

# Evaluación microeconómica de las políticas públicas de empleo: una introducción

Christian Durán  
Departament d'Économia  
Universitat Rovira i Virgili  
Avinguda de la Universitat, 1  
43204 REUS

22 de noviembre de 2004

## Índice

<b>1. Introducción</b>	<b>4</b>
<b>2. Efectos causales de la política pública y sesgo de selección muestral</b>	<b>8</b>
<b>3. Supuestos identificativos y procedimientos de estimación</b>	<b>14</b>
3.1. Experimentos sociales: el papel de la aleatoriedad . . . . .	15
3.2. El estimador antes-después . . . . .	16
3.3. Métodos estadísticos de emparejamiento . . . . .	20
3.4. Modelos econométricos de selección . . . . .	25
3.5. Límites no-paramétricos . . . . .	31
<b>4. La elección del método adecuado: problemas no resueltos de la evaluación</b>	<b>35</b>
<b>5. Conclusión</b>	<b>43</b>
<b>6. Referencias bibliográficas</b>	<b>48</b>

## Resumen

En este trabajo se muestran de forma esquemática los problemas fundamentales de evaluación de las políticas públicas. El impacto de un determinado programa de formación en un participante se mide como la diferencia entre el valor de la variable-resultado (salario, duración del desempleo) alcanzado en su situación actual, en la que ha participado en el programa y el valor de su resultado contrafactual en la situación hipotética de que no hubiese participado en el programa. Sin embargo, el resultado contrafactual no puede observarse por lo que es necesario recurrir a su estimación mediante métodos econométricos y estadísticos.

Palabras clave: Evaluación de políticas activas del mercado de trabajo; modelo de Roy-Rubin; Métodos experimentales y no-experimentales; (JEL 824, 23, 30).

## Resumen

In this paper the fundamental evaluation problem of active labor market policies is outlined. The causal impact of a training program on the trainees is measured as the difference between the value the outcome variable takes (earnings, unemployment duration) in the actual situation with training and the value of the counterfactual outcome in the hypothetical situation of non participation in the program. Nevertheless, the counterfactual outcome is not observable so it becomes necessary to estimate it applying different econometric and statistical methods.

Keywords: Employment Program evaluation; Roy-Rubin Model; Experimental and Non-Experimental Methods; (JEL 824, 23, 30).

## 1. Introducción

El desarrollo económico experimentado por los países industrializados durante los últimos treinta años se ha caracterizado por una rápida innovación tecnológica y unas nuevas estructuras productivas que han provocado un importante incremento de la productividad del trabajo. Esta evolución ha inducido también unos importantes cambios en las necesidades en las empresas de trabajadores cualificados. Por este motivo, los trabajadores tienen que adquirir constantemente nuevas capacidades para poder hacer frente a los requerimientos de cada puesto de trabajo en un entorno empresarial en permanente transformación. Obviamente, la educación primaria, así como la educación secundaria o la universitaria no son suficientes para garantizar una rápida y prolongada inserción en el mercado laboral. El papel de la formación ocupacional a lo largo de la vida laboral de un trabajador adquiere una relevancia cada vez mayor.

Especialmente en aquellos países en los que el desempleo es persistente y se registra una elevada proporción de desempleados de largo plazo –como es el caso de muchos países industrializados– la necesidad de participar en un programa de formación ocupacional parece particularmente importante para los trabajadores con una menor cualificación dado que conforman el grupo social más castigado por el desempleo. Por este motivo, una estrategia común que aplican muchos gobiernos para atajar este problema consiste en el diseño y la puesta en marcha de políticas activas del mercado de trabajo. Una parte importante de los recursos públicos dedicados a financiar las medidas de formación ocupacional se dirigen a los trabajadores menos cualificados. En un escenario ideal, la financiación pública de la formación ocupacional incrementará las capacidades de los participantes, aumentando así las perspectivas futuras de empleo y de su contribución a través de impuestos al erario público y reduciendo su dependencia del sistema de prestaciones por desempleo.

Sin embargo, la evaluación del impacto de los programas de formación financiados por el sector público no es una tarea fácil. La medición del efecto de los programas de formación (u otros programas activos de empleo) sobre las diferentes características laborales del individuo como, por ejemplo, su remuneración, sus episodios de paro, la duración del desempleo o de la ocupación, precisa de una definición clara del concepto

de “impacto causal del programa de formación”. Para un determinado participante en el programa, el impacto se mide como la diferencia entre el valor de su variable-resultado (por ejemplo, salario) en su situación actual con participación y el valor de su resultado contrafactual en la situación hipotética de que no hubiese participado en el programa de formación. En otras palabras, se pretende medir la mejora del resultado inducido exclusivamente por el programa de formación.

Dado que no se conoce el resultado contrafactual es necesario estimarlo a partir de los valores de otros individuos que no han recibido la formación ocupacional objeto de la evaluación. Estos individuos forman parte del denominado grupo de control o de comparación. Cuando se seleccionan los no-participantes para su inclusión en el grupo de control con el objeto de estimar el resultado contrafactual pueden seguirse dos vías: o bien se encuentran individuos idénticos a los participantes con respecto a todas las características relevantes excepto la participación en el programa de formación, o bien se controla por todas las diferencias relevantes que existen entre los participantes y los no-participantes. Si no se tienen en cuenta estas discrepancias entre el grupo de participantes y el grupo de control pueden deducirse unas conclusiones incorrectas sobre el verdadero impacto del programa. Por ejemplo, si aquellos individuos que muestran generalmente unos mejores resultados en su búsqueda de empleo, tienen también una mayor probabilidad de participar en el programa, una comparación de los resultados post-programa entre el grupo de participantes y el grupo de control tenderá a sobreestimar el verdadero efecto del programa de formación. En la literatura econométrica este problema se conoce como sesgo de selección muestral y se estudia en el apartado segundo de este trabajo.

En el apartado tercero se abordan los planteamientos estadísticos y econométricos más relevantes para la evaluación de programas públicos de formación<sup>1</sup>. Se distinguen dos métodos de resolución del problema de selección muestral: el enfoque experimental y el enfoque no-experimental. En el método experimental –al cual se hace referencia en primer lugar– el programa a evaluar se implementa como un experimento social en el que la construcción de un grupo de control adecuado se realiza mediante una asignación aleatoria –al grupo de individuos que reciben formación o al

grupo de control— de todos los individuos que participan en el experimento. Cuando el diseño del experimento es adecuado el enfoque experimental generará estimadores insesgados del impacto del programa.

Sin embargo, no siempre es posible implementar estos experimentos aleatorios por lo que es necesario recurrir a observaciones individuales recogidas para otro cometido de individuos que no tienen nada que ver con el programa a evaluar. La medición del impacto de las políticas de formación a partir de datos no-experimentales tiene que tener en cuenta el proceso no aleatorio de selección de los participantes. Los diferentes procedimientos econométricos y estadísticos utilizados se analizan a continuación de los métodos experimentales y abarcan el estimador antes-después, los métodos de emparejamiento y los modelos econométricos de selección.

El estimador antes-después compara el resultado obtenido por el mismo individuo antes de participar en el programa y después de su participación en la formación. De esta forma los mismos individuos conforman el grupo de tratamiento y el grupo de control aunque la medición del resultado se realiza en instantes diferentes. Los métodos estadísticos de emparejamiento comparan los resultados obtenidos de un individuo del grupo de participantes con un individuo del grupo de control. Ambos individuos deben ser lo más parecidos posible en las características relevantes para el objeto de estudio de forma que el resultado observado para el individuo del grupo de control pueda utilizarse como aproximación al resultado contrafactual del individuo que ha participado en el programa. Los modelos econométricos de selección, finalmente, se componen de dos ecuaciones: la ecuación de resultados que incluye como variable explicativa una variable dummy que refleja la participación de un individuo en el programa de formación y la ecuación de participación que estima la probabilidad de que un individuo sea seleccionado para participar en el programa en función de unas determinadas características del individuo. La estimación conjunta de ambas ecuaciones ofrece una aproximación del impacto del programa que se pretende evaluar. Sin embargo, no siempre es posible obtener un estimador del impacto causal medio de un programa de formación. Por este motivo se hace referencia también al cálculo de los límites no-paramétricos que permiten acotar los valores que puede tomar el estimador del efecto causal.

En el apartado cuarto se formula la pregunta de cuál de los métodos presentados es el más útil para evaluar las políticas públicas. Recientes comparaciones de la bondad de diferentes enfoques no-experimentales con los resultados obtenidos en evaluaciones experimentales no permiten obtener una conclusión clara. Mientras que los procedimientos basados en los modelos econométricos de selección producen estimaciones muy poco fiables y notablemente distintos de los resultados obtenidos mediante experimentos sociales<sup>2</sup>, los métodos de emparejamiento estadístico ofrecen unos mejores resultados, mostrando una mayor capacidad para reproducir los resultados experimentales. Finalmente, el apartado quinto recoge las conclusiones.

## 2. Efectos causales de la política pública y sesgo de selección muestral

La inferencia causal acerca del impacto de un determinado programa público de formación sobre la respuesta o el resultado de un individuo (por ejemplo, la duración del desempleo de un individuo que ha participado en el programa a evaluar) incluye también alguna conjetura sobre cuál hubiese sido su respuesta en el caso de no haber participado en el programa. En la literatura econométrica y estadística es bastante frecuente partir en este tipo de análisis del modelo de ROY (1951) y de RUBIN (1974)<sup>3</sup>. En este modelo se parte de los resultados potenciales ( $Y^1, Y^0$ ) para cada individuo, donde  $Y^1$  denota el resultado alcanzado por el individuo en el caso de participar en el programa de formación, mientras que  $Y^0$  hace referencia al resultado sin participación<sup>4</sup>. Sea  $D$  la variable dummy que indica si un individuo ha participado en el programa de formación que desea evaluarse ( $D = 1$ ) o no lo ha hecho ( $D = 0$ ). Los individuos que participan en el programa conforman el grupo de tratamiento, mientras que los individuos que no han participado, pero que sirven de punto de referencia con el cual se compara el primero, se agrupan en el grupo de control. Para poder llevar a cabo el análisis causal deben ignorarse los efectos cruzados o los efectos de equilibrio general entre los participantes potenciales. De esta forma se garantiza que el resultado potencial de un individuo no esté afectado por el status formativo de los demás individuos. En la literatura estadística este supuesto recibe el nombre de “supuesto de valor estable de la unidad tratada”<sup>5</sup>. El efecto causal de la formación para cada individuo se define

entonces como la diferencia entre sus resultados potenciales:

$$\Delta = Y^1 - Y^0. \quad (1)$$

El problema fundamental a la hora de evaluar el efecto causal de la formación es que el resultado observado para cada individuo depende de si ha recibido o no la formación:

$$Y = D Y^1 + (1 - D) Y^0, \quad (2)$$

Sólo podemos observar o bien  $Y^1$  o bien  $Y^0$ , pero nunca las dos variables resultado simultáneamente. Siempre existe un componente no observable en (1) el cual recibe en la literatura el nombre de resultado contrafactual. Para un participante en el programa de formación este resultado contrafactual es  $Y^0$ , porque es imposible saber cuál hubiese sido el resultado alcanzado sin formación para un individuo que –de hecho– ha participado en el programa. En analogía, el resultado contrafactual para un no participante es  $Y^1$ . Así, el problema de evaluar el efecto causal de la formación es un problema de datos ausentes (missing data) dado que no puede calcularse la diferencia expresada en la ecuación (1) para cada individuo. Es más, el efecto causal de la formación –definido como la diferencia entre dos resultados potenciales– debe distinguirse de la correlación entre el status de formación  $D$  y el resultado observado. Mientras que el efecto causal definido en la ecuación (1) no depende de la participación en la formación, la correlación entre  $D$  e  $Y$  sí lo hace<sup>6</sup>.

Dado que no es posible determinar el impacto de la formación para un individuo en particular, el análisis puede centrarse en la distribución  $F(\cdot)$  de los impactos entre los individuos de toda la población:

$$F(\Delta) = F(Y^1 - Y^0), \quad (3)$$

o en determinadas características de esta distribución. Tal y como apuntaron HECKMAN, SMITH (1993) o HECKMAN, SMITH, CLEMENTS (1997) existen muchos aspectos de la distribución de los impactos de un programa que son de interés económico o político. Esto incluye el impacto medio o el impacto mediano del programa, así como cuantiles concretos de la distribución de los impactos. Por ejemplo, aún cuando el efecto medio o mediano de un programa sea igual a cero, puede ser importante saber

qué segmento del grupo de participantes mejoran sus resultados con el programa. El parámetro que recibe la mayor atención en la literatura y en el que se centra también este trabajo es el de impacto medio. Esto es así porque, exceptuando la media muestral, todos los demás parámetros de la distribución de impactos presentan la desventaja fundamental desde la perspectiva econométrica de que no pueden estimarse de forma consistente a partir de las distribuciones marginales de  $Y^1$  e  $Y^0$ , puesto que dependen de las características de la distribución conjunta de  $(Y^1, Y^0)$ <sup>7</sup>.

Para los individuos extraídos aleatoriamente de una población el impacto esperado (o medio) de la formación (Average Treatment Effect: ATE) se define como:

$$\begin{aligned} E(\Delta) &= E(Y^1 - Y^0) = E(Y^1) - E(Y^0) \\ &= \{E(Y^1|D = 1) - E(Y^0|D = 1)\}P(D = 1) \\ &\quad + \{E(Y^1|D = 0) - E(Y^0|D = 0)\}P(D = 0). \end{aligned} \quad (4)$$

$E(\Delta)$  es la media ponderada de dos términos: el término entre los primeros corchetes es el efecto medio del programa sobre los participantes y el término entre los segundos corchetes es el impacto medio en el grupo de control. Las ponderaciones vienen determinadas por la probabilidad de participar o no participar en el programa, respectivamente. Sin embargo, HECKMAN (1997) apunta que este parámetro no está estrictamente relacionado con los individuos a los que, en principio, está dirigido el programa de formación. Por este motivo no suele ser relevante para la política económica.

Parece más razonable centrar el análisis en el primer componente de la ecuación (4), esto es, en el efecto medio del programa sobre el grupo de participantes. Se analiza, por tanto, el impacto del programa en aquellos individuos que han participado en él (Selected Average Treatment Effect, SATE):

$$E(\Delta|D = 1) = E(Y^1 - Y^0|D = 1) = E(Y^1|D = 1) - E(Y^0|D = 1). \quad (5)$$

Este parámetro responde a la siguiente pregunta: ¿cuál es el beneficio esperado (o medio) de los individuos que han participado en el programa de formación comparado con el caso hipotético de que no hubiesen participado? En la medida en que esta pregunta se centra explícitamente en los

resultados obtenidos por los participantes en el programa de formación, es necesario calcular la mejora bruta causada por la participación en el programa. Esta mejora conforma, junto con los costes del programa, un elemento central para la decisión de continuar o finalizar el programa de formación<sup>8</sup>.

El efecto medio del tratamiento en los participantes puede definirse también para una determinada submuestra de individuos participantes los cuales presentan la característica  $X$ , de forma que diferentes grupos de participantes pueden mostrar diferentes resultados como consecuencia de la formación:

$$E(\Delta|X, D = 1) = E(Y^1 - Y^0|X, D = 1) = E(Y^1|X, D = 1) - E(Y^0|X, D = 1). \quad (6)$$

Para interpretar este parámetro es necesario que –dados los resultados potenciales en ambos estados ( $Y^1, Y^0$ )– la realización actual de  $D$  no cause las características relevantes recogidas en el vector  $X$ , esto es,  $X$  no está determinado por la realización de  $D$ . Si fuera así el efecto del programa estaría condicionado por variables determinadas por el propio programa<sup>9</sup>. El parámetro definido en la ecuación (6) mide el efecto esperado de la formación en los individuos que presentan una determinada realización de la característica  $X$ . Por ejemplo, si  $X$  describe la edad de los individuos, es posible definir el impacto esperado de la formación para un participantes de una determinada edad.

La identificación y estimación del efecto medio del programa en los participantes en su definición más general (ecuación (5)) se enfrenta todavía al problema fundamental de la evaluación, es decir, al problema del sesgo muestral. Generalmente, con datos no-experimentales sólo el primer término  $E(Y^1|D = 1)$  en la definición de  $E(\Delta|D = 1)$  puede identificarse y estimarse de forma consistente a partir de los datos de los individuos de la muestra. Sin embargo, no se dispone de la información necesaria para estimar el término contrafactual  $E(Y^0|D = 1)$ . Un objetivo común de muchos de los enfoques tanto estadísticos, como econométricos, es la estimación del término contrafactual no observable  $E(Y^0|D = 1)$ , utilizando la información contenida en los resultados observados del grupo de control. El estimador del impacto basado en el resultado medio de los no participantes

como variable proxy para  $E(Y^0|D = 1)$  se define entonces como:

$$\tau = E(Y^1|D = 1) - E(Y^0|D = 0). \quad (7)$$

Esta expresión puede descomponerse en dos términos<sup>10</sup>:

$$\begin{aligned} \tau &= \{E(Y^1|D = 1) - E(Y^0|D = 1)\} + \{E(Y^0|D = 1) - E(Y^0|D = 0)\} \\ &= E(\Delta|D = 1) + B, \end{aligned} \quad (8)$$

El primer término entre corchetes refleja el impacto medio del programa de formación en los individuos que han participado en él, mientras que el segundo término entre corchetes describe el sesgo de selección muestral  $B^{11}$ . El impacto estimado a partir de la ecuación (7) sólo representa el verdadero impacto cuando el sesgo de la selección es igual a cero. La condición necesaria para que el sesgo se anule es, por consiguiente:

$$B = 0 \iff E(Y^0|D = 1) = E(Y^0|D = 0), \quad (9)$$

es decir, el sesgo de selección muestral es igual a cero cuando el resultado potencial esperado sin formación es igual para ambos grupos, esto es, para el grupo de participantes ( $D = 1$ ) y el grupo de no participantes ( $D = 0$ ). Sin embargo, en la práctica esta condición no suele cumplirse. Por ejemplo, si aquellos individuos que participan en un programa de formación –diseñado para incrementar la probabilidad individual de encontrar un empleo– muestran de entrada una mayor probabilidad de colocación antes de la formación en comparación con aquellos individuos que no participan en el programa –posiblemente a causa de un mayor nivel formativo o una mayor ambición–, entonces el grupo de control no es comparable al grupo de participantes. Expresado de otra forma, aquellos individuos que participan en el programa de formación tienen a priori unos niveles más elevados de  $Y^0$  (por ejemplo, una mayor probabilidad de ocupación) incluso cuando no participan en un programa de formación, que aquellos que están excluidos del programa. Así, el efecto estimado de la formación sobrevalora el impacto del programa puesto que incluye la mayor probabilidad media de colocación que muestran los participantes incluso antes de que finalice la formación. El sesgo de selección será, por tanto, mayor que cero.

El problema de inferir sobre el término contrafactual a partir de la información observada del grupo de control generó una extensa literatura

econométrica y estadística sobre la evaluación de programas<sup>12</sup>. Un objetivo importante de esta literatura es encontrar unos supuestos plausibles de identificación y unas técnicas de estimación que produzcan estimadores insesgados del resultado contrafactual y, por extensión, del efecto causal de la formación. Obviamente, con independencia de los supuestos identificativos utilizados y de las técnicas econométricas aplicadas, el riesgo de obtener estimadores sesgados está siempre presente puesto que una parte del sesgo de selección –independientemente de la definición de éste– no puede observarse. Puesto que no existe una solución general al problema fundamental de la evaluación, la elección de los supuestos identificativos utilizados y de las técnicas de estimación deberá depender siempre de la situación específica investigada. En otras palabras, es importante reconocer que diferentes circunstancias, por ejemplo, diferentes datos, tipos de programas o resultados observados requieren diferentes enfoques. Además, la elección de un enfoque específico siempre tiene que basarse en consideraciones de plausibilidad puesto que su idoneidad no puede ser contrastada.

### **3. Supuestos identificativos y procedimientos de estimación**

La evaluación de un programa de formación se basa fundamentalmente en la comparación entre los resultados obtenidos por los participantes en el programa y los resultados mostrados por los no-participantes. En este apartado se analizan cuatro procedimientos de estimación del impacto de la formación en los participantes. En los cuatro procedimientos se construye de una forma u otra el resultado contrafactual alcanzado por los no-participantes. Así, en los experimentos sociales los potenciales participantes en un programa de formación se asignan aleatoriamente a un grupo de participación y a un grupo de control que sirve de punto de referencia. En cuanto al estimador antes-después, se toma como resultado contrafactual aquel resultado mostrado por los participantes antes del inicio del programa de formación. En los métodos de emparejamientos el resultado contrafactual se determina a partir de los datos disponibles de un grupo de no-participantes lo más similar posible al grupo de participantes con el objeto de obtener un grupo de control comparable. Finalmente, los modelos econométricos de selección estiman el impacto causal del programa a partir de los datos del grupo de participantes y el grupo de control. Sin embargo,

no siempre es posible estimar el impacto medio del programa aunque, en ocasiones, puede definirse un intervalo en el que se encuentra dicho estimador. Así, en la última parte de este apartado se presenta un método no-paramétrico que permite calcular los límites de este intervalo.

### 3.1. Experimentos sociales: el papel de la aleatoriedad

En un experimento social los individuos a los que se dirige un determinado programa de formación son asignados de forma aleatoria a un grupo de participantes que reciben una determinada formación y un grupo de control que no lo hace. Así, los grupos no difieren sistemáticamente entre ellos con respecto de ninguna característica relevante, a excepción de la participación en el programa. En consecuencia, la diferencia en el resultado después del programa de formación se supone inducido únicamente por el programa en sí, es decir, se mide únicamente el impacto del programa y no debería existir sesgo alguno de selección muestral. En términos formales, el objetivo de la asignación aleatoria de los individuos de la muestra al grupo de participantes y al grupo de control es el de garantizar que los resultados potenciales sean independientes de la asignación a un programa de formación ( $D$ )<sup>13</sup>:

$$Y^1, Y^0 \perp\!\!\!\perp D.$$

De esta forma se cumple que:

$$E(Y^0|D = 1) = E(Y^0|D = 0). \quad (10)$$

con lo que los resultados mostrados por el grupo de control pueden utilizarse para estimar consistentemente el resultado contrafactual y, en consecuencia, también el efecto causal de la formación  $E(\Delta|D = 1)$  definido en la ecuación (5).

### 3.2. El estimador antes-después

A primera vista la solución más sugerente al problema fundamental de evaluación a partir de datos no-experimentales sea probablemente el estimador antes-después. Con este estimador se intenta evitar el problema del sesgo muestral, utilizando únicamente las observaciones de los participantes en el programa de formación. De hecho, los datos recogidos en los

periodos previos a la participación en el programa se utilizan como resultados de referencia, es decir, se comparan los resultados previos y posteriores al programa de los mismos individuos. En consecuencia, para calcular el estimador antes-después es necesario disponer de datos del mismo individuo antes y después de su participación en el programa. Para entender las restricciones identificativas del estimador antes-después es necesario distinguir entre las variables-resultado observables cuyos valores se determinan antes de la participación en el programa (resultados pre-programa  $Y_{-t}$ ) y aquellas variables-resultado cuyos valores se determinan posteriormente (resultados post-programa  $Y_{+t}$ ). Los resultados potenciales pre- y post-tratamiento pueden escribirse como  $\{Y_{-t}^1, Y_{+t}^1\}$  y  $\{Y_{-t}^0, Y_{+t}^0\}$  con y sin tratamiento, respectivamente. Dado que el impacto causal del tratamiento mide el efecto de la participación en el programa en relación a la exclusión, parece obvio que los valores relevantes de las variables-resultado se localizan en el periodo post-programa. Se supone, por tanto, que los valores de las variables-resultado antes del programa no están afectados por el programa. El efecto del tratamiento para cada individuo se define como la diferencia entre sus resultados potenciales después del programa:

$$\Delta = Y_{+t}^1 - Y_{+t}^0. \quad (11)$$

La idea en la que se basa el estimador antes-después es que el resultado observable de los participantes antes del programa ( $Y_{-t} = Y_{-t}^1$ ) representa una descripción válida del resultado contrafactual no observable sin participación ( $Y_{+t}^0$ ). A partir de este supuesto puede obtenerse un estimador del efecto medio del programa sobre los participantes tomando la diferencia entre el resultado medio de los participantes antes y después del programa<sup>14</sup>:

$$\tau^{ead} = E(Y_{+t}^1 | D = 1) - E(Y_{-t}^1 | D = 1). \quad (12)$$

El sesgo potencial resultante del estimador antes-después viene dado por la diferencia entre  $E(\Delta | D = 1)$  y  $\tau^{ead}$ :

$$B^{ead} = \tau^{ead} - E(\Delta | D = 1) = E(Y_{+t}^0 | D = 1) - E(Y_{-t}^1 | D = 1). \quad (13)$$

En consecuencia, el supuesto central de identificación para el estimador antes-después puede escribirse como:

$$B^{ead} = 0 \iff E(Y_{+t}^0 | D = 1) = E(Y_{-t}^1 | D = 1). \quad (14)$$

La validez de este supuesto depende de diferentes condiciones. En primer lugar, es importante que el resultado potencial pre-programa con participación ( $Y_{-t}^1$ ) no esté influido por la participación en el programa. En este caso puede suponerse que  $E(Y_{-t}^1|D = 1) = E(Y_{+t}^0|D = 1)$ . Este supuesto posiblemente no sea válido si los individuos tienden que comportarse de una manera diferente a su forma habitual para entrar en el programa (por ejemplo, registrarse en las oficinas del INEM) o se comportan de forma diferente, anticipando su futura participación en el programa de formación (por ejemplo, reduciendo sus esfuerzos de búsqueda de empleo).

Un segundo supuesto hace referencia a la constancia en el tiempo del resultado potencial, asumiendo implícitamente que  $E(Y_{-t}^0|D = 1) = E(Y_{+t}^0|D = 1)$ . Esto, sin embargo, sólo es cierto cuando no hay efectos que influyan en el resultado potencial de forma diferente en un periodo y en el otro. Si, por ejemplo, los cambios en las características individuales (edad) o en el comportamiento individual (esfuerzos de búsqueda de empleo inducidos por la desmotivación del desempleado) entre  $-t$  y  $+t$  afectan al resultado potencial, entonces se incumple el supuesto de constancia en el tiempo. Además de estos cambios individuales, también las variaciones sistemáticas en las variables agregadas (diferentes niveles de la demanda agregada en  $-t$  y en  $+t$ ) pueden influir a lo largo del tiempo en el resultado potencial. En general puede decirse que el supuesto de constancia es tanto más realista cuanto menor sea la distancia entre las mediciones de las variables-resultado realizadas previa y posteriormente al programa.

Finalmente, un buen ejemplo empírico citado con frecuencia en el que se incumple el supuesto de la anulación del sesgo,  $B^{ead} = 0$ , fue estudiado por ASHENFELTER (1978). Este autor observó que los participantes en los programas de formación experimentaron una reducción de sus ingresos en los periodos anteriores al programa. Esta regularidad se conoce como Ashenfelter's dip o preprogram dip y se observa también en otros estudios empíricos aunque a veces en otras variables distintas como, por ejemplo, la probabilidad de colocación. El problema central asociado al estimador antes-después es saber si el preprogram dip es permanente o transitorio<sup>15</sup>. Si el salto en la variable-resultado es transitorio y se restableciera el valor anterior incluso en ausencia de la participación en la programa,

el estimador antes-después sobreestimaría el impacto del programa de formación cuando la disminución de la variables-resultado antes del programa tiene lugar en el periodo utilizado para medir el resultado pre-programa. Por otro lado, si el pre-program dip es persistente, el mismo estimador antes-después subestimaría el impacto del programa cuando el pre-program dip tiene lugar durante o después del periodo utilizado para medir el resultado pre-programa.

Para tener en cuenta una tendencia generalizada en la variable-resultado se calcula el estimador de diferencias-en-diferencias donde la diferencia media para los participantes se compara con la diferencia media que registran los excluidos. Este estimador se basa en el supuesto de que la tasa de crecimiento de la variable resultado es la misma para ambos grupos<sup>16</sup>. En consecuencia, la variación de la variable-resultado es la misma para el grupo de tratamiento que para el grupo de control en el caso de que ninguno de los dos grupos hubiera participado en el programa. El supuesto crítico de identificación es la constancia del sesgo muestral a lo largo del tiempo por lo que la diferenciación de las diferencias entre participantes y no-participantes elimina el sesgo<sup>17</sup>:

$$E(Y_{+t}^0 - Y_{-t}^0 | D = 1) = E(Y_{+t}^0 - Y_{-t}^0 | D = 0).$$

Si se cumple este supuesto puede identificarse el impacto del programa de formación en los participantes:

$$E(Y_{+t}^1 - Y_{-t}^0 | D = 1) = \{E(Y_{+t} | D = 1) - E(Y_{+t} | D = 0)\} - \{E(Y_{-t} | D = 1) - E(Y_{-t} | D = 0)\}.$$

La figura 1 representa gráficamente la noción del estimador de diferencias-en-diferencias.

### 3.3. Métodos estadísticos de emparejamiento

El método de emparejamiento aplicado a la evaluación no-experimental de programas de gasto público fue propuesto por primera vez por RUBIN (1974, 1977) y se basa en gran medida en la misma filosofía que inspira los experimentos sociales. El procedimiento de emparejamiento tiene como objetivo construir un grupo adecuado de control a partir de datos no-experimentales. Por tanto, es necesario identificar a partir de un amplio grupo de no-participantes aquellos individuos que sean similares en todas

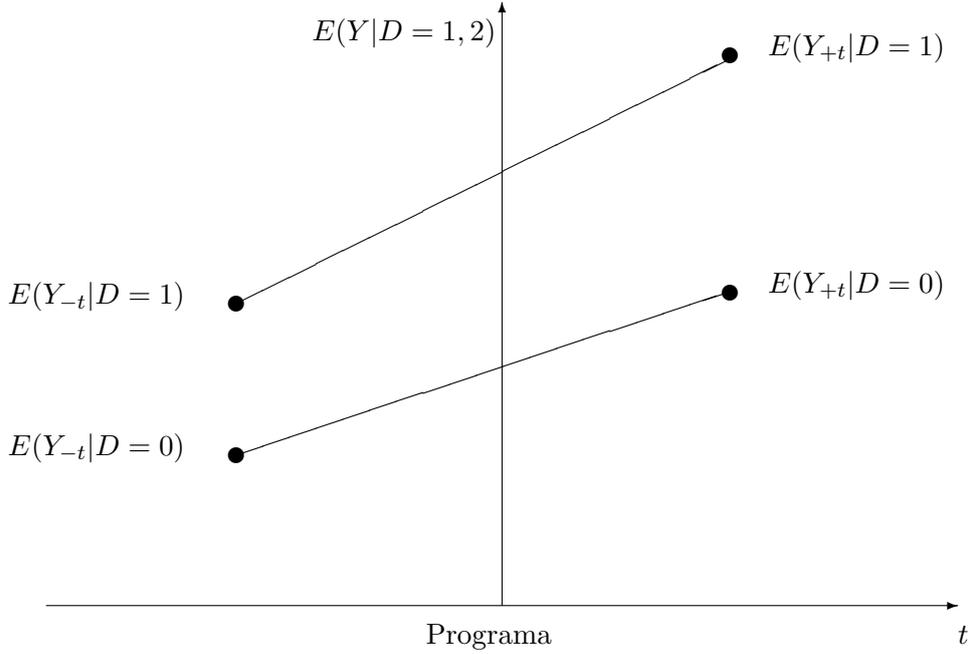


Figura 1: Estimador de diferencias-en-diferencias del impacto medio

las características “relevantes” a los participantes. Las diferencias en los resultados entre los excluidos comparables y los participantes se atribuyen, por consiguiente, al programa de formación. El supuesto necesario para identificar el efecto causal medio del tratamiento es que los resultados potenciales sean independientes de la participación en el programa  $D$  condicionado a unas características dadas  $X$  llamadas covariables. Esta premisa recibe el nombre de supuesto de independencia condicionada<sup>18</sup>:

$$(Y^0, Y^1) \perp\!\!\!\perp D | X. \quad (15)$$

Este supuesto se cumple cuando

$$E(Y^0 | X, D = 1) = E(Y^0 | X, D = 0) = E(Y^0 | X). \quad (16)$$

Aplicando esperanzas iterativas puede reescribirse el resultado contrafactual que interviene en el efecto medio del programa  $E(\Delta | D = 1)$  como:

$$E(Y^0 | D = 1) = E_X (E(Y^0 | X, D = 1) | D = 1).$$

Sustituyendo esta expresión en la ecuación (16) se obtiene

$$E(Y^0|D = 1) = E_X (E(Y^0|X, D = 0)|D = 1) .$$

El estimador que se deriva del supuesto de independencia condicional puede escribirse entonces como:

$$\begin{aligned} \tau &= E(Y^1|D = 1) - E(Y^0|D = 1) \\ &= E(Y^1|D = 1) - E_X (E(Y^0|X, D = 0)|D = 1) \\ &= E_X (E(Y^1|X, D = 1) - E(Y^0|X, D = 0)|D = 1) . \end{aligned} \quad (17)$$

Obsérvese que la esperanza exterior se calcula a partir de la distribución de las covariables relevantes  $X$  en el grupo de participantes ( $D = 1$ ). En consecuencia, la estimación del resultado contrafactual a partir del grupo de control exige tener en cuenta la distribución de las covariables  $X$  en el grupo de participantes<sup>19</sup>. En consecuencia, el efecto causal medio se calcula a través de la ecuación (5).

La implementación de la esperanza condicionada a través de los métodos de emparejamiento significa emparejar cada participante con un excluido que muestre los mismos valores en las covariables<sup>20</sup>. Para garantizar la posibilidad de formar parejas, ROSENBAUM, RUBIN (1983) introducen la siguiente condición de ignorabilidad:

$$0 < P(D = 1|X) \equiv P(X) < 1, \quad (18)$$

es decir,  $X$  no debería contener características para las que la probabilidad de participar en el programa –denominada como propensity score– sea igual a cero o a uno. Si, por ejemplo, la muestra contiene individuos con unas características tal que  $P(X) = 1$ , significaría que estos individuos siempre participarían en el programa de formación. Esto impediría la formación de parejas dado que, por definición, no existirían individuos excluidos con las mismas características  $X$ . Obsérvese que en analogía al proceso aleatorio de asignación en los experimentos sociales, el objetivo de los métodos de emparejamiento es la igualación de las distribuciones de las características relevantes  $X$  entre el grupo de participantes y el grupo de control para alcanzar así la independencia entre los resultados potenciales y la asignación al programa. La expresión “relevante” hace referencia a todas aquellas covariables que influyen tanto en la asignación al programa

como a los resultados potenciales. Además, se supone que las covariables no dependen de la participación o no en la formación puesto que esto distorsionaría el impacto total del programa.

HECKMAN, ICHIMURA, TODD (1998) apuntan que los supuestos (15) y (18) son más restrictivos de lo que realmente es necesario. Para construir el resultado contrafactual  $E(Y^0|X, D = 1)$  es suficiente la condición más débil:

$$Y^0 \perp\!\!\!\perp D|X.$$

En este punto es necesario hacer tres observaciones. En primer lugar, para que el supuesto (18) sea válido,  $X$  debe incluir todas las características exógenas de los individuos, así como las características que determinan la asignación efectiva al grupo de participantes<sup>21</sup>. En realidad, únicamente pueden incluirse los determinantes observables. Por este motivo, el algoritmo de emparejamiento sólo puede tener en cuenta la selección por observables, pero no puede seleccionar por covariables no-observables por lo que es imposible eliminar por completo el sesgo de selección muestral<sup>22</sup>.

En segundo lugar, la idea de considerar la distribución de las covariables  $X$  en la muestra aplicada literalmente supone emparejar cada participante exactamente con un no-participante que presente los mismos valores de las covariables<sup>23</sup>. En las aplicaciones empíricas suele ser difícil encontrar para cada participante un excluido equivalente con unos valores de  $X$  idénticos. El supuesto fundamental (18) sólo se cumple necesariamente cuando las parejas de individuos muestran las mismas características  $X$ . En la medida en que los participantes en el programa no disponen de un equivalente en el grupo de control, el efecto medio del programa para los participantes no puede identificarse para el grupo completo de participantes. De nuevo, si el efecto causal no varía entre los individuos, el efecto identificado es igual al efecto medio del programa de formación para los participantes (*SATE*).

Finalmente, el emparejamiento de participantes y no-participantes a partir de los valores de las covariables suele verse dificultado por la dimensión de  $X$  que impide encontrar a excluidos equivalentes con valores idénticos para cada covariable de  $X$ . ROSENBAUM, RUBIN (1983) sugieren el uso de un balancing score en vez del vector de covariables. Un posible balancing score es el mencionado propensity score, es decir, la probabilidad

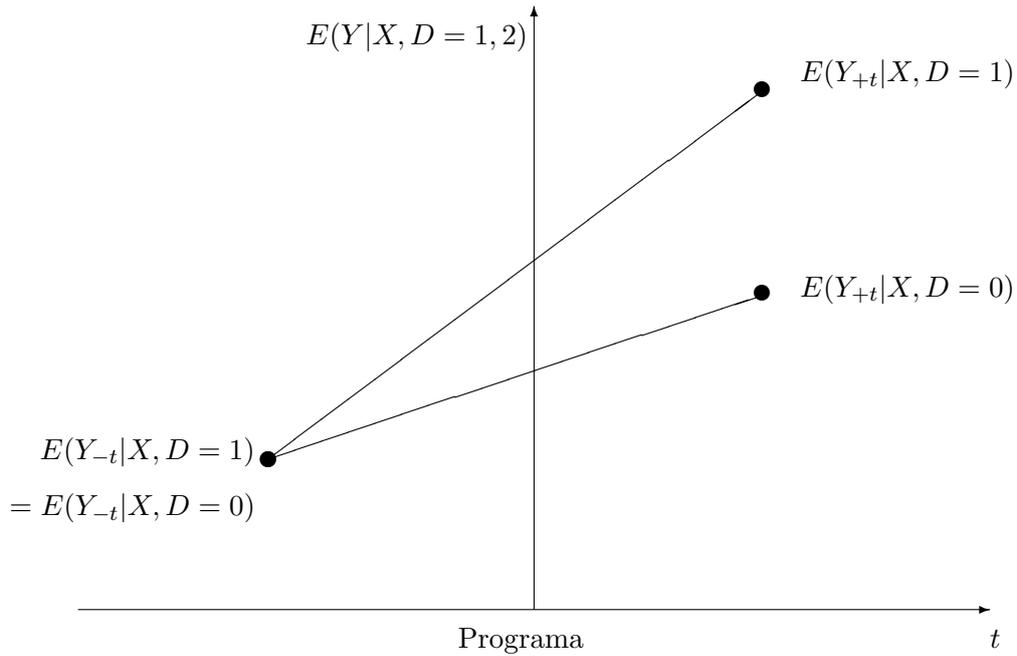


Figura 2: Estimador del impacto medio mediante métodos de emparejamiento

condicional de participar en el programa. Estos autores demuestran que la ignorabilidad fuerte implica

$$Y^0 \perp\!\!\!\perp D|P(X).$$

La utilización de propensity scores tiene la ventaja de que el emparejamiento se realiza a partir de un valor escalar, pero también presenta la desventaja de que este propensity score no puede observarse directamente por lo que debe estimarse<sup>24</sup>.

Los métodos tradicionales de emparejamiento pueden aplicarse también en un contexto de series temporales<sup>25</sup>. Para ello se utiliza una variante del estimador de diferencias-en-diferencias –discutido en el apartado anterior– donde las diferencias se toman condicionadas a  $X$ . Así, el supuesto de iden-

tificación es:

$$E(Y_{+t}^0 - Y_{-t}^0 | X, D = 1) = E(Y_{+t}^0 - Y_{-t}^0 | X, D = 0).$$

La ventaja de esta extensión es que es posible controlar la selección por no-observables mediante la toma de diferencias. El emparejamiento se realiza comparando los resultados de los participantes con los correspondientes resultados de los excluidos tanto en el periodo posterior al programa  $+t$ , como en el periodo previo  $-t$  al programa<sup>26</sup>. La figura 2 ilustra gráficamente los posibles valores de los resultados potenciales.

### 3.4. Modelos econométricos de selección

El enfoque econométrico de la evaluación del impacto de un programa de formación se basa generalmente en el modelo clásico de regresión. El planteamiento de Roy-Rubin puede reescribirse como un modelo econométrico tradicional<sup>27</sup>. Se supone que los resultados potenciales para un individuo están relacionados con los regresores  $X = (X_1, X_2)$  y un término de perturbación a través de las siguientes funciones<sup>28</sup>:

$$Y^0 = X_1' \beta + u \quad (19)$$

$$Y^1 = \alpha + X_1' \beta + u, \quad (20)$$

donde  $\beta$  es el vector de coeficientes asociados a las variables  $X_1$  que influyen en el resultado observado, mientras que  $\alpha$  es el coeficiente que mide el impacto medio del programa de formación. Recuérdese que el efecto observado se define como:

$$Y = DY^1 + (1 - D) Y^0,$$

por lo que las ecuaciones (19) y (20) pueden reescribirse de la siguiente manera:

$$Y = D \alpha + X_1' \beta + u. \quad (21)$$

El modelo descrito es muy restrictivo por lo que se incluye una ecuación de participación. Sea  $D^*$  una variable latente que expresa la mejora neta de la participación en el programa de formación. Se supone que  $D^*$  depende de las covariables  $X_2$  y de un término de perturbación  $v$ . La participación en el programa tiene lugar cuando la mejora neta es positiva<sup>29</sup>:

$$D^* = X_2' \gamma + v \quad (22)$$

$$D = I(D^* > 0), \quad (23)$$

donde  $\gamma$  es un vector de coeficientes. El “supuesto del valor estable del tratamiento” significa que los términos de perturbación  $v$  son independientes entre individuos<sup>30</sup>. Además, se supone que  $E(u) = E(v) = 0$  y  $E(X_1 u) = E(X_2 v) = 0$ .

Utilizando la ecuación (2), el modelo definido por las ecuaciones (21), (22) y (23) puede reescribirse como un modelo lineal de una variable endógena dicotómica:

$$Y = X_1' \beta + D \alpha + u \quad (24)$$

$$D^* = X_2' \gamma + v \quad (25)$$

$$D = I(D^* > 0). \quad (26)$$

Este modelo econométrico considera tanto el proceso de participación (ecuaciones (25) y (26)), como el proceso que determina el resultado (ecuación (24)). Si  $E(D u) = 0$  el efecto medio del programa sobre los participantes puede estimarse de forma consistente, aplicando mínimos cuadrados ordinarios al modelo descrito. No obstante, dado que la asignación al grupo de participantes en los estudios no-experimentales no es aleatoria, se observa una cierta correlación entre  $u$  y  $D$  en la ecuación de resultado (24). Esta correlación puede ser consecuencia de la dependencia estocástica entre  $u$  y  $v$  (selección por no observables) o una dependencia estocástica entre  $u$  y  $X_2$  (selección por observables)<sup>31</sup>.

HECKMAN (1978) discute diferentes estrategias alternativas de estimación de modelos lineales de variables endógenas dicotómicas. Los métodos econométricos que estiman conjuntamente las ecuaciones del modelo tienen en cuenta explícitamente la correlación entre el impacto y la ecuación de participación y, en consecuencia, tanto la selección por observables como por no-observables. Únicamente se distinguen por las condiciones necesarias que afectan a las perturbaciones y a la estructura de su covarianza. Puede suponerse, por ejemplo, que la distribución conjunta de las perturbaciones pertenece a una determinada familia de distribuciones paramétricas como las distribuciones normales multivariantes o que pueden determinarse a partir de métodos no-paramétricos. Procedimientos menos restrictivos como, por ejemplo, métodos de simulación o de estimación no-paramétrica de la función de verosimilitud no son siempre aplicables<sup>32</sup>. No obstante, con una estimación convencional de máxima

verosimilitud, la tratabilidad del modelo depende de forma crucial del número de ecuaciones del sistema. Por ejemplo, el modelo puede incluir diferentes resultados, como la probabilidad de empleo y el nivel salarial, y diferentes ecuaciones que controlan por las condiciones iniciales por lo que se precisan de supuestos identificativos más restrictivos<sup>33</sup>.

Cuanto más restrictivos sean los supuestos necesarios para aplicar la estimación conjunta (simultánea), tanto mayor es la sensibilidad (potencial) de las estimaciones de los coeficientes ante especificaciones alternativas del modelo. Una estrategia diferente y posiblemente más robusta de estimación del modelo econométrico en el caso de  $E(Du) \neq 0$ , es la utilización de variables instrumentales<sup>34</sup>. Se trata básicamente de definir al menos una variable instrumental  $Z$  de  $X_2$  que determine la participación en el programa  $D$ , pero sin que influya directamente en la respuesta  $Y$ <sup>35</sup>. Cualquier influencia de  $Z$  sobre  $Y$  se produce únicamente de forma indirecta a través de  $D$ . En consecuencia,  $Z$  es un elemento de  $X_2$ , pero no forma parte de  $X_1$ , es decir, esta variable cumple la condición de exclusión<sup>36</sup>.

Las variables instrumentales permiten descomponer la varianza del regresor endógeno  $D$  en una componente exógena y otra endógena, utilizándose la componente exógena en la estimación<sup>37</sup>. En este sentido, la utilización de variables instrumentales puede interpretarse como un experimento social en el que la participación en el programa se determina de forma independiente de  $u$  a través de un mecanismo exógeno de asignación. Por otro lado, la asignación aleatoria en un experimento social puede interpretarse como una variable instrumental puesto que crea una variación exógena respecto de la participación en el programa del grupo experimental<sup>38</sup>.

Respecto a la elección de  $Z$ , BOUND, JAEGER, BAKER (1995) aducen que una variable que sólo está débilmente correlacionada con el indicador de participación  $D$  puede generar estimaciones con un elevado error estándar. Además, una correlación baja entre  $Z$  y  $u$  provoca, entre otros problemas, un problema de inconsistencia en las estimaciones con variables instrumentales<sup>39</sup>.

Generalmente los métodos de variables instrumentales no permiten identificar el efecto medio del programa sobre los participantes ni tampoco el efecto medio local del programa (Local Average Treatment Effect, LATE) que se define como el efecto medio causal para aquellos individuos cuya participación en el programa está influida por el instrumento<sup>40</sup>. Supóngase que existe un sólo instrumento  $Z$  que a su vez es un indicador binario que distingue entre individuos que siguen el protocolo del experimento ( $Z = 1$ ) y los que no lo hacen ( $Z = 0$ ) y sea  $\tilde{X}_2$  el vector  $X_2$  sin la variable  $Z$ . Supóngase también que  $E(D|Z = 1, \tilde{X}_2) > E(D|Z = 0, \tilde{X}_2)$ <sup>41</sup>.

La muestra puede subdividirse entonces en las siguientes cuatro submuestra<sup>42</sup>:

1. Never-takers: aquellos individuos que nunca participarán en el programa con independencia del valor de la variable  $Z$  ( $D(Z = 1, \tilde{X}_2) = 0$  y  $D(Z = 0, \tilde{X}_2) = 0$ ).
2. Always-takers: aquellos individuos que siempre participarán en el programa independientemente del valor de  $Z$  ( $D(Z = 1, \tilde{X}_2) = 1$  y  $D(Z = 0, \tilde{X}_2) = 1$ ).
3. Compliers: aquellos individuos que participarán para  $Z = 1$  y no participarán para  $Z = 0$  ( $D(Z = 1, \tilde{X}_2) = 1$  y  $D(Z = 0, \tilde{X}_2) = 0$ ).
4. Defiers: aquellos individuos que participarán para  $Z = 0$  y no participarán para  $Z = 1$  ( $D(Z = 1, \tilde{X}_2) = 0$  y  $D(Z = 0, \tilde{X}_2) = 1$ ).

ANGRIST, IMBENS, RUBIN (1996a, pág. 447) suponen que para todos los individuos en la muestra  $D(Z = 1, \tilde{X}_2) \geq D(Z = 0, \tilde{X}_2)$  por lo que no existen defiers<sup>43</sup>. En los grupos de never-takers y always-takers el efecto causal de  $Z$  sobre  $Y$  es cero. En consecuencia, LATE es el impacto causal medio para el grupo de compliers.

HECKMAN (1997) apunta que el grupo de compliers varía en función de las variables instrumentales incluidas por lo que LATE depende de la elección de  $Z$ . Es más, el grupo de compliers no puede identificarse a partir de los datos observados. Esta imposibilidad de identificación es la mayor desventaja del método de variables instrumentales. Con respecto de las

intervenciones de la política económica, por ejemplo, que no tienen ningún coste social directo y sólo induce a algunos individuos a cambiar de status de participación sin que esto afecte a los demás individuos, este parámetro es, aún así, de una gran relevancia para la política económica<sup>44</sup>. Además, tal y como remarcan ANGRIST, IMBENS, RUBIN (1996b), el grupo de compliers es el único grupo para el cual los datos ofrecen directamente la información relevante y es el único grupo en el que pueden observarse tanto los miembros que participan, como los miembros que no participan en el programa. HECKMAN (1997) discute los supuestos necesarios para identificar el impacto medio del programa o el impacto medio del programa en los participantes mediante LATE. En el caso de que el efecto causal difiera entre individuos, la identificación del impacto medio del programa en los participantes calculado mediante variables instrumentales, requiere de unos supuestos particulares sobre el comportamiento de los individuos, concretamente sobre cómo se autoseleccionan ellos mismos a la hora de decidir su participación en el programa<sup>45</sup>. En el caso especial de que el efecto causal no varíe entre individuos, los tres impactos medios del programa (ATE, SATE, LATE) son idénticos. Este es el caso en el modelo definido por las ecuaciones (24) – (26).

VELLA, VERBEEK (1998, págs. 7 – 9) discuten formas alternativas de implementar métodos basados en variables instrumentales. Una posibilidad es estimar (dado  $X_2$ ) la probabilidad de participación  $P(D = 1|X_2)$  de forma paramétrica a partir de las ecuaciones (25) – (26), utilizando esta estimación en lugar de  $D$  en la posterior estimación de la ecuación de resultados (24). Otra posibilidad de implementar métodos de variables instrumentales es el uso de la predicción lineal  $X_2'\hat{\gamma}$  –obtenida mediante mínimos cuadrados ordinarios– en lugar de  $D$  o la utilización del residuo correspondiente  $D - X_2'\hat{\gamma}$  como un regresor adicional en la ecuación (24). En consecuencia, los métodos de variables instrumentales puede entenderse como unos procedimientos estimación en dos etapas. Así, en el modelo específico discutido aquí no es necesario imponer la condición de exclusión puesto que la forma funcional garantiza la identificación del modelo.

### 3.5. Límites no-paramétricos

MANSKI (1989, 1990) demuestra que puede definirse un intervalo para el impacto del programa cuando los resultados potenciales también están

acotados. En vez de intentar obtener un estimador del efecto medio de la participación, el objetivo de este método es establecer los límites inferior y superior que puede tomar el impacto medio del programa. Para ello es necesario identificar los supuestos necesarios para reducir en la medida de lo posible la amplitud del intervalo definido por estos límites con el objeto de que el intervalo no contenga el valor cero. La principal ventaja de este método es que, a diferencia de los procedimientos no-experimentales discutidos en los apartados anteriores, no es necesario imponer unos estrictos supuestos que afectan al proceso de participación y de resultados.

Supóngase que  $Y^0$  –condicionado a  $X$ – toma valores en un intervalo conocido:

$$[L_X^0, U_X^0], \quad (27)$$

donde  $L_X^0$  denomina el límite inferior y  $U_X^0$  el límite superior del intervalo. Además, se supone que también  $Y^1$  está acotado en un intervalo finito:

$$[L_X^1, U_X^1]. \quad (28)$$

Si  $Y^0$  y  $Y^1$  están acotados por límites finitos se deduce inmediatamente que también la media de  $Y^0$  y la media de  $Y^1$  se encuentran en los respectivos intervalos:

$$E(Y^0|X) \in [L_X^0, U_X^0] \quad E(Y^1|X) \in [L_X^1, U_X^1] \quad (29)$$

$$E(Y^0|X, D = 1) \in [L_X^0, U_X^0] \quad E(Y^1|X, D = 0) \in [L_X^1, U_X^1]. \quad (30)$$

Para el efecto medio del programa se cumple, por consiguiente:

$$E(\Delta|X) \in [L_X^1 - U_X^0, U_X^1 - L_X^0]. \quad (31)$$

Dado que el efecto medio de la participación es igual a la diferencia  $E(Y^1|X) - E(Y^0|X)$ , su límite inferior está dado por la diferencia entre la cota inferior de  $E(Y^1|X)$  y la cota superior de  $E(Y^0|X)$ . El límite superior para  $E(\Delta|X)$  en la ecuación (31) se obtiene de forma análoga. El cálculo de los límites inferior y superior de  $Y^0$  y  $Y^1$  se simplifica notablemente cuando la variable-resultado es binaria. Por ejemplo, la probabilidad individual de encontrar un empleo es como mínimo igual a cero y como máximo igual a la unidad con lo que se cumple  $L_X^0 = L_X^1 = 0$  y  $U_X^0 = U_X^1 = 1$ .

MANSKI (1989) demuestra que los disponibles de la muestra permiten identificar un intervalo para  $E(Y^0|X)$  más estrecho que el intervalo definido en (27). Puesto que  $E(Y^0|X)$  está relacionado con  $E[Y^0|X, D = 1]$  a través de:

$$E(Y^0|X) = E(Y^0|X, D = 1) \cdot P(D = 1|X) + E(Y^0|X, D = 0) \cdot P(D = 0|X).$$

Este intervalo más estrecho es, por tanto:

$$E(Y^0|X) \in [E(Y^0|X, D = 0) \cdot P(D = 0|X) + L_X^0 \cdot P(D = 1|X), E(Y^0|X, D = 0) \cdot P(D = 0|X) + U_X^0 \cdot P(D = 1|X)].$$

El límite inferior se define así como el valor que toma  $E(Y^0|X)$  cuando  $Y^0$  es igual a su cota inferior para todos aquellos individuos que participaron en el programa. El límite superior se determina de forma análoga.

La amplitud de este intervalo se calcula como:

$$w_0(X) = (U_X^0 - L_X^0) \cdot P(D = 1|X),$$

el cual es menor que la amplitud determinada por (27) si  $P(D = 1|X) < 1$ . A partir de la muestra puede estimarse de forma no-parámetrica los valores que toman  $E(Y^0|X, D = 0)$ ,  $P(D = 0|X)$  y  $P(D = 1|X)$  de manera que se obtienen los límites del intervalo estrecho<sup>46</sup>.

Un intervalo más estrecho para  $E(Y^1|X)$  se deriva análogamente:

$$E(Y^1|X) \in [E(Y^1|X, D = 1) \cdot P(D = 1|X) + L_X^1 \cdot P(D = 0|X), E(Y^1|X, D = 1) \cdot P(D = 1|X) + U_X^1 \cdot P(D = 0|X)].$$

El intervalo para el efecto medio del programa puede redefinirse, por consiguiente, como sigue:

$$E(\Delta|X) \in [L_X^1 \cdot P(D = 0|X) + E(Y^1|X, D = 1) \cdot P(D = 1|X) - E(Y^0|X, D = 0) \cdot P(D = 0|X) - U_X^0 \cdot P(D = 1|X), U_X^1 \cdot P(D = 0|X) + E(Y^1|X, D = 1) \cdot P(D = 1|X) - E(Y^0|X, D = 0) \cdot P(D = 0|X) - L_X^0 \cdot P(D = 1|X)].$$

Este intervalo posee una amplitud

$$w(X) = (U_X^1 - L_X^1) \cdot P(D = 0|X) + (U_X^0 - L_X^0) \cdot P(D = 1|X). \quad (32)$$

Cuando se supone que los límites de  $Y^0$  y  $Y^1$  asumen los mismos valores, esto es,  $L_X^0 = L_X^1 = L_X$  y  $U_X^0 = U_X^1 = U_X$ , la amplitud en este último intervalo (ecuación (32)) se reduce a

$$w(X) = U_X - L_X,$$

donde  $w(X)$  asciende sólo la mitad de la amplitud del intervalo original dado por el intervalo definido a partir de (31) para idénticos límites. En consecuencia, el supuesto de que los resultados tienen límites comunes juntamente con el aprovechamiento de la información de la muestra permite acotar el efecto medio del programa a la mitad de su amplitud inicial. La validez de este resultado se cumple especialmente cuando la variable-resultado es binaria dado que el supuesto de límites coincidentes se cumple automáticamente.

El intervalo para el impacto medio del programa  $E(\Delta|X)$ , sin embargo, incluye todavía necesariamente el valor cero por lo que no puede identificarse el signo del efecto<sup>47</sup>. MANSKI (1990) demuestra que el intervalo puede reducirse, introduciendo supuestos adicionales. Estos supuestos son, por ejemplo, que la participación ( $D = 1$ ) tenga lugar cuando dicha participación mejora el resultado obtenidos por los participantes en el programa, esto es, cuando  $Y^1 > Y^0$  o bien si el efecto del programa no depende de  $X$  (level-set restriction). El primer supuesto es muy restrictivo puesto que implica que los individuos o las instituciones que gestionan del programa disponen de una considerable cantidad de información previa a la participación. Si se cumple este supuesto puede identificarse el signo del efecto medio del programa<sup>48</sup>. Bajo el segundo supuesto, el intervalo más estrecho es la intersección de los intervalos correspondientes a las diferentes realizaciones de  $X$  sujetas a la level-set restriction<sup>49</sup>. No obstante, este supuesto no es suficiente para identificar el signo del impacto del programa<sup>50</sup>.

#### 4. La elección del método adecuado: problemas no resueltos de la evaluación

Los experimentos sociales se consideran, con frecuencia, la única base válida para una evaluación fiable del impacto de un programa público de formación. En los Estados Unidos los experimentos sociales controlados se han convertido en la forma estándar de evaluación de los programas sociales. Cuando el Congreso de los Estados Unidos inició en 1988 un nuevo programa de formación llamado Job Opportunities and Basic Skills Training Program se implementó al mismo tiempo un experimento social para evaluar el impacto del programa<sup>51</sup>. En 1986 el Departamento de Empleo de los Estados Unidos encargó un experimento social para evaluar el Job Training Partnership Act (JTPA). Esta evaluación experimental se denominó National JTPA Study (NJS)<sup>52</sup>. Este estudio evidenció que los efectos sobre el empleo y la remuneración de los participantes eran reducidos lo que contribuyó a que el Congreso decidiera reestructurar el programa y recortar notablemente el presupuesto destinado a los programas de formación de jóvenes<sup>53</sup>.

La preferencia muy difundida por los experimentos sociales se debe sustancialmente a los influyentes trabajos de LALONDE (1986) y FRACKER, MAYNARD (1987). Ambos estudios comparan los estimadores obtenidos a partir de experimentos sociales con los estimadores basados en los métodos no-experimentales. La comparación muestra que la utilización de diferentes estimadores no-experimentales produce una gran variabilidad de los resultados. Estos estimadores no fueron capaces, en general, de reproducir los resultados obtenidos mediante los estimadores experimentales. Los autores de ambos estudios concluyen que los métodos no-experimentales no generan estimadores correctos y fiables del efecto causal. No obstante, HECKMAN, HOTZ (1989) remarcan que esta afirmación descansa en la premisa crucial de que los diferentes métodos no-experimentales deben producir las mismas estimaciones del impacto que los métodos experimentales. Dado que en los diferentes métodos no-experimentales se parte de distintos supuestos de identificación, no se obtienen los mismos estimadores. Sin embargo, esto no invalida necesariamente los métodos no-experimentales. Se precisan, por tanto, unos contrastes que permitan elegir el método más

adecuado de entre los diferentes estimadores no experimentales<sup>54</sup>.

Una premisa adicional de los estudios de LALONDE (1986) y FRAKER, MAYNARD (1987) es que suponen que los estimadores basados en los datos experimentales son los estimadores correctos del impacto del programa y, por tanto, configuran el punto de referencia para los estimadores no-experimentales. Sin embargo, esto no tiene que ser necesariamente así. HECKMAN, SMITH (1995) demuestran que incluso los datos experimentales pueden conducir a estimadores sesgados o irrelevantes del efecto causal. Las razones fundamentales son:

1. Sesgo de aleatoriedad: el proceso de una asignación aleatoria puede provocar que los participantes en el programa sean sistemáticamente diferentes del grupo de personas al cual va dirigido normalmente el programa. En el NJS, por ejemplo, muchos de los centros de formación descentralizados geográficamente rehusaron participar en la evaluación del experimento por motivos éticos o de relaciones públicas. Por consiguiente, los centros de formación que sí participaron en la evaluación no constituían una muestra aleatoria de todos los centros. Además, para formar el grupo de control del experimento, el conjunto de personas consideradas aceptables tuvo que ampliarse<sup>55</sup>;
2. Sesgo de sustitución: es posible que los miembros del grupo de control del experimento participen en otros programas por lo que este grupo no está compuesto únicamente por personas no tratadas. Estos programas alternativos pueden ser muy similares e incluso idénticos al programa que se pretende evaluar. En el programa JTPA, por ejemplo, los centros de formación tienen acuerdos con organizaciones de ámbito local como, por ejemplo, institutos o colegios profesionales, que ofrecen cursos de formación a los participantes en el programa. En ocasiones, éstos centros públicos reservan plazas para otros organismos locales que no intervienen en el experimento social de forma que los miembros del grupo de control en el NJS podrían participar en el mismo curso que los seleccionados, aunque a través de otras vías. Incluso se fomentaba esta participación de los miembros del grupo de control en estos programas porque los centros que tomaban parte en la evaluación del experimento estaban autorizados a ofrecer a estas personas una lista de proveedores de servicios alternativos<sup>56</sup>.

Problemas adicionales de los experimentos sociales surgen con la posibilidad del efecto Hawthorne. Esto significa que el comportamiento de los participantes es diferente a su conducta normal simplemente porque saben que el programa es un experimento. También es posible que los participantes no se tomen el programa tan en serio como se tomarían un programa real<sup>57</sup>.

El sesgo de la aleatoriedad invalida el supuesto crucial de que el grupo de participantes aceptados en el experimento de evaluación coincide con el grupo de participantes reales o efectivos en ausencia del experimento. Así, el supuesto identificativo (ecuación (10)) ya no se cumple. El sesgo de sustitución, por otro lado, afecta al supuesto identificativo de que los resultados alcanzados por el grupo de control del experimento representa unas observaciones fiables de los resultados contrafactuales en el caso de no participar en el programa de formación. Sin embargo, los resultados del grupo de control ya no corresponden únicamente al status de no-participación por lo que no representa un grupo de comparación válido. HECKMAN, SMITH (1997b) hallaron que las estimaciones basadas en la evaluación del experimento de la JTPA son muy sensibles a la selección de los centros de formación y están afectados por el sesgo de sustitución. Este último resultado es confirmado por HECKMAN, HOHMANN, KHOO, SMITH (1987). Ambos estudios indican también que los miembros del grupo de tratamiento que abandonan el programa antes de recibir el tratamiento son una fuente adicional de sesgo (dropout bias)<sup>58</sup>.

En ocasiones se ha argumentado que las estimaciones de los impactos del programa basados en la evaluación del experimento son más fáciles de explicar a los responsables políticos de los programas de formación que los basados en la evaluación no-experimental<sup>59</sup>. Sin embargo, la simplicidad de los estimadores experimentales no es tan evidente y también los métodos experimentales se basan en diferentes supuestos restrictivos. La validez de los supuestos depende fundamentalmente del diseño adecuado del experimento. Los experimentos diseñados cuidadosamente, no obstante, pueden facilitar sustancialmente la tarea de evaluación.

Además del sesgo de selección muestral, la medición del impacto del programa está influida por otros problemas que no sólo afectan a los

datos no-experimentales. Hasta ahora la argumentación se ha basado en variables-resultado no censuradas, tomando como punto de referencia el ejemplo de la remuneración de los trabajadores. Sin embargo, en algunos estudios el problema fundamental es consecuencia de la naturaleza de la variable analizada “duración” y del hecho de que el estudio utiliza datos de panel.

En un diseño adecuado del experimento donde la variable analizada no está censurada, una simple comparación entre el resultado medio post-programa del grupo de participantes y el resultado medio post-programa del grupo de control, posiblemente condicionado por las características  $X$ , genera un estimador insesgado del impacto medio del programa en los participantes. Esta simplicidad es uno de los argumentos más poderosos en favor de experimentos bien diseñados y correctamente llevados a término. No obstante, el análisis de duración con datos experimentales es mucho más complicado<sup>60</sup>. La simple comparación de, por ejemplo, la duración media de los episodios de empleo conlleva estimaciones potencialmente sesgadas por los siguientes motivos:

1. Algunas observaciones están censuradas por la derecha. Una comparación entre las medias no tiene en cuenta que estos episodios todavía no han concluido y no pueden compararse directamente con los episodios no censurados.
2. Es necesario considerar todas las demás características observables y no observables que influyen en la tasa de transición dada la naturaleza selectiva de los datos de duración. Sólo pueden observarse las personas en el momento en el que se encuentran en un determinado status cuando la transición no ha tenido lugar con anterioridad. Supóngase que la duración del empleo depende sólo del status de participación y de si el individuo ha abandonado prematuramente la educación secundaria (no ha concluido sus estudios de bachillerato), donde la participación en el programa tiene un efecto positivo y el status de abandono tiene un efecto negativo en la duración del empleo. Dado el carácter experimental de los datos, el status de abandono y el status de tratamiento no están correlacionados. Si todos los episodios de empleo comienzan en el mismo instante, la diferencia entre las proporciones de participantes y no participantes que dejan el trabajo

es una estimación insesgada del impacto del programa en la tasa de transición. Por este motivo, se cumple sólo para el primer intervalo. En los siguientes intervalos la proporción de participantes en el grupo de riesgo –que son los que han abandonado la secundaria– es mayor que la proporción correspondiente de no participantes. Un elevado número de individuos que han abandonado la educación secundaria –que en otro caso perderían su empleo– consiguen permanecer en su puesto de trabajo a causa de su participación en el programa. La diferencia en la tasa de transición refleja ahora tanto el impacto del programa de formación como el efecto del status de abandono de la secundaria. El grupo de participantes contienen ahora un mayor número de personas que han abandonado los estudios y el efecto de la participación en el programa se subestima<sup>61</sup>.

3. Los participantes en el programa se retiran del mercado laboral para participar en el programa de forma que los miembros del grupo de control permanecen más tiempo en el mercado de trabajo. Así, se observan episodios más largos para los no participantes. Además, ambos grupos se enfrentan a diferentes condiciones de demanda en el mercado laboral<sup>62</sup>.
4. Si los seleccionados para el experimento están desempleados cuando la selección tiene lugar, los miembros del grupo de control continuarán desempleados, mientras que los episodios de desempleo de los participantes finalizan con la participación. Los respectivos episodios están ambos censurados por la izquierda al inicio del periodo de observación. No obstante, para observar un nuevo episodio para un excluido, éste tiene que abandonar primero el actual episodio censurado por la izquierda, encontrando un empleo y a continuación abandonarlo para convertirse, de nuevo, en un desempleado. En contraposición, los participantes comienzan automáticamente un nuevo episodio de desempleo después del programa, si no encuentran inmediatamente un puesto de trabajo después de la finalización del programa. En consecuencia, el número de nuevos episodios no censurados por la izquierda será mucho menor para los excluidos que para los participantes. Extendiendo este argumento al argumento 2 expuesto más arriba, supóngase que el status de abandono de los estudios no tiene un efecto relevante en la duración del desempleo. En este caso, es probable

que aquéllos excluidos con nuevos episodios de desempleo tienden a tener un título de bachiller porque muestran una mayor probabilidad de abandonar su episodio de desempleo censurado por la izquierda<sup>63</sup>.

A pesar de estos problemas, la evaluación es mucho más sencilla con datos experimentales puesto que con datos no-experimentales es necesario modelizar también el proceso de selección de los participantes<sup>64</sup>.

Sin embargo, no todos estos problemas surgen necesariamente en un análisis no-experimental. Es más probable que los dos últimos problemas mencionados influyan sistemáticamente en los resultados de la evaluación, si la recolección de datos va ligada a la evaluación de un solo programa. Este no es el caso con datos de panel que generalmente abarcan una amplia gama de cuestiones socioeconómicas. Cuando se dispone de este tipo de datos, los investigadores pueden observar un elevado número de individuos que participan en una multiplicidad de medidas de formación muy heterogéneas durante un número considerable de años. El grupo de no participantes y de participantes se enfrentarán sistemáticamente a diferentes condiciones laborales porque los episodios tienen lugar en diferentes momentos y durante unos periodos más largos. Si el periodo para el que se disponen de datos es bastante largo, los no participantes muestran una mayor probabilidad de iniciar un nuevo episodio de desempleo.

Cuando el número de medidas heterogéneas de formación es muy elevado, es necesario distinguir entre diferentes tipos de formación como, por ejemplo, on-the-job versus off-the-job training, financiado por la empresa versus financiado por el Estado o cursos de larga duración versus cursos cortos. Esto no siempre es posible como consecuencia de la falta de información adicional sobre los distintos programas de formación. Por el mismo motivo, se suele disponer generalmente de poca información sobre el proceso mediante el cual los individuos son seleccionados (por las autoridades responsables del programa o por los propios individuos) para participar en el programa, a pesar de que un conocimiento detallado de este proceso permitiría tener en cuenta la selección no aleatoria. Con un único programa, por ejemplo, el investigador puede familiarizarse, hasta cierto punto, con las reglas que determinan la elegibilidad para el programa de formación.

El problema de las variables resultado censuradas también afecta a los métodos no-experimentales discutidos aquí. Esto limita considerablemente la aplicabilidad del estimador antes-después por los motivos ya discutidos. Tanto, los métodos de emparejamientos, como las estrategias de estimación conjunta y los métodos basados en variables instrumentales pueden tener en cuenta la censura de las observaciones, utilizando modelos econométricos como los modelos de duración<sup>65</sup>.

## 5. Conclusión

En este trabajo se han mostrado de forma esquemática los problemas fundamentales de evaluación de las políticas públicas. Para medir el efecto de un tratamiento debe compararse el resultado después de participar en el programa con el resultado no observable que se hubiese alcanzado en el caso de no haber recibido el tratamiento. El problema reside en construir el resultado contrafactual a partir de las observaciones disponibles. La resolución a este problema de valoración consiste en especificar un grupo adecuado de comparación a partir de la información disponible sobre las personas que no han recibido la formación. En la medida en que la evaluación de los programas públicos no se basa en experimentos sociales adecuados, los grupos de participantes y de control no suelen ser submuestras aleatorias de la población a la que van dirigidos estos programas. La omisión de este problema conduce a un sesgo de selección muestral. Diferentes métodos no experimentales para la evaluación introducen supuestos alternativos para poder identificar el efecto causal. Es necesario tener en cuenta estos supuestos a la hora de elegir el método apropiado de evaluación.

## Notas

<sup>1</sup>La evaluación de programas públicos de formación tiene una larga tradición en los Estados Unidos, discutiéndose los resultados con gran profusión en la literatura académica. La concesión del Premio Nobel a James J. Heckman en el año 2000 evidencia la relevancia de la evaluación de las políticas públicas. Los trabajos de este autor iniciaron una nueva vía en la resolución del problema de selección muestral y, por consiguiente, de la evaluación de las políticas activas del mercado de trabajo.

<sup>2</sup>Véase LALONDE (1986).

<sup>3</sup>Una visión panorámica de los diferentes procedimientos de estimación teniendo en cuenta el sesgo de selección muestral, puede encontrarse en VELLA (1998).

<sup>4</sup>Para extensiones recientes de estos dos modelos a un modelo de tratamiento múltiple con más de dos estados posibles, esto es, donde el conjunto actual de alternativas de un individuo contiene más de dos opciones como, por ejemplo, la participación en diferentes programas de formación, véase IMBENS (1999) y LECHNER (2001).

<sup>5</sup>Stable Unit Treatment Value Assumption: SUTVA. Véase RUBIN (1978, 1991).

<sup>6</sup>Para una discusión extensa de los conceptos de causalidad y los problemas que afectan a la inferencia causal, véase HOLLAND (1986) y SOBEL (1994).

<sup>7</sup>No obstante, ABADIE, ANGRIST, IMBENS (2002) proponen un estimador no-paramétrico para estudiar el impacto de un programa de formación en diferentes cuantiles de la distribución de la remuneración.

<sup>8</sup>Véase HECKMAN, ICHIMURA, TODD (1997).

<sup>9</sup>Véase HECKMAN, ICHIMURA, SMITH, TODD (1998), LECHNER (1998) y HECKMAN, LALONDE, SMITH (1999).

<sup>10</sup>Se suma y resta simultáneamente  $E(Y^0|D = 1)$  en la ecuación (7).

<sup>11</sup>Véase HECKMAN, SMITH (1995).

<sup>12</sup>Para un resumen reciente de esta literatura véase, por ejemplo, ANGRIST, KRUEGER (1999) y HECKMAN, LALONDE, SMITH (1999).

<sup>13</sup>Véase DAWID (1979).

<sup>14</sup>Véase LECHNER (1998).

<sup>15</sup>Véase HECKMAN, SMITH (1997a).

<sup>16</sup>Véase MOFFITT (1991) y ASHENFELTER, CARD (1985). Un procedimiento semi-paramétrico más sofisticado para determinar el estimador de diferencias-en-diferencias se propone en ABADIE (1999). Una aplicación práctica de este método puede encontrarse en el trabajo de CARD (1990) en el que se analiza el impacto de la inmigración en las oportunidades de encontrar un empleo de los trabajadores residentes menos cualificados.

<sup>17</sup>Véase HECKMAN, ICHIMURA, SMITH, TODD (1998).

<sup>18</sup>Véase HECKMAN, ICHIMURA, TODD (1997) y DAWID (1979).

<sup>19</sup>Véase HECKMAN, ICHIMURA, TODD (1998, pág. 266).

<sup>20</sup>Véase HECKMAN, ICHIMURA, TODD (1998).

<sup>21</sup>LECHNER (1998) remarca que  $X$  puede incluir también aquellas variables que sólo influyen en la decisión de participación para incrementar así la eficiencia de la estimación.

<sup>22</sup>Véase HECKMAN, ICHIMURA, SMITH, TODD (1998).

<sup>23</sup>IMBENS (2001) ofrece una panorámica sobre las condiciones que tienen que cumplir las covariables para que la estimación sea eficiente.

<sup>24</sup>ROSENBAUM, RUBIN (1985) y LECHNER (1998) proponen diferentes algoritmos de emparejamiento basados en los balancing score o propensity score.

<sup>25</sup>Véase HECKMAN, ICHIMURA, SMITH, TODD (1998) y HECKMAN, ICHIMURA, TODD (1997). Estos autores utilizan una variante del estimadores de diferencias-en-diferencias.

<sup>26</sup>Véase HECKMAN, ICHIMURA, TODD (1997, pág. 632).

<sup>27</sup>El problema de selección muestral en los modelos tradicionales de regresión se discute, por ejemplo, en HECKMAN, ROBB (1985) o HECKMAN, HOTZ (1989).

<sup>28</sup>Aquí, las covariables observadas se dividen en dos subgrupos,  $X_1$  y  $X_2$ , que pueden contener elementos comunes.

<sup>29</sup>La función  $I$  es la función indicador.

<sup>30</sup>Véase HECKMAN (1996).

<sup>31</sup>Véase HECKMAN, ROBB (1985, 1986).

<sup>32</sup>Para una panorámica de los métodos de simulación, véase STERN (1997).

<sup>33</sup>En ocasiones, se produce el problema de panel attrition. Este problema se da cuando algunos participantes en el experimento deciden abandonarlo. En este caso, se reduce la eficiencia del estimador puesto que se dispone de un menor volumen de datos.

<sup>34</sup>Para una panorámica actual sobre los problemas asociados a la utilización de variables instrumentales, véase HECKMAN, VYTLACIL (2001).

<sup>35</sup>En la literatura el término de variable(s) instrumental(es) se utiliza en ocasiones sólo para designar la(s) variable(s)  $Z$  y otras veces para designar el vector completo  $X_2$ . Aquí se utiliza la primera definición. Para una discusión más profunda, véase ANGRIST, IMBENS (1995).

<sup>36</sup>Esta definición se debe a ANGRIST, IMBENS, RUBIN (1996a) y describe una forma particular del problema de identificación en econometría.

<sup>37</sup>Véase BOUND, JAEGER, BAKER (1995, pág. 443).

<sup>38</sup>Véase HECKMAN, SMITH (1995) y HECKMAN (1996).

<sup>39</sup>Véase también STAIGER, STOCK (1997).

<sup>40</sup>Véase ANGRIST, IMBENS (1995). ANGRIST, IMBENS, RUBIN (1996b) discuten las consecuencias del incumplimiento de los supuestos en los que se basa la aplicación del método de variables instrumentales.

<sup>41</sup>Este supuesto es necesario para poder distinguir entre los cumplidores y los incumplidores a partir del valor del instrumento. Si esta desigualdad se invierte las definiciones tendrían que intercambiarse entre ellas y el signo de la desigualdad

en la condición de monotonicidad debería invertirse.

<sup>42</sup>Véase ANGRIST, IMBENS, RUBIN (1996b, pág. 448).

<sup>43</sup>Se impone así el supuesto de monotonicidad. Sin embargo, este supuesto no puede contrastarse pero se cumple automáticamente si la decisión de participación se formula como un modelo lineal. Para una discusión más profunda, véase IMBENS, ANGRIST (1994).

<sup>44</sup>Véase HECKMAN (1997, pág. 460).

<sup>45</sup>Véase HECKMAN (1997, págs. 448 – 449).

<sup>46</sup>Véase MANSKI (1990, pág. 320).

<sup>47</sup>Véase MANSKI (1990, pág. 320).

<sup>48</sup>El intervalo sólo contiene valores positivos o negativos.

<sup>49</sup>Véase LECHNER (1996).

<sup>50</sup>En este contexto, MANSKI (1989, 1990) discute diferentes métodos para estimar estos límites. LECHNER (1996) considera los límites del efecto medio del programa en los participantes. MANSKI (1995) utiliza datos longitudinales para estimar los límites del efecto de la pertenencia a una familia intacta sobre el nivel de estudios alcanzado.

<sup>51</sup>Véase MANSKI, GERFINKEL (1992).

<sup>52</sup>Véase BLOOM ET AL. (1997).

<sup>53</sup>Véase HECKMAN, HOHMANN, KHOO, SMITH (1997).

<sup>54</sup>Este tipo de contrastes se desarrollan en HECKMAN, HOTZ (1989).

<sup>55</sup>Véase HECKMAN, SMITH (1995, págs. 99 – 101) y HECKMAN, SMITH (1997b).

<sup>56</sup>Véase HECKMAN, SMITH (1995, págs. 105 – 106).

<sup>57</sup>Véase BURTLESS, ORR (1986, pág. 619).

<sup>58</sup>La no-finalización del programa por parte de los participantes no es un problema específico de los estudios experimentales. De hecho, la forma en que la evaluación del experimento de JTPA fue diseñada, provocó que aproximadamente el 40% de los participantes en el grupo de control abandonara el experimento antes de que se iniciara la formación. Para un análisis más detallado, véase HECKMAN, SMITH (1997b, pág. 24).

<sup>59</sup>Véase HECKMAN, SMITH (1995).

<sup>60</sup>Para una argumentación más elaborada, véase HAM, LALONDE (1990, 1996).

<sup>61</sup>Véase HAM, LALONDE (1990, pág. 163).

<sup>62</sup>Véase HAM, LALONDE (1990, pág. 163).

<sup>63</sup>HAM, LALONDE (1996, pág. 180).

<sup>64</sup>Véase HAM, LALONDE (1996, pág. 181).

<sup>65</sup>El resultado será una ecuación no lineal de estimación. En el caso de una

ecuación no lineal de resultado, la utilización de estimadores basados en variables instrumentales no es tan evidente puesto que no queda claro que se obtengan estimaciones consistentes. En el caso general del impacto de tratamientos heterogéneos, ANGRIST (1991) demuestra que para una ecuación resultado no lineal, los métodos basados en variables instrumentales conllevan una estimación consistente del efecto medio del tratamiento, si la esperanza condicionada del indicador del tratamiento  $D$  es separable aditivamente en las covariables observadas y las omitidas. Este autor implementó una simulación de Montecarlo y halló que el sesgo provocado por el incumplimiento de la separabilidad aditiva no representa un problema grave.

## 6. Referencias bibliográficas

- ABADIE, A. (1999) Semiparametric Difference-in-Differences Estimators, John F. Kennedy School of Government, Harvard University.
- ABADIE, A., ANGRIST, J.D., IMBENS, G.W. (2002) “Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings”, Econometrica, 70(1), págs. 91 – 117.
- ANGRIST, J.D. (1991) Linear Instrumental Variables Estimation of Average Causal Effects in Nonlinear Models, Discussion Paper 1542, Harvard Institute of Economic Research, Harvard University.
- ANGRIST, J.D., IMBENS, G.W. (1995) “Two-Stage Least Square Estimation of Average Causal Effects in Models with Variable Treatment Intensity”, Journal of the American Statistical Association, 90, págs. 431 – 442.
- ANGRIST, J.D., IMBENS, G.W., RUBIN, D.B. (1996A) “Identification of Causal Effects Using Instrumental Variables (with discussion)”, Journal of the American Statistical Association, 91, págs. 468 – 472.
- ANGRIST, J.D., IMBENS, G.W., RUBIN, D.B. (1996B) “Rejoinder (to Identification of Causal Effects Using Instrumental Variables)”, Journal of the American Statistical Association, 91, págs. 444 – 468.

- ANGRIST, J.D., KRUEGER, A.B. (1999) "Empirical Strategies in Labor Economics", en ASHENFELTER, O., CARD, D. (EDS.): Handbook of Labor Economics, Vol. III.
- ASHENFELTER, O. (1978) "Estimating the Effects of Training Programs on Earnings", Review of Economics and Statistics, 60, págs. 47 – 57.
- ASHENFELTER, O., CARD, D. (1985) "Using the Longitudinal Structure of Earnings to Estimate the Effects of Training Programs", Review of Economics and Statistics, 67, págs. 648 – 660.
- BLOOM, H.S., ORR, L.L., BELL, S.H., CAVE, G., DOOLITTLE, F., LIN, W., BOS, J.M. (1997) "The Benefits and Costs of JTPA Title II-A Programs – Key Findings from the National Job Training Partnership Act Study", Journal of Human Resources, 32, págs. 549 – 576.
- BOUND, J., JAEGER, D.A., BAKER, R.M. (1995) "Problems with Instrumental Variables Estimation when the correlation Between the Instruments and the Endogeneous Explanatory Variable is Weak", Journal of the American Statistical Association, 90, págs. 443 – 450.
- BURTLESS, G., ORR, L.L. (1986) "Are Classical Experiments Needed Manpower Policy?", Journal of Human Resources, 21, págs. 721 – 728.
- CARD, D. (1990) "The Impact of the Mariel Boatlift on the Miami Labor Market", Industrial and Labor Relations Review, 44, págs. 245 – 257.
- DAWID, A.P. (1979) "Conditional Independence in Statistical Theory", Journal of the Royal Statistical Society (Series B), 41, págs. 1 – 31.
- FRAKER, T., MAYNARD, R. (1987) "The Adequacy of Comparison Group Designs for Evaluations of Employment-Related Programs", Journal of Human Resources, 22, págs. 194 – 227.
- HAM, J.C., LALONDE, R.J. (1990) "Using Social Experiments to Estimate the Effect of Training on Transition Rates", en HARTOG, J., RIDDER, G., THEEUWES, J. (ED.): Panel Data and Labor Market Studies, North Holland, Amsterdam, págs. 157 – 172.

- HAM, J.C., LALONDE, R.J. (1996) “The Effect of Sample Selection and Initial Conditions in Duration Models: Evidence from Experimental Data on Training”, Econometrica, 64, págs. 175 – 205.
- HECKMAN, J.J. (1978) “Dummy endogeneous variables in a simultaneous equation model”, Econometrica, 46, págs. 931 – 956.
- HECKMAN, J.J. (1996) “Randomization as an Instrumental Variable”, Review of Economics and Statistics, 78(2), págs. 336 – 341.
- HECKMAN, J.J. (1997) “Instrumental Variables: A Study of the Implicit Assumption Underlying One Widely Used Estimator For Program Evaluations”, Journal of Human Resources, 32, págs. 441 – 462.
- HECKMAN, J.J., HOHMANN, N., KHOO, M., SMITH, J. (1997) Substitution and Drop Out Bias in Social Experiments: A Study of an Influential Social Experiment, Working Paper, University of Chicago.
- HECKMAN, J.J., HOTZ, J. (1989) “Choosing Among Alternative Nonexperimental Methods for Estimating the Impacts of Social Programs: The Case of Manpower Training”, Journal of the American Statistical Association, 84, págs. 862 – 880.
- HECKMAN, J.J., ICHIMURA, H., SMITH, J.A., TODD, P. (1998) “Characterizing Selection Bias Using Experimental Data”, Econometrica, 66, págs. 1017 – 1098.
- HECKMAN, J.J., ICHIMURA, H., TODD, P. (1997) “Matching as an Econometric Evaluation Estimator: Evidence from Evaluation a Job Training Program”, Review of Economic Studies, 64, págs. 605 – 654.
- HECKMAN, J.J., ICHIMURA, H., TODD, P. (1998) “Matching as an Econometric Evaluation Estimator”, Review of Economic Studies, 65, págs. 261 – 294.
- HECKMAN, J.J., LALONDE, R.J., SMITH, J.A. (1999) “The Economics and Econometrics of Active Labor Market Programs”, en ASHENFELTER, O., CARD, D. (EDS.): Handbook of Labor Economics, Vol. III.

- HECKMAN, J.J., ROBB, R. (1985) “Alternative Methods for Evaluating the Impact of Interventions”, en HECKMAN, J.J., SINGER, B. (EDS.): Longitudinal Analysis of Labor Market Data, Cambridge University Press, Cambridge, págs. 156 – 245.
- HECKMAN, J.J., ROBB, R. (1986) “Alternative Identification Assumptions in Econometric Models of Selection Bias”, en SLOTTJE, S. (EDS): Advances in Econometrics: Innovations in Quantitative Economics, Essays in Honor of Robert L. Basmann, Vol. 5, JAI Press, Greenwich, págs. 242 – 287.
- HECKMAN, J.J., SMITH, J.A. (1993) “Assessing the Case for Randomized Evaluation of Social Programs”, en JENSEN, K., MADSEN, P.K. (EDS.): Measuring Labour Market Measures, Danish Ministry of Labour, Copenhagen, págs. 35 – 95.
- HECKMAN, J.J., SMITH, J.A. (1995) “Assessing the Case for Social Experiments”, Journal of Economic Perspective, 9, págs. 85 – 110.
- HECKMAN, J.J., SMITH, J.A. (1997A) Ashenfelter’s Dip and the Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies, Working Paper, University of Chicago.
- HECKMAN, J.J., SMITH, J.A. (1997B) The Sensivity of Experimental Impact Estimates: Evidence from the National JTPA Study, NBER Working Paper 6105, National Bureau of Economic Research, Cambridge (Mass.).
- HECKMAN, J.J., SMITH, J.A., CLEMENTS, N. (1997) “Making the Most Out of Program Evaluations and Social Experiments: Accounting for Heterogeneity in Program Impacts”, Review of Economic Studies, 64, págs. 421 – 471.
- HECKMAN, J.J., VYTLACIL, E.J. (2001) “Instrumental Variables, Selection Models, and Tight Bounds on the Average Treatment Effects”, en LECHNER, M., PFEIFFER, F. (ED.), Econometric Evaluation of Labour Market Policies, ZEW Economic Studies, Heidelberg.
- HOLLAND, P.W. (1986) “Statistics and Causal Inference”, Journal of the American Statistical Association, 81, págs. 945 – 970.

- IMBENS, G.W. (1999) The Role of the Propensity Score in Estimating Dose-Response Functions, NBER Technical Working Paper, 237.
- IMBENS, G.W. (2001) "Some remarks on instrumental variables", en LECHNER, M., PFEIFFER, L. (ED.), Econometric Evaluation of Labour Market Policies, ZEW Economic Studies, 13, Heidelberg.
- IMBENS, G.W., ANGRIST, J.D. (1994) "Identification and Estimation of Local Average Treatment Effects", Econometrica, 62, págs. 467 – 475.
- LALONDE, R. (1986) "Evaluating the econometric evaluations of training programs with experimental data", American Economic Review, 76, págs. 604 – 651.
- LECHNER, M. (1996) Nonparametric Bounds on Employment and Income Effects of Continuous Vocational Training in East Germany, Discussion Paper 96-31, Zentrum für Europäische Wirtschaftsforschung, Mannheim.
- LECHNER, M. (1998) Training the East German Labor Force. Microeconomic Evaluation of Continuous Vocational Training after Unification, Studies in Contemporary Economics, Heidelberg.
- LECHNER, M. (2001) "Identification and Estimation of Causal Effects of Multiple Treatments under the Conditional Independence Assumption", en LECHNER, M., PFEIFFER, F. (ED.): Econometric Evaluation of Labour Market Policies, ZEW Economic Studies, Heidelberg.
- MANSKI, C.F. (1989) "Anatomy of the Selection Problem", Journal of Human Resources, 24, págs. 343 – 360.
- MANSKI, C.F. (1990) "Nonparametric Bounds on Treatment Effects", American Economic Review, 80, Papers and Proceedings, págs. 319 – 323.
- MANSKI, C.F. (1995) Identification Problems in the Social Sciences, Cambridge: Harvard University Press.

- MANSKI, C.F., GARFINKEL, I. (1992) "Introduction",  
 en MANSKI, C.F., GARFINKEL, I. (EDS.):  
Evaluating Welfare and Training Programs, Harvard University  
 Press, Cambridge (Mass.), págs. 1 – 22.
- MOFFITT, R. (1991) "Program Evaluation with Nonexperimental Data",  
Evaluation Review, 15, págs. 291 – 314.
- ROSENBAUM, P.R., RUBIN, D.B. (1983) "The Central Role of the  
 Propensity Score in Observational Studies for Causal Effects",  
Biometrika, 70, págs. 41 – 55.
- ROSENBAUM, P.R., RUBIN, D.B. (1985) "Constructing a Control Group  
 Using Multivariate Matched Sampling Methods that Incorporate the  
 Propensity Score", The American Statistician, 39, págs. 33 – 38.
- ROY, A.D. (1951) "Some Thoughts on the Distribution of Income",  
Oxford Economic Papers, 2, págs. 135 – 146.
- RUBIN, D.B. (1974) "Estimating Causal Effects of Treat-  
 ment in Randomized and Nonrandomized Studies",  
Journal of Educational Psychology, 66, págs. 688 – 701.
- RUBIN, D.B. (1977) "Assignment to a Treatment Group on the Basis of  
 a Covariate", Journal of Educational Statistics, 2, págs. 1 – 26.
- RUBIN, D.B. (1978) "Bayesian Inference for Causal Effects",  
Annals of Statistics, 6, págs. 34 – 58.
- RUBIN, D.B. (1991) "Practical Implications of Modes of Statistical In-  
 ference for Causal Effects and the Critical Role of the Assignment  
 Mechanism", Biometrika, 47, págs. 1213 – 1234.
- SOBEL, M.E. (1994) "Causal Inference in the Social and Behavioral  
 Sciences", en ARMINGER, G., CLOGG, C.C., SOBEL, M.E. (EDS.):  
Handbook of Statistical Modeling for the Social and Behavioral Sciences,  
 Plenum Press, New York.
- STAIGER, D., STOCK, J.D. (1997) "Instrumental Variables Regression  
 with Weak Instruments", Econometrica, 65, págs. 557 – 586.

STERN, S. (1997) “Simulation Based Estimation”,  
Journal of Economic Literature, 35, págs. 2006 – 2039.

VELLA, F. (1998) “Estimating Models with Sample Selection Bias: A Survey”, Journal of Human Resources, 33, págs. 127 – 169.

VELLA, F., VERBEEK, M. (1998) Estimating and Interpreting Models with Endogeneous Treatment Effects, Working Paper, Rutgers University .