bar{D}_{it})}{\sum_{i,t} (D_{it} - \bar{D}_{it})^2}.$$

(d) ¿Cuándo podría este estimador ser bueno para recuperar el efecto promedio de tratamiento sobre los tratados? ¿Qué problemas podría tener esta especificación?

Este estimador podría ser bueno para recuperar el efecto promedio de tratamiento sobre los tratados cuando todas las unidades tienen el mismo efecto de tratamiento y cuando éste es idéntico independientemente de cuánto tiempo haya pasado desde que comenzó el tratamiento. Entonces, los problemas que podría tener esta especificación es heterogeneidad en efecto de tratamiento en términos de unidades de corte transversal y tiempo transcurrido desde el tratamiento.

(e) Utilizando la base de datos “organ_donations.dta” de Kessler & Roth (2014), estimar el coeficiente β_{post} de esta especificación.

HDFE Linear regression	Number of obs	=	162
Absorbing 2 HDFE groups	F(1, 26)	=	13.42
Statistics robust to heteroskedasticity	Prob > F	=	0.0011
	R-squared	=	0.9793
	Adj R-squared	=	0.9742
	Within R-sq.	=	0.0092
Number of clusters (state) = 27	Root MSE	=	0.0246

(Std. err. adjusted for 27 clusters in state)

	rate	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
D		-.022459	.0061312	-3.66	0.001	-.0350619	-.0098561
_cons		.4454641	.0001135	3923.36	0.000	.4452307	.4456974

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
state	27	27	0 *
quarter	6	1	5

* = FE nested within cluster; treated as redundant for DoF computation

Ejercicio 2: Dynamic Two-Way Fixed Effects.

Con la especificación anterior, se estima un único efecto de tratamiento. Sin embargo, si se posee información de varios períodos, ésta podrá utilizarse para evaluar, por un lado, cómo cambian los efectos en el tiempo y, por otro, evaluar cómo se comportaba la variable de interés previo del otorgamiento del tratamiento. La especificación dinámica de TWFE es:

$$Y_{it} = \alpha_i + \phi_t + \sum_{\substack{r \neq 0 \\ -\underline{T} \leq r \leq \bar{T}}} 1[R_{it} = r] \beta_r + \epsilon_{it}.$$

(a) Escribir cómo quedan las variables si $T = 3$ y $N = 3$ con la unidad 1 sin tratar, la unidad 2 tratada en el período 1, la unidad 3 tratada en el período 2.

id	t	Cuándo	D_i	D_{it}	t(-1)	t(0)	t(+1)	t(+2)
1	1	0	0	0	0	0	0	0
1	2	0	0	0	0	0	0	0
1	3	0	0	0	0	0	0	0
2	1	1	1	1	0	1	0	0
2	2	1	1	1	0	0	1	0
2	3	1	1	1	0	0	0	1
3	1	2	1	0	1	0	0	0
3	2	2	1	1	0	1	0	0
3	3	2	1	1	0	0	1	0

(b) Computar los efectos de tratamiento dinámicos β_r . Notar que se normaliza $\beta_0 = 0$. (¿Por qué?).

HDFE Linear regression
Absorbing 2 HDFE groups
Statistics robust to heteroskedasticity

Number of obs = 162
F(5, 26) = 5.79
Prob > F = 0.0010
R-squared = 0.9793
Adj R-squared = 0.9734
Within R-sq. = 0.0098
Root MSE = 0.0250

Number of clusters (state) = 27

(Std. err. adjusted for 27 clusters in state)

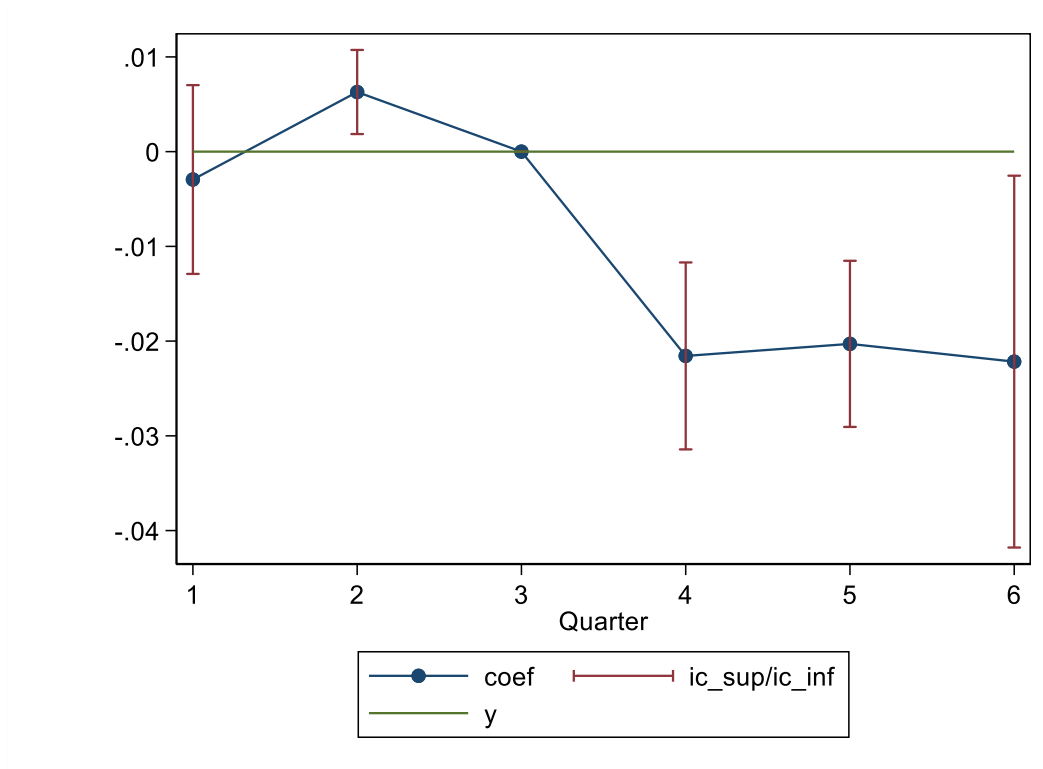
	rate	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
1.California		0	(omitted)				
quarter_num							
Quarter 4 2010		0	(omitted)				
Quarter 1 2011		0	(omitted)				
Quarter 3 2011		0	(omitted)				
Quarter 4 2011		0	(omitted)				
Quarter 1 2012		0	(omitted)				
California#quarter_num							
1#Quarter 4 2010		-.0029423	.0050842	-0.58	0.568	-.013393	.0075084
1#Quarter 1 2011		.0062961	.0022658	2.78	0.010	.0016388	.0109535
1#Quarter 3 2011		-.0215654	.0050337	-4.28	0.000	-.0319124	-.0112184
1#Quarter 4 2011		-.0202923	.0044733	-4.54	0.000	-.0294874	-.0110973
1#Quarter 1 2012		-.0221654	.0100132	-2.21	0.036	-.0427479	-.0015829
_cons		.4454226	.0001052	4232.23	0.000	.4452063	.445639

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs	
state	27	27	0	*
quarter_num	6	1	5	

* = FE nested within cluster; treated as redundant for DoF computation

(c) Graficar los efectos de tratamiento pre y post otorgamiento junto con sus intervalos de confianza.



(d) Utilizar los datos de los períodos anteriores al tratamiento para hacer un “test de placebo”:

- Utilizar sólo los datos que llegaron antes de que el tratamiento entrara en vigor.
- Elegir un período de tratamiento falso.
- Calcular el mismo modelo de diferencias en diferencias que se planeaba usar (por ejemplo), pero crear la variable igual a 1 si está en el grupo tratado y después de la fecha de tratamiento falso que se eligió.
- Si se encuentra un “efecto” para esa fecha de tratamiento donde, realmente, no debería haberlo, eso es evidencia de que hay algo mal con el diseño, lo que puede implicar una violación de tendencias paralelas.

HDFE Linear regression	Number of obs	=	81
Absorbing 2 HDFE groups	F(1, 26)	=	1.43
Statistics robust to heteroskedasticity	Prob > F	=	0.2421
	R-squared	=	0.9938
	Adj R-squared	=	0.9902
	Within R-sq.	=	0.0019
Number of clusters (state)	Root MSE	=	0.0156

(Std. err. adjusted for 27 clusters in state)

	rate	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
FakeTreat1		.0060904	.0050881	1.20	0.242	-.0043684	.0165492
_cons		.4383509	.0001256	3489.15	0.000	.4380926	.4386091

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs	
state	27	27	0	*
quarter	3	1	2	

* = FE nested within cluster; treated as redundant for DoF computation

HDFE Linear regression	Number of obs	=	81
Absorbing 2 HDFE groups	F(1, 26)	=	0.36
Statistics robust to heteroskedasticity	Prob > F	=	0.5540
	R-squared	=	0.9938
	Adj R-squared	=	0.9902
	Within R-sq.	=	0.0001
Number of clusters (state)	Root MSE	=	0.0156

(Std. err. adjusted for 27 clusters in state)

	rate	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
FakeTreat2		-.0016769	.0027968	-0.60	0.554	-.0074259	.004072
_cons		.4385219	.0000345	1.3e+04	0.000	.438451	.4385929

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs	
state	27	27	0	*
quarter	3	1	2	

* = FE nested within cluster; treated as redundant for DoF computation

Por lo tanto, no se encuentra un efecto estadísticamente significativo para estas fechas de tratamiento falsas, por lo que es evidencia de que el diseño está bien hecho.

Ejercicio 3: Callaway & Sant'Anna y csdid.

Una solución a los problemas de TWFE es la que proponen Callaway & Sant'Anna (2020). Ellos proponen computar todos los ATT válidos y ponderarlos adecuadamente. En Stata, esto se puede hacer con el comando csdid. Utilizando la base de datos "mpdta.dta", se busca estimar el impacto de una suba del salario mínimo en el empleo joven.

(a) Estimar todos los ATT (g, t) sin variables explicativas.

Estimación (sin variables explicativas y sin utilizar el grupo de los no tratados todavía):

```
Difference-in-difference with Multiple Time Periods

Number of obs = 2,500

Outcome model : least squares
Treatment model: inverse probability
```

		Coefficient	Std. err.	z	P> z	[95% conf. interval]
-----+-----						
g2004						
t_2003_2004		-.0105032	.023251	-0.45	0.651	-.0560744 .0350679
t_2003_2005		-.0704232	.0309848	-2.27	0.023	-.1311522 -.0096941
t_2003_2006		-.1372587	.0364357	-3.77	0.000	-.2086713 -.0658461
t_2003_2007		-.1008114	.0343592	-2.93	0.003	-.1681542 -.0334685
-----+-----						
g2006						
t_2003_2004		.0065201	.0233268	0.28	0.780	-.0391996 .0522398
t_2004_2005		-.0027508	.0195586	-0.14	0.888	-.0410849 .0355833
t_2005_2006		-.0045946	.0177552	-0.26	0.796	-.0393942 .0302049
t_2005_2007		-.0412245	.0202292	-2.04	0.042	-.0808729 -.001576
-----+-----						
g2007						
t_2003_2004		.0305067	.0150336	2.03	0.042	.0010414 .0599719
t_2004_2005		-.0027259	.0163958	-0.17	0.868	-.0348611 .0294093
t_2005_2006		-.0310871	.0178775	-1.74	0.082	-.0661264 .0039522
t_2006_2007		-.0260544	.0166554	-1.56	0.118	-.0586985 .0065896
-----+-----						

Control: Never Treated

See Callaway and Sant'Anna (2021) for details

(b) Evaluar si es plausible el efecto de tendencias paralelas en base a las tendencias previas al otorgamiento del tratamiento. ¿Puede haber habido factores que hayan afectado la evolución del empleo en todos los estados tratados que no se deba al otorgamiento del tratamiento? Reflexionar acerca del rol de la forma funcional de las variables (por ejemplo, en niveles vs. en logaritmos).

Test para tendencias paralelas (sin variables explicativas y sin utilizar el grupo de los no tratados todavía):

```
Pretrend Test. H0 All Pre-treatment are equal to 0
chi2(5) = 7.7912
p-value = 0.1681
```

(c) *Computar el efecto agregado simple, el efecto agregado por grupos, el efecto agregado por período y el efecto agregado por períodos tras el otorgamiento del tratamiento.*

Efecto agregado simple (sin variables explicativas y sin utilizar el grupo de los no tratados todavía):

Average Treatment Effect on Treated

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
ATT	-.0399513	.012034	-3.32	0.001	-.0635375	-.016365

Efecto agregado por grupos (sin variables explicativas y sin utilizar el grupo de los no tratados todavía):

ATT by group

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
GAverage	-.0310183	.0123872	-2.50	0.012	-.0552967	-.0067399
G2004	-.0797491	.0263678	-3.02	0.002	-.1314291	-.0280692
G2006	-.0229095	.0167033	-1.37	0.170	-.0556475	.0098284
G2007	-.0260544	.0166554	-1.56	0.118	-.0586985	.0065896

Efecto agregado por período (sin variables explicativas y sin utilizar el grupo de los no tratados todavía):

ATT by Calendar Period

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
CAverage	-.0417004	.0159719	-2.61	0.009	-.0730047	-.0103962
T2004	-.0105032	.023251	-0.45	0.651	-.0560744	.0350679
T2005	-.0704232	.0309848	-2.27	0.023	-.1311522	-.0096941
T2006	-.048816	.0201259	-2.43	0.015	-.0882619	-.00937
T2007	-.0370593	.0137471	-2.70	0.007	-.0640031	-.0101156

Efecto agregado por períodos tras el otorgamiento del tratamiento (sin variables explicativas y sin utilizar el grupo de los no tratados todavía):

ATT by Periods Before and After treatment
Event Study:Dynamic effects

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
Pre_avg	.0018283	.007657	0.24	0.811	-.0131791	.0168357
Post_avg	-.0772398	.019965	-3.87	0.000	-.1163705	-.0381092
Tm3	.0305067	.0150336	2.03	0.042	.0010414	.0599719
Tm2	-.0005631	.0132916	-0.04	0.966	-.0266142	.0254881
Tm1	-.0244587	.0142364	-1.72	0.086	-.0523616	.0034441
Tp0	-.0199318	.0118264	-1.69	0.092	-.0431111	.0032474
Tp1	-.0509574	.0168935	-3.02	0.003	-.084068	-.0178468
Tp2	-.1372587	.0364357	-3.77	0.000	-.2086713	-.0658461
Tp3	-.1008114	.0343592	-2.93	0.003	-.1681542	-.0334685

Estimación (con variables explicativas y sin utilizar el grupo de los no tratados todavía):

```
Number of obs = 2,500
```

Treatment model: inverse probability

Control: Never Treated

Test para tendencias paralelas (con variables explicativas y sin utilizar el grupo de los no tratados todavía):

p-value = 0.2327

Efecto agregado simple (con variables explicativas y sin utilizar el grupo de los no tratados todavía):

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
ATT	-.0417518	.0115028	-3.63	0.000	-.0642969	-.0192066

Efecto agregado por grupos (con variables explicativas y sin utilizar el grupo de los no tratados todavía):

ATT by group

		Coefficient	Std. err.	z	P> z	[95% conf. interval]	
GAverage		-.0310183	.0123872	-2.50	0.012	-.0552967	-.0067399
G2004		-.0797491	.0263678	-3.02	0.002	-.1314291	-.0280692
G2006		-.0229095	.0167033	-1.37	0.170	-.0556475	.0098284
G2007		-.0260544	.0166554	-1.56	0.118	-.0586985	.0065896

Efecto agregado por período (con variables explicativas y sin utilizar el grupo de los no tratados todavía):

ATT by Calendar Period

		Coefficient	Std. err.	z	P> z	[95% conf. interval]	
CAverage		-.0441774	.0150382	-2.94	0.003	-.0736516	-.0147031
T2004		-.0145297	.0221292	-0.66	0.511	-.057902	.0288427
T2005		-.0764219	.0286713	-2.67	0.008	-.1326166	-.0202271
T2006		-.0461757	.0212107	-2.18	0.029	-.087748	-.0046035
T2007		-.0395822	.0129299	-3.06	0.002	-.0649242	-.0142401

Efecto agregado por períodos tras el otorgamiento del tratamiento (con variables explicativas y sin utilizar el grupo de los no tratados todavía):

ATT by Periods Before and After treatment
Event Study:Dynamic effects

		Coefficient	Std. err.	z	P> z	[95% conf. interval]	
Pre_avg		-.0000442	.0075204	-0.01	0.995	-.014784	.0146955
Post_avg		-.0803539	.0189576	-4.24	0.000	-.1175101	-.0431978
Tm3		.0267278	.0140657	1.90	0.057	-.0008404	.054296
Tm2		-.0036165	.0129283	-0.28	0.780	-.0289555	.0217226
Tm1		-.023244	.0144851	-1.60	0.109	-.0516343	.0051463
Tp0		-.0210604	.0114942	-1.83	0.067	-.0435886	.0014679
Tp1		-.0530032	.0163465	-3.24	0.001	-.0850417	-.0209647
Tp2		-.1404483	.0353782	-3.97	0.000	-.2097882	-.0711084
Tp3		-.1069039	.0328865	-3.25	0.001	-.1713602	-.0424476

(e) Hasta ahora, se utilizaron los nunca tratados como grupo de control. Repetir los incisos anteriores utilizando el grupo de los no tratados todavía. ¿Se observan cambios?

Difference-in-difference with Multiple Time Periods

Outcome model : least squares
Treatment model: inverse probability

Control: Not yet Treated

Test para tendencias paralelas (sin variables explicativas y utilizando el grupo de los no tratados todavía):

Efecto agregado simple (sin variables explicativas y utilizando el grupo de los no tratados todavía):

	Coefficient	Std. err.	z	P> z	[95% conf. interval]
ATT	-.0397636	.0120524	-3.30	0.001	-.0633859 -.0161413

ATT by group

	Coefficient	Std. err.	z	P> z	[95% conf. interval]
GAverage	-.0304622	.0125031	-2.44	0.015	-.0549678 -.0059566
G2004	-.0836943	.0257016	-3.26	0.001	-.1340685 -.0333201
G2006	-.0182818	.0159222	-1.15	0.251	-.0494888 .0129252
G2007	-.0260544	.0166554	-1.56	0.118	-.0586985 .0065896

Efecto agregado por período (sin variables explicativas y utilizando el grupo de los no tratados todavía):

ATT by Calendar Period

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
CAverage	-.0442671	.0155709	-2.84	0.004	-.0747855	-.0137487
T2004	-.0193724	.0223101	-0.87	0.385	-.0630994	.0243547
T2005	-.0783191	.0303902	-2.58	0.010	-.1378829	-.0187553
T2006	-.0423175	.0190563	-2.22	0.026	-.0796671	-.004968
T2007	-.0370593	.0137471	-2.70	0.007	-.0640031	-.0101156

Efecto agregado por períodos tras el otorgamiento del tratamiento (sin variables explicativas y utilizando el grupo de los no tratados todavía):

ATT by Periods Before and After treatment
Event Study:Dynamic effects

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
Pre_avg	.0011834	.0073393	0.16	0.872	-.0132015	.0155682
Post_avg	-.0773993	.0195602	-3.96	0.000	-.1157366	-.0390621
Tm3	.0305067	.0150336	2.03	0.042	.0010414	.0599719
Tm2	-.0026877	.0134388	-0.20	0.841	-.0290273	.0236519
Tm1	-.0242689	.0144637	-1.68	0.093	-.0526172	.0040794
Tp0	-.0189222	.0120446	-1.57	0.116	-.0425291	.0046847
Tp1	-.0535893	.0169464	-3.16	0.002	-.0868037	-.020375
Tp2	-.1362743	.0354034	-3.85	0.000	-.2056637	-.066885
Tp3	-.1008114	.0343592	-2.93	0.003	-.1681542	-.0334685

Estimación (con variables explicativas y utilizando el grupo de los no tratados todavía):

Difference-in-difference with Multiple Time Periods

Number of obs = 2,500

Outcome model : least squares

Treatment model: inverse probability

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
g2004						
t_2003_2004	-.0211831	.0216482	-0.98	0.328	-.0636128	.0212467
t_2003_2005	-.0816032	.0283415	-2.88	0.004	-.1371516	-.0260548
t_2003_2006	-.1381918	.034228	-4.04	0.000	-.2052775	-.0711061
t_2003_2007	-.1069039	.0328865	-3.25	0.001	-.1713602	-.0424476
g2006						
t_2003_2004	-.0074552	.0218357	-0.34	0.733	-.0502525	.035342
t_2004_2005	-.0045634	.0182914	-0.25	0.803	-.0404138	.0312871
t_2005_2006	.0086607	.0168391	0.51	0.607	-.0243433	.0416647
t_2005_2007	-.0412939	.0197211	-2.09	0.036	-.0799466	-.0026411
g2007						
t_2003_2004	.0267278	.0140657	1.90	0.057	-.0008404	.054296
t_2004_2005	-.0045766	.0157178	-0.29	0.771	-.0353828	.0262297
t_2005_2006	-.0284475	.0181809	-1.56	0.118	-.0640814	.0071864
t_2006_2007	-.0287814	.016239	-1.77	0.076	-.0606091	.0030464

Control: Not yet Treated

See Callaway and Sant'Anna (2021) for details

Test para tendencias paralelas (con variables explicativas y utilizando el grupo de los no tratados todavía):

Pretrend Test. H0 All Pre-treatment are equal to 0
chi2(5) = 6.8655
p-value = 0.2308

Efecto agregado simple (con variables explicativas y utilizando el grupo de los no tratados todavía):

Average Treatment Effect on Treated

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
ATT	-.0413516	.0114278	-3.62	0.000	-.0637498	-.0189535

Efecto agregado por grupos (con variables explicativas y utilizando el grupo de los no tratados todavía):

ATT by group

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
GAverage	-.032264	.0119004	-2.71	0.007	-.0555884	-.0089397
G2004	-.0869705	.024062	-3.61	0.000	-.1341312	-.0398097
G2006	-.0163166	.0161883	-1.01	0.313	-.0480451	.015412
G2007	-.0287814	.016239	-1.77	0.076	-.0606091	.0030464

Efecto agregado por período (con variables explicativas y utilizando el grupo de los no tratados todavía):

ATT by Calendar Period

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
CAverage	-.0456646	.0146983	-3.11	0.002	-.0744727	-.0168566
T2004	-.0211831	.0216482	-0.98	0.328	-.0636128	.0212467
T2005	-.0816032	.0283415	-2.88	0.004	-.1371516	-.0260548
T2006	-.0402901	.0192635	-2.09	0.036	-.0780459	-.0025344
T2007	-.0395822	.0129299	-3.06	0.002	-.0649242	-.0142401

Efecto agregado por períodos tras el otorgamiento del tratamiento (con variables explicativas y utilizando el grupo de los no tratados todavía):

ATT by Periods Before and After treatment
Event Study:Dynamic effects

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
Pre_avg	-.0004609	.0072816	-0.06	0.950	-.0147326	.0138108
Post_avg	-.0799926	.018494	-4.33	0.000	-.1162402	-.043745
Tm3	.0267278	.0140657	1.90	0.057	-.0008404	.054296
Tm2	-.0052499	.0130015	-0.40	0.686	-.0307325	.0202326
Tm1	-.0228606	.0146118	-1.56	0.118	-.0514991	.005778
Tp0	-.0201445	.0116323	-1.73	0.083	-.0429433	.0026544
Tp1	-.0547303	.0164	-3.34	0.001	-.0868738	-.0225869
Tp2	-.1381918	.034228	-4.04	0.000	-.2052775	-.0711061
Tp3	-.1069039	.0328865	-3.25	0.001	-.1713602	-.0424476

(f) Comparar con los resultados que surgen de hacer TWFE estático y dinámico.

TWFE estático:

HDFE Linear regression	Number of obs	=	2,500
Absorbing 2 HDFE groups	F(1, 499)	=	7.59
Statistics robust to heteroskedasticity	Prob > F	=	0.0061
	R-squared	=	0.9932
	Adj R-squared	=	0.9915
	Within R-sq.	=	0.0042
Number of clusters (countyreal) =	500	Root MSE	= 0.1391

(Std. err. adjusted for 500 clusters in countyreal)

	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
D	-.0365489	.0132652	-2.76	0.006	-.0626114	-.0104865
_cons	5.776771	.0015441	3741.28	0.000	5.773737	5.779804

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
countyreal	500	500	0 *
year	5	1	4

* = FE nested within cluster; treated as redundant for DoF computation

TWFE dinámico:

HDFE Linear regression	Number of obs	=	2,500
Absorbing 2 HDFE groups	F(7, 499)	=	3.60
Statistics robust to heteroskedasticity	Prob > F	=	0.0009
	R-squared	=	0.9933
	Adj R-squared	=	0.9915
	Within R-sq.	=	0.0103
Number of clusters (countyreal) =	500	Root MSE	= 0.1388

(Std. err. adjusted for 500 clusters in countyreal)

		Robust				
	lemp	Coefficient	std. err.	t	P> t	[95% conf. interval]
Tm4		.0216933	.0253871	0.85	0.393	-.0281855 .071572
Tm3		.0427674	.0199126	2.15	0.032	.0036446 .0818902
Tm2		.0414987	.0162692	2.55	0.011	.0095342 .0734633
Tm1		.0181439	.0109822	1.65	0.099	-.0034332 .039721
Tp1		-.0253284	.0160191	-1.58	0.114	-.0568016 .0061447
Tp2		-.1136509	.0270505	-4.20	0.000	-.1667978 -.060504
Tp3		-.0741029	.031016	-2.39	0.017	-.1350409 -.0131648
_cons		5.76634	.0040209	1434.09	0.000	5.75844 5.77424

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
countyreal	500	500	0 *
year	5	1	4

* = FE nested within cluster; treated as redundant for DoF computation

Ejercicio 4: Enfoque de Wooldridge y jwddid.

Frente a las críticas de la literatura a la presunta incapacidad de la especificación de TWFE para estimar los efectos promedio de tratamiento sobre los tratados, Wooldridge (2021) propone que el problema no es intrínseco a que la ecuación de estimación sea lineal con efectos fijos ni a que se estime con los métodos tradicionales de datos de panel. En particular, con T períodos temporales y el tratamiento otorgado en cada período desde el período q hasta el último, se propone estimar la siguiente ecuación:

$$Y_{it} = \alpha + \lambda_q d_{iq} + \dots + \lambda_T d_{iT} + \sum_{r=q}^T \sum_{s=r}^T \tau_{rs} (d_{ir} 1\{t = s\}) + \theta_t + \epsilon_{it},$$

donde τ_{rs} representa el ATT en el período s para el grupo r , α es una constante, d_{ir} es una dummy que vale 1 para el grupo tratado en r , con $r = q, \dots, T$. Replicar el ejercicio anterior con el comando `jwddid`.

Estimación (sin variables explicativas y sin utilizar el grupo de los no tratados todavía):

HDFE Linear regression	Number of obs	=	2,500
Absorbing 2 HDFE groups	F(12, 499)	=	2.87
Statistics robust to heteroskedasticity	Prob > F	=	0.0008
	R-squared	=	0.0288
	Adj R-squared	=	0.0213
	Within R-sq.	=	0.0001
Number of clusters (countyreal) = 500	Root MSE	=	1.4926

(Std. err. adjusted for 500 clusters in countyreal)

	lemp	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]
first_treat#year#c._tr						
2004 2004		-.0105032	.0233633	-0.45	0.653	-.0564058 .0353993
2004 2005		-.0704232	.0311344	-2.26	0.024	-.1315938 -.0092525
2004 2006		-.1372587	.0366116	-3.75	0.000	-.2091906 -.0653269
2004 2007		-.1008114	.0345251	-2.92	0.004	-.1686439 -.0329788
2006 2003		-.0037693	.0314934	-0.12	0.905	-.0656452 .0581067
2006 2004		.0027508	.019653	0.14	0.889	-.035862 .0413637
2006 2006		-.0045946	.0178409	-0.26	0.797	-.0396472 .030458
2006 2007		-.0412245	.0203269	-2.03	0.043	-.0811613 -.0012877
2007 2003		.0033064	.0245699	0.13	0.893	-.0449669 .0515796
2007 2004		.033813	.0212312	1.59	0.112	-.0079006 .0755266
2007 2005		.0310871	.0179638	1.73	0.084	-.004207 .0663812
2007 2007		-.0260544	.0167359	-1.56	0.120	-.0589358 .006827
_cons		5.773609	.0666494	86.63	0.000	5.642661 5.904557

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
first_treat	4	0	4
year	5	1	4

Estimación (con variables explicativas y sin utilizar el grupo de los no tratados todavía):

HDFE Linear regression	Number of obs	=	2,500
Absorbing 2 HDFE groups	F(32, 499)	=	260.29
Statistics robust to heteroskedasticity	Prob > F	=	0.0000
	R-squared	=	0.8732
	Adj R-squared	=	0.8712
	Within R-sq.	=	0.8695
Number of clusters (countyreal) =	Root MSE	=	0.5414
500			

(Std. err. adjusted for 500 clusters in countyreal)

	lemp	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
first_treat#year#c._tr_							
2004 2004		-.0149112	.0222514	-0.67	0.503	-.0586293	.0288068
2004 2005		-.0769963	.0278076	-2.77	0.006	-.1316307	-.022362
2004 2006		-.1410801	.0322892	-4.37	0.000	-.2045196	-.0776406
2004 2007		-.1075443	.0329232	-3.27	0.001	-.1722294	-.0428592
2006 2003		.0090343	.0302653	0.30	0.765	-.0504289	.0684976
2006 2004		.0069683	.0182179	0.38	0.702	-.028825	.0427616
2006 2006		.0007655	.0186594	0.04	0.967	-.0358952	.0374263
2006 2007		-.0415356	.0192255	-2.16	0.031	-.0793085	-.0037627
2007 2003		.0068961	.0246894	0.28	0.780	-.0416118	.055404
2007 2004		.0332619	.021331	1.56	0.120	-.0086478	.0751717
2007 2005		.0285021	.0182913	1.56	0.120	-.0074353	.0644395
2007 2007		-.0287895	.0161542	-1.78	0.075	-.0605281	.0029491
first_treat#year#c._tr_#c._x_lpop							
2004 2004		.0005953	.0183817	0.03	0.974	-.0355198	.0367104
2004 2005		.0234096	.018401	1.27	0.204	-.0127434	.0595626
2004 2006		.0482261	.0224513	2.15	0.032	.0041154	.0923369
2004 2007		.0091886	.0271808	0.34	0.735	-.0442144	.0625916
2006 2003		-.0126074	.0243681	-0.52	0.605	-.0604842	.0352693
2006 2004		-.0177865	.0162122	-1.10	0.273	-.0496391	.0140661
2006 2006		.0282074	.0141414	1.99	0.047	.0004234	.0559915
2006 2007		.0277793	.0181101	1.53	0.126	-.0078021	.0633607
2007 2003		.0083787	.0254399	0.33	0.742	-.0416037	.0583611
2007 2004		-.0079105	.0188942	-0.42	0.676	-.0450324	.0292114
2007 2005		-.0025825	.0178553	-0.14	0.885	-.0376633	.0324982
2007 2007		-.0203637	.0162347	-1.25	0.210	-.0522606	.0115331
lpop		1.0634	.0212754	49.98	0.000	1.0216	1.1052
first_treat#c.lpop							
2004		.0530435	.0374929	1.41	0.158	-.02062	.126707
2006		-.030815	.0461109	-0.67	0.504	-.1214104	.0597803
2007		.056047	.0439926	1.27	0.203	-.0303866	.1424806
year#c.lpop							
2004		.0150462	.0092116	1.63	0.103	-.0030521	.0331445
2005		.0224368	.0091008	2.47	0.014	.0045562	.0403173
2006		.0130438	.0132642	0.98	0.326	-.0130168	.0391044
2007		.0229821	.0129513	1.77	0.077	-.0024638	.0484279
_cons		2.152986	.0722075	29.82	0.000	2.011117	2.294854

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
first_treat	4	0	4
year	5	1	4

Estimación (sin variables explicativas y utilizando el grupo de los no tratados todavía):

HDFE Linear regression	Number of obs	=	2,500
Absorbing 2 HDFE groups	F(7, 499)	=	3.81
Statistics robust to heteroskedasticity	Prob > F	=	0.0005
	R-squared	=	0.0288
	Adj R-squared	=	0.0233
	Within R-sq.	=	0.0001
Number of clusters (countyreal) =	Root MSE	=	1.4911

(Std. err. adjusted for 500 clusters in countyreal)

	lemp	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
first_treat#year#c._tr_							
2004 2004		-.0193724	.0223953	-0.87	0.387	-.063373	.0246283
2004 2005		-.0783191	.0305062	-2.57	0.011	-.1382556	-.0183826
2004 2006		-.1360781	.0354769	-3.84	0.000	-.2057806	-.0663756
2004 2007		-.1047075	.0338947	-3.09	0.002	-.1713015	-.0381135
2006 2006		.0025139	.0199448	0.13	0.900	-.0366724	.0417001
2006 2007		-.0391927	.0240232	-1.63	0.103	-.0863919	.0080064
2007 2007		-.043106	.0184423	-2.34	0.020	-.0793401	-.006872
_cons		5.77807	.0665051	86.88	0.000	5.647405	5.908734

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
first_treat	4	0	4
year	5	1	4

Estimación (con variables explicativas y utilizando el grupo de los no tratados todavía):

HDFE Linear regression	Number of obs	=	2,500
Absorbing 2 HDFE groups	F(22, 499)	=	364.06
Statistics robust to heteroskedasticity	Prob > F	=	0.0000
	R-squared	=	0.8732
	Adj R-squared	=	0.8717
	Within R-sq.	=	0.8695
Number of clusters (countyreal) =	Root MSE	=	0.5404
500			

(Std. err. adjusted for 500 clusters in countyreal)

	lemp	Coefficient	Robust std. err.	t	P> t	[95% conf. interval]	
first_treat#year#c.__tr__							
2004 2004		-.021248	.021724	-0.98	0.329	-.0639298	.0214338
2004 2005		-.08185	.0273694	-2.99	0.003	-.1356234	-.0280766
2004 2006		-.1378704	.0307884	-4.48	0.000	-.1983612	-.0773796
2004 2007		-.1095395	.0323153	-3.39	0.001	-.1730302	-.0460487
2006 2006		.0025368	.018879	0.13	0.893	-.0345554	.039629
2006 2007		-.0450935	.0219826	-2.05	0.041	-.0882834	-.0019035
2007 2007		-.0459545	.0179714	-2.56	0.011	-.0812636	-.0106455
first_treat#year#c.__tr__#c._x_lpop							
2004 2004		.0046278	.0175804	0.26	0.792	-.0299129	.0391685
2004 2005		.0251131	.0179003	1.40	0.161	-.0100561	.0602822
2004 2006		.0507346	.0210659	2.41	0.016	.0093457	.0921234
2004 2007		.0112497	.0266118	0.42	0.673	-.0410353	.0635346
2006 2006		.0389352	.0164686	2.36	0.018	.0065789	.0712915
2006 2007		.0380597	.0224724	1.69	0.091	-.0060925	.082212
2007 2007		-.0198351	.0161949	-1.22	0.221	-.0516538	.0119835
lpop		1.065461	.0218238	48.82	0.000	1.022583	1.108339
first_treat#c.lpop							
2004		.0509824	.0377558	1.35	0.178	-.0231975	.1251622
2006		-.0410954	.0473896	-0.87	0.386	-.1342031	.0520122
2007		.0555184	.0392124	1.42	0.157	-.0215233	.1325601
year#c.lpop							
2004		.0110137	.0075537	1.46	0.145	-.0038274	.0258548
2005		.0207333	.0081044	2.56	0.011	.0048103	.0366564
2006		.0105354	.0108157	0.97	0.330	-.0107145	.0317853
2007		.020921	.0118084	1.77	0.077	-.0022793	.0441212
_cons		2.1617	.0699859	30.89	0.000	2.024197	2.299204

Absorbed degrees of freedom:

Absorbed FE	Categories	- Redundant	= Num. Coefs
first_treat	4	0	4
year	5	1	4

Microeconometría II
Práctica 5
Variables instrumentales

1. Estimador de Wald

Suponga un modelo de regresión simple

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

donde x_i es potencialmente endógena. Además, suponga que el instrumento, z_i es una variable binaria. Muestre que el estimador IV en este caso es

$$\beta_1^{IV} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0}$$

donde \bar{y}_1, \bar{x}_1 (\bar{y}_0, \bar{x}_0) representan las medias cuando $z = 1$ ($z = 0$).

2. Estimador de Wald con datos simulados

En este ejercicio se propone extender la simulación del Problem Set 1 a un marco en el que la asignación del tratamiento y quienes resultan tratados no son iguales.

1. Inicialice una muestra con 100 observaciones. Genere resultados potenciales de no recibir el tratamiento como

$$Y_0 \sim \mathcal{N}(100, 30)$$

.

2. Genere ahora un efecto de tratamiento constante e igual a 20 para todos, es decir, $TE_i = 20$ para todo $i = 1, \dots, n$. Genere una variable aleatoria normal estándar. Genere una variable de tratamiento D_i igual a 1 para aquellas observaciones que poseen un valor positivo en la variable aleatoria normal.
3. Genere una variable aleatoria uniforme en el intervalo $[0, 1]$. Con ella, genere variables que indiquen el tipo de individuo. Utilice: *always taker* si la variable es menor a 0.25, *never taker* si la variable está entre 0.25 y 0.5, *defier* si la variable está entre 0.5 y 0.75 y *complier* si la variable es mayor a 0.75. Genere la variable de si los individuos toman el tratamiento o no dependiendo del grupo en el que están.
4. Genere la variable Y observada como $Y = DY_1 + (1 - D)Y_0$.
5. Estime el LATE y compare con el ATE.

3. Galiani & Schargrodsky (2010)

Lea el artículo “*Property rights for the poor: Effects of land titling*” de Galiani & Schargrodsky.

1. ¿Qué efectos intentan estimar en el paper?
2. ¿Cuál es la estrategia de identificación? ¿Por qué no funciona la diferencia de medias simple?
3. Replique los resultados del paper.

Trabajo Práctico N° 5: **Variables Instrumentales.**

Ejercicio 1.

Suponer un modelo de regresión simple:

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

donde x_i es, potencialmente, endógena. Además, suponer que el instrumento, z_i , es una variable binaria. Mostar que el estimador IV, en este caso, es:

$$\beta_1^{IV} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0},$$

donde \bar{y}_1 , \bar{x}_1 (\bar{y}_0 , \bar{x}_0) representan las medias cuando $z=1$ ($z=0$).

$$\beta^{IV} = (Z'X)^{-1} Z'y$$

$$\beta^{IV} = \left(\begin{bmatrix} 1_n \\ z \end{bmatrix} \begin{bmatrix} 1_n & x \end{bmatrix} \right)^{-1} \begin{bmatrix} 1_n \\ z \end{bmatrix} y$$

$$\beta^{IV} = \begin{bmatrix} 1_n' 1_n & 1_n' x \\ z' 1_n & z' x \end{bmatrix}^{-1} \begin{bmatrix} 1_n' y \\ z' y \end{bmatrix}$$

$$\beta^{IV} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n z_i & \sum_{i=1}^n z_i x_i \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n z_i y_i \end{bmatrix}$$

$$\beta^{IV} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ n_1 & \sum_{z_i=1} x_i \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{z_i=1} y_i \end{bmatrix}$$

$$\beta^{IV} = \frac{1}{n \sum_{z_i=1} x_i - n_1 \sum_{i=1}^n x_i} \begin{bmatrix} \sum_{z_i=1} x_i & -\sum_{i=1}^n x_i \\ -n_1 & n \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{z_i=1} y_i \end{bmatrix}$$

$$\beta^{IV} = \begin{bmatrix} \frac{\sum_{z_i=1} x_i}{n \sum_{z_i=1} x_i - n_1 \sum_{i=1}^n x_i} & \frac{-\sum_{i=1}^n x_i}{n \sum_{z_i=1} x_i - n_1 \sum_{i=1}^n x_i} \\ \frac{-n_1}{n \sum_{z_i=1} x_i - n_1 \sum_{i=1}^n x_i} & \frac{n}{n \sum_{z_i=1} x_i - n_1 \sum_{i=1}^n x_i} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{z_i=1} y_i \end{bmatrix}.$$

$$\beta_1^{IV} = \frac{-n_1}{n \sum_{z_i=1} x_i - n_1 \sum_{i=1}^n x_i} \sum_{i=1}^n y_i + \frac{n}{n \sum_{z_i=1} x_i - n_1 \sum_{i=1}^n x_i} \sum_{z_i=1} y_i$$

$$\beta_1^{IV} = \frac{n \sum_{z_i=1} y_i - n_1 \sum_{i=1}^n y_i}{n \sum_{z_i=1} x_i - n_1 \sum_{i=1}^n x_i}$$

$$\beta_1^{IV} = \frac{n \sum_{z_i=1} y_i - n_1 (\sum_{z_i=1} y_i + \sum_{z_i=0} y_i)}{n \sum_{z_i=1} x_i - n_1 (\sum_{z_i=1} x_i + \sum_{z_i=0} x_i)}$$

$$\beta_1^{IV} = \frac{n \sum_{z_i=1} y_i - n_1 \sum_{z_i=1} y_i - n_1 \sum_{z_i=0} y_i}{n \sum_{z_i=1} x_i - n_1 \sum_{z_i=1} x_i - n_1 \sum_{z_i=0} x_i}$$

$$\beta_1^{IV} = \frac{(n - n_1) \sum_{z_i=1} y_i - n_1 \sum_{z_i=0} y_i}{(n - n_1) \sum_{z_i=1} x_i - n_1 \sum_{z_i=0} x_i}$$

$$\beta_1^{IV} = \frac{n_0 \sum_{z_i=1} y_i - n_1 \sum_{z_i=0} y_i}{n_0 \sum_{z_i=1} x_i - n_1 \sum_{z_i=0} x_i}$$

$$\beta_1^{IV} = \frac{n_0 n_1 \bar{y}_1 - n_1 n_0 \bar{y}_0}{n_0 n_1 \bar{x}_1 - n_1 n_0 \bar{x}_0}$$

$$\beta_1^{IV} = \frac{n_0 n_1 (\bar{y}_1 - \bar{y}_0)}{n_0 n_1 (\bar{x}_1 - \bar{x}_0)}$$

$$\beta_1^{IV} = \frac{\bar{y}_1 - \bar{y}_0}{\bar{x}_1 - \bar{x}_0}.$$

Ejercicio 2.

En este ejercicio, se propone extender la simulación del Problem Set 1 a un marco en el que la asignación del tratamiento y quienes resultan tratados no son iguales.

(a) *Inicializar una muestra con 100 observaciones. Generar resultados potenciales de no recibir el tratamiento como: $Y_0 \sim \mathcal{N}(100, 30)$.*

Stata.

(b) *Generar, ahora, un efecto de tratamiento constante e igual a 20 para todos, es decir, $TE_i = 20$, para todo $i = 1, \dots, n$. Generar una variable aleatoria normal estándar. Generar una variable de tratamiento D_i igual a 1 para aquellas observaciones que poseen un valor positivo en la variable aleatoria normal.*

Stata.

(c) *Generar una aleatoria uniforme en el intervalo $[0, 1]$. Con ella, generar variables que indiquen el tipo de individuo. Utilizar: *always taker* si la variable es menor a 0,25, *never taker* si la variable está entre 0,25 y 0,5, *defier* si la variable está entre 0,5 y 0,75 y *complier* si la variable es mayor a 0,75. Generar la variable de si los individuos toman el tratamiento o no dependiendo del grupo en el que están.*

Stata.

(d) *Generar la variable Y observada como $Y = DY_1 + (1 - D)Y_0$.*

Stata.

(e) *Estimar el LATE y comparar con el ATE.*

Stata.

Ejercicio 3.

Leer el artículo “Property Rights for the Poor: Effects of Land Titling” de Galiani & Schargrotsky.

(a) ¿Qué efectos intentan estimar en el paper?

(b) ¿Cuál es la estrategia de identificación? ¿Por qué no funciona la diferencia de medias simple?

(c) Replicar los resultados del paper.

Stata.