Microeconometría II

Lecture 3

Microeconometría II

1. Estimador de diferencia en diferencias
   - Modelo Canónico
2. Extensiones al Modelo Canónico de DID
   - Cortes Transversales Repetidos
   - Refinamiento: Diferencia en Diferencia en Diferencias
3. Múltiples Períodos Temporales
   - Asignación del tratamiento en el mismo período y efecto tratamiento homogéneo en el tiempo
   - Asignación del tratamiento en el mismo período y efecto tratamiento heterogéneo en el tiempo
   - Múltiples Períodos Temporales y Asignación Escalonada
4. Referencias
5. Modelos de Datos de Panel
   - Modelos de Efectos Fijos y Aleatorios

Agenda

Estimador de diferencia en diferencias

- Existe otro estimador del impacto de una política/programa que se utiliza en la práctica cuando se tiene un diseño cuasi-experimental y la variable de resultado puede medirse en dos puntos del tiempo, antes de la implementación de la política y después de la implementación de la misma.
- Este estimador se conoce con el nombre de estimador de diferencia en diferencias (DID).
- El estimador de DID compara los cambios en la variable de interés en el grupo de tratamiento antes y después de la intervención con los cambios en la variable de interés en el grupo de control.
- El cambio en la variable de interés en el grupo de control es una estimación del contrafáctico verdadero. Esto es, **qué es lo que hubiera ocurrido en el grupo de tratamiento si no hubiera habido intervención.**

Estimador de diferencia en diferencias

- Comparando los cambios en ambos grupos se controla por el efecto de factores observables y no-observables constantes en el tiempo.
- Los cambios en la variable de interés en el grupo de tratamiento controlan por características fijas y los cambios en la variable de interés en el grupo de control controlan cambios en factores que varían en el tiempo y que son comunes a ambos grupos.
- Pensemos en un ejemplo.
- En 2002 Aguas Argentinas, en colaboración con el gobierno implementó un programa para expandir el servicio de agua potable a barrios carenciados denominado Modelo Participativo de Gestión (MPG).
  - El programa comenzaba con el pedido de agua potable por parte de los vecinos del barrio.
  - Aguas Argentinas evaluaba la factibilidad técnica de brindar el servicio.
  - La municipalidad donde el barrio estaba localizado debía aprobar la expansión del servicio.

Estimador de diferencia en diferencias

- El objetivo de la evaluación fue determinar si el programa ayudaba a mejorar la salud de los habitantes de los barrios que recibían el servicio.
- En particular se analizó si el servicio de agua potable disminuyó los episodios de diarrea en niños menores a seis años.
- Se construyeron dos grupos de barrios, el de tratamiento compuesto por seis barrios que recibieron el MPG y 3 barrios que no lo recibieron.
- Una posibilidad para analizar el impacto del MPG es comparar el porcentaje de episodios de diarrea en el grupo de tratamiento con ese porcentaje en el grupo de control después de implementado el programa:

$$y_i = \beta_0 + \beta_1 T_i + \epsilon_i$$

donde $y_i = 1$ si el niño i tuvo un episodio de diarrea y $T_i = 1$ si el niño i vive en un barrio del grupo de tratamiento.

Estimador de diferencia en diferencias

- Calculando las esperanzas matemáticas,

$$\begin{aligned} E(y_i|T_i = 0) &= \beta_0 \\ E(y_i|T_i = 1) &= \beta_0 + \beta_1 \Longrightarrow \\ E(y_i|T_i = 1) - E(y_i|T_i = 0) &= \beta_1 \end{aligned}$$

- La estimación del parámetro β_1 mide la diferencia entre la proporción de episodios de diarrea en el grupo de tratamiento y en el control.
- Qué podría amenazar la validez de nuestra estimación?

Estimador de diferencia en diferencias

- Que pasaría si los padres de los chicos que viven en los barrios del grupo de tratamiento tienen mayor educación que los padres que viven en los barrios del grupo de control?
- El problema es que al tener mayor educación pueden informar a los chicos acerca de las enfermedades transmisibles por el consumo de agua no potable. Es términos estadísticos es lo mismo que decir que en la ecuación anterior ϵ_i está relacionada con T_i.
- En otras palabras, aún si el programa MPG no se implementara, esperaríamos que el porcentaje de episodios de diarrea fuera menor en los barrios del grupo de tratamiento.
- Lo que amenaza la validez de nuestra estimación son las diferencias entre los dos grupos que no medimos.
- La diferencia entre los episodios de diarrea antes y después de la implementación del MPG en el grupo de tratamiento evitaría estas diferencias no medidas...

Estimador de diferencia en diferencias

- El impacto del programa se puede estimar entonces con una regresión en un panel de dos períodos ($t = 1, 2$):

$$y_{it} = \beta_0 + \beta_1 \, I(t = 2) + \epsilon_{it}$$

donde $t = 2$ indica el período temporal después de la implementación del programa e $I(t = 2) = 1$ cuando $t = 2$.

- Calculando las esperanzas matemáticas,

$$\begin{aligned} E(y_{it}|I(t=2)=0) &= \beta_0 \\ E(y_{it}|I(t=2)=1) &= \beta_0 + \beta_1 \Longrightarrow \\ E(y_{it}|I(t=2)=1) - E(y_{it}|I(t=2)=0) &= \beta_1 \end{aligned}$$

- La estimación del parámetro β_1 mide la diferencia entre la proporción de episodios de diarrea en el grupo de tratamiento después de implementada la política y antes de la implementación de la misma.

Estimador de diferencia en diferencias

- Qué podría amenazar la validez de nuestra estimación?
- La preocupación fundamental aquí son otros factores que afecten a los episodios de diarrea en el tiempo.
- Por ejemplo, una campaña del gobierno informando a la población sobre las enfermedades que se transmiten a través del consumo de agua no potable.
- Es decir, aún cuando no se implemente el programa, uno esperaría que la campaña provoque una disminución en los episodios de diarrea (en el tratamiento y en el control).
- El estimador DID combina los dos estimadores anteriores: Los cambios en la variable de interés en el grupo de tratamiento controlan por características fijas y los cambios en la variable de interés en el grupo de control controlan cambios en factores que varían en el tiempo y que son comunes a ambos grupos.

Estimador de diferencia en diferencias

- Estimamos la siguiente ecuación de datos de panel por POLS:

$$y_{it} = \beta_0 + \beta_1 \, T_{it} + \beta_2 \, I(t=2) + \beta_{DiD} \, T_{it} \times I(t=2) + \epsilon_{it}, \quad t = 1, 2. \tag{1}$$

  Donde $T_{it} = 1$ si $i \in T$ e $I(t=2) = 1$ si $t = 2$. El parámetro de interés es β_{DiD}.

- Calculando las esperanzas matemáticas,

$$\begin{aligned}
E(y_{it}|T_{it} = 1, I(t=2) = 1) &= \beta_0 + \beta_1 + \beta_2 + \beta_{DiD} \\
E(y_{it}|T_{it} = 1, I(t=2) = 0) &= \beta_0 + \beta_1 \\
E(y_{it}|T_{it} = 0, I(t=2) = 1) &= \beta_0 + \beta_2 \\
E(y_{it}|T_{it} = 0, I(t=2) = 0) &= \beta_0
\end{aligned}$$

- Tomando la diferencia de la esperanza matemática de la veriable de resultado después de la implementación del programa y antes en el grupo de tratamiento y restándole la misma diferencia en el grupo de control se obtiene β_{DiD}.

Estimador de diferencia en diferencias

- Esto es:

$$
\begin{aligned}
\{E(y_{it}|T_{it}=1, I(t=2)=1) - E(y_{it}|T_{it}=1, I(t=2)=0)\} &= \beta_2 + \beta_{DiD} \\
-\{E(y_{it}|T_{it}=0, I(t=2)=1) - E(y_{it}|T_{it}=0, I(t=2)=0)\} &= \beta_2 \\
&= \beta_{DiD}
\end{aligned}
$$

Esperanzas Matemáticas Condicionales

	Después	Antes	Diferencia
Tratamiento	$\beta_0 + \beta_1 + \beta_2 + \beta_{DiD}$	$\beta_0 + \beta_1$	$\beta_2 + \beta_{DiD}$
Control	$\beta_0 + \beta_2$	β_0	β_2
Diferencia	$\beta_1 + \beta_{DiD}$	β_1	$\boldsymbol{\beta_{DiD}}$

- Gráficamente:

Estimador de diferencia en diferencias

- Esto es:

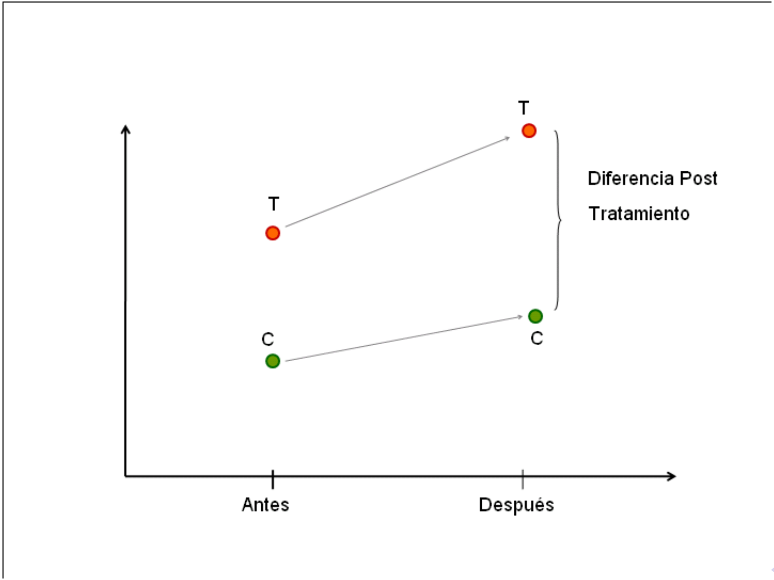
$$
\begin{aligned}
\{E(y_{it}|T_{it}=1, I(t=2)=1) - E(y_{it}|T_{it}=1, I(t=2)=0)\} &= \beta_2 + \beta_{DiD} \\
-\{E(y_{it}|T_{it}=0, I(t=2)=1) - E(y_{it}|T_{it}=0, I(t=2)=0)\} &= \beta_2 \\
&= \beta_{DiD}
\end{aligned}
$$

Esperanzas Matemáticas Condicionales

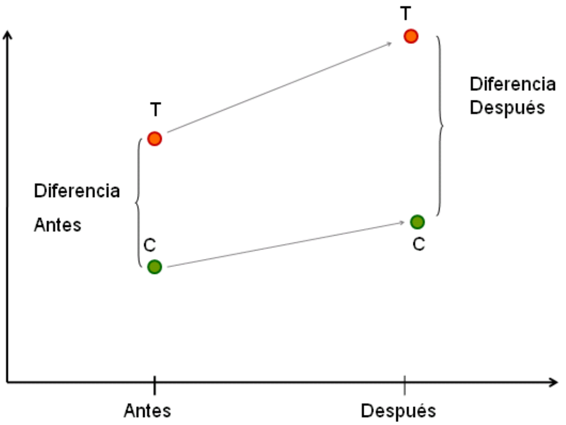
	Después	Antes	Diferencia
Tratamiento	$\beta_0+\beta_1+\beta_2+\beta_{DiD}$	$\beta_0+\beta_1$	$\beta_2+\beta_{DiD}$
Control	$\beta_0+\beta_2$	β_0	β_2
Diferencia	$\beta_1+\beta_{DiD}$	β_1	$\boldsymbol{\beta_{\mathbf{DiD}}}$

- Gráficamente:

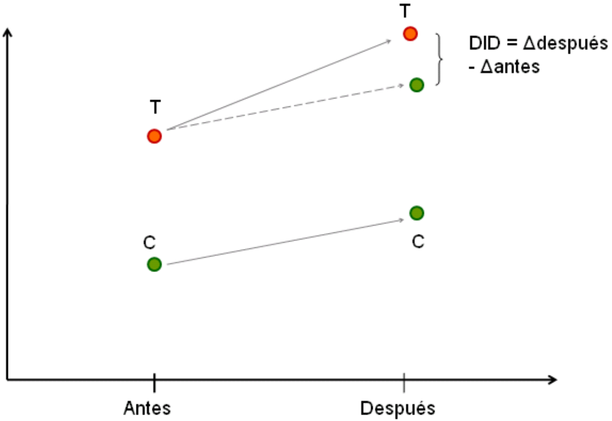
Estimador de diferencia en diferencias



Estimador de diferencia en diferencias



Estimador de diferencia en diferencias



Estimador de diferencia en diferencias

- Qué efecto identifica β_{DiD}?
- Recodemos que

$$\begin{aligned}
\beta_{DiD} &= \{E(y_{it}|T_{it}=1, I(t=2)=1) - \mathbf{E(y_{it}|T_{it}=1, I(t=2)=0)}\} \\
&- \{E(y_{it}|T_{it}=0, I(t=2)=1) - E(y_{it}|T_{it}=0, I(t=2)=0)\} \qquad (2) \\
&= \{E(Y^1_{it}|T_{it}=1, I(t=2)=1) - \mathbf{E(Y^0_{it}|T_{it}=1, I(t=2)=0)}\} \\
&- \{E(Y^0_{it}|T_{it}=0, I(t=2)=1) - E(Y^0_{it}|T_{it}=0, I(t=2)=0)\} \qquad (3)
\end{aligned}$$

- La ecuación anterior utiliza el supuesto de **no anticipación**. Este supuesto nos dice que *antes de la implementación del programa* los resultados potenciales son iguales.
- **Supuesto de NA:** $E(Y^0_{it}|T_{it}=1, I(t=2)=0) = E(Y^1_{it}|T_{it}=1, I(t=2)=0)$
- Este supuesto se conoce en la literatura como **supuesto de no anticipación condicional.**

Estimador de diferencia en diferencias

- sumando y restando $E(Y_{it}^0|T_{it}=1, I(t=2)=1)$ en la expresión anterior tenemos

$$\begin{aligned} \beta_{DiD} &= ATT + \left\{E(Y_{it}^0|T_{it}=1, I(t=2)=1) - E(Y_{it}^0|T_{it}=1, I(t=2)=0)\right\} \\ &- \left\{E(Y_{it}^0|T_{it}=0, I(t=2)=1) - E(Y_{it}^0|T_{it}=0, I(t=2)=0)\right\} \\ &= ATT + Sesgo \end{aligned} \tag{4}$$

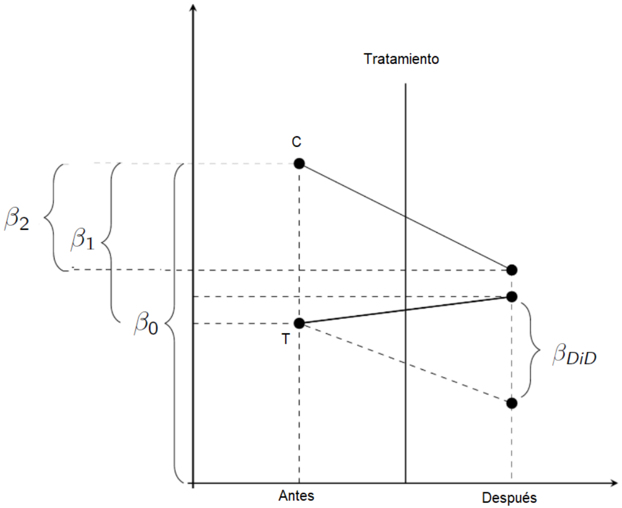
donde $ATT = E(Y_{it}^1|T_{it}=1, I(t=2)=1) - E(Y_{it}^0|T_{it}=1, I(t=2)=1) = E(Y_{it}^1 - Y_{it}^0|T_{it}=1, I(t=2)=1)$

Estimador de diferencia en diferencias

- Como se observa en la ecuación (4) β_{DiD} identifica el ATT siempre y cuando el término que llamamos *sesgo* sea cero.
- Este término se denomina en la literatura sesgo de tendencias no paralelas (non-parallel trends bias).
- Note que la primera esperanza en este sesgo no se observa en la práctica (es un contrafáctico) y por eso no se puede contrastar.
- Esto es lo que en la literatura se conoce como el supuesto de tendencias paralelas (parallel trends assumption), si las tendencias entre los grupos son paralelas el sesgo es cero y β_{DiD} identifica el ATT.

Estimador de diferencia en diferencias

- Usemos la ecuación (1) y veámoslo en una figura



Estimador de diferencia en diferencias

- **Remark 1:** El coeficiente β_{DiD} mide el ATT porque es la diferencia entre el valor de la variable de resultado observado después del tratamiento (círculo negro de arriba) y el valor del contrafáctico (el círculo negro de abajo en el extremo de la línea punteada).
- **Remark 2:** DID solo identificará este efecto siempre que la pendiente de la línea del control sea exactamente igual a la pendiente contrafáctica del tratamiento.
- **Remark 3:** Piense que DID siempre estima un coeficiente β_{DiD} usando como contrafáctico del tratamiento el cambio en el tiempo del control, independientemente del cambio en el contrafáctico del tratamiento (i.e. por ejemplo que tenga otra pendiente).
- **Remark 4:** Solo en el caso que se cumpla el supuesto de tendencias (pendientes) paralelas DID estimará correctamente el ATT.

Estimador de diferencia en diferencias

- Se puede recuperar el estimador de β_{DiD} estimando una ecuación de componentes no observados:
$$y_{it} = \beta_{DiD} D_{it} + \alpha_i + \lambda_t + u_{it}, \quad i = 1, 2, \ldots, N, \; t = 1, 2. \tag{5}$$
donde $D_{it} = T_{it} \times I(t = 2)$, α_i es la heterogeneidad de corte transversal y λ_t son efectos fijos temporales.
- El estimador de β_{DiD} en (5) se obtiene:
$$\min \sum_{i=1}^{N} \sum_{t=1}^{2} (y_{it} - \beta_{DiD} D_{it} - \alpha_i - \lambda_t)^2 \tag{6}$$
- El estimador de β_{DiD} que resulta de esta minimización se conoce en la literatura como Two-way fixed effects estimator (TWFE): $\hat{\beta}_{FE}$ y coincide con el estimador de POLS, $\hat{\beta}_{DiD}$, en la ecuación (1).

Equivalencia entre $\hat{\beta}_{DiD}$ y $\hat{\beta}_{FE}$

- Recuerde que para estimar (5) necesitamos transformar las variables restándoles su media temporal en cada corte transversal y su media de corte transversal para cada corte transversal y luego sumarle la media total de la variable.
- Esto es $\ddot{D}_{it} = D_{it} - \bar{D}_{i\cdot} - \bar{D}_{\cdot t} + \bar{D}$ donde:
  $\bar{D}_{i\cdot} = (1/2)\sum_{t=1}^{2} D_{it}$, $\bar{D}_{\cdot t} = (1/N)\sum_{i=1}^{N} D_{it}$ y la media total
  $\bar{D} = 1/(N \times 2)\sum_{i=1}^{N}\sum_{t=1}^{2} D_{it} = (1/N)\sum_{i=1}^{N} \bar{D}_{i\cdot} = (1/2)\sum_{t=1}^{2} \bar{D}_{\cdot t}$
- Estimando por MCC: y_{it} sobre $\ddot{D}_{it}$, $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$ se obtiene $\hat{\beta}_{FE}$:

$$\hat{\beta}_{FE} = \sum_{i=1}^{N}\sum_{t=1}^{2} y_{it}\left(\frac{\ddot{D}_{it}}{\sum_{i=1}^{N}\sum_{t=1}^{2}\ddot{D}_{it}^{2}}\right) \tag{7}$$

  es decir que el estimador $\hat{\beta}_{FE}$ es un promedio ponderado de los valores de la variable de resultado a lo largo de todas las observaciones.

Equivalencia entre $\hat{\beta}_{DiD}$ y $\hat{\beta}_{FE}$

- En términos generales para $N \times T$ observaciones tenemos

$$y = D\beta_{DiD} + (I_N \otimes i_T)\gamma + (i_N \otimes I_T)\delta + u = D\beta_{DiD} + Z_\gamma\gamma + Z_\delta\delta + u. \tag{8}$$

  donde y es un vector $NT \times 1$ de observaciones de la variable dependiente, D es un vector $NT \times 1$ de observaciones de la variable D_{it}, i_T e i_N son vectores de unos T dimensionales y N dimensionales, respectivamente. I_T e I_N son matrices identidades de orden T y N respectivamente.
- $Z'_\gamma Z_\gamma = I_N \otimes J_T$ con J_T una matriz de unos de orden T y $Z'_\delta Z_\delta = J_N \otimes I_T$ con J_N una matriz de unos de orden N.
- $P_\gamma = Z_\gamma(Z'_\gamma Z_\gamma)^{-1}Z'_\gamma = I_N \otimes \bar{J}_T$ es la matriz proyección, con $\bar{J}_T = J_T/T$ y $P_\delta = Z_\delta(Z'_\delta Z_\delta)^{-1}Z'_\delta = \bar{J}_N \otimes I_T$, con $\bar{J}_N = J_N/N$

Equivalencia entre $\hat{\beta}_{DiD}$ y $\hat{\beta}_{FE}$

- Defina la matriz

$$Q = (I_N - \bar{J}_N) \otimes (I_T - \bar{J}_T) = I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T \tag{9}$$

- Q es la matriz que elimina los efectos fijos de corte transversal y temporales. Un elemento típico de esta matriz en el contexto de nuestra ecuación es $\ddot{D}_{it} = D_{it} - \bar{D}_{i\cdot} - \bar{D}_{\cdot t} + \bar{D}$
- El estimador de efectos fijos de β_{DiD} en (5) es

$$\hat{\beta}_{FE} = (D'QD)^{-1}D'Qy \tag{10}$$

- Baltagi (2001) muestra la equivalencia entre este estimador y el que se obtiene de (8) usando D, Z_γ y Z_δ como variables explicativas y estimando por MCC.

Equivalencia entre $\hat{\beta}_{DiD}$ y $\hat{\beta}_{FE}$

- Cuando $T = 2$

$$\hat{\beta}_{FE} = \hat{\beta}_{DiD} = N_1^{-1} \sum_{i=1}^{N} s_i \times \Delta \bar{y}_i - N_0^{-1} \sum_{i=1}^{N} (1 - s_i) \times \Delta \bar{y}_i \qquad (11)$$

donde

$$\Delta \bar{y}_i \equiv \bar{y}_{i,post} - \bar{y}_{i,pre} \qquad (12)$$

Estimador de diferencia en diferencias

- En Stata:
  `xtset` *var. indicadora de panel* *var. indicadora de tiempo*
  `xtreg` *var. de resultado* D_{it} $I(t = 2)$, *fe*
- Datos de Panel

- En Excel: Ejemplo DiD.xlxs

Estimador de diferencia en diferencias con variables explicativas

- La ecuación (1) puede aumentarse para incluir variables explicativas observadas:

$$y_{it} = \beta_0 + \beta_1 T_{it} + \beta_2 I(t=2) + \beta_{DiD} T_{it} \times I(t=2) + X_{it}\gamma + \epsilon_{it}, \quad t = 1, 2. \qquad (13)$$

- Sin embargo, para que β_{DiD} en este modelo identifique el ATT necesitamos imponer más supuestos.
- Para ver esto, consideremos un ejemplo de (13) con una sola variable explicativa θx_{it}.
- Supongamos que se cumple el supuesto de tendencias paralelas de forma que en el caso canónico de DID sin variables explicativas el coeficiente β_{DiD} identificaría el efecto causal.
- Ahora

$$E(Y^1_{it} | T_{it} = 1, I(t=2) = 1, x_{it}) = \beta_0 + \beta_1 + \beta_2 + \beta_{DiD} + \theta x_{it}$$

Estimador de diferencia en diferencias con variables explicativas

- El supuesto de tendencias paralelas condicional implica que:

$$
\begin{aligned}
\left\{E(Y^0_{it}|T_{it}=1, I(t=2)=1, x_{it}) - E(Y^0_{it}|T_{it}=1, I(t=2)=0, x_{it})\right\} &= \\
\left\{E(Y^0_{it}|T_{it}=0, I(t=2)=1, x_{it}) - E(Y^0_{it}|T_{it}=0, I(t=2)=0, x_{it})\right\} &\Longrightarrow \\
E(Y^0_{it}|T_{it}=1, I(t=2)=1, x_{it}) = E(Y^0_{it}|T_{it}=1, I(t=2)=0, x_{it}) &+ \\
E(Y^0_{it}|T_{it}=0, I(t=2)=1, x_{it}) - E(Y^0_{it}|T_{it}=0, I(t=2)=0, x_{it}) & \qquad (14)
\end{aligned}
$$

- En el lado derecho de la ecuación (14) tenemos:

$$E(Y^0_{it}|T_{it}=1, I(t=2)=0, x_{it}) = \beta_0 + \beta_1 + \theta x_{it}$$

$$E(Y^0_{it}|T_{it}=0, I(t=2)=1, x_{it}) = \beta_0 + \beta_2 + \theta x_{it}$$

$$E(Y^0_{it}|T_{it}=0, I(t=2)=0, x_{it}) = \beta_0 + \theta x_{it}$$

- Entonces

$$
\begin{aligned}
E(Y^0_{it}|T_{it}=1, I(t=2)=1, x_{it}) &= \beta_0 + \beta_1 + \theta x_{it} + \beta_0 + \beta_2 + \theta x_{it} - \beta_0 - \theta x_{it} \\
&= \beta_0 + \beta_1 + \beta_2 + \theta x_{it} \qquad (15)
\end{aligned}
$$

Estimador de diferencia en diferencias con variables explicativas

- Ahora, en términos del modelo de resultados potenciales

$$E(Y_{it}^1 | T_{it} = 1, I(t = 2) = 1, x_{it}) = \beta_0 + \beta_1 + \beta_2 + \beta_{DiD} + \theta_2 x_{it}$$

$$E(Y_{it}^0 | T_{it} = 1, I(t = 2) = 1, x_{it}) = \beta_0 + \beta_1 + \beta_2 + \theta_1 x_{it}$$

- Por lo tanto

$$\begin{aligned} ATT &= E(Y_{it}^1 | T_{it} = 1, I(t = 2) = 1, x_{it}) - E(Y_{it}^0 | T_{it} = 1, I(t = 2) = 1, x_{it}) \\ &= \beta_{DiD} + (\theta_2 x_{it} - \theta_1 x_{it}) \end{aligned} \tag{16}$$

- Si $\theta_1 x_{it} \neq \theta_2 x_{it}$ se abre la posibilidad de que el estimador de datos de panel no identifique el verdadero ATT
- Sabemos que el estimador de datos de panel impone mecánicamente esta igualdad.

Estimador de diferencia en diferencias con variables explicativas

- Entonces para que el estimador de datos de panel con variables explicativas identifique el ATT correctamente necesitamos imponer el siguiente supuesto
- **Supuesto TWFE1: Efecto de tratamiento homogeneo en X**

$$E(Y_{it}^1 - Y_{it}^0 | T_{it} = 1, I(t = 2) = 1, x_{it}) = E(Y_{it}^1 - Y_{it}^0 | T_{it} = 1, I(t = 2) = 1)$$

- Esto es porque como mostramos arriba necesitamos que los términos θx_{it} se cancelen de forma tal que β_{DiD} identifique el ATT.

Estimador de diferencia en diferencias con variables explicativas

- En realidad el estimador de datos de panel también impone restricciones sobre las tendencias en los dos grupos.
- Tomemos esperanzas condicionales en (13)

$$E(Y_{i1}^{1}|T_{it}=1, I(t=2)=1, x_{it}) = \beta_0+\beta_1+\beta_2+\beta_{DiD}+\theta x_{i1}^{(T)}$$

$$E(Y_{i0}^{0}|T_{it}=1, I(t=2)=0, x_{it}) = \beta_0+\beta_1+\theta x_{i0}^{(T)}$$

$$E(Y_{i1}^{0}|T_{it}=0, I(t=2)=1, x_{it}) = \beta_0+\beta_2+\theta x_{i1}^{(C)}$$

$$E(Y_{it}^{0}|T_{it}=0, I(t=2)=0, x_{it}) = \beta_0+\theta x_{i0}^{(C)}$$

- El estimador de DID es:

$$\begin{aligned} DID &= [(\beta_0+\beta_1+\beta_2+\beta_{DiD}+\theta x_{i1}^{(T)})-(\beta_0+\beta_1+\theta x_{i0}^{(T)})]-[(\beta_0+\beta_2+\theta x_{i1}^{(C)})-(\beta_0+\theta x_{i0}^{(C)})] \\ &= \beta_{DiD}+(\theta x_{i1}^{(T)}-\theta x_{i0}^{(T)})-(\theta x_{i1}^{(C)}-\theta x_{i0}^{(C)}) \end{aligned}$$

Estimador de diferencia en diferencias con variables explicativas

- La última linea de la ecuación requiere que la tendencia en las X en el tratamiento sea igual a la del control para que el estimador del modelo de datos de panel identifique correctamente el ATT.
- **Supuesto TWFE2:** **Tendencias iguales en las X en los dos grupos**. Para $D = 1, 0$

$$E(Y_{it}^{1} - Y_{it}^{0} | D = d, x_{it}) = E(Y_{it}^{1} - Y_{it}^{0} | D = d)$$

- Las Xs son variables que varian en el tiempo y no se cancelan dentro de cada grupo.
- Sin estos supuestos el estimador de datos de panel no identifica correctamente el ATT.

Extensiones al modelo canónico de DID

- Cortes transversales repetidos
- Refinamiento: DDD
- Múltiples períodos temporales
  - Asignación del tratamiento en el mismo período y efecto tratamiento homogéneo en el tiempo.
  - Asignación del tratamiento en el mismo período y efecto tratamiento heterogéneo en el tiempo.
  - Asignación escalonada del tratamiento y efecto tratamiento homogéneo.
  - Asignación escalonada del tratamiento y efecto tratamiento heterogéneo.

Agenda

Estimador DID en Cortes Transversales Repetidos

- Una buena característica del procedimiento de DID es que no necesariamente requiere observar a los mismos individuos en el grupo de tratamiento y de control en cada período.
- Es posible realizar la estimación con cortes transversales repetidos de diferentes individuos.
- Cortes transversales repetidos son muestras de corte transversal que corresponden a diferentes períodos:
  - Por ejemplo, restaurantes de más de 100 metros cuadrados de la CABA y la PBA antes y después de la implementación de la ley de ambientes libres de humo en CABA.
- La idea es que, si los individuos (unidades) son aleatoriamente elegidos de la misma población, los seleccionados en el primer corte transversal pueden funcionar como sustitutos de aquellos seleccionados en el segundo corte transversal (algunos de los cuales recibirán el tratamiento).

Estimador DID en Cortes Transversales Repetidos

- El estimador de DID sigue siendo la diferencia entre las esperanzas matemáticas condicionales antes y después en los dos grupos.
- En términos de un modelo de regresión ahora estimamos:

$$y_{it} = \beta_0 + \beta_1 \, T_i + \beta_2 \, I(t=2) + \beta_{DiD} \, T_i \times I(t=2) + \epsilon_{it}, \quad t = 1, 2. \tag{17}$$

  donde $T_i = 1$ si $i \in T$ e $I(t=2) = 1$ si $t = 2$. El parámetro de interés es β_{DiD}.
- El estimador de DID puede extenderse para incluir variables explicativas adicionales.

Agenda

Refinamiento: diferencia en diferencia en diferencias (DDD)

- El estimador de DID puede extenderse también para considerar un refinamiento de la definición de grupo de tratamiento y de contro, con el obejtivo de analizar en mayor profundidad la aplicación de determinada política.
- Recuerde el ejemplo de la Ley de Ambientes Libres de Humo de CABA. Esta ley convirtió a los locales gastronómicos de más de 100 m^2 en locales exclusivos para no fumadores
- Supongamos que la variable de resultado es la facturación de los restaurantes.
- Una posibilidad es usar a los restaurantes de CABA antes y después de la aplicación de la ley y considerar a los de más de 100 m^2 como grupo de tratamiento.
- El problema potencial es que puede haber factores no relacionados con la política que afecten la facturación de los restaurantes de más de 100 m^2 en relación a los de menos de 100 m^2.

Refinamiento: diferencia en diferencia en diferencias (DDD)

- Otro análisis de DID podría ser utilizar a los restaurantes de más de 100 metros cuadrados de la PBA como unidades del grupo de control.
- El problema potencial que surge es que cambios en la facturación de los restaurantes pueden ser sistemáticamente diferentes entre CABA y PBA debido a, por ejemplo, diferencias en ingresos más que a la aplicación de la política.
- Un análisis más robusto que los dos anteriores se puede obtener comparando el estimador de DID para CABA (donde se implementó la política) con el mismo estimador pero en la PBA.
- Entonces, una versión extendida de la ecuación (1) es

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 T_i + \beta_2 R_i + \beta_3 T_i \times R_i + \delta_0 I(t=2) + \delta_1 T_i \times I(t=2) \\ &+ \delta_2 R_i \times I(t=2) + \delta_3 T_i \times R_i \times I(t=2) + \epsilon_{it}, \quad t = 1, 2. \end{aligned} \tag{18}$$

donde $R_i = 1$ si el restaurante tiene más de 100 m^2, $T_i = 1$ si el restaurante i está en CABA (o donde se implemente la política).

Refinamiento: diferencia en diferencia en diferencias (DDD)

- El parámetro de interés es δ_3.
- Note que

$$
\begin{aligned}
\delta_3 &= \{[E(y|T=1,R=1,I(t_2=1)=1)-E(y|T=1,R=1,I(t_2=1)=0)] \\
&- [E(y|T=1,R=0,I(t_2=1)=1)-E(y|T=1,R=0,I(t_2=1)=0)]\} \\
&- \{[E(y|T=0,R=1,I(t_2=1)=1)-E(y|T=0,R=1,I(t_2=1)=0)] \\
&- [E(y|T=0,R=0,I(t_2=1)=1)-E(y|T=0,R=0,I(t_2=1)=0)]\}
\end{aligned}
$$

- MCC aplicado en (18) nos dará un estimador consistente, $\hat{\delta}_3$, de δ_3.
- $\hat{\delta}_3$ recibe el nombre de **estimador de diferencia en diferencia en diferencias (DDD)**.

Refinamiento: diferencia en diferencia en diferencias (DDD)

- El primer término en $\{\cdot\}$ es el estimador DID tomando como grupo de control a los restaurantes de menos de 100 m^2 y los perídos temporales antes y después de la implementación de la política.
- Para evitar que este estimador capture tendencias diferentes entre la facturación de restaurantes de más y menos de 100 m^2 el estimador de DDD le resta la misma estimación pero para el grupo de control (PBA) (segundo término en $\{\cdot\}$).
- El estimador DDD puede extenderse fácilmente para incluir variables explicativas adicionales.
- La inferencia en el modelo es estándar y se puede hacer robusta a la presencia de heterocedasticidad con los estadísticos usuales.

Múltiples Períodos Temporales

- La extensión a múltiples períodos temporales se divide en dos grupos:
  - Las unidades que reciben el tratamiento lo reciben en un mismo período temporal y hay múltiples períodos post-tratamiento
  - Las unidades que reciben el tratamiento lo reciben en forma escalonada en el tiempo y además hay múltiples períodos post-tratamiento
- A su vez, el efecto de la política puede ser constante o variable en el tiempo.

Agenda

Asignación única y Tratamiento Homogéneo

- Como ejemplo, suponga un programa que se implementa en $t = 0$ y tiene 4 períodos post-tratamiento $(t = 1, 2, 3, 4)$.
- s_i es una variable binaria que adopta el valor unitario si la unidad i eventualmente es beneficiaria del programa.
- Denotemos por $d_t = 1(t = 1, 2, 3, 4)$ a una variable indicadora de cada período post-tratamiento y con $p_t = d_1 + d_2 + d_3 + d_4$.
- Con estas definiciones la generalización de la ecuación de componentes no observados es

$$y_{it} = \alpha_i + \delta_t + \beta_3 (s_i \times p_t) + \epsilon_{it} \tag{19}$$

  Donde α_i son efectos fijos de corte transversal y δ_t son efectos fijos temporales.
- El coeficiente β_3 mide el impacto del programa.

Múltiples Períodos Temporales y Tratamiento Homogéneo

- Calculando las esperanzas matemáticas condicionales tenemos,

$$\begin{aligned}
E\left[y_{it} \mid p_t = 0, s_i = 0\right] &= \alpha_i + \delta_0 && (20)\\
E\left[y_{it} \mid p_t = 1, s_i = 0\right] &= \alpha_i + \delta_1 + \delta_2 + \delta_3 + \delta_4 && (21)\\
E\left[y_{it} \mid p_t = 0, s_i = 1\right] &= \alpha_i + \delta_0 && (22)\\
E\left[y_{it} \mid p_t = 1, s_i = 1\right] &= \alpha_i + \delta_1 + \delta_2 + \delta_3 + \delta_4 + \beta_3 && (23)
\end{aligned}$$

- Entonces,

$$\begin{aligned}
\tau^{DID} &= \left\{E\left[y_{it} \mid p_t = 1, s_i = 1\right] - E\left[y_{it} \mid p_t = 0, s_i = 1\right]\right\} && (24)\\
&- \left\{E\left[y_{it} \mid p_t = 1, s_i = 0\right] - E\left[y_{it} \mid p_t = 0, s_i = 0\right]\right\} && (25)\\
&= \left\{\left[\alpha_i + \delta_1 + \delta_2 + \delta_3 + \delta_4 + \beta_3\right] - \left[\alpha_i + \delta_0\right]\right\} && (26)\\
&- \left\{\left[\alpha_i + \delta_1 + \delta_2 + \delta_3 + \delta_4\right] - \left[\alpha_i + \delta_0\right]\right\} && (27)\\
&= \beta_3 && (28)
\end{aligned}$$

Múltiples Períodos Temporales y Tratamiento Homogéneo

- En términos generales, suponga T períodos temporales: $t = 1, 2, \ldots, T$.
- Defina el indicador del tratamiento como: $D_{it} = s_i \times p_t$, donde $s_i = 1$ si la unidad i eventualmente recibe el tratamiento y p_t es una variable indicadora de los períodos post-tratamiento (i.e. $p_t = 0$ para $t = 1, 2, \ldots q - 1$ y $p_t = 1$ para $t = q, q + 1, \ldots T$).
- Defina variables indicadoras de los períodos temporales.
- Para el período t: $\{\lambda s_t, \;\; t = 2, \ldots T\}$ con $\lambda s_t = 1$ si $s = t$ y $\lambda s_t = 0$ si $s \neq t$.
- Note que se $p_t = \lambda q_t + \lambda q + 1_t + \cdots + \lambda T_t$.
- Entonces, la ecuación (19) queda:

$$y_{it} = \alpha_i + \delta_t + \beta_3 D_{it} + \epsilon_{it} \tag{29}$$

Múltiples períodos temporales y asignación común

- Caso 1: Asignación del tratamiento en un mismo período y efecto tratamiento homogéneo en el tiempo
- En los períodos $q-1$ y anteriores ninguna unidad recibe la política o programa. En $t=q$ algunas unidades reciben el tratamiento y lo mantienen de ese período en adelante (el tratamiento es un estado absorvente).
- Para estimar el impacto de la política o programa se estima

$$y_{it} = \beta D_{it} + \alpha_i + g_t + u_{it}, \quad i = 1, 2, \ldots, N, \; t = 1, 2, \ldots, T. \tag{30}$$

- $\hat{\beta}_{FE}$ coincide con el estimador de diferencia en diferencias $\hat{\beta}_{DiD}$ de la regresión de y_{it} sobre una constante, D_{it}, s_i, p_t, $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$:

$$y_{it} = \alpha_0 + \beta_{DiD} D_{it} + \alpha_1 s_i + \alpha_2 p_t + v_{it}, \quad i = 1, 2, \ldots, N, \; t = 1, 2, \ldots, T. \tag{31}$$

donde:

$$\begin{aligned} E(y_{it}|s_i = 1, p_t = 1) &= \alpha_0 + \beta_{DiD} + \alpha_1 + \alpha_2; \; E(y_{it}|s_i = 1, p_t = 0) = \alpha_0 + \alpha_1 \\ E(y_{it}|s_i = 0, p_t = 1) &= \alpha_0 + \alpha_2; \; E(y_{it}|s_i = 0, p_t = 0) = \alpha_0 \end{aligned}$$

Múltiples períodos temporales y asignación común

- Entonces

$$\begin{aligned}\beta_{DiD} &= \{E(y_{it}|s_i = 1, p_t = 1) - E(y_{it}|s_i = 1, p_t = 0)\} \\ &- \{E(y_{it}|s_i = 0, p_t = 1) - E(y_{it}|s_i = 0, p_t = 0)\}\end{aligned}$$

- La estimación se hace reemplazando las esperanzas matemáticas por sus análogos muestrales.

$$\hat{\beta}_{DiD} = N_1^{-1}\sum_{i=1}^{N} s_i \times \Delta\bar{y}_i - N_0^{-1}\sum_{i=1}^{N}(1 - s_i) \times \Delta\bar{y}_i$$

donde

$$\Delta\bar{y}_i \equiv (T - q + 1)^{-1}\sum_{t=q}^{T} y_{it} - (q-1)^{-1}\sum_{t=1}^{q-1} y_{it} = \bar{y}_{i,post} - \bar{y}_{i,pre} \tag{32}$$

Múltiples períodos temporales y asignación común

- El estimador de $\hat{\beta}_{DiD}$ en (31) no se altera si en lugar de p_t se utilizan variables binarias temporales λq_t, $\lambda q + 1_t$, ..., λT_t.
- El estimador tampoco se altera si introducimos variables observables adicionales constantes en el tiempo $\mathbf{x}_i$, e incluso si las interactuamos con el indicador del tratamiento s_i.
- En términos matemáticos estimamos
  y_{it} sobre una constante, D_{it}, s_i, λq_t, $\lambda q + 1_t$, ..., λT_t, $\mathbf{x}_i$, $s_i \times \mathbf{x}_i$ $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$:

Múltiples períodos temporales y asignación común

- La estimación del impacto del programa tampoco cambia si interactuamos $\mathbf{x}_i$ con p_t o con D_{it}:

$$\begin{aligned} y_{it} &= \beta_{DiD} D_{it} + [D_{it} \times (\mathbf{x}_i - \bar{x}_1)]\gamma + \mathbf{x}_i\xi + \zeta s_i + (s_i \times \mathbf{x}_i)\rho + \theta p_t + (p_t \times \mathbf{x}_i)\pi \\ &+ \alpha_i + g_t + u_{it}, \;\; i = 1, \ldots, N; t = 1, \ldots, T. \end{aligned} \tag{33}$$

donde $\bar{x}_1$ es el promedio de las variables observables en el grupo de tratamiento.

Múltiples períodos temporales y asignación común

- La estimación de β_{DiD}, $\hat{\beta}_{FE}$, en (33) mide el impacto promedio de la política sobre los tratados (ATT) (Wooldridge, 2021):

$$\hat{\beta}_{FE} = \hat{\beta}_{DiD} = N_1^{-1} \sum_{i=1}^{N} s_i \times \Delta\bar{y}_i - N_0^{-1} \sum_{i=1}^{N} (1 - s_i) \times \Delta\bar{y}_i \tag{34}$$

donde

$$\Delta\bar{y}_i \equiv \bar{y}_{i,post} - \bar{y}_{i,pre} \tag{35}$$

Agenda

Múltiples períodos temporales y asignación común

- Caso 2: Asignación del tratamiento en un mismo período y efecto tratamiento heterogéneo en el tiempo
- Para estimar el impacto de la política o programa se estima

$$y_{it} = \sum_{j=q}^{T} \beta_j (D_{it} \times \lambda j_t) + \alpha_i + g_t + u_{it}, \quad i = 1, 2, \ldots, N, \; t = 1, 2, \ldots, T. \tag{36}$$

- Esta especificación permite que el impacto del tratamiento sea diferente en cada período tratado.
- Note que una especificación equivalente es:

$$y_{it} = \sum_{j=q}^{T} \beta_j (s_i \times \lambda j_t) + \alpha_i + g_t + u_{it}, \quad i = 1, 2, \ldots, N, \; t = 1, 2, \ldots, T. \tag{37}$$

Múltiples períodos temporales y asignación común

- Los estimadores de TWFE de estas ecuaciones son equivalentes a estimar las siguientes ecuaciones por POLS

$$
\begin{aligned}
y_{it} &= \alpha + \beta_q \left(s_i \cdot \lambda q_t\right) + \cdots + \beta_T \left(s_i \cdot \lambda T_t\right) + \zeta s_i + \theta_q \lambda q_t + \cdots + \theta_T \lambda T_t + e_{it} \\
&= \alpha + \beta_q \left(D_{it} \cdot \lambda q_t\right) + \cdots + \beta_T \left(D_{it} \cdot \lambda T_t\right) + \zeta s_i + \theta_q \lambda q_t + \cdots + \theta_T \lambda T_t + e_{it}
\end{aligned} \tag{38}
$$

- Estimar la ecuación (38) permite contrastar si los β_r son constantes en el tiempo.
- La ecuación permite incluir controles que no varíen en el tiempo, $\mathbf{x}_i$, sin que afecten el estimador de impacto.
- También se puede flexibilizar la especificación para considerar que los efectos de la política cambien con $\mathbf{x}_i$

Múltiples períodos temporales y asignación común

- Para esto se puede estimar:

$$\begin{aligned} y_{it} &= \beta_q \left(D_{it} \cdot \lambda q_t\right) + \cdots + \beta_T \left(D_{it} \cdot \lambda T_t\right) + \left[D_{it} \cdot \lambda q_t \cdot \left(\mathbf{x}_i - \bar{x}_1\right)\right] \gamma_q + \cdots \\ &+ \left[D_{it} \cdot \lambda T_t \cdot \left(\mathbf{x}_i - \bar{x}_1\right)\right] \gamma_T + \left(\lambda q_t \cdot \mathbf{x}_i\right) \boldsymbol{\delta}_q + \cdots + \left(\lambda T_t \cdot \mathbf{x}_i\right) \boldsymbol{\delta}_T \\ &+ c_i + g_t + u_{it} \end{aligned} \tag{39}$$

- Qué parámetros estiman consistentemente estos estimadores?
- Considere la ecuación (37) sin variables observables pero permitiendo efectos tratamiento que varíen en el tiempo:

$$E\left(y_t \mid s\right) = \alpha + \beta_q \left(s \cdot \lambda q_t\right) + \cdots + \beta_T \left(s \cdot \lambda T_t\right) + \zeta s + \theta_q \lambda q_t + \cdots + \theta_T \lambda T_t . \tag{40}$$

Múltiples períodos temporales y asignación común

- Haciendo $s = 0$ tenemos:

$$\begin{aligned} E(y_t \mid s=0) &= \alpha, t = 1, \ldots, q-1 \\ &= \alpha + \theta_t, t = q, \ldots, T \end{aligned}$$

- y haciendo $s = 1$

$$\begin{aligned} E(y_t \mid s=1) &= \alpha + \zeta, t = 1, \ldots, q-1 \\ &= \alpha + \zeta + \theta_t + \beta_t, t = q, \ldots, T \end{aligned}$$

- Por lo tanto para cada $r \geq q$ y cualquier $t < q$:

$$\begin{aligned} \beta_r &= [E(y_r \mid s=1) - E(y_r \mid s=0)] - [E(y_t \mid s=1) - E(y_t \mid s=0)] \\ &= [E(y_r \mid s=1) - E(y_t \mid s=1)] - [E(y_r \mid s=0) - E(y_t \mid s=0)] \qquad (41) \end{aligned}$$

que es la expresión poblacional de diferencia en diferencias.

Múltiples períodos temporales y asignación común

- En términos del modelo de resultados potenciales considere una muestra aleatoria en el corte transversal i y tomemos los T períodos temporales $t = 1, 2, \ldots, T$ como dados.
- En el primer período, como mínimo, nadie recibe el tratamiento.
- La intervención ocurre en el período q para todas las unidades tratadas y estas unidades mantienen el tratamiento hasta el período T.
- Los resultados potenciales son $\{[y_t(0), y_t(1)] : t = 1, \ldots, T\}$, donde $y_t(0)$ es el resultado potencial en el estado del control y $y_t(1)$ es el resultado potencial en el tratamiento.
- $s = 0$ indica que la unidad es asignada al control y $s = 1$ al tratamiento.
- Observamos $y_t(0)$ ó $y_t(1)$ pero no ambos.

Múltiples períodos temporales y asignación común

- Para cada t el efecto del tratamiento para una intervención ocurrida en $t = q$ es:

$$te_t = y_t(1) - y_t(0) \tag{42}$$

- Estamos interesados en el efecto promedio del tratamiento sobre los tratados:

$$\tau_t \equiv E\left[y_t(1) - y_t(0) \mid s = 1\right], t = q, q+1, \ldots, T \tag{43}$$

- **Supuesto: No anticipación.** Los resultados potenciales antes de $t = q$ están definidos en términos de una política que se implementa más adelante, esto significa que las unidades no cambian su comportamiento en anticipación del tratamiento de maneras que afectarían el resultado,

$$y_t(1) = y_t(0), \quad t < q \tag{44}$$

Múltiples períodos temporales y asignación común

- Una versión más débil de este supuesto es
- **Supuesto NA: No anticipación.**

$$E\left[y_t(1) - y_t(0) \mid s = 1\right] = 0, \quad \forall t < q \tag{45}$$

- **Supuesto TP: Tendencias paralelas o comúnes.**

$$E\left[y_t(0) - y_1(0) \mid s\right] = E\left[y_t(0) - y_1(0)\right] \equiv \theta_t, \quad t = 2, \ldots, T. \tag{46}$$

La primera igualdad es la que define el supuesto y dice que la tendencia promedio en el estado de control, en cada período relativo al período inicial, no depende de cual es la asignación del tratamiento. Una vez que asumimos la tendencia común la definimos como θ_t.

- Escribamos el resultado observado como,

$$y_t = y_t(0) + s \times \left[y_t(1) - y_t(0)\right] = y_t(0) + s \times te_t \tag{47}$$

Múltiples períodos temporales y asignación común

- Tomando esperanzas matemáticas condicionales en s

$$\begin{aligned} E\left(y_t \mid s\right) &= E\left[y_t(0) \mid s\right] + s \times E\left(te_t \mid s\right) \\ &= E\left[y_t(0) \mid s\right] + s \times \tau_t \end{aligned} \tag{48}$$

porque
$s \times E\left(te_t \mid s\right) = s \times \left[(1-s) \times E\left(te_t \mid s=0\right) + s \times E\left(te_t \mid s=1\right)\right] = s \times E\left(te_t \mid s=1\right).$

- $y_t(0) = y_t(0) + y_1(0) - y_1(0) = y_1(0) + \left[y_t(0) - y_1(0)\right] = y_1(0) + g_t(0)$
- Imponiendo el **Supuesto TP:** $E\left[g_t(0) \mid s\right] = E\left[g_t(0)\right] \equiv \theta_t,\ \ t = 2, \ldots, T$
- Como s es una variable binaria siempre podemos escribir: $E\left[y_1(0) \mid s\right] = \rho + \xi s$
- Combinando estos resultados junto con $\theta_1 = 0$ tenemos:

$$E\left(y_t \mid s\right) = \rho + \xi s + \theta_t + s \times \tau_t,\ \ t = 1, \ldots, T \tag{49}$$

Múltiples períodos temporales y asignación común

- Bajo el **Supuesto NA** $\tau_t = 0 \; \forall t < q$, por lo tanto

$$\begin{aligned} E\left(y_t \mid s\right) &= \rho + \xi s + \theta_t, \;\; t < q && (50) \\ &= \rho + \xi s + \theta_t + s \times \tau_t, \;\; t = q, \ldots, T && (51) \end{aligned}$$

  y estas ecuaciones identifican τ_t.

- Podemos escribir la ecuación a estimar en función de las variables binarias temporales como:

$$E\left(y_t \mid s\right) = \rho + \xi s + \theta_2 \lambda 2_t + \cdots + \theta_T \lambda T_t + \tau_q \left(s \times \lambda q_t\right) + \cdots + \tau_T \left(s \times \lambda T_t\right), \;\; t = 1, \ldots, T \quad (52)$$

- Escribiendo la ecuación para una observación genérica de corte transversal i tenemos

$$E\left(y_{it} \mid s_i\right) = \rho + \xi s_i + \theta_2 \lambda 2_t + \cdots + \theta_T \lambda T_t + \tau_q \left(D_{it} \times \lambda q_t\right) + \cdots + \tau_T \left(D_{it} \times \lambda T_t\right), \;\; \forall i, t \quad (53)$$

  donde $D_{it} = s_i \times p_t$.

Múltiples períodos temporales y asignación común

- Una estimación consitente de τ_r viene dada por POLS en

$$y_{it} = \rho + \zeta s_i + \theta_q \lambda q_t + \cdots + \theta_T \lambda T_t + \tau_q \left(D_{it} \times \lambda q_t\right) + \cdots + \tau_T \left(D_{it} \times \lambda T_t\right) + e_{it} \tag{54}$$

Note la equivalencia con la ecuación (38).

- Si queremos agregar como controles variables explicativas que no varíen en el tiempo, $\mathbf{x}_i$, tenemos que modificar los supuestos NA y TP.
- **Supuesto NAC: no anticipación condicional.** Para el indicador de tratamiento s y las variables $\mathbf{x}$,

$$E\left[y_t(1) - y_t(0) \mid s = 1, \mathbf{x}\right] = 0, \; \forall t < q \tag{55}$$

- **Supuesto TPC: tendencias paralelas condicional.** Para el indicador de tratamiento s y las variables $\mathbf{x}$,

$$E\left[y_t(0) - y_1(0) \mid s, \mathbf{x}\right] = E\left[y_t(0) - y_1(0) \mid \mathbf{x}\right], \; t = 2, \ldots, T \tag{56}$$

Múltiples períodos temporales y asignación común

- Tomemos la ecuación (48) pero con esperanzas condicionales en $(s, \mathbf{x})$

$$E\left(y_t \mid s, \mathbf{x}\right) = E\left[y_t(0) \mid s, \mathbf{x}\right] + s \times E\left(te_t \mid s, \mathbf{x}\right) \tag{57}$$

- Definamos el ATT condicional en $\mathbf{x}$ como

$$\tau_t(\mathbf{x}) \equiv E\left(te_t \mid s = 1, \mathbf{x}\right) \tag{58}$$

  tal que

$$E\left(y_t \mid s, \mathbf{x}\right) = E\left[y_t(0) \mid s, \mathbf{x}\right] + s \times \tau_t(\mathbf{x}) \tag{59}$$

- Usando la ley de espectativas iteradas, $\tau_t = E\left[\tau_t(\mathbf{x}) \mid s = 1\right]$

Múltiples períodos temporales y asignación común

- Nuevamente, $y_t(0) = y_t(0) + y_1(0) - y_1(0) = y_1(0) + [y_t(0) - y_1(0)] = y_1(0) + g_t(0)$ y

$$E\left[y_t(0) \mid s, \mathbf{x}\right] = E\left[y_1(0) \mid s, \mathbf{x}\right] + E\left[g_t(0) \mid s, \mathbf{x}\right], \;\; t = 2, \ldots, T \tag{60}$$

- Por el Supuesto TPC $E\left[g_t(0) \mid s, \mathbf{x}\right] = E\left[g_t(0) \mid \mathbf{x}\right]$ y la ecuación queda

$$E\left[y_t(0) \mid s, \mathbf{x}\right] = E\left[y_1(0) \mid s, \mathbf{x}\right] + E\left[g_t(0) \mid \mathbf{x}\right], \;\; t = 2, \ldots, T \tag{61}$$

- Por lo tanto,

$$E\left(y_t \mid s, \mathbf{x}\right) = E\left[y_1(0) \mid s, \mathbf{x}\right] + E\left[g_t(0) \mid \mathbf{x}\right] + s \times \tau_t(\mathbf{x}) \tag{62}$$

Múltiples períodos temporales y asignación común

- Asumiendo linealidad en las esperanzas matemáticas tenemos

$$
\begin{aligned}
E\left[y_1(0) \mid s, \mathbf{x}\right] &= \eta + \lambda s + \dot{\mathbf{x}}\boldsymbol{\kappa} + (s \times \dot{\mathbf{x}})\zeta \\
E\left[g_t(0) \mid \mathbf{x}\right] &= \theta_t + \dot{\mathbf{x}}\pi_t, \;\; t = 2, \ldots, T \\
\tau_t(\mathbf{x}) &= \tau_t + \dot{\mathbf{x}}\rho_t, \;\; t = q, \ldots, T \\
\text{con: } \dot{\mathbf{x}} &\equiv \mathbf{x} - E(\mathbf{x} \mid s = 1) \equiv \mathbf{x} - \mu_1.
\end{aligned}
$$

- Note que centrando $\mathbf{x}$ alrededor de $E(\mathbf{x} \mid s = 1)$ forzamos a que la ordenada al origen de la ecuación de $\tau_t(\mathbf{x})$ sea el parámetro que nos interesa estimar τ_t.
- Reemplazando estas esperanzas en (62) tenemos

$$
E\left(y_t \mid s, \mathbf{x}\right) = \eta + \lambda s + \dot{\mathbf{x}}\kappa + (s \times \dot{\mathbf{x}})\zeta + \theta_t + \dot{\mathbf{x}}\pi_t + \tau_t s + (s \times \dot{\mathbf{x}})\boldsymbol{\rho}_t, \;\; t = 1, \ldots, T \quad (63)
$$

Múltiples períodos temporales y asignación común

- Para $t < q$, el **Supuesto NAC** implica $\tau_t = 0$ y $\boldsymbol{\rho}_t = 0$ y la ecuación que estima, usando POLS, consistentemente el impacto de la política es,

$$
\begin{aligned}
y_{it} &= \eta + \lambda s_i + \dot{\mathbf{x}}_i \boldsymbol{\kappa} + (s_i \times \dot{\mathbf{x}}_i)\,\zeta + \theta_2 \lambda 2_t + \cdots + \theta_T \lambda T_t \\
&+ (\lambda 2_t \times \dot{\mathbf{x}}_i)\,\pi_2 + \cdots + (\lambda T_t \times \dot{\mathbf{x}}_i)\,\pi_T + \tau_q\,(s_i \times \lambda q_t) + \cdots + \tau_T\,(s_i \times \lambda T_t) \\
&+ (s_i \times \lambda q_t \times \dot{\mathbf{x}}_i)\,\rho_q + \cdots + (s_i \times \lambda T_t \times \dot{\mathbf{x}}_i)\,\rho_T + v_{it} \qquad (64)
\end{aligned}
$$

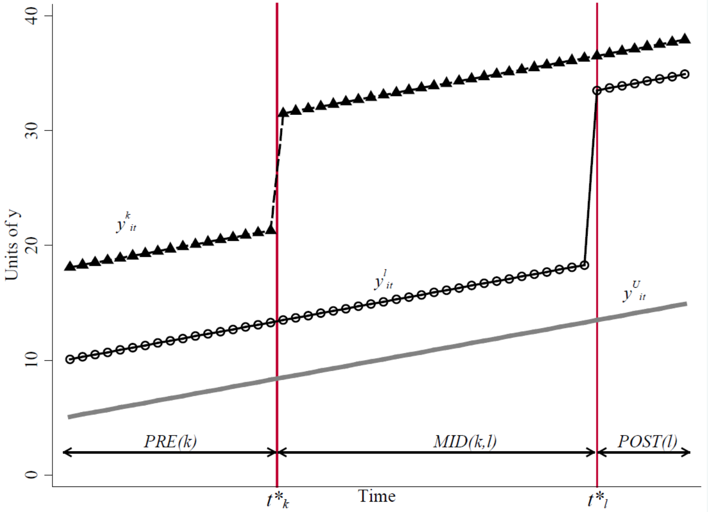
la implementación práctica se hace con $\dot{\mathbf{x}}_i = \mathbf{x}_i - \overline{\mathbf{x}}_1$.

Agenda

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- El caso típico de heterogeneidad se da cuando las unidades reciben el tratamiento en diferentes períodos temporales.
- Para ilustrar esto considere un panel con T períodos temporales (t) y N unidades de corte transversal (i) que pertenecen a un grupo que nunca recibe el tratamiento (U); a un grupo que recibe el tratamiento "temprano" (k) en t_k^* y a un grupo que recibe el tratamiento más tarde (l) en $t_l^* > t_k^*$.
- La siguiente figura ejemplifica esta estructura:

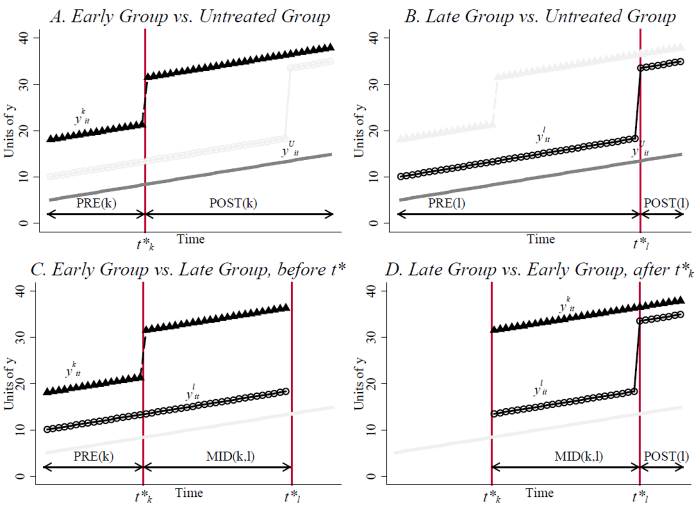
Múltiples Períodos Temporales y Tratamiento Heterogéneo



Múltiples Períodos Temporales y Tratamiento Heterogéneo

- En esta estructura la pregunta relevante es que efecto estima el coeficiente $\hat{\beta}_3$ en (29).
- En el caso canónico 2×2 sabemos que representa el cambio promedio después-antes de la variable de resultado en el grupo de tratamiento menos el mismo cambio en el grupo no tratado.
- En esta nueva estructura tenemos varias comparaciones como esta.
- La siguiente figura las representa:

Múltiples Períodos Temporales y Tratamiento Heterogéneo



Múltiples Períodos Temporales y Tratamiento Heterogéneo

- Goodman-Bacon (2021) muestra que el estimador de β_3 en (29) es un pomedio ponderado de todas las comparaciones potenciales de dos grupos como las que muestra la figura anterior.
- Si el efecto del tratamiento varía en el tiempo $\hat{\beta}_3$ estima en forma sesgada el efecto promedio del tratamiento sobre los tratados.
- Los ponderadores, en el promedio ponderado de $\hat{\beta}_3$, dependen de la varianza del tiempo que pasan en el tratamiento las unidades tratadas.
- En general, ser tratado en "el medio del panel" influencia el valor numérico estimado.
- En consecuencia alargar o acortar el panel puede cambiar la estimación puntual simplemente cambiando la varianza del tiempo pasado en el tratamiento y nada más.
- **Problema:** qué criterio usaríamos para determinar el largo óptimo del panel?

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- de Chaisemartin y D'Haultfœuille (2021) escriben el estimando como un promedio ponderado de ciertos efectos causales y en esta representación algunos ponderadores pueden ser negativos.
- Considere heterogenidad en el tiempo, $\tau_{it}(g) = \sum_{s \geqslant 0} \tau_s 1[t - g = s]$ es el impacto del programa para la unidad i en el período t si recibió el tratamiento en g.
- Todas las unidades tienen un efecto tratamiento τ_s en el s-ésimo período después de haber recibido el tratamiento.
- En este caso β_3 es un promedio ponderado potencialmente no convexo de los parámetros τ_s: $\beta_3 = \sum_s \omega_s \tau_s$ donde los ponderadores ω_s suman uno pero pueden ser negativos.
- Esto es problemático porque los τ_s pueden ser todos positivos pero β_3 puede ser negativo!
- De la misma manera, si los efectos de tratamiento difieren entre unidades i pero son constantes en el tiempo, $\tau_{it}(g) = \tau_i \; \forall \, t \geqslant g$, β_3 puede ser un promedio ponderado potencialmente no convexo de los efectos individuales τ_i.

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- **Intuición matemática:** por el teorema de Frisch-Waugh-Lovell el coeficiente $\hat{\beta}_3$ en (29) se puede recuperar de la estimación de y_{it} sobre $D_{it} - \hat{D}_{it}$, donde $\hat{D}_{it}$ es la predicción de la regresión de D_{it} sobre α_i y δ_t ($\hat{D}_{it} = \hat{\alpha}_i + \hat{\delta}_t$).
- Esto es,

$$\hat{\beta}_3 = \frac{\sum_{i,t}\left(D_{it} - \widehat{D}_{it}\right) y_{it}}{\sum_{i,t}\left(D_{it} - \widehat{D}_{it}\right)^2}. \tag{65}$$

- Sabemos que los valores de $\hat{D}_{it}$ pueden salir del intervalo cero-uno. Si esto ocurre, el ponderador $\frac{\left(D_{it}-\widehat{D}_{it}\right)}{\sum_{i,t}\left(D_{it}-\widehat{D}_{it}\right)^2} < 0$ y $\hat{\beta}_3$ es decreciente en Y_{it}.
- Pero como el impacto del programa $\tau_{it}(g)$ depende de y_{it}, se sigue que $\tau_{it}(g)$ tiene ponderadores negativos en $\hat{\beta}_3$.

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- Supuestos y Ecuaciones para estimar
- Definimos variables binarias de tratamiento por grupo, $\mathbf{d} = (d_q, \ldots, d_T)$ que indican el primer período en el que una unidad (un grupo) recibe el tratamiento.
- Asumiendo que hay un grupo de unidades que nunca recibe el tratamiento, hay $T - q + 2$ grupos que reciben la política.
- Hay múltiples niveles de tratamiento dependiendo de cuando una unidad inicialmente recibe el tratamiento.
- Defina para $r \in \{q, q+1, \ldots, T\}$, $y_t(r)$ denota el resultado potencial en el período t de una unidad que recibió la política en r.
- Defina $y_t(\infty)$ al resultado potencial en el período t de una unidad i que no recibe el tratamiento en ningún período, $t = 1, 2, \ldots, T$.

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- Defina el efecto del tratamiento en el período t como la comparación entre el resultado potencial del grupo que recibe la política en r y el grupo que nunca la recibe,

$$te_t(r) = y_t(r) - y_t(\infty), \; r = q, \ldots, T \tag{66}$$

- Los efectos del tratamiento que queremos identificar son los ATT de los períodos en los que los grupos reciben el tratamiento,

$$\tau_{rt} \equiv E\left[te_t(r) \mid d_r = 1\right], \; r = q, \ldots, T; \; t = r, \ldots, T \tag{67}$$

  el requerimiento de tener al menos un período sin unidades tratadas significa que $q \geq 2$.
- La ecuación dice que si una unidad entra en el tratamiento en r esperamos estimar los ATT de las unidades de este grupo en los períodos $\{r, r+1, \ldots, T\}$.

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- **Supuesto NAE (No Anticipación) Escalonado.** Para los grupos de tratamiento $r = q, q+1, \ldots, T$,

$$E\left[y_t(r) - y_t(\infty) \mid \mathbf{d}\right] = 0, \;\; t < r. \tag{68}$$

  La forma fuerte de este supuesto es que $y_t(r) = y_t(\infty)$ para $t < r$ que implica que no importa cuando la unidad recibe el tratamiento, los resultados potenciales son iguales antes de la implementación.
- Por ejemplo, suponga $T = 5$ y $q = 3$, entonces $y_t(3) = y_t(4) = y_t(5) = y_t(\infty)$ para $t < 3$; $y_t(4) = y_t(5) = y_t(\infty)$ para $t < 4$ y $y_t(5) = y_t(\infty)$ para $t < 5$.
- Este supuesto es importante para definir los grupos de control relevantes. Por ejemplo, si el primer período de la intervención es $t = 3$ entonces los grupos $r \in \{4, 5, \infty\}$ contienen unidades de control válidas

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- **Supuesto TPE (Tendencias Paralelas) Escalonadas.** Para variables las binarias de tratamiento por grupo, $d_q, \ldots, d_T$,

$$E\left[y_t(\infty) - y_1(\infty) \mid d_q, \ldots, d_T\right] = E\left[y_t(\infty) - y_1(\infty)\right] \equiv \theta_t, \; t = 2, \ldots, T \tag{69}$$

- Esto es lo mismo que asumir que las tendencias promedio en períodos adyacentes no cambian con la intervención:

$$E\left[y_t(\infty) - y_{t-1}(\infty) \mid d_q, \ldots, d_T\right] = E\left[y_t(\infty) - y_{t-1}(\infty)\right], \; t = 2, \ldots, T \tag{70}$$

- Asumimos en principio que hay un grupo que nunca recibe el tratamiento y que hay una probabilidad positiva de recibir el tratamiento en cada período.
- Esto implica que τ_{rt} está identificado para $r \in \{q, \ldots, T\}$ y $t \in \{r, \ldots, T\}$.

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- Escribamos la variable de resultado observada en cualquier período t como:

$$\begin{aligned} y_t &= y_t(\infty) + d_q \times [y_t(q) - y_t(\infty)] + d_{q+1} \times [y_t(q+1) - y_t(\infty)] + \\ \cdots & +d_T \times [y_t(T) - y_t(\infty)] \\ &= y_t(\infty) + d_q \times te_t(q) + \cdots + d_T \times te_t(T) \end{aligned} \tag{71}$$

  subyacente a esta ecuación es que $d_q + \cdots + d_T$ no es identicamente igual a uno y por lo tanto $d_\infty \equiv 1 - (d_q + \cdots + d_T)$ no es identicamente cero.
- Tomando esperanzas condicionales tenemos

$$E(y_t \mid \mathbf{d}) = E[y_t(\infty) \mid \mathbf{d}] + d_q \times E[te_t(q) \mid \mathbf{d}] + \cdots + d_T \times E[te_t(T) \mid \mathbf{d}] \tag{72}$$

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- Recuerde que las variables binarias de tratamiento por grupo son ortogonales, solo contienen un valor unitario en cada grupo y cero en el resto:

$$d_r E\left[te_t(r) \mid d_q, \ldots, d_t\right] = d_r E\left[te_t(r) \mid d_r = 1\right] \tag{73}$$

- Por lo tanto

$$E\left(y_t \mid \mathbf{d}\right) = E\left[y_t(\infty) \mid \mathbf{d}\right] + d_q \times E\left[te_t(q) \mid d_q = 1\right] + \cdots + d_T \times E\left[te_t(T) \mid d_T = 1\right] \tag{74}$$

- Para $r < t$, el Supuesto NAE implica que $E\left[te_t(r) \mid d_r = 1\right] = 0$ por lo que para $t < q$:

$$E\left(y_t \mid \mathbf{d}\right) = E\left[y_t(\infty) \mid \mathbf{d}\right] \tag{75}$$

- Para $t \geq q$

$$\begin{aligned} E\left(y_t \mid \mathbf{d}\right) &= E\left[y_t(\infty) \mid \mathbf{d}\right] + d_q \times E\left[te_t(q) \mid d_q = 1\right] + \cdots + d_T \times E\left[te_t(T) \mid d_T = 1\right] \\ &\equiv E\left[y_t(\infty) \mid \mathbf{d}\right] + d_q \tau_{qt} + \cdots + d_T \tau_{Tt} \end{aligned} \tag{76}$$

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- Escribamos el resultado potencial $y_t(\infty)$ en términos del resultado potencial inicial para los que nunca reciben el tratamiento más el cambio con relación al primer período:

$$y_t(\infty) = y_1(\infty) + g_t(\infty), \;\; t = 2, \ldots, T \tag{77}$$

- Imponiendo el Supuesto TPE tenemos,

$$E\left[y_t(\infty) \mid \mathbf{d}\right] = E\left[y_1(\infty) \mid \mathbf{d}\right] + E\left[g_t(\infty) \mid \mathbf{d}\right] \equiv \eta + \psi_q d_q + \cdots + \psi_T d_T + \theta_t \tag{78}$$

donde $\theta_1 = 0$ porque ya tenemos la contante η.

- Combinando estas ecuaciones para $1 \leq t < q$,

$$E\left[y_t \mid \mathbf{d}\right] = \eta + \psi_q d_q + \cdots + \psi_T d_T + \theta_t \tag{79}$$

y para $q \leq t \leq T$,

$$E\left(y_t \mid \mathbf{d}\right) = \eta + \psi_q d_q + \cdots + \psi_T d_T + \theta_t + \tau_{qt} d_q + \cdots + \tau_{Tt} d_T \tag{80}$$

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- El Supuesto TPE impone coeficientes comunes sobre d_q para $t < q$.
- Para imponer estas restricciones en la ecuación a estimar es útil escribir la ecuación para una observación genérica de corte transversal i y cualquier t usando variables binarias temporales como sigue:

$$E\left(y_{it} \mid \mathbf{d}_i\right) = \eta + \psi_q d_{iq} + \cdots + \psi_T d_{iT} + \sum_{s=2}^{T} \theta_s \lambda s_t + \sum_{r=q}^{T} \sum_{s=r}^{T} \tau_{rs} \left(d_{ir} \times \lambda s_t\right) \tag{81}$$

- Como antes, esta expresión es equivalente a:

$$E\left(y_{it} \mid \mathbf{d}_i\right) = \eta + \psi_q d_{iq} + \cdots + \psi_T d_{iT} + \sum_{s=2}^{T} \theta_s \lambda s_t + \sum_{r=q}^{T} \sum_{s=r}^{T} \tau_{rs} \left(D_{it} \times d_{ir} \times \lambda s_t\right) \tag{82}$$

- Los coeficientes de esta ecuación se pueden estimar consistentemente por POLS regresando y_{it} sobre
una constante, $d_{iq}, \ldots, d_{iT}, \lambda 2_t, \ldots, \lambda T_t, d_{iq} \times \lambda q_t, \ldots, d_{iq} \times \lambda T_t, \ldots, d_{iT} \times \lambda T_t$

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- Una vez estimados los τ_{rt} se pueden agregar para obtener un solo valor, por ejemplo, la agregación simple es:

$$\bar{\tau} \equiv \frac{1}{(T-q+1)(T-q+2)/2} \sum_{r=q}^{T} \sum_{t=r}^{T} \tau_{rt} \tag{83}$$

- También se puede calcular el efecto tratamiento por grupo:

$$\widehat{\bar{\tau}}_r \equiv \frac{1}{(T-r+1)} \sum_{t=r}^{T} \hat{\tau}_{rt}, \tag{84}$$

- Siguiendo la misma lógica que la de asignación común del tratamiento en esta estimaciones se pueden incluir variables explicativas que no varíen en el tiempo sin alterar las estimaciones de los ATT.

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- Callaway y Sant'Anna (2021) sugieren la siguiente estimación.
- Defina $G_i = \min\{t : D_{it} = 1\}$ como el período más temprano en el que la unidad i recibió el tratamiento. Si nunca recibió el tratamiento denotamos $G_i = \infty$.
- Suponga que el tratamiento es un estado absorbente, esto es $D_{it} = 1 \; \forall t \geq G_i$.
- Defina $R_{it} = t - G_i + 1$ como el tiempo transcurrido desde que recibió el tratamiento tal que $R_{it} = 1$ es el primer período en el que la unidad i recibió el tratamiento.
- **Resultados Potenciales:** defina con $\mathbf{0}_s$ y $\mathbf{1}_s$ a vectores s-dimensionales de ceros y unos, respectivamente.

Múltiples Períodos Temporales y Tratamiento Homogéneo

- Denote el resultado potencial de la unidad i en el período t si recibió el tratamiento en g por $Y_{it}\left(\mathbf{0}_{g-1}, \mathbf{1}_{T-g+1}\right)$, y denote por $Y_{it}\left(\mathbf{0}_{T}\right)$ su resultado potencial si nunca recibió el tratamiento.
- Como asumimos que el tratamiento es un estado absorbente el sendero entero de resultados potenciales se puede resumir con la primera fecha en que recibió el tratamiento, g, simplificando la notación con la fecha en la que inició el tratamiento: $Y_{it}(g) = Y_{it}\left(\mathbf{0}_{g-1}, \mathbf{1}_{T-g+1}\right)$ y $Y_{it}(\infty) = Y_{it}\left(\mathbf{0}_{T}\right)$.
- **Supuesto de Tendencias Paralelas:** para todo $t \neq t'$ y $g \neq g'$,

$$\mathbb{E}\left[Y_{it}(\infty) - Y_{it'}(\infty) \mid G_i = g\right] = \mathbb{E}\left[Y_{it}(\infty) - Y_{it'}(\infty) \mid G_i = g'\right]. \tag{85}$$

  este supuesto dice que en el contrafáctico donde el tratamiento no se recibió, los resultados potenciales para todos los grupos que recibieron el tratamiento habrían evolucionado de forma paralela.

Múltiples Períodos Temporales y Tratamiento Homogéneo

- **Supuesto de No Anticipación:** $Y_{it}(g) = Y_{it}(\infty)$ para todo $t < g$
- Intuitivamente, este supuesto dice que si una unidad no recibió el tratamiento en t, su resultado potencial no depende de en que período de tiempo lo reciba en el futuro.
- **Remark:** cuando no hay heterogeneidad en el tratamiento, $\tau_{it}(g) = Y_{it}(g) - Y_{it}(\infty) = \tau^{DID}, \; \forall t \geq g$ y se cumplen los supuestos de tendencias paralelas y no anticipación el coeficiente β_3 en (29) coincide con τ^{DID}.
- Los problemas surgen cuando el tratamiento es heterogéneo en el tiempo o entre unidades de la población.

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- Una forma de tomar en cuenta estos problemas cuando **el tratamiento es heterogéneo en el tiempo y constante entre unidades** es estimar la versión dinámica de la ecuación (29):
$$Y_{it} = \alpha_i + \delta_t + \sum_{r \neq 0} 1\left[R_{it} = r\right] \beta_r + \epsilon_{it}. \tag{86}$$
donde $1\left[R_{it} = r\right]$ son variables indicadoras del tiempo transcurrido desde el tratamiento.
- Borusyak, Jaravel y Spiess (2022) y Sun y Abraham (2021) muestran que si $\tau_{it}(g) = \sum_{s \geqslant 0} \tau_s 1[t-g=s]$ (es decir, todas las unidades tienen un efecto tratamiento τ_s en el s-ésimo período después de recibir el tratamiento) y se cumple el supuesto de tendencias paralelas: $\beta_s = \tau_s$.
- Sun y Abraham (2021) muestran que si el tratamiento es heterogéneo entre unidades este resultado no se mantiene.

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- Definen el efecto tratamiento promedio sobre los tratados por grupo-tiempo como:

$$ATT(g,t) = \mathbb{E}\left[Y_{it}(g) - Y_{it}(\infty) \mid G_{it} = g\right] \tag{87}$$

  esto es, el efecto tratamiento promedio en el período t para el grupo que recibió el tratamiento en el período g.

- Se puede identificar el $ATT(g,t)$ comparando el cambio esperado en la variable de resultado para el grupo g entre los períodos $g-1$ y t con el mismo cambio para un grupo de control que todavía no recibió el tratamiento al período t:

$$ATT(g,t) = \mathbb{E}\left[Y_{it} - Y_{i,g-1} \mid G_i = g\right] - \mathbb{E}\left[Y_{it} - Y_{i,g-1} \mid G_i = g'\right], \text{ para todo } g' > t \tag{88}$$

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- Esta identificación se mantiene para cualquier grupo de comparación $g' > t$ y por lo tanto se mantiene si promediamos sobre el conjunto de grupos comparables $\mathcal{G}$ tal que $g' > t$ para todo $g' \in \mathcal{G}$,

$$ATT(g,t) = \mathbb{E}\left[Y_{it} - Y_{i,g-1} \mid G_i = g\right] - \mathbb{E}\left[Y_{it} - Y_{i,g-1} \mid G_i \in \mathcal{G}\right] \tag{89}$$

- La estimación se hace reemplazando las esperanzas matemáticas por sus análogos muestrales:

$$\widehat{ATT}(g,t) = \frac{1}{N_g} \sum_{i:G_i=g} \left[Y_{it} - Y_{i,g-1}\right] - \frac{1}{N_{\mathcal{G}}} \sum_{i:G_i \in \mathcal{G}} \left[Y_{it} - Y_{i,g-1}\right] \tag{90}$$

- Callaway y Sant'Anna (2021) sugieren usar en $\mathcal{G}$ las unidades que nunca reciben el tratamiento ($\mathcal{G} = \{\infty\}$).

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- En este caso, $\widehat{ATT}(g,t)$ se puede recuperar restringiendo la muestra a observaciones en el momento t y $g-1$ para unidades que recibieron el tratamiento en g y unidades que nunca lo recibieron y estimando:

$$Y = \alpha_1^{g,t} + \alpha_2^{g,t} \cdot G_g + \alpha_3^{g,t} \cdot 1\{T=t\} + \beta^{g,t} \cdot (G_g \times 1\{T=t\}) + \epsilon^{g,t} \tag{91}$$

  en esta ecuación $\beta^{g,t} = ATT(g,t)$.
- Cuando hay un número relativamente pequeño de períodos y grupos tiene sentido reportar $\widehat{ATT}(g,t)$ para todos los g y t relevantes.
- Si hay muchos períodos post-tratamiento y/o grupos se puede calcular un promedio ponderado del efecto del tratamiento l períodos después de recibido a lo largo de los diferentes grupos que recibieron el tratamiento:

$$ATT_l^w = \sum_g w_g ATT(g, g+l). \tag{92}$$

  los w_g son los ponderadores que incluyen el promedio simple de los grupos considerados.

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- En términos generales el efecto agregado surge de

$$ATT = \sum_{g \in \mathcal{G}} \sum_{t=2}^{T} w_{g,t} ATT(g,t). \tag{93}$$

donde $w_{g,t}$ son los ponderadores.

- **Efecto Agregado Simple:** considera el promedio ponderado de los efectos $ATT(g,t)$ de cada cada grupo en cada período post-tratamiento.
- Analíticamente, los ponderadores se expresan como:

$$\mathrm{w}^{\mathrm{S}}(\mathrm{g},\mathrm{t}) = \frac{1\{\mathrm{t} \geq \mathrm{g}\}\mathrm{P}(\mathrm{G} = \mathrm{g} \mid \mathrm{G} \leq T)}{\kappa} \tag{94}$$

donde $\mathrm{P}(\mathrm{G} = \mathrm{g} \mid \mathrm{G} \leq T)$ es la probabilidad condicional de ser tratado por primera vez en g y $\kappa = \sum_{\mathrm{g} \in \mathcal{G}} \sum_{\mathrm{t}=2}^{T} 1\{\mathrm{t} \geq \mathrm{g}\}\mathrm{P}(\mathrm{G} = \mathrm{g} \mid \mathrm{G} \leq T)$ asegura que los ponderadores sumen 1.

Múltiples Períodos Temporales y Tratamiento Heterogéneo

- **Efecto Agregado por Grupos:** utiliza como ponderadores para las unidades tratadas en período $\tilde{g}$
$$w(g,t) = \mathbf{1}\{t \geq g\}1\{g = \tilde{g}\}/(T - \tilde{g} + 1) \tag{95}$$
donde $\mathbf{1}\{t \geq g\}$ es una indicadora para cuando el período t es mayor o igual a g, $1\{g = \tilde{g}\}$ es una indicadora de si las unidades del grupo g son iguales a las del grupo $\tilde{g}$ y $T - \tilde{g} + 1$ solo cuenta el número de períodos en los que se observa la unidad.
- Como ejemplo piense en un programa con tres períodos temporales y unidades tratadas en el primer y tercer período.
- El efecto promedio del tratamiento para las unidades tratadas el primer período es el promedio simple de los efectos $ATT(\widetilde{g} = 1, t = 1)$, $ATT(\widetilde{g} = 1, t = 2)$ y $ATT(\widetilde{g} = 1, t = 3)$, y el efecto promedio para las firmas tratadas en el último período es simplemente $ATT(\widetilde{g} = 3, t = 3)$.

Referencias

- Kirill Borusyak, Xavier Jaravel, Jann Spiess (2022). Revisiting Event Study Designs: Robust and Efficient Estimation. *https://arxiv.org/abs/2108.12419*.
- Callaway, B., & Sant'Anna, P. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*. Volume 225, Issue 2, December 2021, Pages 200-230.
- Liyang Sun & Sarah Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*. Volume 225, Issue 2, December 2021, Pages 175-199.
- Jonathan Roth, Pedro H. C. Sant'Anna, Alyssa Bilinski, John Poe (2022). What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature. *https://arxiv.org/abs/2201.01194*
- Jeffrey M. Wooldridge (2021). Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators *https://economics.princeton.edu/wp-content/uploads/2021/08/two_way_mundlak-Wooldridge.pdf*

Agenda

Modelos de Efectos Fijos y Aleatorios

- Consideremos un modelo para variables observadas a través de unidades de corte transversal durante varios períodos de tiempo. β es el parámetro que estamos interesados en estimar y α_i es un efecto no observado, invariante en el tiempo, denominado heterogeneidad no observada:

$$\begin{aligned} y_{it} &= \alpha_i + x_{it}\beta + u_{it} \\ i &\in \{1, ..., N\} = \mathcal{N} \\ t &\in \{1, ..., T\} = \mathcal{T} \end{aligned}$$

- Este modelo puede ser escrito apilando las observaciones en el tiempo como:

$$y_i = X_i\beta + \alpha_i J_T + u_i$$

- Este modelo se conoce en la literatura con el nombre de modelo de componentes no observados.

Modelos de Efectos Fijos y Aleatorios

- donde:

$$y_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix} \quad ; \quad X_i = \begin{bmatrix} x_{i1}^{'} \\ \vdots \\ x_{iT}^{'} \end{bmatrix} ; \quad J_T = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

- Los métodos tradicionales de datos de panel para estimar β son los denominados modelos de efectos fijos y de efectos aleatorios.
- La diferencia entre estos métodos se basa en el supuesto acerca de la relación entre la heterogeneidad no observada y las variables explicativas observadas.
- El modelo de efectos aleatorios asume que $cov(\alpha_i, x_{it}) = 0 \; \forall \, t$ y deja a la heterogeneidad no observada dentro del término de error:

$$y_{it} = \beta^{'} x_{it} + v_{it} \tag{96}$$

donde $v_{it} = \alpha_i + u_{it}$ y $E(v_{it}|X_i) = 0$.

Modelo de Efectos Aleatorios

- Los supuestos tradicionales para estimar este modelo son:
  - $E(u_{it}|X_i, \alpha_i) = 0$ y
  - $E(\alpha_i|X_i) = 0$ con $X_i = (x_{i1}, x_{i2}, \cdots, x_{iT})$
- La estimación se hace por MCGE.
- Apilando las observaciones en el tiempo tenemos:

$$y_i = X_i\beta + v_i \tag{97}$$

  con $v_i = \alpha_i J_T + u_i$.
- Bajo homocedasticidad y ausencia de correlación serial en u_{it}:

$$\Omega = E(v_i v_i') = \sigma_\alpha^2 J_T J_T' + \sigma_u^2 I_T \tag{98}$$

  y
-

$$\widehat{\beta}^{RE} = \left(\sum_{i=1}^{N} X_i'\widehat{\Omega}^{-1}X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i'\widehat{\Omega}^{-1}y_i\right) \tag{99}$$

Estimation

Modelo de Efectos Fijos

- El modelo de efectos fijos asume que $cov(\alpha_i, x_{it}) \neq 0$ y por lo tanto $E(v_{it}|X_i) \neq 0$. En otras palabras, una o más variables explicativas están correlacionadas con el error compuesto y MCGE dará estimaciones sesgadas e inconsistentes.
- **Solución:** transformar el modelo de forma de eliminar α_i.
- Denotemos a las variables en desvíos de sus medias temporales con ¨ por ejemplo, $\ddot{y} = y_{it} - \bar{y}_i$ con $\bar{y}_i = (1/T)\sum_{t=1}^{T} y_{it}$.
- Entonces el estimador de efectos fijos de β, $\widehat{\beta}^{FE}$ es:

$$\begin{aligned} \widehat{\beta}^{FE} &= \left(\sum_{i=1}^{N} \ddot{X}_i' \ddot{X}_i\right)^{-1} \left(\sum_{i=1}^{N} \ddot{X}_i' \ddot{y}_i\right) \\ &= \left(\ddot{X}'\ddot{X}\right)^{-1} \ddot{X}'\ddot{y} \qquad (100) \end{aligned}$$

Modelo de Efectos Fijos

- (cont.) El estimador de efectos fijos de β, $\widehat{\beta}^{FE}$ es:

$$
\begin{aligned}
\widehat{\beta}^{FE} &= \left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}'\ddot{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} \ddot{x}_{it}'\ddot{y}_{it}\right) \\
&= \left(\sum_{i=1}^{N} X_i' Q_T X_i\right)^{-1}\left(\sum_{i=1}^{N} X_i' Q_T y_i\right) \\
&= \left(\ddot{X}'\ddot{X}\right)^{-1}\ddot{X}'\ddot{y} \qquad (101) \\
&= \left[X'(I_N \otimes Q_T)X\right]^{-1} X'(I_N \otimes Q_T)y
\end{aligned}
$$

donde $Q_T = I_T - J_T(J_T'J_T)^{-1}J_T'$ es la matriz time-demeaning que es una matriz simétrica e idempotente de rango $T-1$.

- Note que $Q_T \times J_T = 0$; $Q_T \times y_i = \ddot{y}_i$ y $Q_T \times X_i = \ddot{X}_i$.

Modelo de Efectos Fijos

- Transformar el modelo restándole a cada variable su media temporal elimina la heterogeneidad no observada α_i. Esta transformación recibe el nombre de transformación de efectos fijos.
- Existen otras transformaciones para eliminar la heterogeneidad no observada. Una de las más utilizadas es la transformación de diferencias finitas.
- Considere el modelo de componentes no observados escrito para los períodos t y $t-1$,

$$
\begin{aligned}
y_{it} &= \alpha_i + x_{it}\beta + u_{it} && (102)\\
y_{it-1} &= \alpha_i + x_{it-1}\beta + u_{it-1} && (103)
\end{aligned}
$$

- Restando miembro a miembro,

$$
\begin{aligned}
y_{it} - y_{it-1} &= (x_{it} - x_{it-1})\beta + (u_{it} - u_{it-1}) && (104)\\
\Delta y_{it} &= \Delta x_{it}\beta + \Delta u_{it} && (105)
\end{aligned}
$$

Modelo de Efectos Fijos

- El estimador de diferencias finitas de β, $\hat{\beta}^{FD}$ es:

$$
\begin{aligned}
\widehat{\beta}^{FD} &= \left(\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta x_{it}'\Delta x_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T}\Delta x_{it}'\Delta y_{it}\right) \\
&= \left(\sum_{i=1}^{N}(DX_i)'DX_i\right)^{-1}\left(\sum_{i=1}^{N}(DX_i)'Dy_i\right) \\
&= \left(\sum_{i=1}^{N}X_i'D'DX_i\right)^{-1}\left(\sum_{i=1}^{N}X_i'D'Dy_i\right) \qquad (106)
\end{aligned}
$$

donde

$$
D = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ddots & -1 & 1 \end{bmatrix} \qquad (107)
$$

Modelo de Efectos Fijos

- Note que para que este estimador sea consistente necesitamos que se cumpla que:
  Supuesto $FD.1: \; E(u_{it}|X_i, \alpha_i) = 0, \;\; t = 1, 2, \ldots, T$
- Bajo el supuesto $FD.1$, MCC en (100) y (106) será consistente porque $E(\ddot{x}'_{it}\ddot{u}_{it}) = 0, t = 1, \ldots, T$ y $E(\Delta x'_{it}\Delta u_{it}) = 0, t = 2, \ldots, T$, respectivamente. Recuerde que además de este supuesto necesitamos una condición de rango estándar.
- Si agregamos el supuesto:
  Supuesto $FD.2: \; E(u_i u'_i|X_i, \alpha_i) = \sigma_u^2 I_T.$
  el estimador de efectos fijos es el más eficiente de los dos estimadores.
- Note que bajo $FD.2$, $E(Du_i u'_i D'|X_i, \alpha_i) = DE(u_i u'_i|X_i, \alpha_i)D' = \sigma_u^2 DD'$.
- Lo que muestra que los errores estarán correlacionados para períodos adjacentes.

Modelo de Efectos Fijos

- Bajo argumentos estándar de MCG la estimación más eficiente con la transformación de diferencias finitas es:

$$\widehat{\beta}^{FD} = \left(\sum_{i=1}^{N} X_i' D'(DD')^{-1} DX_i \right)^{-1} \left(\sum_{i=1}^{N} X_i' D'(DD')^{-1} Dy_i \right) \tag{108}$$

- Otro punto interesante es que, la matriz idempotente $D'(DD')^{-1}D$ también puede escribirse como:
  $D'(DD')^{-1}D \equiv I_T - J_T(J_T' J_T)^{-1} J_T' = Q_T$
  donde Q_T es la matriz time demeaning de la transformación de efectos fijos que vimos antes.
- Esto muestra que $\widehat{\beta}^{FE} \equiv \widehat{\beta}^{FD}$.
- Back

Modelo de Efectos Aleatorios

- Implementación: Necesitamos un estimador consistente de Ω.
- Para poder implementar FGLS necesitamos las estimaciones consistentes de σ_c^2 y σ_u^2. Para esto definamos $\sigma_v^2 = \sigma_c^2 + \sigma_u^2$.
- $\sigma_v^2 = T^{-1} \sum_{t=1}^{T} E(v_{it}^2)$ para todo i.
- Por lo tanto, un estimador consistente de σ_v^2 es:

$$\sigma_v^2 = \frac{1}{NT - K} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\hat{v}}_{it}^2$$

Donde $\hat{\hat{v}}_{it}$ son los residuos de POLS.

Modelo de Efectos Aleatorios

- Para encontrar un estimador consistente de σ_c^2, recuerde que $\sigma_c^2 = E(v_{it}v_{is})$, para todo $t \neq s$.
- Un estimador consistente es entonces:

$$\sigma_c^2 = \frac{1}{[NT(T-1)-1]/2-K} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{\hat{v}}_{it} \hat{\hat{v}}_{is} \tag{109}$$

- Con estos resultados podemos estimar consistentemente: $\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_c^2$.

Back