

"Evaluación de programas públicos de formación a través de métodos observacionales: el estimador de diferencias en diferencias. Un ejercicio de simulación"

José Manuel Cansino Muñoz-Repiso¹

(Universidad de Sevilla y Centra)

Antonio Sánchez Braza²

(Universidad de Sevilla)

Draft paper

(No citar sin permiso de los autores)

¹ Dirección para comentarios: Departamento de Teoría Económica y Economía Política. Facultad de CC. Económicas y Empresariales. Universidad de Sevilla. Avda. Ramón y Cajal, 1. 41018 SEVILLA. Correo electrónico: jmcansino@us.es. Tfno : + 34 954 55 75 28. Fax : + 34 954 55 76 29.

² Dirección para comentarios: Departamento de Teoría Económica y Economía Política. Facultad de CC. Económicas y Empresariales. Universidad de Sevilla. Avda. Ramón y Cajal, 1. 41018 SEVILLA. Correo electrónico: asb@us.es. Tfno : + 34 954 55 75 29. Fax : + 34 954 55 76 29.

RESUMEN:

El estimador de diferencias en diferencias es un método de estimación de la inferencia causal estadística apropiado en el contexto de estudios observacionales. Su utilización en la evaluación de programas públicos de formación ha mostrado resultados ciertamente interesantes.

Este trabajo desarrolla un ejercicio de evaluación basado en la utilización del estimador de diferencias en diferencias sobre unos datos ficticios de la variable respuesta (ingresos laborales), generados y asignados aleatoriamente entre individuos beneficiarios del programa e individuos de control.

El resultado del ejercicio es un valor positivo del estimador de diferencias en diferencias que permite evaluar favorablemente los efectos del programa de formación sobre la población beneficiaria.

Adicionalmente, el estimador demuestra ser sensible a la manipulación de los datos.

Palabras clave: Evaluación de programas públicos de formación, estudios observacionales, identificación, estimadores de diferencias en diferencias.

ABSTRACT:

Difference-in-difference estimator is a method of causal inference estimation appropriate in the context of observational studies. Difference-in-difference estimator applied to evaluation of public training programs has shown interesting results.

This paper implements an evaluation based on the application of difference-in-difference estimator over a sample of non real values of the potential outcome (labour earnings) which have been generated also assignment in a random way between participants and non participants in the training program.

The result of the simulation became a positive value of the difference-in-difference estimator. This implies a favourable evaluation for the training program.

Additionally, difference-in-difference estimator became sensitive to changes in the data used.

Key words: Evaluation of public training programs, observational studies, identification, difference-in-difference estimators.

JEL Classification: H52, J38

Í N D I C E

1.- Introducción.	4
2.- Consideraciones preliminares.	6
3.- El estimador de diferencias en diferencias en los estudios observacionales.....	7
3.1.- Las características no observables.....	7
3.2.- Planteamiento analítico del método de diferencias en diferencias. Los datos longitudinales.	8
3.3.- La evolución paralela de los individuos.....	9
4.- Modelización del cálculo del estimador de diferencias en diferencias.....	10
5.- Resultados obtenidos y sensibilidad.	12
5.1.- Resultados.	12
5.2.- La sensibilidad de los resultados.....	13
6.- Conclusiones.	15
Bibliografía.....	16
Apéndice 1.	
Cálculo del estimador de diferencias en diferencias.	19
Apéndice 2.	
Tres ejercicios adicionales.....	20

"Evaluación de programas públicos de formación a través de métodos observacionales: el estimador de diferencias en diferencias. Un ejercicio de simulación"

1.- Introducción.

La inferencia causal estadística ha sido empleada en campos científicos muy diversos de entre los que la epidemiología, la criminología o la planificación urbanística son sólo algunos ejemplos. El desarrollo de la evaluación económica de políticas públicas también se ha beneficiado del uso de los métodos de inferencia causal.

Un desarrollo más específico se observa en la evaluación de las políticas públicas de empleo, receptoras de una parte considerable del gasto público. Ejemplos de este desarrollo son los trabajos realizados en EE.UU. por Card y Sullivan (1988) y Manski y Garfinkel (1992), el realizado en Francia por Bonnal, Fougère y Sérandon (1997), los trabajos de Andrews, Bradley y Upward (1999) y Blundell et al. (2002) para el Reino Unido, el de Bergemann, Fitzenberger y Speckesser (2002) para Alemania y el de Park et al. (1996) para Canadá.

Los gobiernos destinan una parte considerable de su presupuesto de gasto en financiar este tipo de programas, bien de forma autónoma, bien de forma conjunta con otros niveles de gobierno que también se muestran interesados en conocer el grado de efectividad³ de los mismos. Los métodos de inferencia causal aplicados a la evaluación económica de programas públicos de formación valoran dicha efectividad.

³ Esta modalidad de financiación pública a través de transferencias condicionadas compensatorias es especialmente importante para los países miembros de la UE con regiones beneficiarias de fondos estructurales. Entre las actuaciones públicas que se financian con estos fondos se encuentran los programas de formación.

Particularmente, están interesados en evaluar el efecto causal⁴ de un programa de formación sobre algunas variables que se consideran relevantes en relación con la efectividad de la actuación pública⁵.

Los valores individualizados de las variables pueden extraerse de experimentos aleatorios sobre los que se diseña la evaluación, o pueden resultar de una muestra de observaciones tomadas como consecuencia del desarrollo de métodos observacionales o cuasi-experimentales. Ambos entornos condicionarán notablemente la evaluación.

La evaluación de los efectos causales de las actuaciones públicas mediante experimentos aleatorios resulta impracticable en el mayor número de casos. En ocasiones limitaciones de índole moral impiden este tipo de experimentos⁶. En otras ocasiones, el elevado coste de su realización es la causa de que los experimentos aleatorios resulten impracticables. En el caso concreto de la evaluación de programas públicos de formación, esta última es la limitación más relevante.

Con estas limitaciones, la estimación del efecto causal debe hacerse sobre inferencias hechas a partir de datos observacionales, en el contexto de lo que se denominan "estudios observacionales"⁷. Los métodos cuasi-experimentales son métodos observacionales que tratan de reproducir escenarios experimentales⁸. Estos métodos cuasi-experimentales aplicados a la determinación de la inferencia causal permiten llevar a cabo la evaluación de los efectos de un programa de formación teniendo en

⁴ Una referencia amplia del enfoque teórico de la causalidad y su consideración en los experimentos aleatorios y diseñados puede encontrarse en Cox (1992).

⁵ Trabajos pioneros en esta materia son los de Rubin (1974 y 1990) y Heckman (1990).

⁶ Supongamos que queremos investigar los efectos del tabaco sobre la mortalidad. Un experimento aleatorio con personas para explorar los efectos del tabaco sería contrario a la ética.

⁷ Algunos trabajos iniciales en el desarrollo de métodos observacionales han sido los de Kiefer (1979), Bassi (1984) y Rosenbaum (1999). Los trabajos de Kiefer (1979) y Bassi (1984) fueron comentados por LaLonde (1986) en su estudio comparativo entre los métodos experimentales y no experimentales.

⁸ Los estudios pioneros en este ámbito se desarrollaron en Medicina. Particularmente deben destacarse los trabajos de Cameron y Pauling (1976), Billewicz (1965) y Cochran (1968). Un comentario interesante sobre estos primeros desarrollos se encuentra en Rosenbaum (1995).

cuenta nuestro conocimiento institucional del problema. Asimismo, presentan la ventaja de permitir la consideración de muestras de mayor tamaño, frente a los métodos experimentales que trabajan con muestras relativamente pequeñas, y admiten también la posibilidad de aplicar las nuevas técnicas y los nuevos estimadores que se han ido desarrollando en el ámbito no experimental, como así argumentan Heckman y Smith (1995) en su réplica al trabajo de LaLonde (1986)⁹.

De entre los métodos observacionales deben destacarse tres¹⁰. El primero de ellos se denomina método de “selección sobre variables observables” o “selección sobre observables” y engloba tres tipos de procedimientos diferenciados: la subclasificación, los estimadores “matching” y el cálculo del “propensity score”. El segundo de los métodos observacionales es el método del estimador de “diferencias en diferencias”. Finalmente, el tercer método es el de las variables instrumentales. En este trabajo aplicamos el segundo de los métodos a la evaluación de programas públicos de formación.

Para ello hemos estructurado el trabajo en seis apartados. Tras la introducción, el apartado 2 realiza unas consideraciones preliminares necesarias para ubicar el método aplicado. El apartado 3 resume el estimador de diferencias en diferencias. El apartado 4 explica los datos utilizados y el tratamiento al que se han sometido. Los resultados se muestran en el apartado 5 mientras que las conclusiones se presentan en el apartado 6. Los apéndices 1 y 2 detallan aspectos de interés que se han preferido presentar separadamente.

⁹ Trabajo en el cual LaLonde (1986) defiende la supremacía de los métodos experimentales sobre los métodos observacionales, trabajo que ha sido replicado en estudios posteriores, entre ellos, el mencionado de Heckman y Smith (1995).

¹⁰ Para una visión más amplia de estos métodos observacionales y su clasificación véanse las exposiciones realizadas por Angrist y Krueger (1999) y por Meyer (1995).

2.- Consideraciones preliminares.

A nivel individual, conocer el efecto causal de un programa de formación exige, inicialmente, definir un indicador de tratamiento D_i en forma de variable binaria para cualquier individuo i -ésimo de los que, potencialmente, pueden participar en el programa.

Así,

$D_i = 1$, indicará que el individuo i ha participado en el programa.

O bien, $D_i = 0$, indicará que el individuo i no ha participado en dicho programa.

Definido el indicador D_i , los métodos de inferencia causal utilizan la noción de "respuestas potenciales" de los beneficiarios del programa para definir la (s) variable (s) respuesta (s) de cada individuo al tratamiento o programa de formación.

De esta forma, si consideramos una única variable respuesta Y_i que representa los ingresos laborales obtenidos por el individuo i -ésimo, las respuestas potenciales resultan ser:

Y_{0i} = Valor de la variable respuesta en el caso de que el individuo i -ésimo no se beneficie del programa de formación.

Y_{1i} = Valor de los ingresos laborales del individuo i -ésimo (variable respuesta) en caso de beneficiarse del programa público de formación.

Definidas las respuestas potenciales, el efecto causal del programa de formación sobre los ingresos laborales del individuo i -ésimo vendría determinado por la diferencia $Y_{1i} - Y_{0i}$ a partir de cuya magnitud podría evaluarse la efectividad del programa.

Naturalmente, no podemos observar simultáneamente Y_{0i} e Y_{1i} ya que se trata de efectos contrafactuales. Para cada individuo sólo puede observarse la respuesta realizada Y_i definida como¹¹:

$$Y_i = D_i \cdot Y_{1i} + (1 - D_i) \cdot Y_{0i}$$

Ecuación 1

Este problema, conocido como el problema fundamental de identificación en los estudios de causalidad, impide calcular los efectos causales individuales del programa de formación.

El resultado general anterior, basado en el carácter contrafactual de las respuestas potenciales, obliga a buscar soluciones de "second best". Estas soluciones podrían derivarse de la búsqueda de efectos causales agregados.

3.- El estimador de diferencias en diferencias en los estudios observacionales.

3.1.- Las características no observables.

Con frecuencia, el desarrollo de la evaluación deja traslucir que los individuos beneficiarios y los de control difieren en características que no son observables. En estos casos, los métodos observacionales de selección sobre observables no resultan del todo adecuados para estimar los efectos causales del programa de formación evaluado. Los estimadores basados en el método de "diferencias en diferencias" permiten salvar esta dificultad.

En los estudios observacionales los datos observados permiten detectar variaciones en el valor de las variables relevantes en la evaluación de forma similar a lo que ofrecen los datos experimentales. Particularmente, estas variaciones pueden resultar de las

¹¹ Véase Angrist e Imbens (1991).

observaciones tomadas de los mismos individuos, antes y después de someterse a un tratamiento o, más concretamente, beneficiarse de un programa de formación. En este contexto y bajo ciertas condiciones, el efecto causal de la política pública se puede determinar mediante un estimador de diferencias en diferencias.

Así, una vez detectada la existencia de características diferenciadoras no observables entre los individuos beneficiarios y los asignados al grupo de control, se comparan los resultados observados de la variable respuesta Y para los beneficiarios, tanto antes de someterse al programa como después de finalizarlo. En definitiva, se supone que las observaciones pre-programa de Y así como las posteriores al desarrollo del programa de formación estarán igualmente contaminadas a causa de las variables diferenciadoras no observables.

Dado que las variables contaminantes son ajenas a la naturaleza del programa, se supone que sus valores se mantendrán invariables antes y después de la ejecución del programa. No obstante, la mera comparación de los valores observados de Y para los individuos beneficiarios, antes y después de la ejecución del programa, puede resultar sesgada como consecuencia de dos circunstancias.

Por una parte, han de tenerse en cuenta las tendencias temporales de la variable respuesta Y . Por otra parte, es posible que durante el desarrollo del programa aparezcan variables contaminantes que afecten al valor final de Y que, sin embargo, no estaban presentes en el momento previo al inicio del programa.

Para salvar estas complicaciones, se recurre al grupo de control que, en este caso, va a ser utilizado para detectar la variación temporal ajena al programa de los resultados o la existencia de hechos contaminantes situados entre los dos periodos de observación.

En definitiva, como resume Heckman et al. (1998, p. 1020), el cálculo del estimador de diferencias en diferencias consiste en calcular la diferencia de la variable respuesta de

los individuos beneficiarios entre sus valores antes y después de dicho programa, y volver a calcular la diferencia con respecto a la variación producida en la variable respuesta de los individuos de control¹².

3.2.- Planteamiento analítico del método de diferencias en diferencias. Los datos longitudinales.

Comparado con otros métodos observacionales, el método de diferencias en diferencias presenta como novedad el uso de observaciones hechas en dos momentos diferentes del tiempo. Esto obliga a referenciar los datos utilizados al momento de su obtención. Así, siguiendo la notación de Abadie (2003, pp. 7 y ss.) definimos¹³:

$Y_{1i}(t)$: Valores de la variable respuesta para el individuo i en un momento posterior a la finalización del programa (t) cuando i es beneficiario del mismo.

$Y_{0i}(t)$: Valores de la variable respuesta para el individuo i en un momento posterior a la finalización del programa (t) cuando i no ha participado en el mismo (i es un individuo asignado al grupo de control).

$Y_{1i}(t-1)$: Valores de la variable respuesta para el individuo i en el momento previo al inicio del programa ($t-1$) cuando i es beneficiario del mismo.

$Y_{0i}(t-1)$: Valores de la variable respuesta para el individuo i en el momento previo al inicio del programa (t) cuando i no ha participado en el mismo (i es un individuo asignado al grupo de control).

Con esta notación, el efecto causal del programa de formación sobre el individuo i resultaría del siguiente cálculo:

¹² Nos remitimos también a la exposición del estimador de diferencias en diferencias en Heckman y Robb (1985).

¹³ En relación con el desarrollo metodológico de este estimador, véanse también los trabajos de Card y Krueger (1994), Imbens, Liebman y Eissa (1998) y Angrist y Krueger (2000).

$$Y_{1i}(t) - Y_{0i}(t)$$

Ecuación 2

No obstante, la expresión anterior nos conduce de nuevo al problema fundamental de la inferencia estadística ya que los valores $Y_{1i}(t)$ e $Y_{0i}(t)$ no pueden observarse simultáneamente para el mismo individuo i . Efectivamente, se trata de sucesos contrafactuales.

Si, por un lado, para simplificar las expresiones, eliminamos el subíndice i , indicador del individuo, y, por otro lado, consideramos el periodo $t-1$ como el momento 0 y t como el momento 1, podremos denotar las expresiones anteriores como $Y_1(1)$, $Y_0(1)$, $Y_1(0)$, $Y_0(0)$. Con esta nueva notación sabemos que:

$Y_1(0) = Y_0(0) = Y(0)$, ya que los resultados en el momento 0 son independientes de que con posterioridad el individuo se convierta en beneficiario del programa o no.

$Y_1(1) = Y_0(1) (1-D) + Y_1(1)D$, siendo $D = D(1)$, lo que indica que se trata de una variable que sólo toma valores (0, 1) en el momento 1.

Ahora es posible estimar el efecto promedio del programa de formación sobre los individuos beneficiarios del mismo en el momento 1 según la expresión

$$E[Y_1(1) - Y_0(1) | D = 1]$$

Ecuación 3

3.3.- La evolución paralela de los individuos.

Athey e Imbens (2002, p. 1) señalan que en el método de diferencias en diferencias subyace el supuesto de que la tendencia temporal registrada por los individuos pertenecientes al grupo de control sirve, a modo de variable “proxy”, para conocer la

evolución que hubieran seguido los beneficiarios del programa en el caso de que no lo hubieran seguido. Analíticamente,

$$E[Y_0(1) - Y_0(0) | D = 1] = E[Y_0(1) - Y_0(0) | D = 0]$$

Ecuación 4

A partir de la Ecuación 4, la diferencia entre los momentos 0 y 1 de los resultados potenciales para un individuo del grupo de los beneficiarios ($D=1$) en el caso de no haber recibido el programa de formación (Y_0) será igual a la diferencia entre los momentos 0 y 1 de los resultados potenciales obtenidos por un individuo del grupo de control ($D=0$).

De cumplirse lo anterior, podría obtenerse un estimador del efecto promedio del programa de formación sobre los beneficiarios a partir de la expresión:

$$\begin{aligned} \alpha_{EMPS} &= E[Y_1(1) - Y_0(1) | D = 1] = \\ &= \{E[Y(1) | D = 1] - E[Y(1) | D = 0]\} - \{E[Y(0) | D = 1] - E[Y(0) | D = 0]\} \end{aligned}$$

Ecuación 5

En la Ecuación 5 el primer término muestra la diferencia entre los resultados promedios observados después de la ejecución del programa (momento 1) para los beneficiarios y para los individuos adscritos al grupo de control.

Por su parte, el segundo término de la Ecuación 5 recoge la diferencia entre los valores promedio observados antes de la ejecución del programa (momento 0) también para los beneficiarios y para los individuos adscritos al grupo de control.

A partir de los datos contenidos en la muestra, es posible construir un estimador del efecto causal promedio del programa de formación sobre los beneficiarios.

Así, tenemos la siguiente ecuación:

$$\begin{aligned}
\alpha_{EMPS DID} &= \left\{ \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i(1) - \frac{1}{n_0} \sum_{i=1}^{n_0} Y_i(1) \right\} - \left\{ \frac{1}{n_1} \sum_{i=1}^{n_1} Y_i(0) - \frac{1}{n_0} \sum_{i=1}^{n_0} Y_i(0) \right\} = \\
&= \underbrace{\frac{1}{n_1} \sum_{i=1}^{n_1} \{Y_i(1) - Y_i(0)\}}_{1^a \text{ Diferencia}} - \underbrace{\frac{1}{n_0} \sum_{i=1}^{n_0} \{Y_i(1) - Y_i(0)\}}_{2^a \text{ Diferencia}} = \\
&\qquad\qquad\qquad \underbrace{\hspace{10em}}_{\text{Diferencias en Diferencias}}
\end{aligned}$$

Ecuación 6

donde n es el tamaño de la muestra, n_1 el número de individuos beneficiarios y n_0 el número de individuos adscritos al número de control, cumpliéndose que $n = n_1 + n_0$.

El estimador α_{EMPS} queda reducido a una diferencia de diferencias. La primera de ellas estima la variación en la variable respuesta (ingresos laborales) para los individuos beneficiarios. Este resultado muestra la variación bruta o total registrada en los ingresos laborales (Y) de los individuos que se han beneficiado del programa.

La segunda diferencia recoge la variación registrada en la variable respuesta, Y , para los individuos de control. Dado que estos individuos no se han beneficiado del programa, la variación de sus ingresos (Y) recogerá los efectos de variables ajenas a la propia ejecución del programa y que actuarían como contaminantes.

La tercera diferencia resulta de la resta de las dos anteriores y es la que permite estimar el efecto promedio producido por el programa sobre los individuos beneficiarios. Esta tercera diferencia permite calcular la variación neta producida en los ingresos laborales de los beneficiarios debida exclusivamente al programa de formación, ya que descuenta de la variación bruta de Y , la parte que se debe al efecto de las variables contaminantes.

4.- Modelización del cálculo del estimador de diferencias en diferencias.

Expuesto analíticamente el estimador de diferencias en diferencias, procedemos a su aplicación sobre los datos de un ejercicio de simulación diseñado a tal efecto. En dicho ejercicio generamos los datos necesarios de las variables que intervienen y, a partir de ellos, obtenemos el valor de dicho estimador.

Como se ha expuesto, en el método de diferencias en diferencias es necesario conocer los valores de la variable respuesta –nivel de ingresos- de los individuos referidos tanto a un momento anterior al programa como a un momento posterior.

Asimismo, necesitaremos contar, además de con una muestra obtenida de entre los individuos que han recibido el programa de formación –individuos beneficiarios–, con una muestra de individuos que, aun cumpliendo unas características similares a los anteriores, no han participado en dicho programa –individuos de control–.

En nuestro ejercicio, consideraremos una muestra de 200 individuos beneficiarios y de 100 individuos de control. Así,

$$\begin{aligned}n_1 &= 200 \\n_2 &= 100 \\n &= n_1 + n_2 = 300\end{aligned}$$

Los datos referidos a la variable respuesta Y –nivel de ingresos– se generan mediante un procedimiento aleatorio¹⁴.

Los valores de $Y_1(0)$ e $Y_0(0)$ corresponden al nivel de ingresos de los individuos beneficiarios y de control, respectivamente, antes de que tenga lugar el programa de

¹⁴ El procedimiento seguido está basado en la generación aleatoria de un valor dentro del intervalo de variación fijado, valor que se irá asignando a cada individuo como dato registrado por su variable respuesta Y correspondiente.

formación. Estos valores han sido generados y asignados a los individuos aleatoriamente tomando como extremos los valores 500000 y 1000000¹⁵.

Los valores de $Y_1(1)$ y de $Y_0(1)$ corresponden al nivel de ingresos de los individuos beneficiarios y de control, respectivamente, en un periodo posterior a la realización del programa de formación. Para estas variables, igualmente, se han asignado a los individuos de forma aleatoria valores comprendidos entre 1000000 y 2000000¹⁶.

Una vez generados los datos, calculamos el estimador de diferencias en diferencias según la Ecuación 6. En el apéndice 1 se recoge detalladamente el cálculo.

5.- Resultados obtenidos y sensibilidad.

5.1.- Resultados.

En el ejercicio desarrollado, una vez realizada la generación y asignación de los distintos valores de la variable respuesta Y a los individuos, se obtienen los resultados en el cálculo del estimador de diferencias en diferencias, recogidos en la Tabla 1.

¹⁵ Este intervalo de variación elegido de la variable respuesta para el momento previo al programa está fijado de manera arbitraria, teniendo en cuenta que su establecimiento entre estos límites u otros cualesquiera no influye en la obtención del estimador en la medida en que ambos grupos, beneficiarios y de control, cuentan con el mismo intervalo de variación.

¹⁶ Este intervalo de extremos mayores de variación de los valores de la variable respuesta para el momento posterior al programa se ha elegido para poder tener en cuenta e incluir en el modelo posibles tendencias temporales de la variable respuesta o la aparición de alguna variable contaminante sobrevenida en un momento intermedio, entre los momentos previo y posterior al programa, efectos ambos que el estimador de diferencias en diferencias aísla al comparar el grupo de beneficiarios con el de control, tal y como se expuso en el desarrollo teórico previo. Los valores están expresados en unidades monetarias.

Proceso de Simulación 1			
Grupo de Beneficiarios		Grupo de Control	
n_1	200	n_0	100
$\sum_{\substack{i=1 \\ D=1}}^{n1} \{Y_i(1) - Y_i(0)\}$	176981000	$\sum_{\substack{i=1 \\ D=0}}^{n0} \{Y_i(1) - Y_i(0)\}$	77004000
1ª Diferencia: Diferencias en Beneficiarios		2ª Diferencia: Diferencias Grupo de Control	
$\frac{1}{n_1} \sum_{\substack{i=1 \\ D=1}}^{n1} \{Y_i(1) - Y_i(0)\}$	884905	$\frac{1}{n_0} \sum_{\substack{i=1 \\ D=0}}^{n0} \{Y_i(1) - Y_i(0)\}$	770040
Diferencias en Diferencias			
$\alpha_{EMPS DID} = \frac{1}{n_1} \sum_{\substack{i=1 \\ D=1}}^{n1} \{Y_i(1) - Y_i(0)\} - \frac{1}{n_0} \sum_{\substack{i=1 \\ D=01}}^{n0} \{Y_i(1) - Y_i(0)\}$			
$\alpha_{EMPS DID} =$		114865 (14,92 %)	

Tabla 1

Como vemos en esta Tabla 1, el estimador del efecto del programa sobre los individuos beneficiarios toma un valor positivo. Este signo positivo del estimador permite concluir la existencia de un efecto favorable del programa de formación sobre el nivel de ingresos (variable respuesta Y considerada) de los beneficiarios. La evaluación del programa de formación sería, por tanto, favorable.

Este estimador, que en nuestro ejemplo alcanza un valor absoluto de 114865, tomado en términos relativos supone un incremento, entre los periodos previo y posterior al programa, del nivel de ingresos de los individuos beneficiarios un 14,92% superior al incremento en el nivel de ingresos producido entre ambos periodos para el caso de los individuos de control.

Como el proceso de cálculo de este estimador está basado en un método de generación automática de datos, podemos repetir este cálculo cuantas veces sea necesario, para ello basta generar nuevos valores para la variable respuesta Y para calcular posteriormente las diferencias. El Apéndice 2 incluye los resultados que arrojan tres escenarios alternativos al que nos ha servido de ejemplo.

5.2.- La sensibilidad de los resultados.

Como podemos extraer del ejemplo anterior y de los ejemplos recogidos en el Apéndice 2, si repetimos el proceso de simulación planteado, al estar los valores de $Y_1(1)$ e $Y_0(1)$ generados y asignados de forma aleatoria, obtendremos resultados diferentes que se irán alternando entre valores positivos (evaluación favorable del efecto del programa sobre los beneficiarios) y negativos (evaluación desfavorable) del valor del estimador, e incluso, valores muy próximos a cero, que denotarían un efecto prácticamente nulo del programa.

Sin embargo, si queremos verificar cómo el estimador calculado basado en el método de diferencias en diferencias detecta un determinado efecto sobre el nivel de ingresos, podemos introducir premeditadamente ese efecto sobre los datos y comprobar si el estimador calculado acusa dicho efecto.

Para ello, procedemos de forma similar al caso anterior. Consideramos una muestra de beneficiarios formada por 200 individuos y una muestra de 100 individuos de control.

Damos valores a las variables $Y_1(0)$ e $Y_0(0)$ de igual forma que en el caso anterior, mediante generación aleatoria, asignando a cada individuo un nivel de ingresos en el momento previo al programa que oscilará entre los valores 500000 y 1000000.

Sin embargo, para el caso de las variables referidas al periodo posterior al desarrollo del programa, $Y_1(1)$ e $Y_0(1)$, introducimos una modificación sobre el ejercicio original, asignando aleatoriamente valores comprendidos entre 1000000 y 2000000 para el caso del grupo de control (igual en este caso que el supuesto inicial) y valores comprendidos entre 10000000 y 22000000 para los individuos beneficiarios (ampliando, por tanto, el rango de variación).

De esta forma, al oscilar dentro de un intervalo acotado por un extremo superior mayor, se permite que los posibles valores de $Y_1(1)$ puedan alcanzar importes mayores.

Es decir, el nivel de ingresos de los beneficiarios se ve influenciado por un efecto que podemos considerar propio del programa de formación que se evalúa, ya que, como quedó expuesto en el planteamiento teórico del método de diferencias en diferencias, al calcularse la variación del nivel de ingresos de los beneficiarios entre un momento posterior y uno previo al programa y la diferencia de esa variación con respecto a la variación entre ambos momentos del nivel de ingresos de los individuos de control, se están excluyendo todos los posibles efectos producidos por características no observables, aislándose así el efecto del programa.

De esta forma, calculando ahora el estimador, deberíamos obtener un valor positivo del mismo, lo cual demostraría cómo dicho estimador es sensible a la modificación introducida.

Realizamos la prueba repitiendo el proceso de manera que, tras realizar 20 procesos consecutivos, los resultados son los recogidos en la Tabla 2.

Se comprueba que, en todos los supuestos resulta siempre un estimador de signo positivo, mostrándose, por tanto, sensible a la manipulación de los datos de los individuos beneficiarios.

Procesos de Simulación con Efecto Inducido					
Proceso n°	$\alpha_{EMPS DID}$	%	Proceso n°	$\alpha_{EMPS DID}$	%
1	69490	9,18	11	86465	11,40
2	61185	8,14	12	39020	4,74
3	104945	14,24	13	163905	22,79
4	146815	21,74	14	45850	5,79
5	20050	2,42	15	76225	9,62
6	118285	16,06	16	75485	9,96
7	121470	16,43	17	54860	7,20
8	153370	21,72	18	43370	5,78
9	108360	14,79	19	117565	15,94
10	85585	11,93	20	115690	15,20

Tabla 2

6.- Conclusiones.

Los resultados del ejercicio central de este artículo permiten obtener un valor positivo del estimador de diferencias en diferencias. Expresado en términos porcentuales, el estimador muestra que, en promedio, el incremento en el nivel de ingresos laborales de los individuos beneficiarios del programa de formación es un 14,92 % superior al incremento que registran los individuos de control entre los momentos previo y posterior al desarrollo del programa de formación.

La evaluación del programa basada en el valor del estimador de diferencias en diferencias sería favorable.

Finalmente, ante una manipulación en el rango de valores de la variable respuesta para los individuos beneficiarios, el estimador demuestra ser sensible y arroja valores coherentes con el sentido de la manipulación.

Bibliografía.

- Abadie, A. (2003): “Semiparametric Difference-in-Differences Estimators”, Working Paper, abril 2003. Universidad de Harvard y NBER.
- Andrews, M., S. Bradley y R. Upward (1999): “Estimating Youth Training Wage Differentials During and After Training”, *Oxford Economic Papers*, vol 51(3), 517-544.
- Angrist, J. D. y A. B. Krueger (1999): “Empirical Strategies in Labor Economics”, in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card. North Holland: Elsevier, sec. 1 y 2.
- Angrist, J. D. and A. B. Krueger (2000): “Empirical Strategies in Labor Economics” in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card. North Holland: Elsevier, 1277-1366.
- Angrist, J. D. y G. W. Imbens (1991): “Sources of Identifying Information in Evaluation Models”. NBER Technical Working Paper 117.
- Athey, S. y G. Imbens (2002): “Identification and Inference in Nonlinear Difference-in-Differences Models”. Working Paper, mayo 2002. Universidad de Standfor, Universidad de Berkeley y NBER.
- Bassi, L. (1984): “Estimating the Effect of Training Programs with Non-Random Selection”, *Review of Economics and Statistics*, vol. 66, 36-43.
- Bergemann, A., B. Fitzenberger y S. Speckesser (2004): “Evaluating the Dynamic Employment Effects of Training Programs in East Germany Using Conditional Difference-in-Differences”. Discussion Paper n° 04-41. ZEW Centre for European Economic Research, Abril, 2004.
- Billewicz, W. Z. (1965): “The Efficiency of Matched Samples: an Empirical Investigation”, *Biometrics*, vol. 21, 623-643.

- Blundell, R., M. Costa, C. Meghir y J. Van Reenen (2002): “Evaluating the Employment Impact of a Mandatory Job Search Program”. Working Paper 01-20. ZEW Centre for European Economic Research, Abril 2004. The Institute for Fiscal Studies, Diciembre 2002.
- Bonnall, L., D. Fougère y A. Sérandon (1997): “Evaluating the Impact of French Employment Policies on Individual Labour Market Histories”, *Review of Economic Studies*, vol. 64, 683-713.
- Cameron, E. y L. Pauling (1976): “Supplemental Ascorbate in the Supportive Treatment of Cancer: Prolongation of Survival Times in Terminal Human Cancer”, *Proceedings of the National Academy of Sciences(USA)*, 73, 3685-3689.
- Card, D. y A. Krueger (1994): “Minimum wages and employment: A case study of the fast food industry”. *American Economic Review*, 84, 772-793.
- Card, D. y D. Sullivan (1988): “Measuring the Effects of Subsidized Training Programs on Movements In and Out of Employment”, *Econometrica*, vol. 56, 497-530.
- Cochran, W. G. (1968): “The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies”, *Biometrics*, vol. 24, 295-313.
- Cox, D. R., (1992): “Causality: Some Statistical Aspects,” *Journal of the Royal Statistical Society, Series A*, 155, part 2, 291-301.
- Heckman, J. J. (1990): “Varieties of Selection Bias”, *American Economic Review, Papers and Proceedings*, 80, 313-338.
- Heckman, J. J., H. Ichimura y P. E. Todd (1997): “Matching As an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme”, *Review of Economics Studies*, 64, 605-654.

- Heckman, J. J. y R. Robb (1985): "Alternative Methods for Evaluating the Impact of Interventions", in *Longitudinal Analysis of Labour Market*, ed. by J. Heckman y B. Singer. New York: Cambridge University Press, 156-245.
- Heckman, J. J. y J. A. Smith (1995): "Assessing the Case for Social Experiments", *Journal of Economic Perspectives*, vol. 9 (2), 85-100.
- Imbens, G., J. B. Liebman y N. Eissa (1998): "The Econometrics of Difference in Difference". Working Paper, Enero 1998. Universidad de Harvard.
- Kiefer, N. (1979): "Economic Benefits from Four Manpower Training Programs, Garland Series of Outstanding Dissertations in Economics". Garland Press, New York.
- LaLonde, R. (1986): "Evaluating the Econometric Evaluation of Training Programas with Experimental Data", *American Economic Review*, vol. 76, 604-620.
- Manski, C. y I. Garfinkel (1992): "Evaluating Welfare and Training Programs", Cambridge, Mass.: Harvard University Press.
- Meyer, B. D. (1995): "Natural and Quasi-Experiments in Economics", *Journal of Business & Economic Statistics*, vol. 13, 151-161.
- Park, N., B. Power, W. C. Riddell y G. Wong (1996): "An assessment of the Impact of Government-Sponsored Training", *Canadian Journal of Economics*, vol. 29, S93-S98.
- Rosenbaum, P. R. (1995): "Observational Studies". Nueva York: Srpingger-Verlag.
- Rubin, D. B (1974): "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies", *Journal of Educational Psychology*, 66, 688-701.
- Rubin, D. B. (1990): "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies", *Statistical Science*, 5, 472-480.

Apéndice 1. Cálculo del estimador de diferencias en diferencias.

En primer lugar, se procede al cálculo del primer bloque de diferencias, aquellas producidas entre los momentos previo y posterior al programa para todos los beneficiarios, tal y como recoge la Ecuación 7.

1ª Diferencia

$$\text{Diferencias en Beneficiarios} = \frac{1}{n_1} \sum_{\substack{i=1 \\ D=1}}^{n_1} \{ Y_i(1) - Y_i(0) \}$$

Ecuación 7

Posteriormente, se realiza el cálculo del segundo bloque de diferencias, aquellas producidas entre los momentos previo y posterior al programa para los individuos de control considerados, tal y como recoge la Ecuación 8.

2ª Diferencia

$$\text{Diferencias en Grupo de Control} = \frac{1}{n_0} \sum_{\substack{i=1 \\ D=0}}^{n_0} \{ Y_i(1) - Y_i(0) \}$$

Ecuación 8

Finalmente, se realiza el cálculo de la diferencia entre las calculadas para ambos bloques.

Diferencias en Diferencias

$$\text{Diferencias en Beneficiarios} - \text{Diferencias en Grupo de Control}$$

Ecuación 9

Apéndice 2. Tres ejercicios adicionales.

A modo de ejemplo, recogemos a continuación los resultados obtenidos en la realización de tres procesos de simulación generados de manera similar al primer ejemplo que plantean tres escenarios alternativos.

Proceso de Simulación 2			
Grupo de Beneficiarios		Grupo de Control	
n_1	200	n_0	100
$\sum_{\substack{i=1 \\ D=1}}^{n_1} \{Y_i(1) - Y_i(0)\}$	153811000	$\sum_{\substack{i=1 \\ D=0}}^{n_0} \{Y_i(1) - Y_i(0)\}$	73985000
1ª Diferencia: Diferencias en Beneficiarios		2ª Diferencia: Diferencias Grupo de Control	
$\frac{1}{n_1} \sum_{\substack{i=1 \\ D=1}}^{n_1} \{Y_i(1) - Y_i(0)\}$	769055	$\frac{1}{n_0} \sum_{\substack{i=1 \\ D=0}}^{n_0} \{Y_i(1) - Y_i(0)\}$	739850
Diferencias en Diferencias			
$\alpha_{EMPS DID} = \frac{1}{n_1} \sum_{\substack{i=1 \\ D=1}}^{n_1} \{Y_i(1) - Y_i(0)\} - \frac{1}{n_0} \sum_{\substack{i=1 \\ D=01}}^{n_0} \{Y_i(1) - Y_i(0)\}$			
$\alpha_{EMPS DID} =$		29205 (3,95 %)	

Tabla 3

Proceso de Simulación 3			
Grupo de Beneficiarios		Grupo de Control	
n_1	200	n_0	100
$\sum_{\substack{i=1 \\ D=1}}^{n_1} \{Y_i(1) - Y_i(0)\}$	151513000	$\sum_{\substack{i=1 \\ D=0}}^{n_0} \{Y_i(1) - Y_i(0)\}$	77501000
1ª Diferencia: Diferencias en Beneficiarios		2ª Diferencia: Diferencias Grupo de Control	
$\frac{1}{n_1} \sum_{\substack{i=1 \\ D=1}}^{n_1} \{Y_i(1) - Y_i(0)\}$	757565	$\frac{1}{n_0} \sum_{\substack{i=1 \\ D=0}}^{n_0} \{Y_i(1) - Y_i(0)\}$	775010
Diferencias en Diferencias			
$\alpha_{EMPS DID} = \frac{1}{n_1} \sum_{\substack{i=1 \\ D=1}}^{n_1} \{Y_i(1) - Y_i(0)\} - \frac{1}{n_0} \sum_{\substack{i=1 \\ D=01}}^{n_0} \{Y_i(1) - Y_i(0)\}$			
$\alpha_{EMPS DID} =$		-17445 (-2,25 %)	

Tabla 4

Proceso de Simulación 4			
Grupo de Beneficiarios		Grupo de Control	
n_1	200	n_0	100
$\sum_{\substack{i=1 \\ D=1}}^{n_1} \{Y_i(1) - Y_i(0)\}$	176981000	$\sum_{\substack{i=1 \\ D=0}}^{n_0} \{Y_i(1) - Y_i(0)\}$	75271000
1ª Diferencia: Diferencias en Beneficiarios		2ª Diferencia: Diferencias Grupo de Control	
$\frac{1}{n_1} \sum_{\substack{i=1 \\ D=1}}^{n_1} \{Y_i(1) - Y_i(0)\}$	754410	$\frac{1}{n_0} \sum_{\substack{i=1 \\ D=0}}^{n_0} \{Y_i(1) - Y_i(0)\}$	752710
Diferencias en Diferencias			
$\alpha_{EMPS DID} = \frac{1}{n_1} \sum_{\substack{i=1 \\ D=1}}^{n_1} \{Y_i(1) - Y_i(0)\} - \frac{1}{n_0} \sum_{\substack{i=1 \\ D=01}}^{n_0} \{Y_i(1) - Y_i(0)\}$			
$\alpha_{EMPS DID} =$		1700 (0, 32 %)	

Tabla 5

En el proceso de simulación recogido en la Tabla 3 se obtiene de nuevo como resultado la aceptación de un efecto favorable del programa de formación sobre el nivel de ingresos de los individuos beneficiarios, al resultar el valor del estimador con signo positivo. Sin embargo, su valor es bastante inferior al obtenido en el primer supuesto.

El valor del estimador para el caso recogido en la Tabla 4 toma un valor negativo, lo cual nos plantea una evolución más favorable de la variable respuesta, nivel de ingresos, para el caso de los individuos de control que para los individuos beneficiarios.

Por último, en el supuesto recogido en la Tabla 5, el valor del estimador es muy bajo, casi nulo, con lo cual prácticamente podíamos desestimar el efecto del programa sobre el nivel de ingresos de los individuos beneficiarios.