Semana 11. Estudios de Evento

Equipo Econometría Avanzada

Universidad de los Andes

28 de octubre de 2022

Contenido

Estudios de evento

Los estudios de evento son una generalización del método de diferencias en diferencias cuando las unidades son tratadas en distintos momentos del tiempo.

En esta clase nos centraremos en el caso donde el tratamiento constituye un estado absorbente (en la literatura se conoce como "staggered adoption"). Más precisamente, nos interesa el caso donde algunas unidades reciben un tratamiento (dicótomo), una única vez en el tiempo, y desde entonces permanecen siempre tratadas.

Existen, sin embargo, generalizaciones a este escenario:

- **Tratamiento continuo:** Callaway, Goodman-Bacon & Sant'Anna (2021)
- **El status de tratamiento se "apaga" y se "prende" en el tiempo:** de Chaisemartin & D'Haultfœuille (2021) e Imai y Kim (2021).
- **Fuzzy DiD:** Dos grupos que son tratados simultáneamente, pero la proporción de unidades expuestas incrementa diferencialmente en uno de los grupos. de Chaisemartin and D'Haultfœuille (2018).

Contexto y pregunta

Callaway & Sant'Anna (2020) estudian el efecto del salario mínimo en el empleo juvenil en E.E.U.U.

- Los autores estudian el efecto de incrementos en el salario mínimo en el empleo juvenil a nivel condado.
- En el 2001 el salario mínimo estaba fijo en $5.15 USD la hora.
- La mayoría de los condados se acogían a la norma federal, pero algunos incrementaron el pago mínimo entre 2001-2007. **Estos incrementos ocurrieron en distintos momentos del tiempo.**
- El grupo de tratados está compuesto por aquellos condados que aumentaron su salario mínimo.

Notación: Sean T periodos indexados por $t = 1, \ldots, T$ y N unidades indexadas por i. Consideramos las siguientes variables:

- D_{it}: Dicótoma que toma el valor de uno si la unidad i recibe el tratamiento en el periodo t
- $G_i = \min\{t : D_{it} = 1\}$: Variable que denota el primer periodo en el que la unidad i recibe el tratamiento. Si el individuo nunca es tratado, notamos $G_i = \infty$.
- $Y_{it}(g)$: Resultado potencial del individuo i en el periodo t en caso de haber sido tratado por primera vez en el periodo $G_i = g$.
- $Y_{it}(\infty)$: Resultado potencial del individuo i en el periodo t en caso de nunca haber sido tratado.
- $ATT(g, t) = \mathbb{E}[Y_{it}(g) - Y_{it}(\infty) | G_i = g]$: Efecto del tratamiento promedio en el momento t para los tratados en el periodo g.

Unidades que comparten el mismo momento de tratamiento g se conocen como una *cohorte*.

El modelo de TWFE

Intuición:

Supondremos que el resultado potencial de nunca ser tratado depende de las características de la unidad y de un componente temporal:

$$Y_{it}(\infty) = \alpha_i + \phi_t + \epsilon_{it}$$

Por otro lado, ser tratado en la cohorte g añade un efecto "$\tau_{it}(g)$" para los periodos posteriores al tratamiento. De manera que

$$Y_{it}(g) = \begin{cases} Y_{it}(\infty) = \alpha_i + \phi_t + \epsilon_{it} & \text{si } t < G_i \\ Y_{it}(\infty) + \tau_{it}(g) = \alpha_i + \phi_t + \tau_{it}(g) + \epsilon_{it} & \text{si } t \geq G_i \end{cases}$$

Entonces, si para cada unidad sólo observamos un estado del mundo, la variable dependiente se escribe como $Y_{it} = \sum_g \mathbb{1}[G_i = g] Y_{it}(g)$, de manera que:

$$Y_{it} = \alpha_i + \phi_t + \left(\sum_g \mathbb{1}[G_i = g] \tau_{it}(g) \right) \mathbb{1}[t \geq G_i] + \varepsilon_{it} = \alpha_i + \phi_t + \tau_{it}(G_i) D_{it} + \varepsilon_{it}$$

Estimación-TWFE por MCO

Modelo estático - TWFE

$$Y_{it} = \alpha_i + \phi_t + \tau D_{it} + \varepsilon_{it}$$

- Bajo los supuestos de DiD, si estimamos este modelo por OLS, **la intuición** nos dice que tendríamos que

$$\mathbb{E}[\hat{\tau}_{OLS}] \approx \mathbb{E}[\tau_{it}] = \sum_{g,t} \mathbb{P}_t(G_i = g)\mathbb{E}_t[Y_{it}(g) - Y_{it}(\infty)|G_i = g] = \sum_{g,t} \omega(g,t) ATT(g,t)$$

esto es, estimamos un promedio ponderado de los efectos causales de la política en cada cohorte.

Supuestos de identificación (?):

1. **Supuesto 1 - Tendencias paralelas generalizadas:** Hay tendencias paralelas entre cualesquiera pares de cohortes y momentos del tiempo. Para todo $t \neq t'$ y $g \neq g'$:

$$\mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty)|G_i = g] = \mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty)|G_i = g']$$

2. **Supuesto 2 - No anticipación:** Para todo $t < g$:

$$Y_{i,t}(g) = Y_{i,t}(\infty)$$

Resumen ejectutivo: problemas de TWFE

Si bien la estimación por MCO de los modelos TWFE para estudios de eventos es una práctica común en diseños de diferencias en diferencias y sus distintas variaciones, **los últimos años hemos aprendido que no permite estimar consistentemente efectos causales en el caso de estudios de evento.**

- Goodman-Bacon (2021): **El estimador de MCO de un modelo de TWFE no converge a una cantidad interpretable como efecto causal, incluso bajo el cumplimiento de tendencias paralelas generalizadas y no anticipación.** Sólo lo logra cuando hay efectos homogéneos o no existen efectos dinámicos. No obstante, incluso en este caso, los efectos tienen una interpretación confusa.

- Sun & Abraham (2021): Los anteriores problemas se mantienen aun cuando se usa el modelo dinámico. Esto es problemático pues: (1) los efectos dinámicos no tienen interpretación causal; (2) **la validación del supuesto de tendencias paralelas es inválido**.

- de Chaisemartin & D'Haultfœuille (2021), Wooldridge (2021): **El problema no es MCO, es que el modelo TWFE está mal especificado.**

Estimando por MCO un modelo de TWFE:

Goodman-Bacon (2021) se preguntan **¿qué estamos estimando exactamente al utilizar MCO en un modelo de TWFE?**

Respuesta: Un promedio de todos los diseños 2×2 que se pueden construir con los datos.

Descomposición de Goodman-Bacon

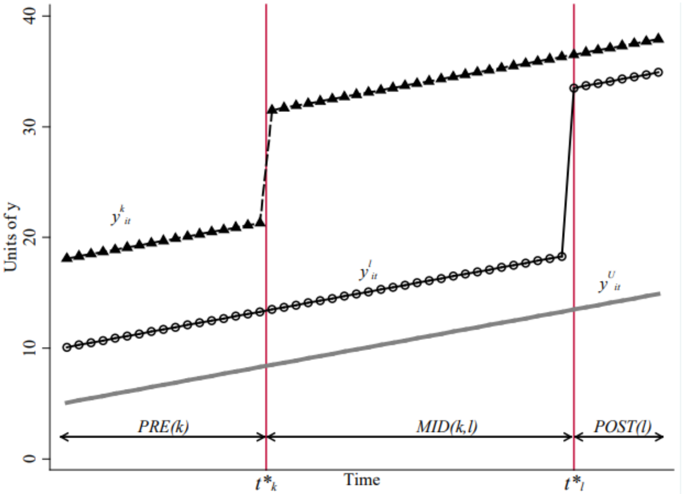
El estimador de MCO de TWFE se puede escribir como

$$\hat{\tau}_{OLS} = \sum_{g \neq g', t < t'} v_{g,g',t,t'} \; \hat{\delta}^{2\times 2}_{g,g',t,t'}$$

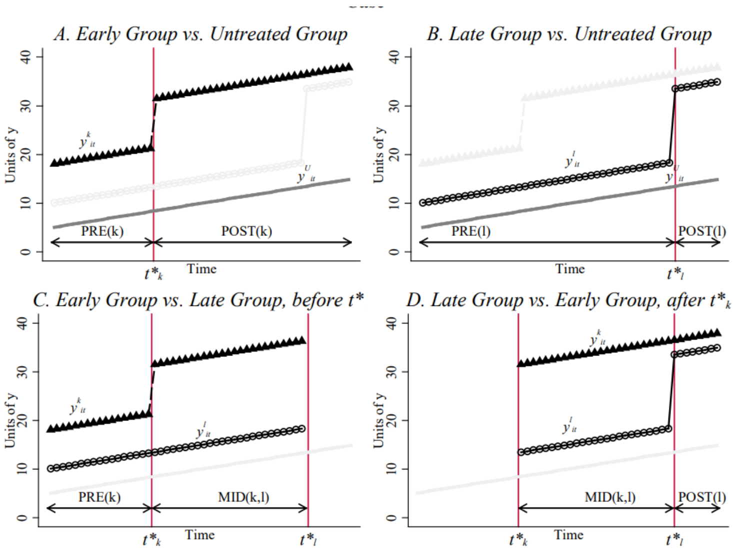
Donde $\hat{\delta}^{2\times 2}_{g,g',t,t'}$ es el estimador de diferencias en diferencias cuando el grupo de control es g, el grupo de tratamiento es g', el periodo pre es t y el post t'; y $v_{g,g',t,t'}$ son pesos no negativos, que suman uno, y tales que $v_{g,g',t,t'} > 0$ si y sólo si g' cambia de tratamiento entre t y t' pero g no.

Ejemplo sencillo

3 grupos: Nunca tratado (u), tratado temprano (k) y tratado tardío (l) - 4 comparaciones posibles.



Comparaciones 2×2



Implicaciones de Goodman-Bacon (2021)-I

Se puede mostrar que

$$\underset{N\to\infty}{\text{p-lim}}\ \hat{\tau}_{OLS} = \text{VWATT} + \text{VWPT} - \Delta\text{ATT}$$

donde

- **VWATT- Variance weighted ATT:** es un promedio ponderado de los efectos causales (ATT de cohorte-tiempo) por algunos pesos no negativos y que suman 1. Los pesos dependen del tamaño de la cohorte y la varianza de la variable de tratamiento (es máxima en la mitad del panel), por lo que no siguen una lógica clara (✓pero a la vez ✗):
  - ¿Por qué nos importaría tener más en cuenta a aquellas cohortes que fueron tratadas en la mitad de la cohorte?
  - ¿Por qué el parámetro de interés dependería de la longitud del panel?
- **VWPT- Variance weighted Parallel Trend Bias:** Es el sesgo que se produce por las posibles diferencias en tendencias de cada grupo 2×2. Bajo tendencias paralelas generalizadas, este término debería ser 0. (✓)

Implicaciones de Goodman-Bacon (2021)-II

Pero queda un término todavía... el **heterogeneity bias**

$$\Delta\text{ATT} \approx \sum_{g\neq\infty, t<t'} \chi(g,t,t')\left[\text{ATT}(g,t') - ATT(g,t)\right]$$

donde los pesos $\chi(\cdot)$ suman cero. Este término sólo desaparece en dos casos

- **Supuesto 3.1- Hay efectos constantes:** En este caso $\text{ATT}(g,t) = \text{ATT}$. Luego $\Delta ATT = 0$ y $VWATT = ATT$. (teóricamente ✓pero en la práctica nunca ocurre ✗)
- **Supuesto 3.2- Los efectos varían entre cohortes pero no en el tiempo (no efectos dinámicos):** $\text{ATT}(g,t) = \text{ATT}(g)$, luego $\Delta ATT = 0$. Logramos estimar $VWATT$, pero $VWATT$ tiene esos pesos raros. (✓pero a su vez ✗)

Resumen: Si no se cumplen 3.1 o 3.2, el estimador de MCO del modelo TWFE es *sesgado incluso bajo el cumplimiento de tendencias paralelas generalizadas.* Además, ¡si la heterogeneidad es muy grande, el estimador puede tomar el signo contrario al VWATT!

Intuición del problema

El problema viene de las comparaciones "prohibidas" que describe la descomposición de Goodman-Bacon. En general, hay 3 tipos de comparaciones 2×2:

- Tratados vs. Aún no tratados (✓)
- Tratados vs. Nunca tratados (✓)
- Tratados vs. Ya tratados ( )

Pero estas últimas no son comparaciones deseables: usan de control unidades ya tratadas, lo cual contamina el "contrafactual". ¡Estas comparaciones provocan el heterogeneity bias!

¿Y los problemas se mantienen si permitimos efectos dinámicos?

Modelo Dinámico - TWFE

$$Y_{it} = \alpha_i + \phi_t + \sum_{\substack{r \neq -1 \\ -T \leq r \leq \bar{T}}} \beta_r D_{it}^r + \varepsilon_{it}$$

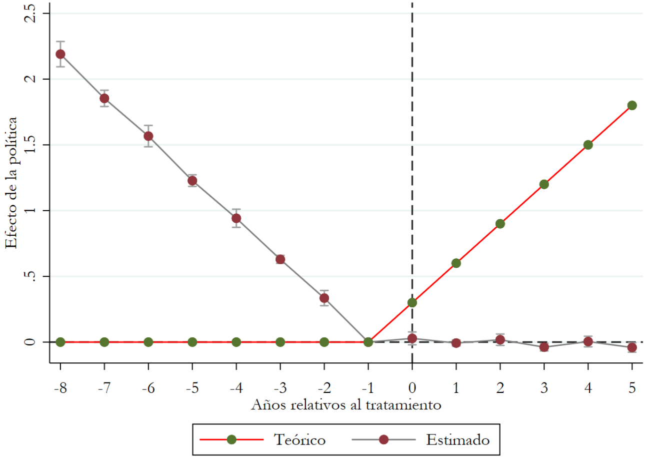
Donde

- D_{it}^r es una variable dicótoma que es igual a 1 si el año t está a r periodos desde el momento del tratamiento de i, t_i. Es decir, $D_{it}^r = \mathbb{1}\{t - t_i = r\}$.
- β_r: Para $r \geq 0$, **se piensa como** el efecto del tratamiento r periodos después de ser tratado. Para $r < -1$, **se piensa como** el efecto anticipatorio del tratamiento $|r|$ periodos antes.

Sun & Abraham (2021) muestran que, incluso bajo tendencias paralelas, los β_r no capturan los efectos dinámicos si hay efectos heterogéneos entre cohortes r periodos después del tratamiento. Así las cosas:

- Los efectos dinámicos son incorrectos.
- **La validación de tendencias paralelas puede ser incorrecta.**

¡Las cosas pueden salir muy mal! - Simulación



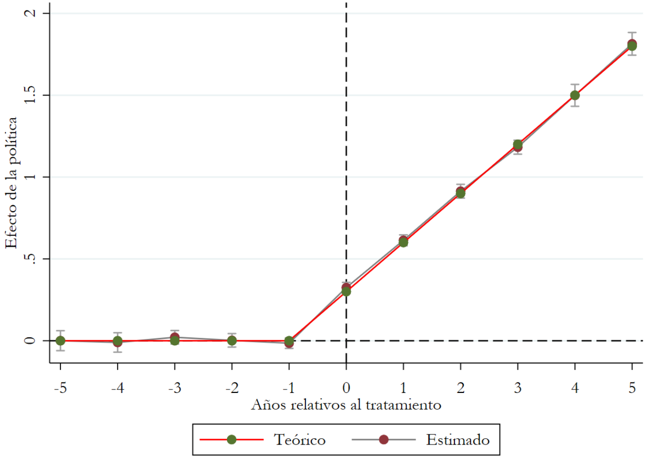
Resumen ejecutivo: nuevos estimadores

El problema de estimar un modelo de TWFE por MCO es que de forma automática hacen todas las posibles comparaciones, independientemente de si tienen sentido o no. La solución es **hacer manualmente las comparaciones correctas y luego agregar**.

- Callaway & Sant'Anna (2021): Método en dos etapas:
  - Estimar de manera no-paramétrica cada uno de los "bloques" $ATT(g,t)$. Los estimadores utilizados se conocen como estimadores **doblemente-robustos**.
  - Agregar manualmente los efectos usando pesos predefinidos dependiendo del efecto de interés: agregados, dinámicos, calendario, etc ... ¡Veremos esto qué significa en el do-file!
- Sun & Abraham (2021): Estimadores parametricos de los $ATT(g,t)$ usando modelos de regresion saturados.
- Boyusak, Jaravel & Spiess (2021): Estimador por imputación: extrapolamos el contrafactual usando los supuestos del modelo, luego hacemos diferencias en diferencias.
- Wooldridge (2021). Modifica el modelo de TWFE agregando términos adicionales lo cual correctamente identifica el efecto causal. Luego se estima por MCO.

Volviendo al ejemplo desastroso... ¡pero ahora con CS!

Volviendo al ejemplo desastroso... ¡pero ahora con CS!



Otros estimadores

Panel B: Estimators allowing dynamic effects
Can be used when outcome affected by past treatments

Treatment	*Estimators available*	*Stata commands*	*See:*
Binary and staggered	Callaway and Sant'Anna (2021)	`csdid`	3.2.1
	Sun and Abraham (2021)	`eventstudyinteract`	3.2.2
	Borusyak et al. (2021)	`did_imputation`	3.2.3
	de Chaisemartin and D'Haultfœuille (2021a)	`did_multiplegt`	3.3
Binary or discrete, non-staggered	de Chaisemartin and D'Haultfœuille (2021a)	`did_multiplegt`	3.3
Continuous and staggered	de Chaisemartin and D'Haultfœuille (2021a)	`did_multiplegt`	3.3
	Callaway et al. (2021)		3.3
Continuous and non-staggered, with stayers	de Chaisemartin et al. (2022)	See paper	3.3
Continuous and non-staggered, without stayers	No estimator available yet		
Several treatments	de Chaisemartin and D'Haultfœuille (2021b)	`did_multiplegt`	3.3

Tabla de estimadores para ES. de Chaisemartin & D'Haultfœuille (2022)

Conclusiones y prácticas recomendadas

- "The results discussed above show that while conventional TWFE specifications make sensible comparisons of treated and untreated units in the canonical two-period DiD model, in the staggered case they typically make "forbidden comparisons" between already-treated units. As a result, treatment effects for some units and time periods receive negative weights in the TWFE estimand. In extreme cases, this can lead the TWFE estimand to have the "wrong sign" — e.g., the estimand may be negative even if all the treatment effects are positive. Even if the weights are not so extreme as to create sign reversals, it may nevertheless be difficult to interpret which comparisons the TWFE estimator is making, as the "control group" is not transparent, and the weights it chooses are unlikely to be those most relevant for economic policy." (Roth et al.,2021)
- "The most direct remedy for this problem is to use methods that explicitly specify the comparisons to be made between treatment and control groups, as well as the desired weights in the target parameter." (Roth et al.,2021)

Conclusiones y prácticas recomendadas

However...

"It is also important to stress that at this stage, it is still unclear whether researchers should systematically abandon TWFE estimators. [...] While there are examples where TWFE and heterogeneity robust DID estimators are economically and statistically different (see, e.g., the empirical examples in de Chaisemartin and D'Haultfœuille, 2020, 2021a,b; Baker et al., 2022), the previous section also shows a data set where TWFE and heterogeneity-robust DID estimators lead to very similar conclusions. Understanding the circumstances where TWFE and heterogeneity-robust DID estimators are more likely to differ is an important question." (de Chaisemartin & D'Haultfœuille ,2022)

Context-specific knowledge will often be needed to choose the right identifying assumptions and accompanying methods.

Anexo: Detalles estimadores

Callaway & Sant'Anna (2021)

- Estimadores no-paramétricos de los $\mathrm{ATT}(g,t)$. Son estimadores doblemente-robustos, en tanto son consistentes siempre que se cumpla una de dos condiciones. (Ahorita hablaremos de eso)
- Permite recuperar efectos causales en presencia de heterogeneidad entre cohortes y de efectos dinámicos. Se pueden estimar tanto efectos causales dinámicos como en otros "niveles de agregación".
- Permite validar el supuesto de tendencias paralelas.
- Permite incluir controles (pre-determinados) para suavizar el supuesto de tendencias paralelas.
- Sólo es válido cuando el tratamiento es absorbente y el tratamiento binario.
- Es válido tanto para datos panel como para cortes transversales.
- Implementable en R y en Stata.

Supuestos de identificación

1) **Anticipación limitada al tratamiento:** Existe $\delta \geq 0$ conocido tal que, para todo $g, t \in \{1, \cdots, T\}$ tal que $t < g - \delta$

$$\mathbb{E}[Y_{it}(g)|X, G_i = g] = \mathbb{E}[Y_{it}(\infty)|X, G_i = g] \text{ c.s.}$$

2) **Tendencias paralelas:** Sea $\delta \geq 0$ el que determina el grado de anticipación. Entonces para todo $g \neq \infty$ y $t \in \{2, \cdots, T\}$ tales que $t \geq g - \delta$

$$\mathbb{E}[Y_{it}(\infty) - Y_{i,t-1}(\infty)|X, G_i = g] = \mathbb{E}[Y_{it}(\infty) - Y_{i,t-1}(\infty)|X, G_i = \infty] \text{ c.s.}$$

3) **Soporte comúm:** Para cada $t \in \{1, \cdots, T\}$ y g, existe un $\epsilon > 0$ tal que $\mathbb{P}(G_i = g) > \epsilon$ y $\mathbb{P}(D_{i,t} = 1|X) < 1 - \epsilon$ c.s.

4) Condición de consistencia del modelo no-paramético. (Continúo con el suspenso)

Nota: Se puede cambiar supuesto de tendencias paralelas: en vez de nunca tratados, considerar como control no tratados todavía ("not-yet treated").

Idea:

Bajo los supuestos $1) - 3)$, el ATT(g,t) se puede identificar como

$$ATT(g,t) = \mathbb{E}_{X|G_i=g}\Bigg[\mathbb{E}[Y_{it} - Y_{i,g-1}|G_i = g, X_i = x]$$
$$- \mathbb{E}[Y_{it} - Y_{i,g-1}|G_i \in \mathcal{G}, X_i = x]\Bigg]$$

donde $\mathcal{G}$ son un grupo de cohortes usadas como controles. Una opción es usar el grupo nunca tratado, otra es usar las cohortes que no han sido todavía tratadas, y otra usar ambas.

Con (MUCHA) paciencia, podemos manipular esta expresión de manera que podamos expresar la resta en cantidades que se puedan fácilmente estimar..

Idea 1 - The outcome regression approach

Una primera posibilidad es expresarlo de esta manera (esto viene de Heckman et al. (1998)):

$$\text{ATT}^{OR}(g,t;\delta) = \mathbb{E}\left[\frac{G_g}{\mathbb{E}[G_g]}(Y_t - Y_{g-\delta-1} - m_{g,t,\delta}(X))\right]$$

donde G_g es el tamaño del grupo $g \neq \infty$ y $m_{g,t,\delta}(X)) = \mathbb{E}[Y_t - Y_{g-\delta-1}|X, G_i \in \mathcal{G}]$ es la función esperada condicional del grupo de control.Entonces

$$\widehat{\text{ATT}}^{OR}(g,t;\delta) = \hat{\mathbb{E}}\left[\frac{G_g}{\hat{\mathbb{E}}[G_g]}(Y_t - Y_{g-\delta-1} - \widehat{m}_{g,t,\delta}(X))\right]$$

El estimador $\widehat{\text{ATT}}^{OR}(g,t;\delta)$ será un estimador consistente de $\text{ATT}^{OR}(g,t;\delta)$ siempre que $\widehat{m}_{g,t,\delta}(X)$, la función de la esperanza condicional estimada (sobre el grupo de control), esté correctamente especificada.

- Usualmente se usa regresiones lineales, incluyendo dummies de grupo y tiempo, junto con sus interacciones con X_i.
- Se puden recurrir a métodos no paramétricos como vecinos más cercanos, polinomios locales, etc...

Idea 2 - The inverse probability weighting approach

Otra posibilidad es expresarlo como un estimador del tipo Horvitz-Thompson (esto viene de Abadie(2005)):

$$\mathrm{ATT}^{IPW}(g,t;\delta) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\dfrac{p_g(X)\mathbb{1}[G_i \in \mathcal{G}]}{1-p_g(X)}}{\mathbb{E}\left[\dfrac{p_g(X)\mathbb{1}[G_i \in \mathcal{G}]}{1-p_g(X)}\right]}\right)(Y_t - Y_{g-\delta-1})\right]$$

donde $p_g(X) = \mathbb{P}(G_i = g|X)$ es el p-score. Entonces

$$\widehat{\mathrm{ATT}}^{IPW}(g,t;\delta) = \widehat{\mathbb{E}}\left[\left(\frac{G_g}{\widehat{\mathbb{E}}[G_g]} - \frac{\dfrac{\hat{p}_g(X)\mathbb{1}[G_i \in \mathcal{G}]}{1-\hat{p}_g(X)}}{\widehat{\mathbb{E}}\left[\dfrac{\hat{p}_g(X)\mathbb{1}[G_i \in \mathcal{G}]}{1-\hat{p}_g(X)}\right]}\right)(Y_t - Y_{g-\delta-1})\right]$$

El estimador $\widehat{\mathrm{ATT}}^{IPW}(g,t;\delta)$ será un estimador consistente de $\mathrm{ATT}^{IPW}(g,t;\delta)$ siempre que el modelo del p-score $\hat{p}_g(X)$ esté bien identificado.

Idea 3 - ¿Why not both?

Una última posibilidad es expresarlo como:

$$\text{ATT}^{DR}(g,t;\delta) = \mathbb{E}\left[\left(\frac{G_g}{\mathbb{E}[G_g]} - \frac{\dfrac{p_g(X)\mathbb{1}[G_i \in \mathcal{G}]}{1-p_g(X)}}{\mathbb{E}\left[\dfrac{p_g(X)\mathbb{1}[G_i \in \mathcal{G}]}{1-p_g(X)}\right]}\right)(Y_t - Y_{g-\delta-1} - m_{g,t,\delta}(X))\right]$$

Entonces

$$\widehat{\text{ATT}}^{DR}(g,t;\delta) = \widehat{\mathbb{E}}\left[\left(\frac{G_g}{\widehat{\mathbb{E}}[G_g]} - \frac{\dfrac{\hat{p}_g(X)\mathbb{1}[G_i \in \mathcal{G}]}{1-\hat{p}_g(X)}}{\widehat{\mathbb{E}}\left[\dfrac{\hat{p}_g(X)\mathbb{1}[G_i \in \mathcal{G}]}{1-\hat{p}_g(X)}\right]}\right)(Y_t - Y_{g-\delta-1} - \hat{m}_{g,t,\delta}(X))\right]$$

El estimador $\widehat{\text{ATT}}^{DR}(g,t;\delta)$ será un estimador consistente de $\text{ATT}^{DR}(g,t;\delta)$ siempre que el modelo del p-score $\hat{p}_g(X)$ **o** la CEF $\hat{m}_{g,t,\delta}(X)$ estén bien identificados. De ahí que sean **doblemente robustos.**

¿Y ahora qué? Pesamos los efectos de una manera controlada

Habiendo estimado los $ATT(g,t)$, lo que tenemos es que escoger pesos razonables para agregar los efectos causales. En general queremos estimar cosas de la forma:

$$\theta = \sum_{g} \sum_{t=2}^{T} w(g,t)\text{ATT}(g,t)$$

Ejemplos:

- **Overall effects:** Promedio de los efectos por cohorte:

$$\theta_{Agg} = \sum_{g} \theta_{sel}(g)\mathbb{P}(G_i = g | G_i \leq T); \quad \theta_{sel}(g) = \frac{1}{T-g+1} \sum_{t=g}^{T} \widehat{\text{ATT}}(g,t)$$

- **Relative time:** El efecto l unidades luego del tratamiento.

$$\theta_{es}(l) = \sum_{h} \mathbb{1}[g+l \leq T]\mathbb{P}(G_i = g | G_i + l \leq T)\widehat{\text{ATT}}(g,g+l)$$

Sun and Abraham (2021)

- Estimadores paramétricos de los ATT(g, t) usando un modelo de regresión saturados.
- Permite recuperar efectos causales en presencia de heterogeneidad entre cohortes y de efectos dinámicos. Se pueden estimar tanto efectos causales dinámicos como en otros niveles de agregación.
- Permite validar el supuesto de tendencia paralelas.
- Permite incluir controles que no varían en el tiempo.
- Sólo es válido para el cuando el tratamiento es absorbente y el tratamiento binario.
- Es válido tanto para datos panel como para cortes transversales.
- Implementable en Stata.

Supuestos de identificación

1. **Supuesto 1 - Tendencias paralelas generalizadas:** Hay tendencias paralelas entre cualesquiera pares de cohortes y momentos del tiempo. Para todo $t \neq t'$ y $g \neq g'$:

$$\mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty)|G_i = g] = \mathbb{E}[Y_{i,t}(\infty) - Y_{i,t'}(\infty)|G_i = g']$$

2. **Supuesto 2 - No anticipación:** Para todo $t < g$:

$$Y_{i,t}(g) = Y_{i,t}(\infty)$$

Nota: Tendencias paralelas puede debilitarse a tendencias paralelas condicionales si se incluyen variables pre-determinadas.

El problema del modelo dinámico

Siguiendo a Sun & Abraham (2021), los parámetro del modelo dinámico no recuperan los efectos dinámicos por dos razones:

- El coeficiente β_r puede asignarle a algunas unidades pesos negativos al efecto tratamiento r periodos luego del tratamiento.
- El coefficiente β_r puede asignarle a algunas unidades pesos postivos a lo que sucede en otros leads/lags $r' \neq r$, causando contaminación cruzada.

Notación:

- Dos individuos que son tratados al mismo tiempo, pertenecen a la misma cohorte g. El efecto del tratamiento l periodos después del tratamiento (cohort-specific average treatmente effect) se define como

$$\text{CATT}_{gl} = \mathbb{E}[Y_{i,g+l}(g) - Y_{i,g+l}(\infty)|G_i = g] = \text{ATT}(g, g+l)$$

- Sea $\mathcal{B}$ una colección disjunta de conjuntos $\{b_i\}_{i=1}^{n}$ de periodos relativos:

$$\mathcal{B} = \dot{\bigcup_{1 \leq i \leq n}} b_i \subseteq [-T, T] \cap \mathbb{Z}$$

Especificaciones usuales del modelo dinámico

- Usualmente los investigadores agrupan ("bin") o eliminan algunos periodos relativos ("trim"). Por ende, el modelo dinámico a estimar se ecribe de manera general como:

$$Y_{it} = \alpha_i + \phi_t + \sum_{b \in \mathcal{B}} \beta_b \mathbb{1}\{t - t_i \in b\} + \varepsilon_{it}$$

- Notemos por $b^{excl} = \{l \in [-T, T] : l \notin \mathcal{B}, l \in \mathbb{Z}\}$ los periodos omitidos de la estimación.

Ejemplos:

- **Fully dynamic specification:**

$$Y_{it} = \alpha_i + \phi_t + \sum_{l=-K}^{-2} \beta_l D_{i,t}^l + \sum_{l=0}^{L} \beta_l D_{i,t}^L + \nu_{i,t}$$

donde $b^{excl} = \{-T, \cdots, -K-1, -1, K, L+1, \cdots, T\}$

- **Binned+Trimmed specification:**

$$Y_{it} = \alpha_i + \phi_t + \tau \sum_{l<-K} D_{i,t}^l + \sum_{l=-K}^{-2} \beta_l D_{i,t}^l + \sum_{l=0}^{L} \beta_l D_{i,t}^L + \gamma \sum_{l>L} D_{i,t}^l + \nu_{i,t}$$

donde $b^{excl} = \{-1\}$ (Cuando no hay grupo nunca tratado, se deben omitir al menos dos periodos Borusyak and Jaravel (2017))

¿Entonces qué son los β_b?

Descomposición de Sun & Abraham (2021):

Bajo los supuestos de tendencias paralelas y no anticipación:

$$\beta_b = \underbrace{\sum_{l' \in b, l' \geq 0} \sum_{g} \omega^b_{g,l'} \text{CATT}_{g,l'}}_{\text{Deseable-Promedio de los efectos dinámicos en la caja } b}$$

$$+ \underbrace{\sum_{b' \neq b, b' \in \mathcal{B}} \sum_{l' \in b', l' \geq 0} \sum_{g} \omega^b_{g,l'} \text{CATT}_{g,l'}}_{\text{Indeseable - Contaminación de otras cajas}}$$

$$+ \underbrace{\sum_{l' \in b^{excl}, l' \geq 0} \sum_{g} \omega^b_{g,l'} \text{CATT}_{g,l'}}_{\text{Indeseable -Contaminación por periodos omitidos}}$$

Propiedades de la descomposición

Caso	**Pesos**
Periodos dentro de una misma "caja"	$\sum_{l' \in b, l' \geq 0} \sum_{g} \omega^{b}_{g,l'} = 1$
Periodos dentro de otra "caja" b':	$\sum_{l' \in b', l' \geq 0} \sum_{g} \omega^{b}_{g,l'} = 0$
Para periodos excluidos	$\sum_{l' \in b^{excl}, l' \geq 0} \sum_{g} \omega^{b^{excl}}_{g,l'} = -1$

Entonces

- β_b puede no preservar el signo de los efectos causales. (Invalida validación de tendencias paralelas e interpretación de efectos dinámicos)
- Si todos los elementos de una misma caja tienen el mismo efecto ($\text{CATT}_{g,l'} = \text{CATT}$ para todo $l' \in b$ y $g \neq \infty$), entonces el segundo término desaparece pues los pesos suman cero. Se logra, por ejemplo, si las cajas son un periodo y hay efectos homogéneos.
- En el caso "fully-dynamic" con efectos homogéneos ($\text{CATT}_{g,l'} = \text{CATT}_{l'} = \text{ATT}_{l'}$ para todo l'), entonces

$$\beta_l = \text{ATT}_l + \sum_{l' \in b^{excl}} \omega^{l}_{l'} \, \text{ATT}_{l'}$$

- Podemos eliminar el efecto de los periodos excluidos asegurándonos de excluir sólo periodos donde el efecto de tratamiento es cero.

Ejemplo Mickey-Mouse

Panel balanceado, $T = 2$, con cohortes $G_i \in \{1, 2\}$. Excluimos dos periodos para evitar multicolinealidad. Se incluyen los periodos $\{-2, 0\}$ y omitimos $\{-1, 1\}$. Entonces:

$$\mu_{-2} = \underbrace{\text{CATT}_{2,-2}}_{\text{Periodo de interés}} + \underbrace{\frac{1}{2}\text{CATT}_{1,0} - \frac{1}{2}\text{CATT}_{2,0}}_{\text{Contaminación por otras cajas}}$$

$$+ \underbrace{\frac{1}{2}\text{CATT}_{1,1} - \text{CATT}_{1,-1} - \frac{1}{2}\text{CATT}_{2,-1}}_{\text{Contaminación por periodos excluidos}}$$

Note

- No anticipación elimina el quinto y sexto término,
- Si hubiese homogeneidad entre cohortes, el segundo y tercer término se cancelan.
- Pero queda $\frac{1}{2}\text{CATT}_{1,1}$- Esto es un problema porque excluimos un grupo con efecto de tratamineto.

Moraleja: Sólo excluir leads. Si se excluyen los lags, se contamina el efecto a través de los periodos excluidos.

The intraction-weighted estimator

Es un estimdor en tres etapas:

1) El estimador se base en estimar por MCO el modelo de regresión saturado, excluyendo las indicadores de cohortes de un conjunto C:

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{g \notin C} \sum_{l \neq -1} \delta_{g,l} \left(\mathbb{1}[G_i = g] \times D^l_{i,t} \right) + \epsilon_{i,t}$$

donde $C = \{\infty\}$ cuando hay unidades nunca tratadas y $C = \text{máx}\{G_i\}$ de lo contrario. El coeficiente $\hat{\delta}_{g,l}$ es un estimador de $\text{CATT}_{g,l}$.

2) Estimar los tamaños de los grupos $\hat{\mathbb{P}}(G_i = g | G_i \in [-l, T - l])$.

3) Calcular el efecto para la caja b usando:

$$\hat{\beta}_b = \frac{1}{|b|} \sum_{l \in b} \sum_{g} \hat{\delta}_{g,l} \, \hat{\mathbb{P}}(G_i = g | G_i \in [-l, T - l])$$

Wooldridge (2021), WP

- Estimar por MCO un modelo de regresión lineal con efectos fijos de unidad y tiempo más flexible corrije el problema descrito por otros autores.
- Permite recuperar efectos causales en presencia de heterogeneidad entre cohortes y de efectos dinámicos. Se pueden estimar tanto efectos causales dinámicos como entros niveles de agregación.
- Permite validar el supuesto de tendencia paralelas.
- Permite incluir controles que no varían en el tiempo.
- Sólo es válido para el cuando el tratamiento es absorbente y el tratamiento binario.
- Es válido tanto para datos panel como para cortes transversales. De hecho, el autor argumenta que este es la aproximación más robusta a algunos problemas de datos faltantes y paneles no balanceados.
- El Teorema de Gauss-Markov hace que los estimadores sean lo más eficiente posibles. Sin embargo, los estimadores dependen del modelo dado a la función de esperanza condicional. Si esto no se cumple, el estimador puede producir un sesgo mayor al de Callaway and Sant'Anna (2021) o el de Chaisemartin and D'Haultfœuille (2020).
- Permite calcular efectos heterogéneos.(Extiende a Sun & Abraham (2021))
- No hay paquetes todavía.

Supuestos de identificación

- **No anticipación condicional:** Para todas las cohortes tratadas $g = q, \cdots, T$:

$$\mathbb{E}[y_t(g) - y_t(\infty)|G_i = g, X] = 0, \ \forall t < g$$

- **Tendencias paralelas condicionales:**

$$\mathbb{E}[y_t(\infty) - y_t(\infty)|G_i = g, X] = \mathbb{E}[y_t(\infty) - y_t(\infty)|X], \ \forall t = 2, \cdots, T$$

- **CEF lineal:** Para variables x_i, la CEF $\mathbb{E}[y_{it}(\infty)|x_i]$ es lineal en los parámetros. (Pero en general uno puede proponer cualquier modelo paramétrico arbitrario)

La idea general: Es un problema de flexibilidad del modelo

La propuesta de Wooldridge (2021), WP se resume en la siguiente frase:

"The important takeaway from the analysis is that there is nothing inherently wrong with TWFE: we simply need to apply the estimation method to equations that allow sufficient heterogenety in the treatment effects."

El método de Wooldridge se basa en una equivalencia algebraica que se remonta a Mundlak (1978): El estimador de efectos fijos de la regresión y_{it} on $1, c_2, \cdots, c_N, x_{it}$ es equivalente a un estimador de mínimos cuadrados generalizados de y_{it} on $1, \bar{x}_{i.}, x_{it}$, donde

$$\bar{x}_{i.} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$$

Esta última ecuación se conoce como la regresión de Mundlak. Esto se puede extender al caso de dos efectos fijos: "Two-way Mundlak regression."

Two-way Mundlak regression.

Two-way Mundlak regression.

Para un panel de dimensión $N \times T$ y algunas condiciones de no-multicolinealiadad perfecta, sea $\hat{\beta}_{FE}$ el estimador por efectos fijos del modelo:

$$y_{it} = \beta x_{it} + c_i + f_t + \epsilon_{it}$$

y sea $\hat{\beta}_M$ el estimador por MCO del modelo

$$y_{it} = \beta x_{it} + \gamma \bar{x}_{i.} + \psi \bar{x}_{.t} + \epsilon_{it}$$

entonces $\hat{\beta}_{FE} = \hat{\beta}_M$.

Corolario:

- El resultado no cambia si se agregan controles invarientes en el tiempo o entre unidades. Ejemplo: Si corremos el modelo

$$y_{it} = \beta x_{it} + \gamma \bar{x}_{i.} + \psi \bar{x}_{.t} + \xi z_i + \theta m_t + \epsilon_{it}$$

entonces el beta resultante es igual al del teorema.

Caso Mickey-Mouse 2: Modelo de Diferencias en Diferencias tradicional

Consideremos el caso estándar donde hay un dos grupos y T periodos de tiempo. Notemos por d_i la dummy de pertenecer al grupo de tratamiento y p_t una dummy de estar en un periodo post-tratamiento. Entonces el modelo usual de diferencias en diferencias de TWFE es:

$$y_{it} = \alpha + \beta w_{it} + c_i + g_t + u_{it}$$

donde w_{it} es una dummy de si el individuo i se encuentra tratado en el tiempo t. Claramente:

$$w_{it} = d_i \times p_i$$

Note que el teorema de Mundlak dice que el $\hat{\beta}_{FE}$ es equivalente a aquel de correr

$$y_{it} = \alpha + \beta w_{it} + \gamma \bar{w}_{i.} + \psi \bar{w}_{.t} + u_{it}$$

Mundlak y DiD

Pero:

$$w_{i.} = \frac{1}{T}\sum_{t=1}^{T} w_{it} = \frac{1}{T}\sum_{t=1}^{T} d_i \times p_i = d_i \bar{p}$$

y similarmente

$$w_{.t} = \frac{1}{N}\sum_{i=1}^{N} w_{it} = \frac{1}{N}\sum_{i=1}^{N} d_i \times p_i = \bar{d} p_t$$

Entonces el modelo de Mundlak es

$$\begin{aligned} y_{it} &= \alpha + \beta w_{it} + \gamma \bar{x}_{i.} + \psi \bar{x}_{.t} + u_{it} \\ &= \alpha + \beta w_{it} + \gamma \bar{p} d_i + \psi \bar{d} p_t + u_{it} \\ &= \alpha + \beta w_{it} + \tilde{\gamma} d_i + \tilde{\psi} p_t + u_{it} \end{aligned}$$

que es el modelo de diferencias en diferencias. Entonces

$$\hat{\beta}_{FE} = \hat{\beta}_M = \hat{\beta}_{DiD} \xrightarrow{P} ATT$$

Mundlak y DiD: ¿Y si tenemos controles?

Consideremos el modelo por TWFE más flexible posible:

$$y_{it} = \sum_{s=q}^{T} \beta_s w_{it} \mathbb{1}[s=q] + \sum_{s=q}^{T} \gamma_s w_{it} (\boldsymbol{x}_i - \mu_1) \mathbb{1}[s=q] + \sum_{s=q}^{T} \gamma_s (\boldsymbol{x}_i - \mu_1) \mathbb{1}[s=q] + c_i + g_t + u_{it}$$

donde $\mu_1 = \mathbb{E}[\boldsymbol{x}_i | d_i = 1]$. El análogo de Mundlak es

$$y_{it} = \alpha + \psi d_i + \varphi \boldsymbol{x}_i + \phi(d_i \times \boldsymbol{x}_i) + \sum_{s=q}^{T} \delta_s \mathbb{1}[s=q] +$$
$$\sum_{s=q}^{T} \gamma_s \boldsymbol{x}_i + \sum_{s=q}^{T} \beta_s w_{it} \mathbb{1}[s=q] + \sum_{s=q}^{T} \gamma_s w_{it} (\boldsymbol{x}_i - \mu_1) + u_{it}$$

¿Y esto de qué nos ayuda?

Resulta que bajo Tendencias Paralelas, No anticipación y Linealidad, **la única forma funcional posible de la CEF es**:

$$\mathbb{E}[y_{it}|\boldsymbol{x}_i] = \alpha + \psi d_i + \varphi \boldsymbol{x}_i + \phi(d_i \times \boldsymbol{x}_i) + \sum_{s=q}^{T} \delta_s \mathbb{1}[s = q] +$$

$$\sum_{s=q}^{T} \gamma_s \boldsymbol{x}_i + \sum_{s=q}^{T} \beta_s w_{it} \mathbb{1}[s = q] + \sum_{s=q}^{T} \gamma_s w_{it}(\boldsymbol{x}_i - \mu_1)$$

que es la ecuación de Mundlak. Luego estimaríamos correctamente la CEF por TWFE **si, y solo si** usamos la expresión dual en términos de efectos fijos:

$$y_{it} = \sum_{s=q}^{T} \beta_s w_{it} \mathbb{1}[s = q] + \sum_{s=q}^{T} \gamma_s w_{it}(\boldsymbol{x}_i - \mu_1)\mathbb{1}[s = q] + \gamma_s(\boldsymbol{x}_i - \mu_1) + c_i + g_t + u_{it}$$

Además, por construcción, estimar por efectos fijos este modelo (y reemplazar μ_1 por su análogo muestral) resulta en los parámetros correctos:

$$\hat{\beta}_s \xrightarrow{P} \text{“El efecto de tratamiento en el tiempo } s\text{''}$$

Entonces, si omitiéramos algunos de los términos, o no permitimos la suficiente heterogeneidad, el modelo estadístico estaría mal por definición.

¿Y en event studies?

Bajo los supuestos, sabemos que la **única** forma de la CEF es

$$E(y_{it}|d_{iq},\ldots,d_{iT},\mathbf{x}_i) = \eta + \sum_{r=q}^{T}\lambda_r d_{ir} + \mathbf{x}_i\boldsymbol{\kappa} + \sum_{r=q}^{T}(d_{ir}\cdot\mathbf{x}_i)\boldsymbol{\zeta}_r + \sum_{s=2}^{T}\theta_s fs_t + \sum_{s=2}^{T}(fs_t\cdot\mathbf{x}_i)\boldsymbol{\pi}_s \qquad (6.33)$$
$$+ \sum_{r=q}^{T}\sum_{s=r}^{T}\tau_{rs}(w_{it}\cdot d_{ir}\cdot fs_t) + \sum_{r=q}^{T}\sum_{s=r}^{T}(w_{it}\cdot d_{ir}\cdot fs_t\cdot\dot{\mathbf{x}}_{ir})\boldsymbol{\rho}_{rs},$$

donde $\dot{\boldsymbol{x}}_i = \boldsymbol{x}_i - \mu_i$.

Luego basta estimar su "dual" en TWFE: Todas las interacciones entre efectos fijos de cohortes, efecto fijo de tiempo, variables de control y la variable de tratamiento w_{it}. (Muy grande para siquiera escribirlo)

¿Y qué estimamos?

Observación: En el caso sin controles, el modelo correcto de TWFE es el de Sun and Abraham (2021):

$$Y_{i,t} = \alpha_i + \lambda_t + \sum_{g \notin C} \sum_{l \neq -1} \delta_{g,l} \left(\mathbb{1}[G_i = g] \times D_{i,t}^{l} \right) + \epsilon_{i,t}$$

luego los $\delta_{g,l}$ son los $ATT(g, g+l)$. Luego podemos agregar estos efectos de alguna manera que tenga sentido como en Callaway & Sant'Anna.

Dos moralejas:

- ¡Estimar por efectos fijos está bien si especificamos bien el modelo! Esto es Sun & Abraham (2021)
- Pero esto es mucho más general, porque podemos capturar efectos heterogéneos.

Gracias!