

Documento de trabajo

E2004/86

**Propuestas metodológicas para la evaluación
de programas públicos de formación. Una re-
visión crítica**

**José Manuel Cansino Muñoz-Repiso
Antonio Sánchez Braza**



Centro de Estudios Andaluces
CONSEJERÍA DE LA PRESIDENCIA

Propuestas metodológicas para la evaluación de programas públicos de formación. Una revisión crítica.

José Manuel Cansino Muñoz-Repiso ¹. U. de Sevilla y Centro de Estudios Andaluces
Antonio Sánchez Braza ². Universidad de Sevilla.

Resumen: La evaluación de programas públicos de formación ha experimentado en las últimas décadas un considerable desarrollo, gracias a la aplicación de métodos de estimación de inferencia causal. Este trabajo ofrece una panorámica de los métodos de estimación disponibles del efecto causal de los programas públicos de formación tanto en el marco de los experimentos aleatorios como de estudios observacionales. Adicionalmente se valora la operatividad de estos métodos de evaluación. Desde una perspectiva teórica, estos métodos pueden aplicarse en experimentos aleatorios si bien, a menudo, están diseñados sobre muestras poblacionales pequeñas. La explicación de ello se encuentra en los elevados costes asociados al desarrollo del experimento. Como consecuencia de lo anterior, la evaluación de los efectos causales de las actuaciones públicas mediante experimentos aleatorios resulta impracticable en el mayor número de casos. Por estas razones, la estimación del efecto causal debe hacerse sobre inferencias hechas a partir de datos observacionales en el contexto de los métodos cuasi-experimentales.

La evaluación de los programas públicos de formación puede desarrollarse a partir de métodos observacionales de entre los que se han destacado tres. 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”. En nuestra opinión se trata del método más operativo de los analizados. Finalmente, el tercer método observacional es el de las variables instrumentales y ha quedado fuera de esta investigación.

Palabras clave: Evaluación de programas públicos de formación, experimentos aleatorios, estudios observacionales, identificación, selección sobre observables, subclasificación, estimadores “matching”, “propensity score”, estimadores de diferencias en diferencias.

Abstract: Evaluation of public training programs has stronger developed in the last decades because of the use of causal inference methods. This paper shows an overview of the disposable methods of public training programs evaluation in the context of randomized experiments also in observational studies. Additionally, we evaluate the real possibilities of these methods. From a theoretical point of view, these methods can be developed in randomized experiments, although they are often designed over a small size of the population. From another point of view, public training programs evaluation can be implemented with observational methods. We can remark the three main observational methods. The first one is named selection on observables and includes three types of differentiated procedures: subclassification, matching estimators and the propensity score. The second one is the difference in difference estimator. From our point of view, this is the more appropriate method when we evaluate public training programs. Finally, the last one use instrumental variables to develop the evaluation. This third method is out from this paper.

Key words: Evaluation of public training programs, randomized experiments, observational studies, identification, selection on observables, subclasiffication, matching estimators, propensity score, difference-in-difference estimators.

JEL classification: H52, J38.

¹ Departamento de Teoría Económica y Economía Política. Facultad de CC. Económicas y Empresariales. Universidad de Sevilla. Avda. Ramón y Cajal, nº 1, 41018 Sevilla (Spain). Tfno : + 34 954 55 75 28 , Fax : + 34 954 55 76 29, Correo electrónico: jmcansino@us.es

² Departamento de Teoría Económica y Economía Política. Facultad de CC. Económicas y Empresariales. Universidad de Sevilla. Avda. Ramón y Cajal, nº 1, 41018 Sevilla (Spain). Tfno : + 34 954 55 75 29 , Fax : + 34 954 55 76 29, Correo electrónico: asb@us.es

"Propuestas metodológicas para la evaluación de programas públicos de formación. Una revisión crítica"

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 Bonnall, 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. Los métodos de inferencia causal aplicados a la evaluación económica de programas públicos de formación valoran la efectividad de los mismos. 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 como tendremos oportunidad de mostrar.

Los objetivos de este trabajo son dos. En primer lugar ofrecer una panorámica de los métodos de estimación disponibles del efecto causal de los programas públicos de formación tanto en el marco de los experimentos aleatorios como de estudios observacionales. Este primer objetivo incluye una revisión de la literatura sobre la materia. El segundo objetivo consiste en valorar la operatividad de estos métodos de evaluación.

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.

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).

Hemos estructurado este documento en 8 apartados. Tras la introducción y especificación de los objetivos del trabajo, el apartado 2 incorpora algunas consideraciones preliminares necesarias para el desarrollo posterior. El apartado 3 aborda el problema fundamental de la identificación. La identificación de los efectos causales de los programas de formación en experimentos aleatorios se estudia en el apartado 4, mientras que la misma identificación a partir de métodos cuasi-experimentales se plantea, de manera genérica, en el apartado 5. Los apartados 6 y 7 están dedicados a desarrollar dos métodos concretos de estimación de efectos causales en experimentos observacionales: el método basado en la selección sobre variables, que contempla los procedimientos de subclasificación, "matching" y "propensity score", y el método basado en los estimadores de diferencias en diferencias. Por último, las conclusiones se presentan en el apartado 8.

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 como⁶:

⁶ Véase Angrist e Imbens (1991).

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

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.

3.- Buscando soluciones al problema fundamental de la identificación.

3.1.- Homogeneidad y efectos causales agregados.

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. Así, un elevado grado de homogeneidad entre los beneficiarios potenciales del programa de formación que implicase que los mismos tuviesen idénticas respuestas potenciales, permitiría resolver el problema. De esta forma, para algunos individuos "j" observaríamos "Y_{1j}" mientras que para otros individuos "k" -idénticos a los anteriores- observaríamos "Y_{0k}".

La homogeneidad permitiría calcular el efecto causal del programa para cada individuo "i" (Y_{1i}-Y_{0i}) mediante la comparación del valor de las variables respuesta entre un individuo que ha seguido el programa y otro que no lo ha hecho.

El elevado grado de homogeneidad en las respuestas es una característica habitual en experimentos realizados en disciplinas científicas como la Física o la Química. Sin embargo, en el ámbito de los experimentos sociales suele darse un amplio grado de heterogeneidad en las respuestas individuales a la participación (y no participación) en programas públicos.

3.2.- Efectos medios de los programas de formación.

Aunque no es posible calcular los efectos causales individuales de los programas de formación, puede resultar factible el cálculo de algún efecto promedio. En este sentido, podemos definir el efecto causal medio del programa como:

$$\alpha_{EMP} = E[Y_1 - Y_0]$$

Ecuación 2

así, se obtendría el efecto medio sobre una muestra extraída de la población diana del programa de formación.

Alternativamente, es posible calcular el efecto medio únicamente sobre los individuos que han seguido el programa de formación, esto es, los individuos beneficiarios

$$\alpha_{EMPS} = E[Y_1 - Y_0 | D = 1]$$

Ecuación 3

En este caso, el cálculo del efecto medio se restringe únicamente a los individuos seleccionados.

Sin embargo, las comparaciones de las variables respuesta entre los beneficiarios y excluidos del programa de formación no suelen ofrecer habitualmente una respuesta correcta. Este resultado puede concluirse de la siguiente expresión:

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

Ecuación 4

La Ecuación 4 compara la media de la variable respuesta de los individuos que han seguido el programa de formación con la media de la misma variable para los individuos que han permanecido al margen. El primer sumando coincide con el valor del efecto medio del programa sobre los beneficiarios, mientras que el segundo representa el sesgo. Este sesgo viene determinado por la diferencia entre la media de la variable respuesta de los individuos que han seguido el programa en el caso de no haberlo hecho, y la media de la variable respuesta de los individuos que no han seguido el programa.

En la mayoría de los casos, el término error será distinto de cero. La explicación se encuentra en los problemas que plantea el proceso de selección de los individuos beneficiarios del programa. Es frecuente que en los programas públicos de formación, los individuos seleccionados suelen tener unos ingresos laborales previos a su participación en el programa inferiores a los individuos no participantes. Ashenfelter (1978) demostró que esta realidad explicaba que el término error en la Ecuación 4 suela ser negativo. En definitiva, dado que existe una fuerte inercia en la evolución de los ingresos laborales individuales, puede inferirse que los ingresos individuales después de participar en un programa de formación suelen ser inferiores a los ingresos de los individuos no participantes en ausencia de programa. Lo anterior viene subrayado por el hecho de que los beneficiarios son seleccionados (e incluso auto-seleccionados) de entre una subpoblación de individuos en situaciones laborales desfavorables asociadas a bajos ingresos laborales. Por todo ello, cabe esperar que

$$E[Y_0|D=1] - E[Y_0|D=0] < 0$$

Ecuación 5

En general, los métodos de inferencia causal estiman los efectos causales de un tratamiento o programa a partir de datos experimentales y no experimentales. La Ecuación 5 sugiere que el establecimiento de los efectos causales se aborde mediante la modelización de la distribución de Y_0 condicionada al valor del indicador D . Sin embargo, esta forma de proceder carece de operatividad pues resulta imposible observar Y_0 cuando $D = 1$.

En cambio, puede resultar más adecuado para tratar el problema de la estimación de los efectos causales, estudiar el mecanismo de asignación de la población diana del programa de formación entre dos grupos posibles. Un grupo integrado por los individuos seleccionados para recibir el programa de formación y otro formado por los individuos que resultan excluidos⁷. En la práctica, muchos de los modelos estadísticos de inferencia causal intentan la identificación de los efectos causales de un programa a través de la restricción, en algún sentido, del mecanismo de asignación.

4.- Los efectos causales en experimentos aleatorios.

4.1.- Las ventajas de los estudios aleatorios.

Cuando la evaluación de los efectos causales de un programa de formación utiliza un mecanismo de asignación aleatorio, los individuos beneficiarios potenciales son asignados aleatoriamente a uno de los siguientes grupos. Un grupo -denominado grupo de tratamiento- en el que se incluyen las personas que finalmente acceden al programa de formación y que nosotros denominaremos beneficiarios. Alternativamente, los individuos pueden ser asignados al llamado grupo de control que agrupa a los excluidos del programa de formación. En definitiva, en un estudio aleatorio, el mecanismo de asignación es aleatorio.

Esta forma de asignar los individuos a los grupos de tratamiento o control garantiza que ambos grupos sean comparables. De esta forma, la Ecuación 4 vuelve a ser útil, ya que, en este caso el término del sesgo sería igual a cero.

Adicionalmente, la aleatoriedad del mecanismo de selección garantiza que el mismo sea independiente de las respuestas potenciales de los individuos, esto es⁸,

$$(Y_1, Y_0) \perp D$$

Ecuación 6

Cuando la Ecuación 6 se cumple puede afirmarse, como expone Holland (1986, pp. 948 y 949), que $E[Y_0 | D=1] = E[Y_0 | D=0]$, por lo que

⁷ La distribución aleatoria de individuos entre el grupo de tratamiento y el de control permite diferenciar el efecto causal entre variables de la mera asociación estadística.

⁸ Seguimos aquí la notación utilizada por David (1979).

$$\alpha_{EMP} = E[Y_1 - Y_0] = E[Y_1|D=1] - E[Y_0|D=0]$$

Ecuación 7

Además, dado que también es cierto que $E[Y_1 | D=1] = E[Y_1 | D=0]$, puede escribirse que

$$\alpha_{EMP} = E[Y_1 - Y_0] = E[Y_1 - Y_0 | D=1] = E[Y | D=1] - E[Y | D=0]$$

Ecuación 8

Por tanto, ante el cumplimiento de la mencionada condición de independencia recogida en la Ecuación 6, podemos igualar la expresión de ambos indicadores α_{EMP} y α_{EMPS} :

$$\alpha_{EMP} = E[Y_1 - Y_0 | D=1] = E[Y | D=1] - E[Y | D=0] = \alpha_{EMPS}$$

Ecuación 9

Los métodos de inferencia causal basados en mecanismos de selección aleatorios permiten obtener resultados aún más precisos. Así, aunque la determinación del efecto causal en α_{EMPS} y α_{EMP} esté basada en los efectos medios del programa de formación, la aleatorización identifica toda la distribución marginal de Y_0 e Y_1 . Por tanto,

$$F_{Y_0}(y) = P(Y_0 \leq y) = E[1\{Y_0 \leq y\}] = E[1\{Y_0 \leq y\} | D=0] = E[1\{Y_0 \leq y\} | D=0]$$

Ecuación 10

De manera similar,

$$F_{Y_1}(y) = E[1\{Y \leq y\} | D=1]$$

Ecuación 11

En definitiva, se trata de una variable cuyo valor es 1 si $Y_1 \leq y$. Esto permite comparar el efecto causal del programa de formación no sólo a nivel de media sino también para cualquier cuantil⁹ $Q_\theta(Y_1) - Q_\theta(Y_0)$, donde θ es el índice del cuantil que toma valores entre 0 y 1. De manera análoga, $F_{Y_d}(Q_\theta(Y_t)) = \theta$ para $d=0,1$.

Supongamos entonces que dirigimos una prueba aleatoria con n individuos. Para cada individuo i , la cara de una moneda determina si será asignado al grupo de tratados ($D_i=1$) o al grupo de control ($D_i=0$). Después de que el grupo de tratamiento siga el programa de formación, se recopila la información sobre los ingresos laborales de cada individuo, Y_i (variable respuesta). La estimación del efecto causal del programa se

⁹ Debe señalarse, no obstante, que la aleatorización no identifica los cuantiles del efecto causal del programa de formación, $Q_\theta(Y_1 - Y_0)$, pues, como es sabido y contrariamente a lo que ocurre con las medias, la diferencia de cuantiles no es igual al cuantil de la diferencia.

realiza a partir de una muestra análoga de los resultados de la población. Por ejemplo, sea $\alpha_0 = E [Y_1 - Y_0]$ ($=\alpha_{EMPS} = \alpha_{EMP}$) el efecto medio del programa. La Ecuación 8 sugiere estimar α_0 utilizando una diferencia simple de medias muestrales extraídas de los grupos de tratados y de control, respectivamente. En este sentido, si definimos - como el estimador de α_0 , entonces

$$\hat{\alpha} = \hat{Y}_1 - \hat{Y}_0$$

Ecuación 12

donde

$$\bar{Y}_1 = \frac{\sum Y_i \cdot D_i}{\sum D_i} = \frac{1}{n_1} \sum_{D_i=1} Y_{1i}$$

Ecuación 13

$$\bar{Y}_0 = \frac{\sum Y_i \cdot (1 - D_i)}{\sum (1 - D_i)} = \frac{1}{n_0} \sum_{D_i=0} Y_{0i}$$

Ecuación 14

siendo $n_1 = \sum_i D_i$ y $n_0 = n - n_1$. Es fácil demostrar que - es un estimador insesgado y consistente de α_0 .

4.2.- Los contrastes de hipótesis.

En principio, 't' nos permite plantear un contraste de hipótesis sobre los efectos causales del programa de formación¹⁰. Naturalmente, la hipótesis nula implica que el efecto causal del programa de formación sobre la población diana es, en media, igual a 0 lo que conduciría directamente al resultado de la evaluación de la efectividad del programa.

No obstante, los experimentos aleatorios, a menudo, están diseñados sobre muestras poblacionales pequeñas. La explicación de ello se encuentra en los elevados costes asociados al desarrollo del experimento. Como consecuencia de lo anterior, los test basados en resultados asintóticos no resultan ser adecuados.

¹⁰ En este contexto, rechazaremos la hipótesis nula $H_0: \alpha_0=0$ frente a la hipótesis alternativa $H_1: \alpha_0 \neq 0$ al nivel asintótico de significación si $|t| > 1.96$.

Un test más adecuado para experimentos menos ambiciosos es el Test Exacto de Fisher¹¹. Este test permitiría la comparación de las medias de los valores observados en las dos poblaciones planteando la posible igualdad de las varianzas.

En primer lugar tendríamos que calcular el valor del estadístico t para contrastar la hipótesis nula de que las medias de dos poblaciones (grupos de tratados y de control) son iguales¹².

$$t = \frac{\frac{n_1}{n_1 - 1} \cdot \delta_1^2}{\frac{n_0}{n_0 - 1} \cdot \delta_0^2} \xrightarrow{d} F(n_1 - 1, n_0 - 1)$$

Ecuación 15

El criterio para aceptar la hipótesis nula será que el valor del estadístico t no supere el valor asignado para esa $F(n_1 - 1, n_0 - 1; \beta / 2)$ para un nivel de significación de β , con un planteamiento de las hipótesis similar al anterior. Es decir, $H_0: Y_{1i} - Y_{0i} = 0$, (luego $\alpha_0=0$); o bien, $H_1: Y_{1i} - Y_{0i} \neq 0$, (luego $\alpha_0 \neq 0$).

En el caso de que:

$$|t| < Q_{F(n_1 - 1, n_0 - 1)(\beta / 2)}$$

Ecuación 16

aceptaríamos la hipótesis nula y, por tanto, admitiríamos la no existencia de efecto del programa sobre los perceptores ya que estaríamos aceptando que el efecto promedio que medíamos a partir de α_0 es igual a cero.

¹¹ Los fundamentos teóricos del test quedaron expuestos en Fisher (1935). La idea del Test Exacto de Fisher es derivar una distribución aleatoria para cualquier test estadístico mediante la imposición de una hipótesis nula más fuerte.

El Test de Fisher es aplicado normalmente como instrumento idóneo para la comparación de los valores promedio (media) y las medidas de dispersión (varianza) de dos muestras. Frente a dos subpoblaciones como las establecidas en nuestro experimento, una población de participantes en el programa que evaluamos y otra de control, se miden los niveles alcanzados por una variable determinada como variable significativa o variable respuesta. De estos valores observados obtendremos su valor promedio y su medida de dispersión.

Lo esperado en nuestro caso sería que el nivel promedio de la variable respuesta fuera superior para el grupo de participantes que para el de control, lo cual implicaría que el programa de formación ha tenido efectos positivos. La cuestión clave sería comprobar si se observa esa diferencia del nivel promedio de un grupo con respecto al otro y determinar si la diferencia observada entre los grupos es lo suficientemente grande como para poder rechazar la hipótesis nula planteada y poder aceptar la hipótesis alternativa que supondría afirmar la existencia del efecto de la participación en el programa sobre el nivel de la variable respuesta.

Nos remitimos también a las especificaciones sobre el contraste de hipótesis recogidas en Neyman (1990), especificaciones realizadas originariamente en 1923 y que fueron replicadas por Fischer (1935).

¹² Como expresa la Ecuación 15, el estadístico t contrasta la igualdad de ambas medias mediante el cociente entre las varianzas, cociente que va a seguir la distribución de una ley de Fisher con n_1-1 y n_0-1 grados de libertad.

Y rechazaríamos la hipótesis nula aceptando la hipótesis alternativa y, por tanto, admitiríamos la existencia de un efecto del programa sobre los perceptores en el caso contrario.

5.- Estimación de los efectos causales a partir de métodos cuasi-experimentales.

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 experimentos¹³. 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"¹⁴. Como es sabido, los métodos cuasi-experimentales son métodos observacionales que tratan de reproducir escenarios experimentales¹⁵. Estos métodos cuasi-experimentales aplicados a la determinación de la inferencia causal permiten realizar la evaluación de los efectos de un programa de formación teniendo en 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, así lo argumentan Heckman y Smith (1995) en su réplica al trabajo de LaLonde (1986)¹⁶.

De entre los métodos observacionales consideramos que deben destacarse tres¹⁷. 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 estudiamos sólo los dos primeros dejando para una investigación posterior el análisis del método de variables instrumentales.

¹³ 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.

¹⁴ 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.

¹⁵ 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).

¹⁶ Trabajo en el cual LaLonde (1986) defiende la supremacía de los métodos experimentales sobre los métodos observacionales, trabajo que ha sido replicado en múltiples estudios posteriores, entre ellos, el mencionado de Heckman y Smith (1995).

¹⁷ 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).

6.- Estudios observacionales (I): la selección sobre variables observables.

En estos estudios puede esperarse que los individuos beneficiarios y los pertenecientes al grupo de control difieran en características (distintas de las propias del programa) que afectan a los ingresos laborales individuales. No obstante, si ambos tipos de individuos difieren sólo en las características observables, es posible controlar estas diferencias a través de lo que se denomina "selección sobre observables". Esta forma de abordar las dificultades que plantean los estudios observacionales justifica el uso del procedimiento de subclasificación, de los estimadores de "matching" y del cálculo del "propensity score" para calcular el efecto causal de programas como los de formación.

6.1.- Identificación: Selección sobre variables observables.

En ausencia de datos experimentales, la condición de independencia

$$(Y_1, Y_0) \perp D$$

Ecuación 17

es difícilmente plausible. La condición de independencia exigiría que beneficiarse o no de un programa de formación no estuviese relacionado con los ingresos laborales potenciales de los individuos.

Decimos que la variable X está predeterminada con respecto al programa de formación, D , si para cada individuo i , $X_{1i} = X_{0i}$. Esto es, el valor de X_i no depende del valor de D_i . Frente a este tipo de variables, llamamos variables respuesta a cualquier variable Y cuyo valor puede depender del valor de D_i . Para las variables respuesta y para cualquier individuo i , $Y_{1i} \neq Y_{0i}$. En la medida en que estas características ajenas al programa de formación puedan ser observadas y que los individuos beneficiarios y los del grupo de control presenten diferencias sólo en estas características que pueden ser observadas, se podrán controlar tales diferencias. Sobre esta base se fundamenta el método de "selección sobre observables"¹⁸.

Con frecuencia, las variables predeterminadas se denominan covariables¹⁹. No obstante, el hecho de que X esté predeterminada no implica que X y D sean independientes en la población²⁰.

En lo sucesivo, X representará normalmente un vector de covariables predeterminadas.

¹⁸ Para mayor abundamiento nos remitimos a Barnow, Cain y Goldberger (1981).

¹⁹ Para una exposición más detallada sobre la naturaleza de las variables contaminantes o covariables, véase Rubin (1978).

²⁰ Por ejemplo, podemos estudiar los efectos de una titulación universitaria sobre los salarios. Es posible definir X = "raza" como una covariable (en el lenguaje estadístico, las covariables se denominan variables de confusión o contaminantes). Obviamente el título universitario no afecta a la raza. Sin embargo, la raza y los resultados académicos de los individuos no son independientes en la población.

Como ya hemos expuesto, de entre los métodos observacionales, nos referimos en primer lugar al método de "selección sobre variables observables" o "selección sobre observables"²¹.

Decimos que hay selección sobre observables cuando existe un vector de covariables predeterminadas observadas, X , tal que los ingresos laborales potenciales son independientes del programa de formación condicionado por las covariables

$$(Y_1, Y_0) \perp\!\!\!\perp D | X$$

Ecuación 18

Para un mismo valor de X tendríamos una especie de experimento aleatorio. La condición anterior permite comparar a individuos que pertenecen al grupo de tratados y al grupo de control para un mismo valor de X como puede ser la edad, la raza, etc.

Como exponen Heckman y Hotz (1989, p. 865), la selección sobre observables se da cuando la dependencia entre Y y D se debe a la covariable X que influye en el proceso de selección de los individuos. En este caso, controlando la variable X se soluciona el problema del sesgo en la selección, esto es, se elimina la dependencia entre Y y D .

Bajo selección sobre observables, el efecto causal del programa de formación puede expresarse, siguiendo a Dehejia y Wahba (1999, p. 1057), como

$$E[Y_1 - Y_0 | X] = E[Y_1 - Y_0 | X, D = 1] = E[Y | X, D = 1] - E[Y | X, D = 0]$$

Ecuación 19

Por tanto, el efecto causal medio del programa de formación puede expresarse como

$$\alpha_{EMP} = E[Y_1 - Y_0] = \int E[Y_1 - Y_0 | X] dP(X) = \int (E[Y | X, D = 1] - E[Y | X, D = 0]) dP(X)$$

Ecuación 20

De manera similar, el efecto causal medio del programa de formación sobre los individuos beneficiarios resulta

$$\alpha_{EMPS} = E[Y_1 - Y_0 | D = 1] = \int (E[Y | X, D = 1] - E[Y | X, D = 0]) dP(X | D = 1)$$

Ecuación 21

La expresión anterior muestra que el efecto causal promedio del programa de formación sobre los beneficiarios, es igual a la diferencia entre los valores promedio observados de

²¹ El concepto de "selección sobre observables" hace referencia al procedimiento mediante el cual se eligen las variables predeterminadas que, además, son observables. Este procedimiento se contrapone a los casos en los que las variables predeterminadas no son observables aunque se tenga constancia de su existencia.

los ingresos laborales de los beneficiarios y de los individuos del grupo de control para cada uno de los valores posibles de X cuando $D = 1$.

6.2.- Métodos de selección sobre variables observables.

6.2.1.- La subclasificación.

Realizada la selección sobre observables desaparecen las diferencias sistemáticas previas al programa de formación entre el grupo de individuos beneficiarios y el grupo de control²². Este resultado permitirá calcular los efectos causales del programa mediante la construcción de estimadores de α_{EMP} y de α_{EMPS} a partir de la Ecuación 20 y la Ecuación 21.

Identificada la variable predeterminada X , habrán de especificarse los valores que puede tomar. En función de los valores posibles de X se procede a la subclasificación de la población diana. Lo anterior supone dividir los grupos de tratamiento y control en diferentes subclases.

Los valores que pueda tomar X conducen a dos situaciones:

- Si X es una variable discreta existirá una subclase de individuos para cada valor que pueda tomar X . El número total de subclases es finito.
- Si X es una variable continua, se tendrá que construir un número finito de intervalos que incluyan todo el rango de valores posibles de X . Cada uno de estos intervalos constituirá una subclase de individuos.

La subclasificación de la población origina un conjunto (X^1, X^2, \dots, X^j) de j subclases de individuos²³. Así, podemos denotar por n el número de individuos que participan en la evaluación del programa de formación, de los cuales n_1 se benefician del programa ($D=1$) y los n_0 restantes se incluyen en el grupo de control ($D=0$). Siendo j el número de subclases en que se ha dividido la población, habrá que determinar para cada una el número de individuos que la componen. De manera genérica tendremos:

n_j : Número total de individuos que componen la subclase j .

n_1^j : Número total de individuos beneficiarios del programa de formación que pertenecen a la subclase j .

n_0^j : Número total de individuos que integran el grupo de control en la subclase j .

En cada momento se cumplirá que:

²² Véase, en este sentido, Rubin (1977).

²³ Ejemplos de subclasificaciones de la población diana pueden encontrarse en Cochran (1968), LaLonde (1986) y Krueger (1999). En concreto, el trabajo de LaLonde (1986) evalúa los programas de formación a partir de datos experimentales.

$$n_1^j + n_0^j = n^j$$

$$n^1 + n^2 + \dots + n^j = n$$

n^j/n será la ponderación atribuible a la subclase j sobre el total de la población y n_1^j/n^j el peso que los individuos que se han beneficiado del programa de formación tienen sobre el conjunto de individuos de cada subclase.

Al mismo tiempo, la \bar{Y} mide la respuesta media esperada, distinguiendo entre:

$$\bar{Y}_1^j$$

Ecuación 22

que representa los ingresos laborales medios de los beneficiarios pertenecientes a la subclase j , y

$$\bar{Y}_0^j$$

Ecuación 23

que representa los ingresos laborales medios de los individuos del grupo de control pertenecientes a la subclase j .

Dentro de cada subclase, el grupo de beneficiarios y de control estarán formados por individuos similares. Es posible estimar los efectos causales α_{EMP} y α_{EMPS} mediante la suma, para cada una de las subclases construidas, de las diferencias entre los valores medios observados en los ingresos laborales de los beneficiarios y de los individuos de control. Estas diferencias estarán ponderadas por el peso que corresponde a cada subclase. Analíticamente,

$$\alpha_{EMP} = \sum_{j=1}^j (\bar{Y}_1^j - \bar{Y}_0^j) \cdot \left(\frac{n^j}{n} \right)$$

Ecuación 24

$$\alpha_{EMPS} = \sum_{j=1}^j (\bar{Y}_1^j - \bar{Y}_0^j) \cdot \left(\frac{n_1^j}{n_1} \right)$$

Ecuación 25

6.2.2.- Los estimadores "matching" o estimadores de coincidencia²⁴.

Los estimadores "matching" suponen un procedimiento especial para determinar el efecto causal de un programa de formación sobre los beneficiarios del mismo (α_{EMPS}). Rubin (1973, p. 159), en un desarrollo pionero de este método, expuso que el

²⁴ Para una revisión teórica de los estimadores "matching" véanse los trabajos de Heckman, Ichimura y Todd (1997 y 1998).

procedimiento de "matching" es un método de recogida y tratamiento de datos diseñado para reducir el riesgo y aumentar la precisión de los estudios observacionales.

El método de "matching" aplicado a la evaluación de programas de formación implica que una vez identificada una variable predeterminada X , se elimina su efecto sobre los ingresos laborales por un procedimiento más estricto que el ofrecido por la subclasificación. De esta forma, en lugar de comparar las subclases de individuos, se asigna a cada individuo participante en la evaluación un individuo de control con el mismo valor de la variable predeterminada X .

Dehejia y Wahba (2002) argumentan que la técnica de "matching" permite hacer pares de individuos beneficiarios y de control que son similares en relación a sus características observables. Cuando todas las diferencias relevantes entre los dos miembros del par quedan recogidas por las covariables observables, el método "matching" podrá proporcionar un estimador insesgado de los efectos del tratamiento.

Resumidamente, el procedimiento sería el siguiente;

- Para un individuo i , beneficiario del programa y que muestra un valor X_i para la variable predeterminada X , se procede a buscar un individuo m , que ha participado en el programa de formación y con un valor X_m para la variable X , de manera que $X_i = X_m$ o, en una versión menos estricta, $X_i \sim X_m$. Si la igualdad o similitud se logra, denotaremos al individuo m como $m_{(i)}$.
- Se calcula la diferencia entre los valores observados de los ingresos laborales para el individuo i y para la pareja (match) asignada $m_{(i)}$.
- Se realiza la suma de las diferencias anteriores para todos y cada uno de los individuos que componen el grupo de beneficiarios del programa y se divide por el número de individuos que componen dicho grupo (n_1).

Analíticamente,

$$\alpha_{EMPS}^{matching} = \frac{1}{n_1} \sum_{i=1}^{n_1} (Y_i - Y_{m_{(i)}})$$

Ecuación 26

La Ecuación 26 permite calcular el estimador del tipo "matching" para los individuos beneficiarios, esto es, el efecto promedio del programa de formación sobre los seleccionados.

También es posible calcular el estimador de "matching" del efecto promedio del programa sobre todos los individuos participantes en la evaluación. En este caso, se trataría de calcular un estimador de "matching" del α_{EMP} . La manera de proceder sería análoga a la expuesta anteriormente con la salvedad de que habría de asignarse una pareja a todos los individuos observados.

A cada individuo beneficiario del programa ($D = 1$) se le asignaría como pareja un individuo que no participe en el programa y que presente un valor igual o similar para la

variable predeterminada X . Igualmente, a cada individuo no participante ($D = 0$) se le asignaría como pareja un individuo beneficiario que cumple la misma condición de igualdad o similitud en su valor de X .

A continuación se calcularía la diferencia entre el valor de los ingresos laborales de cada beneficiario con el de su pareja para luego sumar todas las diferencias. Finalmente, el resultado obtenido se dividiría por el número total de individuos observados²⁵ (n).

Analíticamente,

$$\alpha_{EMP\ matching} = \frac{1}{n} \sum_{i=1}^n (Y_i - Y_{m(i)})$$

Ecuación 27

6.2.3.- El método del "propensity score".

La principal utilidad que muestran los métodos de subclasificación y "matching" está en su aplicación a los estudios observacionales dentro del marco de la selección sobre observables. Su operatividad, en cambio, se limita a los casos en los que el número de variables predeterminadas a tener en cuenta sea pequeño²⁶.

En estudios donde el número de variables predeterminadas es elevado resultan adecuados los métodos basados en el "propensity score" o la "propensión a participar". Según la definición de Rosenbaum y Rubin (1983, p. 42) y (1984), la "propensión a participar" o "propensity score" sería la probabilidad de participar en un programa de formación (probabilidad de que $D = 1$) condicionada a los valores que adopte un vector X de variables predeterminadas ($X = X_1, X_2, \dots, X_n$). Así, si denotamos $\xi(X)$ como la propensión a la participación en el programa de los individuos observados, ésta puede expresarse como

$$\xi(X) = \Pr(D = 1 | X)$$

Ecuación 28

quedando de manifiesto que el "propensity score" $\xi(X)$, es una función de X . La forma funcional de ξ será normalmente desconocida y deberá estimarse a partir de los datos muestrales.

²⁵ Una exposición simplificada de este planteamiento puede encontrarse en Abadie e Imbens (2002, pp. 7 y 46) y en Abadie e Imbens (2004).

²⁶ Rosenbaum (1995, pp. 69 y 70) señala que al emplear el método de subclasificación con un número elevado de variables predeterminadas, nos podemos ver obligados a diseñar un número excesivamente amplio de subclases. Más aún, puede ocurrir que para algunas subclases no existan observaciones para individuos beneficiarios del programa o para individuos de control. Este problema resultaría insalvable. También con el método de "matching" las dificultades de aplicación en evaluaciones donde el número de covariables es elevado, pueden ser graves. Esencialmente, resultaría muy difícil encontrar, para cada individuo observado, otro que presentase unos valores similares para cada una de esas covariables.

Rosenbaum y Rubin (1978) exponen también la proposición de independencia del "propensity score" según la cual, si $\varepsilon(X)$ es la probabilidad de entrar a formar parte de un programa de formación condicionada sobre X, puede afirmarse que²⁷

$$(Y_1, Y_0) \perp\!\!\!\perp D \mid \varepsilon(X)$$

Ecuación 29

Basándonos en esta proposición podemos plantear que, para todas las observaciones manejadas que muestren un mismo "propensity score", la distribución del vector X de variables predeterminadas, será igual. Así, resulta posible comparar los resultados observados para los individuos participantes en el programa de formación y para los individuos de control con el mismo valor del "propensity score".

Procediendo de manera análoga a la selección sobre observables, es posible comparar individuos beneficiarios y de control condicionando el valor del "propensity score", de forma que el efecto contaminante de las variables predeterminadas del vector X queda aislado²⁸.

Adicionalmente, al trabajar con el "propensity score", actuamos como en el caso de una covariable unidimensional frente al extenso número de covariables que puede recoger el vector X. Esto permite resolver los problemas planteados para el caso de que el número de variables predeterminado sea muy elevado, ya que al sustituir los distintos valores posibles de cada una de las covariables por el valor unidimensional del "propensity score", el problema se soluciona.

Lo anterior permite calcular el efecto promedio del programa para todos los individuos con un mismo valor para el "propensity score". El efecto promedio vendría establecido por la diferencia del promedio de los resultados registrados por los individuos de los grupos de participantes en el programa y de control en los que el "propensity score" es el mismo.

Analíticamente, el efecto promedio de un programa de formación sobre el conjunto de individuos (participantes y de control) con un mismo valor ε^i del "propensity score" será²⁹:

$$\begin{aligned} \alpha_{EMP} > \varepsilon^i &= E[Y \mid D = 1, \varepsilon(X) = \varepsilon^i] - E[Y \mid D = 0, \varepsilon(X) = \varepsilon^i] = \\ &= E[Y_1 - Y_0 \mid \varepsilon(X) = \varepsilon^i] \end{aligned}$$

Ecuación 30

²⁷ La Ecuación 29 indica que se cumplirá la hipótesis de independencia entre el suceso D y la variable respuesta Y (ingresos laborales) condicionada sobre el vector de variables predeterminadas X. De esta manera se vuelve a forzar la condición de independencia propia de los experimentos aleatorios.

²⁸ Hahn (1998, p. 16) subraya que este método resulta crucial para controlar el sesgo que puede afectar al cálculo de un estimador de los efectos del programa.

²⁹ Un desarrollo amplio del "propensity score" puede encontrarse en Hirano, Imbens y Ridder (2002).

Todo lo anterior permite resumir el proceso para estimar los efectos de un programa de formación, bajo la selección sobre variables y aplicando el método del "propensity score", mediante la aplicación de un procedimiento en dos etapas:

- La primera etapa consiste en determinar el "propensity score" $\varepsilon(X)$.
- La segunda etapa vendría marcada por la utilización del método de subclasificación o de "matching", si bien no utilizando los posibles valores de las variables predeterminadas sino los valores estimados del "propensity score".

Una posibilidad adicional -aunque menos frecuente- que puede utilizarse en la segunda etapa permitiría calcular el efecto promedio del programa a través de la ponderación mediante el "propensity score". Este tercer método propone el cálculo de estimadores de los efectos del programa α_{EMP} y α_{EMPS} mediante la utilización de los valores medios muestrales observados para la variable respuesta Y (ingresos laborales) y los valores estimados para el "propensity score" en la primera etapa, mediante la aplicación de las siguientes expresiones:

$$\hat{\alpha}_{EMP} = \frac{1}{n} \sum_{i=1}^n Y_i \frac{D_i - \hat{\varepsilon}(X_i)}{\hat{\varepsilon}(X_i) \cdot [1 - \hat{\varepsilon}(X_i)]}$$

Ecuación 31

$$\hat{\alpha}_{EMPS} = \frac{1}{n_1} \sum_{i=1}^n Y_i \frac{D_i - \hat{\varepsilon}(X_i)}{1 - \hat{\varepsilon}(X_i)}$$

Ecuación 32

donde $\hat{\varepsilon}(X_i)$ es el valor estimado del "propensity score" para el vector X_i de variables predeterminadas.

7.- Los estudios observacionales (II): los estimadores de diferencias en diferencias.

7.1.- Las características no observables.

Hasta este momento se ha supuesto que las variables predeterminadas son observables. En muchas ocasiones, sin embargo, el desarrollo de la evaluación deja traslucir que los individuos beneficiarios y los de control difieren en características que nos son observables. En estos casos, los métodos de selección sobre observables anteriormente descritos 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.

De manera general hemos tenido oportunidad de comentar que 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 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 estimar mediante un coeficiente 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 como las posteriores al desarrollo del programa de formación, estarán igualmente contaminadas por 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 variables 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 diferenciar la variable respuesta de los individuos participantes con respecto a sus valores antes y después de dicho programa, y volver a diferenciar de nuevo con respecto a la variación producida en la variable respuesta de los individuos de control³⁰.

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

Comparado con los métodos expuestos hasta aquí, el método de diferencias en diferencias presenta como novedad el uso de observaciones hechas en dos momentos

³⁰ Nos remitimos también a la exposición del estimador de diferencias en diferencias en Heckman y Robb (1985).

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.) definimos³¹:

$Y_{1i}(t)$: 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

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

Ecuación 33

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 periodo 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 variables 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 34

³¹ 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).

7.3.- 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 hubieran seguido. Analíticamente,

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

Ecuación 35

A partir de la Ecuación 35, 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 (Y₀) 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 expresión³²

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

Ecuación 36

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,

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

Ecuación 37

³² En la Ecuación 36 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 36 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.

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.

7.4. El cálculo del estimador de diferencias en diferencias vía regresión.

El estimador de diferencias en diferencias también puede ser calculado mediante la realización de una regresión, recurriendo a un modelo semi-paramétrico.

Para exponer el desarrollo a seguir para la determinación de este estimador vía regresión seguiremos el modelo inicial de estimación simple planteado por Ashenfelter y Card (1985), junto con los desarrollos posteriores de Abadie (2003).

Estos autores parten de un modelo lineal con la siguiente expresión:

$$Y = \mu + \lambda \cdot D + \omega \cdot T + \alpha \cdot (D \cdot T) + \varepsilon$$

Ecuación 38

En esta expresión, las variables que intervienen son:

Y : como variable dependiente, que será la variable referente a los resultados potenciales de los individuos.

D : como variable explicativa, de carácter binario, recogerá el hecho de si el individuo participa en el programa ($D = 1$) o es un individuo de control ($D = 0$).

T : como variable explicativa, de carácter binario, indicará el momento del tiempo en el que el individuo es observado, que podrá ser bien en un momento posterior al programa ($T = 1$) o bien en un momento previo ($T = 0$).

El resto son los parámetros que intervienen en el modelo, siendo:

λ y ω : los parámetros que acompañan a las variables explicativas D y T.

μ : el parámetro indicador de efectos fijos en el modelo.

ε : el error aleatorio que incluye el modelo. El valor medio de este error aleatorio será cero: $E[\varepsilon | D, T] = 0$.

α : el parámetro que determina el efecto conjunto de las variable o interacción de las variables explicativas D y T.

El parámetro α será el estimador de diferencias en diferencias del EPSP. Su obtención mediante mínimos cuadrados a partir de este modelo lineal sigue una fórmula similar a la recogida en la Ecuación 37.

La aplicación de este modelo lineal y el cálculo del estimador de EPSP de diferencias en diferencias por el procedimiento de la regresión por mínimos cuadrados presenta, además de la posibilidad de calcular el estimador de una manera más sencilla y rápida, la posibilidad de incluir en el modelo la consideración de variables predeterminadas o contaminantes (X).

$$Y = \mu + \lambda \cdot D + \omega \cdot T + \alpha \cdot (D \cdot T) + \beta \cdot X + \varepsilon$$

Ecuación 39

De todas formas, hay que ser muy cautos en la introducción de variables predeterminadas en este modelo.

Para el caso de que X no varíe en relación con el tiempo (no variará su valor entre los periodos 0 y 1) y para poblaciones relativamente amplias, la incorporación de covariables al modelo no supone ningún problema.

Para el de que X sí varíe a lo largo del tiempo (entre los periodos 0 y 1), esta variable probablemente podrá estar afectada por el tratamiento, con lo cual si la introducimos en el modelo afectará el resultado final del estimador del EPSP introduciendo endogenidad en el modelo.

Una posible forma de solventar este problema y eliminar los problemas de endogenidad incorporados por una covariable que se ve afectada por el factor tiempo sería introducir esta variable en el modelo pero haciéndola interactuar con la variable T, variable indicadora del tiempo.

La expresión del modelo quedaría de la siguiente forma:

$$Y = \mu + \lambda \cdot D + \omega \cdot T + \alpha \cdot (D \cdot T) + \beta_1 \cdot (T \cdot X) + \beta_0 \cdot ((1-T) \cdot X) + \varepsilon$$

Ecuación 40

8.- Conclusiones.

La evaluación de políticas públicas de formación constituye una materia de indudable interés científico que en las últimas décadas ha experimentado un considerable desarrollo gracias, a la aplicación de métodos de estimación de inferencia causal.

Desde una perspectiva teórica, estos métodos pueden aplicarse en experimentos aleatorios si bien, a menudo, están diseñados sobre muestras poblacionales pequeñas. La explicación de ello se encuentra en los elevados costes asociados al desarrollo del experimento.

Como consecuencia de lo anterior, 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 experimentos. 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. Por estas razones, la estimación del efecto causal debe hacerse sobre inferencias hechas a partir de datos observacionales en el contexto de los métodos cuasi-experimentales.

La evaluación de los programas públicos de formación puede desarrollarse a partir de métodos observacionales de entre los que se han destacado tres.

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”.

La principal utilidad que muestran los métodos de subclasificación y "matching" está en su aplicación a los estudios observacionales dentro del marco de la selección sobre observables. Su operatividad, en cambio, se limita a los casos en los que el número de variables predeterminadas a tener en cuenta sea pequeño.

En estudios donde el número de variables predeterminadas es elevado resultan adecuados los métodos basados en el "propensity score" o la "propensión a participar".

El segundo de los métodos observacionales es el método del estimador de “diferencias en diferencias”. En nuestra opinión se trata del método más operativo de los analizados. Ello se debe a que resulta frecuente que los individuos beneficiarios y los de control difieran en características que nos son observables. En estos casos, los métodos de selección sobre observables anteriormente descritos 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.

Finalmente, el tercer método observacional es el de las variables instrumentales y ha quedado fuera de esta investigación.

Bibliografía.

Abadie, A. (2003): “Semiparametric Difference-in-Differences Estimators”, Working Paper, abril 2003. Universidad de Harvard y NBER.

Abadie, A., J. D. Angrist y G. W. Imbens (2001): “Instrumental Variables Estimation of the Effect of Subsidized Training on the Quantiles of Trainee Earnings”, *Econometrica*, vol. 70, 91-117.

Abadie, A. y G. Imbens (2004): “Large Sample Properties of Matching Estimators for Average Treatment Effects”. Working Paper, Enero 2004. Universidad de Harvard, Universidad de Berkeley y NBER.

Abadie, A. y G. Imbens (2002): “Simple and Bias-Corrected Matching Estimators for Average Treatment Effects”, Working Paper, junio 2002. Universidad de Harvard, Universidad de Berkeley 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. y G. W. Imbens (1991): “Sources of Identifying Information in Evaluation Models”. NBER Technical Working Paper 117.

Ashenfelter, O. (1978): “Estimating the Effect of Training Programs on Earnings”, *Review of Economics and Statistics*, vol. 60, 47-57.

Ashenfelter, O. y D. Card (1985): “Using the Longitudinal Structure of Earnings to Estimate the Effects of Training Programs”, *Review of Economics and Statistics*, vol. 67, 648-660.

- Barnow, B., G. Cain y A. Goldberger (1981): "Selection on Observables". *Evaluation Studies Review Annual* 5, Beverly Hills, Sage, 43-59.
- 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.
- Campbell, D. T. and J.C. Stanley (1966): "Experimental and Quasi-Experimental Designs for Research". Rand McNally, Chicago.
- 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.
- Dawid, A. P. (1979): "Conditional Independence in Statistical Theory", *Journal of the Royal Statistical Society*, Series B 41, 1-31.
- Fisher, R. A. (1935): "The design of experiments", chapter 2, "The principles of experimentation, illustrated by a psycho-physical experiment." Oliver and Boyd, Edinburgh.
- Hahn, J. (1998): "On the Role of the Propensity Score in Efficient Semiparametric Estimation of Average Treatment Effects", *Econometrica*, vol. 66, 315-331.
- 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., H. Ichimura y P. E. Todd (1998): "Matching As an Econometric Evaluation Estimator", *Review of Economics Studies*, 65, 261-294.
- Heckman, J. J. y J. A. Smith (1995): "Assessing the Case for Social Experiments", *Journal of Economic Perspectives*, vol. 9 (2), 85-100.
- Hirano, K., G. Imbens, and G. Ridder, (2002), "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score". Research Paper n° 02-13. USC for Law, Economics and Organization.
- Holland, P. W. (1986): "Statistics and Causal Inference (with discussion)", *Journal of the American Statistical Association*, vol. 81, 945-970.

- 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.
- Neyman, J. (1990): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles. Section 9", *Statistical Science* 5 (4), 465-480.
- 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: Srpinger-Verlag.
- Rosenbaum, P. R. (1999): "Choice As an Alternative to Control in Observational Studies", *Statistical Science*, 14, 259-304.
- Rosenbaum, P. R. y D. B. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects", *Biometrika*, 70, 41-55.
- Rubin, D. B. (1973): "Matching to remove bias in observational studies". *Biometrics*, vol. 29, 159-183.
- 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., (1977), "Assignment to a Treatment Group on the Basis of a Covariate," *Journal of Educational Statistics*, vol. 2, 1-26.

Rubin, D. B., (1978), “Bayesian Inference for Causal Effects the Role of Randomization,” *Annals of Statistics*, vol. 6, 34-58.

Rubin, D. B. (1990): “Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies”, *Statistical Science*, 5, 472-480.