

# Diferencias en Diferencias

Gustavo A. García

[ggarci24@eafit.edu.co](mailto:ggarci24@eafit.edu.co)

Econometría II

Programa de Economía

Universidad EAFIT

Link slides formato **html**

Link slides formato **PDF**

## En este tema

- La evaluación de impacto
- Evaluación prospectiva versus evaluación retrospectiva
- Inferencia causal y contrafactuales
- Métodos de evaluación de impacto
- Diferencias en diferencias
- Ejemplo 1 del modelo 2x2 DD
- Ejemplo 2 del modelo 2x2 DD
- El supuesto de tendencias paralelas
- Ejercicio 1 aplicado en R: efecto de la formación en el empleo sobre los salarios
- Ejercicio 2 aplicado en R: efecto de Empleando Futuro sobre la formalidad

# Lecturas

- Gertler, P. J., Martínez, S., Premand, P., Rawlings, L. B., Vermeersch, C. M. (2011). *La evaluación de impacto en la práctica*. Banco Mundial.
- Roth, J., Sant'Anna, P., Bilinski, A., Poe, J. (2023). *What's trending in difference-in-differences? A synthesis of the recent econometrics literature*, *Journal of Econometrics*, 235: 2218-2244.
- Wing, C., Simon, K., Bello-Gomez, R. (2018). *Designing Difference in difference studies: best practices for public health policy research*, *Annual Review of Public Health*, 39: 453-469.
- Algunas webs: [DS4PS](#), [Tidy Finance](#), [RE\\_STAT](#), [DiDs in R](#)

# La evaluación de impacto

- La evaluación de impacto sirve para determinar **si un programa ha logrado o no los resultados previstos**, y para explorar si existen estrategias alternativas con las que alcanzar mejor dichos resultados
- Las evaluaciones de impacto forman parte de un programa más amplio de **formulación de políticas basadas en evidencias**. Esta tendencia mundial creciente se caracteriza por un cambio de enfoque: **de los insumos a los resultados**
- En términos sencillos, una evaluación de impacto **evalúa los cambios en el bienestar de las personas que pueden atribuirse a un proyecto, programa o política particular**
- El principal desafío para llevar a cabo evaluaciones eficaces de impacto es identificar **la relación causal entre el proyecto**, el programa o la política y los resultados de interés
- Las evaluaciones de impacto son un tipo particular de evaluación que intenta responder a preguntas sobre **causa y efecto**
- Las evaluaciones de impacto se preocupan por saber cuál es el **impacto o efecto causal** de un programa sobre un resultado de interés. Solo interesa el impacto del programa: **el efecto directo que tiene en los resultados**

# La evaluación de impacto

- Para estimar el efecto causal o el impacto de un programa sobre los resultados, cualquier método elegido debe estimar el denominado **contrafactual**  $\implies$  **cuál habría sido el resultado para los participantes en el programa si no hubieran participado en él**
- La evaluación de impacto requiere que el evaluador encuentre **un grupo de comparación** para estimar lo que habría ocurrido con los participantes sin el programa
- La pregunta fundamental de una evaluación de impacto puede formularse así:

**¿Cuál es el impacto (o efecto causal) de un programa sobre un resultado de interés?**

- Por ejemplo:
  - ¿Cuál es el efecto causal de las becas sobre la asistencia escolar y el logro académico?
  - ¿Cuál es el impacto de la formación para trabajo sobre la formalidad?
  - ¿La mejora de las carreteras mejora el acceso a mercados laborales y aumenta el ingreso de los hogares?
  - ¿Incrementos en el salario mínimo tienen efectos negativos o positivos sobre el empleo?

# Evaluación prospectiva versus evaluación retrospectiva

Las evaluaciones de impacto pueden dividirse en dos categorías: **prospectivas** y **retrospectivas**

## Evaluaciones prospectivas

- Se realizan al mismo tiempo que se diseña el programa y forman parte de la implementación del programa
- Para ello, se recolectan datos de **línea de base de los grupos de tratamiento y de comparación** antes de la implementación del programa

## Evaluaciones retrospectivas

- Examinan el impacto del programa después de su implementación, lo que genera grupos de tratamiento y de comparación ex post

En general, las evaluaciones prospectivas de impacto tienen más probabilidades de producir resultados más sólidos y creíbles, por tres motivos:

1. Levantamiento de una línea base que permite tener información antes de la intervención e identificar que los grupos de tratamiento y comparación sean similares
2. Aclarar los objetivos del programa e identificar la medida de éxito
3. Los grupos de tratamiento y de comparación se identifican antes de la implementación del programa

# Inferencia causal y contrafactuales

## Inferencia causal

- La pregunta básica de la evaluación de impacto es esencialmente un problema de **inferencia causal**
- Evaluar el impacto de un programa sobre una serie de resultados es equivalente a **evaluar el efecto causal del programa sobre ellos**
- Aunque las preguntas sobre causa y efecto son habituales, determinar que **una relación es causal no es tan sencillo**
- Las evaluaciones de impacto nos ayudan a **atribuir causalidad al establecer empíricamente en qué medida cierto programa, y solo ese programa, ha contribuido a cambiar un resultado**
- Para atribuir causalidad entre un programa y un resultado se usan **los métodos de evaluación de impacto, que descartan la posibilidad de que cualquier factor distinto del programa de interés explique el impacto observado**
- La respuesta a la pregunta básica de la evaluación de impacto, **¿Cuál es el impacto o efecto causal de un programa  $P$  sobre un resultado de interés  $Y$ ?**, se obtiene mediante la fórmula básica de la evaluación de impacto:

$$\alpha = (Y|P = 1) - (Y|P = 0)$$

- Esta formula implica medir  $Y$  en un momento para la misma unidad de observación (persona, firma, región...), pero en dos realidades diferentes, lo cual no es posible



# Inferencia causal y contrafactuales

## Contrafactuales

- Como se mencionó, es imposible medir a la misma persona en dos situaciones diferentes en el mismo momento: una persona habrá participado en un programa, o no lo habrá hecho; no se la puede observar simultáneamente en dos estados diferentes
- Este es el problema contrafactual: ¿cómo se mide lo que habría ocurrido si hubiera prevalecido la otra circunstancia?
- Aunque se puede observar y medir el resultado ( $Y$ ) para los participantes en el programa ( $Y|P = 1$ ), no se cuenta con datos para establecer cuál habría sido el resultado en ausencia del programa ( $Y|P = 0$ )
- En la fórmula básica de la evaluación de impacto, la expresión ( $Y|P = 0$ ) representa el contrafactual, que puede entenderse como lo que habría pasado si un participante no hubiera participado en el programa  $\implies$  El contrafactual es lo que habría sido el resultado ( $Y$ ) en ausencia de un programa ( $P$ )
- ( $Y|P = 0$ ) no se puede observar directamente, y de ahí la necesidad de completar esta información mediante la estimación del contrafactual
- Para ello se recurre habitualmente a grupos de comparación o grupos de control
- La identificación de esos grupos de comparación es el quid de cualquier evaluación de impacto  $\implies$  sin una estimación válida del contrafactual, no se puede conocer el impacto de un programa

# Inferencia causal y contrafactuales

## Contrafactuales

- Un objetivo clave de una evaluación de impacto es identificar a un grupo de participantes en el programa (el grupo de tratamiento) y a un grupo de no participantes (el grupo de comparación) estadísticamente idénticos en ausencia del programa
- Si los dos grupos son iguales, a excepción de que uno de ellos participa en el programa y el otro no, cualquier diferencia en los resultados deberá provenir del programa
- El reto principal es identificar a un grupo de comparación con las mismas características que el grupo de tratamiento. Al menos en tres aspectos deben ser iguales:
  - ambos grupos deben ser idénticos en ausencia del programa
  - los grupos deben reaccionar de la misma manera al programa
  - ambos grupos no pueden estar expuestos de manera diferente a otras intervenciones durante el período de la evaluación
- Cuando el grupo de comparación para una evaluación es inválido, la estimación del impacto del programa también será inválida: no estimará el impacto real del programa. En términos estadísticos, estará sesgada

# Métodos de evaluación de impacto

Los métodos o estrategias en la evaluación de impacto consisten en **identificar grupos de comparación válidos que reproduzcan o imiten exactamente el contrafactual**

Los métodos de evaluación de impacto más comúnmente utilizados son:

- asignación aleatoria
- promoción aleatoria
- diseño de regresión discontinua
- diferencias en diferencias (DD)
- pareamiento

Los tres primeros métodos estiman el contrafactual a través de reglas explícitas de **asignación del programa**, que el evaluador conoce y entiende. Estos métodos ofrecen **estimaciones creíbles del contrafactual haciendo relativamente pocas suposiciones e imponiendo pocas condiciones**.

Los dos últimos métodos, ofrecen al evaluador herramientas adicionales que pueden aplicarse cuando las **reglas de asignación del programa sean menos claras o cuando no sea factible ninguno de los tres métodos antes descritos**

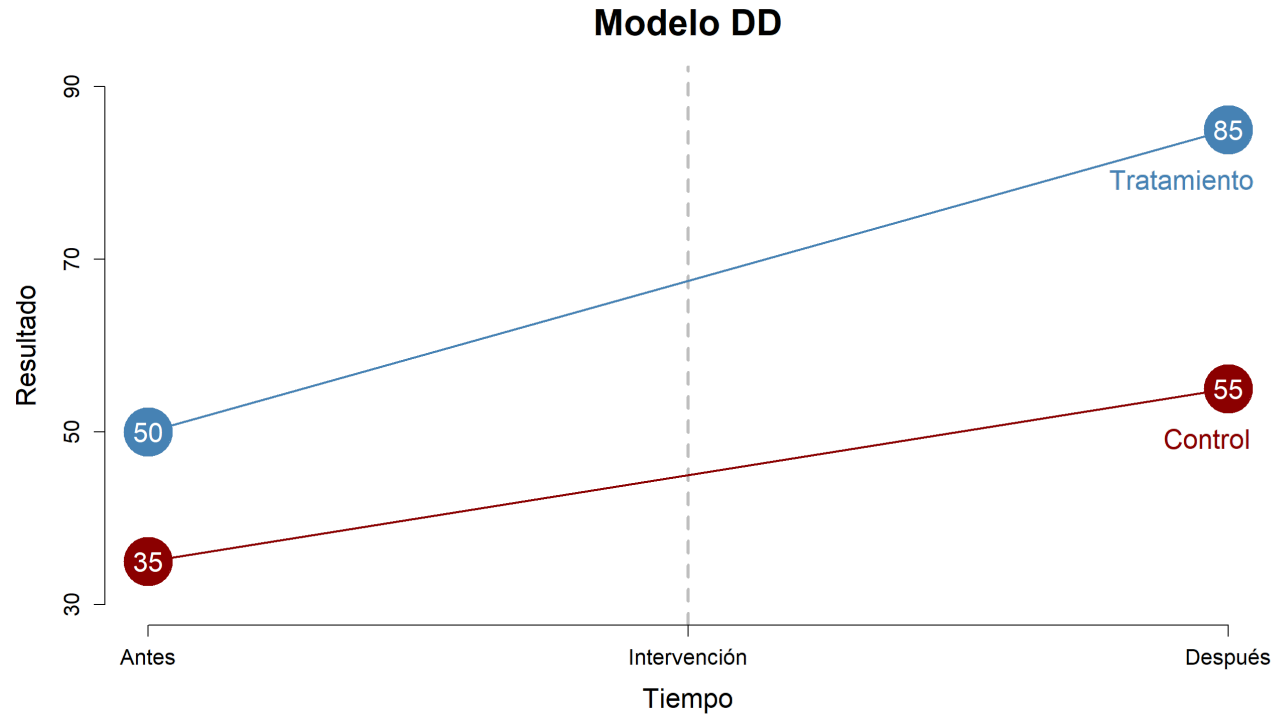
# Diferencias en diferencias

- El método de DD consiste, como indica su nombre, en aplicar una doble diferencia. Compara los cambios a lo largo del tiempo en la variable de interés entre una población inscrita en un programa (el grupo de tratamiento) y una población no inscrita (el grupo de comparación)
- DD estima el contrafactual del cambio en el resultado para el grupo de tratamiento calculando el cambio del resultado para el grupo de comparación. Este método nos permite tener en cuenta cualquier diferencia constante en el tiempo entre los grupos de tratamiento y de comparación
- Para aplicar DD, solo hace falta medir los resultados del grupo que recibe el programa (el grupo de tratamiento) y del grupo que no lo recibe (el grupo de comparación) antes y después del programa

# Diferencias en diferencias

El estimador DD considera:

- cómo cambia la media de un grupo antes y después de una intervención política (grupo de tratamiento)
- comparar éste cambio con la media a lo largo del tiempo de un grupo similar que no se sometió al tratamiento (grupo de control)



# Diferencias en diferencias

Como sugiere su nombre, el estimador DD es la **diferencia de sus diferencias medias**. Más claramente, el estimador DD toma:

- la diferencia en el grupo de tratamiento antes y después del tratamiento (el efecto del tratamiento) y resta
- la diferencia en el grupo de control antes y después del tratamiento (la tendencia en el tiempo) como en la siguiente fórmula:

$$(\text{Tratamiento después} - \text{Tratamiento antes}) - (\text{Control después} - \text{Control antes}) = \text{estimación DD}$$
$$(85 - 50) - (55 - 35) = 15$$

Este efecto es lo que en la literatura de evaluación de impacto se llama el **efecto medio del tratamiento sobre los tratados** o **Average Treatment Effect on Treated (ATT)**

En la siguiente tabla puede verse un cálculo más detallado:

	Después	Antes	Diferencia
Tratamiento	85	50	35
Control	55	35	20
Diferencias	30	15	15

# Diferencias en diferencias

- Se podría estar tentado a pensar que la primera diferencia (Tratamiento después – Tratamiento antes) o la comparación de tratados y no tratados, pueden medir el efecto del programa
- La diferencia de los resultados antes-después del grupo tratado (la primera diferencia) considera factores constantes en el tiempo para dicho grupo, ya que se compara al grupo con sí mismo. Sin embargo, todavía nos quedan los factores externos que varían con el tiempo
- Una manera de observar estos factores variables en el tiempo es medir el cambio antes-después de los resultados de un grupo que no se trató, pero estuvo expuesto a las mismas condiciones ambientales (la segunda diferencia)
- Si se “limpia” la primera diferencia de otros factores variables en el tiempo que afectan al resultado de interés sustrayéndole la segunda diferencia, se habrá eliminado la principal causa de sesgo que nos preocupaba en el caso de las comparaciones simples de antes-después.
- Por lo tanto, el método de DD combinaría los dos falsos contrafactuales (comparaciones antes-después y comparaciones tratados-no tratados) para generar una mejor estimación del contrafactual

# Diferencias en diferencias

## El clásico 2x2 DD

El modelo clásico 2x2 DD es la forma más simple del diseño DD y es el caso especial en el cual hay sólo dos grupos observados en dos periodos de tiempo, y cómo vimos es fácilmente representable en la tabla anterior 2x2

El modelo clásico 2x2 DD tiene la siguiente estructura:

$$y_{it} = \beta_1 + \beta_2 \text{Tratamiento}_i + \beta_3 \text{Post}_{it} + \beta_4 \text{Tratamiento}_i * \text{Post}_{it} + \epsilon_{it}$$

donde

$y$  : variable resultado

$\text{Tratamiento}$  : variable dummy indicando el grupo de tratamiento (= 1) y control (= 0)

$\text{Post}$  : variable dummy indicando pre-tratamiento (= 0) y el pos-tratamiento (= 1)

$\text{Tratamiento} * \text{Post}$  : variable dummy indicando si la variable de resultado se observó en el grupo de tratamiento y esto fue observado después de la intervención (= 1) o en cualquier otro caso (= 0)



# Diferencias en diferencias

## Múltiples grupos y periodos de tiempo (Two-way Fixed Effects - TWFE)

El diseño DD de dos grupos-dos periodos es muy intuitivo, pero no se ajusta adecuadamente a la realidad, donde pueden haber tratamientos en múltiples grupos y múltiples periodos de tiempo

El [modelo general DD](#) con múltiples tratamientos y periodos de tiempo sigue la forma general:

$$y_{it} = \alpha_i + \gamma_t + \beta D_{it} + \epsilon_{it}$$

donde  $\alpha_i$  y  $\alpha_t$  son efectos fijos de individuo y tiempo, y  $D_{it} = \text{Tratamiento}_i * \text{Post}_{it}$

A este modelo le llaman el *Two-way Fixed Effects* (TWFE) al incluir doble efecto fijo asociado a  $i$  y  $t$ . Note que las variables *Tratamiento* y *Post* individuales del 2x2 DD se subsumen por la estructura de los efectos fijos

La versión dinámica del TWFE para medir el efecto del tratamiento en el tiempo, es la especificación [TWFE event study](#):

$$y_{it} = \beta_1 + \sum_{k=T_0}^T \beta_k (\text{Tratamiento}_i * t_k) + \alpha_i + \gamma_t + \epsilon_{it}$$

donde  $t_k$  es una variable binaria para cada año  $k$  y la intervención empieza en  $T_0$  para el grupo de tratados

# Ejemplo 1 del modelo 2x2 DD

Supongamos que se tiene la siguiente información donde dos individuos son tratados y dos son no tratados, y se registra información sobre la variable de resultado antes y después de la intervención:

Individuo	Y	Tratamiento	Post	Trat*Post
1	74	1	1	1
1	46	1	0	0
2	96	1	1	1
2	54	1	0	0
3	50	0	1	0
3	30	0	0	0
4	60	0	1	0
4	40	0	0	0

# Ejemplo 1 del modelo 2x2 DD

	Y
$\beta_1$	35.00 <sup>***</sup> (0.00)
Tratamiento	15.00 <sup>***</sup> (0.00)
Post	20.00 <sup>***</sup> (0.00)
Trat*Post	15.00 <sup>***</sup> (0.00)

Note:       \* $p < 0.1$ ; \*\* $p < 0.05$ ; \*\*\* $p < 0.01$

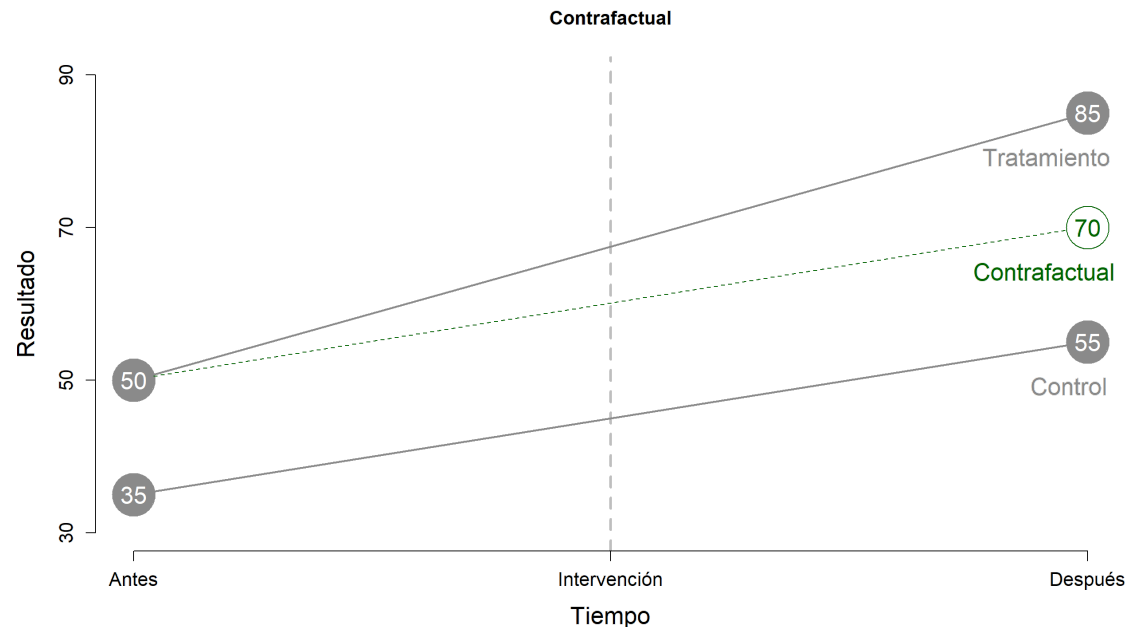
Errores estándar en ()

$\beta$	Interpretación
$\beta_1 = 35$	Promedio de $Y$ en el grupo de control antes de la intervención
$\beta_2 = 15$	Diferencia entre el tratamiento y el control antes de la intervención
$\beta_3 = 20$	Cambio en $Y$ del grupo de control después de la intervención
$\beta_4 = 15$	<b>Parámetro de interés.</b> Cambio en $Y$ del grupo tratado después de la intervención, en comparación con lo que le habría ocurrido al mismo grupo si no se hubiera producido la intervención

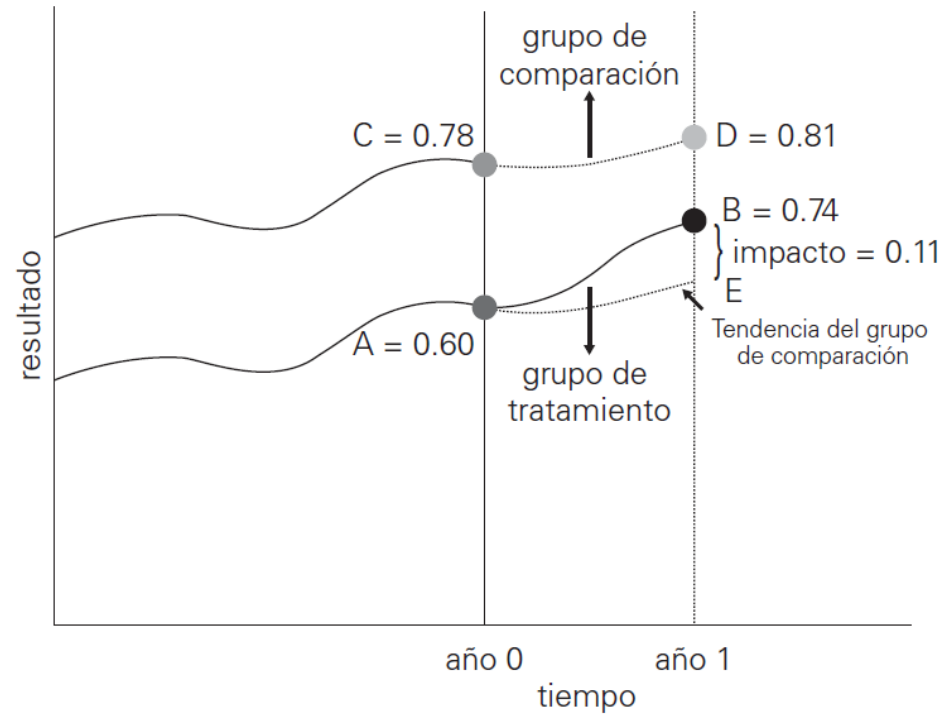
# Ejemplo 1 del modelo 2x2 DD

## El contrafactual

- Para entender mejor  $\beta_4$ , debemos examinar mejor el contrafactual. El contrafactual es lo que le habría ocurrido a  $Y$ , si no se hubiera producido la intervención
- En el modelo DD, el contrafactual es el resultado del grupo de intervención si no se hubiera producido la intervención
- $\beta_4$  representa la diferencia entre el resultado contrafactual y el resultado medio del grupo de tratamiento en el periodo posterior al tratamiento



## Ejemplo 2 del modelo 2x2 DD



Aquí los dos falsos contrafactuales serían:

- diferencia antes y después de la intervención en los tratados:  $(B - A) = 0.74 - 0.60 = 0.14$
- diferencia después de la intervección entre tratados y control:  $(B - D) = 0.74 - (-0.81) = 1.55$

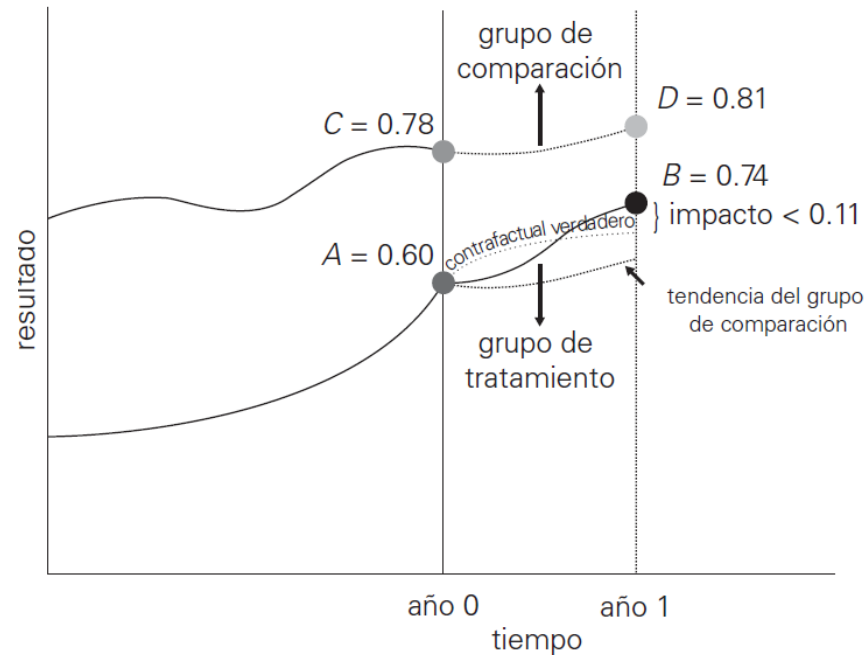
Con DD el impacto o efecto del programa está definido por:

$$(B - A) - (D - C) = (B - E) = (0.74 - 0.60) - (0.81 - 0.78) = 0.11$$

# El supuesto de tendencias paralelas

- El método de DD supone que muchas características de las unidades o las personas son constantes (o invariables en el tiempo)
- Para que DD genere una estimación válida del contrafactual, se debe suponer la inexistencia de diferencias que aparecen a lo largo del tiempo entre los grupos de tratamiento y de comparación
- Es decir que en ausencia del programa, las diferencias en los resultados entre los grupos de tratamiento y de comparación tendrían que evolucionar paralelamente  $\implies$  sin el tratamiento, los resultados aumentarían o disminuirían al mismo ritmo en ambos grupos; es decir, es necesario que los resultados reflejen tendencias iguales en ausencia del tratamiento
- Lamentablemente, no existe una manera de demostrar que las diferencias entre los grupos de tratamiento y de comparación habrían evolucionado paralelamente en ausencia del programa. Esto se debe a que no se puede observar lo que le habría pasado al grupo de tratamiento en ausencia del tratamiento  $\implies$  ¡no se puede observar el contrafactual!
- Por lo tanto, cuando se usa el método de DD, se debe suponer que, en ausencia del programa, el resultado del grupo de tratamiento habría evolucionado paralelamente al grupo de control

# El supuesto de tendencias paralelas



- Si las tendencias de los resultados son diferentes para los grupos de tratamiento y de comparación, **el efecto estimado de tratamiento obtenido mediante métodos de DD será inválido o sesgado**. Esto se debe a que **la tendencia del grupo de comparación no es una estimación válida de la tendencia del contrafactual**
- Como se ve en el gráfico, los resultados del grupo de comparación crecen más lentamente que los resultados del grupo de tratamiento en ausencia del programa
- El uso de la tendencia del grupo de comparación como estimación de la “tendencia contrafactual” del grupo de tratamiento provoca una sobreestimación del impacto del programa

# El supuesto de tendencias paralelas

A pesar de que no se puede demostrar el supuesto fundamental de igualdad de tendencias, se puede comprobar. En la literatura existen cuatro formas regularmente utilizadas para corroborar el supuesto:

## *Testing for pre-trends*

Una técnica comúnmente utilizada para probar tendencias paralelas antes del tratamiento, es el estudio de eventos (*event-study*)

La idea de este método consiste en incluir la interacción entre las variables binarias de tiempo antes y después de la fecha de la intervención y el indicador de tratamiento como variables explicativas de la variable resultado a analizar

Para comprobar la igualdad de las tendencias antes de la intervención, se necesitan al menos dos rondas de datos para los grupos de tratamiento y de comparación antes del comienzo del programa

Se estima el siguiente *event-study*, con  $t$  siendo una variable binaria de tiempo:

$$y_{it} = \beta_1 + \underbrace{\sum_{k=T_0^*}^{T_0-1} \theta_k (\text{Tratamiento}_i * t_k)}_{\text{Pretratamiento}} + \underbrace{\sum_{\tau=T_1}^T \delta_\tau (\text{Tratamiento}_i * t_\tau)}_{\text{Postratamiento}} + \alpha_i + \gamma_t + \epsilon_{it}$$

Para corroborar que la tendencia entre los grupos de tratamiento y control es la misma, se debe probar que los coeficientes asociados al pretratamiento no son estadísticamente significantes



# El supuesto de tendencias paralelas

## *Placebo test*

los investigadores pueden probar si se observó un efecto en una muestra diferente que no debería haberse visto afectada por la intervención

En esta prueba placebo,  $\theta$  es el cambio en el resultado en el grupo que no debería verse afectado por la intervención

Esto podría implicar el uso del mismo modelo de DD en una población diferente, y ver si hay un efecto medio significativo del tratamiento

# Ejercicio 1 aplicado en R: efecto de la formación en el empleo sobre los salarios

La *National Supported Work* (NSW) en los Estados Unidos, llevó acabo en los 70 un análisis sobre el impacto de la formación en el empleo sobre los salarios. El programa fue realizado entre 1976 y 1977. Hay 185 tratados y un grupo de control de 2490. La variable de resultado a analizar son los salarios reales, y se cuenta con información en el pretratamiento en 1974 y 1975, y en el postratamiento en 1978 (Lalonde, 1986; Dehejia y Wahba, 1999, 2002). La base de datos [nswpsid.dta](#) cuenta con las siguientes principales variables:

Variable	Definición
treat	1=tratados, 0=control
age	Edad en años
educ	Número de años de educación
black	1=afrodescendiente, 0=no afro
re74	Salario real en 1974
re75	Salario real en 1975
re78	Salario real en 1978

En el siguiente link se encuentra el código utilizado en R:

- [Código en R](#)

## Ejercicio 2 aplicado en R: efecto de Empleando Futuro sobre la formalidad

En este ejercicio se va a realizar una aplicación del modelo de Diferencias en Diferencias (DD) para analizar el efecto causal de **Empleando Futuro**. Este programa fue el primer **Bono de Impacto Social (BIS)** implementado en Colombia y Latinoamérica, y tenía como objetivo fomentar la inserción laboral y retención de la población vulnerable.

Empleando Futuro se ejecutó entre marzo de 2017 y diciembre de 2018 en Cali, Pereira y Bogotá. Este ejercicio se basa en el documento de **Chaparro, García y Cardona (2020)** donde se encuentra más detallada la evaluación de impacto de Empleando Futuro.

Todo el ejercicio, junto con el código, se encuentra en el siguiente link:

- **Empleando Futuro**