

Introducción a los métodos experimentales y cuasi experimentales para la evaluación de programas de capacitación

*Rafael Diez de Medina**

1. Introducción

En este capítulo se presentan los principales métodos de tipo cuantitativo que en los últimos años han sido crecientemente utilizados a la hora de evaluar los programas de empleo en general y de capacitación, en particular. Estos métodos se edifican en base a los avances metodológicos y econométricos de la última mitad del siglo XX y principios del XXI.

Los métodos de evaluación clásicos basados en estudios costo-beneficio, costo-eficiencia o de estudios de impacto, han sido utilizados principalmente por gobiernos y organismos internacionales en su necesidad de evaluar intervenciones de política, con la idea de brindar elementos objetivos sobre la racionalidad de los programas y evitar su fracaso. Uno de los campos en donde estas evaluaciones han sido más utilizadas es precisamente el campo de las políticas activas de empleo y, en particular, en los programas de capacitación y formación para el trabajo en grupos objetivo vulnerables, como pueden ser los jóvenes de bajos ingresos o los desempleados de larga duración.

Sin embargo, desde medios académicos y, posteriormente, desde su adopción por organismos multilaterales y consultores independientes, ha sido creciente la utilización de métodos más elaborados que implican el uso de técnicas estadísticas y econométricas más avanzadas que los métodos más clásicos. La localización de problemas serios de simultaneidad y de sesgos importantes en las estimacio-

* **Rafael Diez de Medina** se desempeña como especialista en el Departamento de Políticas de Integración de la Oficina Internacional del Trabajo.

nes de los indicadores que medían el impacto de los programas, por parte de académicos y estudiosos del tema, llamaron la atención hacia problemas antes ignorados. La necesidad de contar con métodos rigurosos de validación y, a la vez, los avances en la teoría estadística y econométrica, pusieron al alcance de la práctica una cantidad de opciones metodológicas que prometían ser métodos que aportaran mayor rigurosidad a la hora de discutir la propia existencia de varios programas de empleo y formación.

En ese sentido, los trabajos pioneros de J. Heckman (Premio Nobel de Economía) desde la década de los setenta hasta la actualidad, marcaron la irrupción de estos métodos de evaluación, los cuales se han multiplicado exponencialmente mediante numerosos aportes de otros investigadores y evaluadores, principalmente en su aplicación en países como Estados Unidos, Canadá y Europa. Sin embargo, los mismos han sido aplicados también en América Latina, y de ahí el interés de introducirlos en este Manual de Evaluación.

En este capítulo se introducen los métodos basados en diseños experimentales y los métodos cuasi o no experimentales que son aquellos en donde el avance en la rigurosidad ha sido mayor en los últimos años para la evaluación de programas de capacitación o de políticas activas del mercado de trabajo.

2. El problema fundamental de la evaluación

La literatura sobre evaluación de programas toma prestada la terminología utilizada principalmente por ciencias médicas o biológicas, donde se plantea un *tratamiento* nuevo a implementarse (políticas o intervenciones) que se administra a un grupo de personas (*“grupo de tratamiento”*). Por otra parte existe otro grupo que no es tratado que constituye el llamado *“grupo de control”*. Se trata, por tanto, de estudiar el o los impactos de este tratamiento en diferentes variables denominadas *“resultados”*. De la diferencia entre los resultados que se muestran luego de la intervención y los prevalecientes antes de la misma, surge la evaluación. En el caso de las políticas de empleo que se basan en capacitación para el trabajo, los resultados pueden ser, por ejemplo, los ingresos salariales del individuo, su estatus laboral, un índice de competencia laboral o de suficiencia en los conocimientos o destrezas, entre otros muchos.

Naturalmente que lo anterior es una simplificación de un fenómeno más complejo. Siguiendo a Heckman, LaLonde y Smith (1999); Smith (2000); Wooldridge (2002); y Stock y Watson (2003) se pueden clasificar los efectos de un programa activo de empleo, como el de capa-

citación en varias dimensiones. En primer término, los efectos directos del programa se dan primeramente en los individuos que son los participantes del mismo, pero las políticas también impactan en la población en su conjunto (dependiendo de la magnitud de la intervención). Quizá impactan más en algunos grupos especiales de población, como los más vulnerables o los menos calificados. También hay efectos indirectos o de “*equilibrio general*” que llevan a que el resultado de una política pueda influir tanto a los participantes como a los no participantes, a la vez que puede llegar a tener impactos de corto o largo plazo en el Producto Bruto Interno de un país o en los mercados de trabajo locales o regionales.

Además de los efectos directos e indirectos mencionados anteriormente, es interesante notar que, en todo programa, hay diferentes actores (individuos del grupo de control, del grupo de tratamiento, administradores del proyecto, políticos, etc.) que juegan con diferentes expectativas y características influyendo notoriamente en su impacto. Es por ello que los métodos han venido incorporando cada vez más elementos de comportamiento. Así, es común que los métodos propongan algún modelo causal que trate de explicar la decisión de participación en un programa de capacitación, por ejemplo, en función de variables como la edad, el sexo, la etnia, el estado de los mercados de trabajo de la región, el ingreso familiar, la experiencia laboral, entre otras. Muchos de ellos incorporan también la lógica del punto de vista de los administradores y ejecutores del programa (por ejemplo, las decisiones de las entidades de capacitación en los programas de formación descentralizados y sus determinantes) o de los políticos u organismos externos evaluadores. Otros, atienden al impacto macroeconómico del programa: cómo afecta a la totalidad de la economía o cuáles son los efectos directos e indirectos en el bienestar de los hogares de un programa de capacitación.

El llamado “*problema fundamental*” en la evaluación de programas parte de hacer reflexionar sobre el hecho de que una misma persona no puede simultáneamente participar y no participar en un programa de capacitación. Así, se puede definir una variable dicotómica o binaria P_i de participación del individuo i , de la siguiente manera:

$$P_i = \begin{cases} = 1 & \text{si } i \text{ participa} \\ = 0 & \text{si } i \text{ no participa} \end{cases}$$

La comparación fundamental que es entre la participación en el programa de formación y la situación hipotética de que el participante no hubiera recibido la misma, no es observable directamente para cada individuo y eso es precisamente lo crucial en la evaluación.

Si el resultado del tratamiento se expresa como Y_{1i} (ingreso salarial posterior al tratamiento del individuo i), el mismo se debería comparar con Y_{0i} que sería el resultado obtenido, en el mismo período de duración del tratamiento, por el mismo individuo, de no haber pasado por la capacitación. Lo ideal, por lo tanto, sería conocer la diferencia $\Delta_i = Y_{1i} - Y_{0i}$, denominada “impacto” o “efecto del tratamiento”, calculada para todos los participantes. Sin embargo, ello es imposible, ya que hay necesariamente falta de información en alguno de los dos términos.

Los métodos de evaluación cuantitativa suponen generalmente que la terna (Y_{0i}, Y_{1i}, P_i) es un vector aleatorio, es decir, formada por variables que, antes de ser observadas son aleatorias y por tanto tienen una distribución de probabilidad conjunta. Los resultados Y_{ji} pueden ser de tipo continuo (como los ingresos) o discretos (estatus laboral).²²

Se supone que la muestra es independiente e idénticamente distribuida, es decir, que las extracciones se hacen en forma aleatoria e independiente, lo que implica que el tratamiento de un individuo no tiene influencia en el resultado de otros (no existen efectos de equilibrio general) y que el vector X representa las variables explicativas de los resultados Y , entonces se puede afirmar lo siguiente:²³

$$[1] \quad Y_{0i} = X_i \beta_0 + u_{0i}$$

$$[2] \quad Y_{1i} = X_i \beta_1 + u_{1i}$$

donde $E(u_{ji}/X_i) = 0$

Por otra parte existiría un conjunto Z de variables que explicarían la probabilidad de que un individuo participe en un programa de capacitación, donde pueden existir variables coincidentes con las incluidas en el conjunto X :

$$[3] \quad \text{Prob}(P_i = 1) = f(Z_i) + v_i$$

La X_i representa el vector que incluye variables de características personales, familiares y condiciones del mercado laboral local del individuo. Estas variables deben ser preexistentes a la participación, ya que P no debería influir en las X .

En realidad, no es posible conocer todas las X y Z para todos los individuos. Con los participantes se puede estimar Y_{1i} (ecuación [2]) y con los no participantes se puede estimar Y_{0i} (ecuación [1]) y con todos se puede estimar un modelo que explique la participación en función de Z (ecuación [3]).

²² Una variable es continua cuando entre dos valores de la misma se puede encontrar siempre otro posible valor (conjunto denso), y es discreta cuando la misma adopta valores finitos o infinitos numerables. Así, el ingreso es una variable continua y estar desocupado o no, es discreta.

²³ Se supone que $E(Y_{1i}/X_i) = g_1(X_i) = X_i \beta_1$ y lo mismo para Y_{0i} . El operador E es la esperanza.

Sin embargo en las tres ecuaciones hay variables no observadas e información no disponible y por ello existen los términos de error u_{0i} , u_{1i} y v_i . La estimación se hará en forma condicional al conocimiento de la información sobre X y Z disponible por el evaluador.

El impacto de los programas es diferente según los individuos aunque analíticamente se puede ir suponiendo un abordaje crecientemente real a partir de supuestos simplificadores. Estos tres supuestos van a condicionar el tipo de características que se estudiarán en la evaluación de los programas.

- Uno podría comenzar suponiendo que el efecto es igual para todos los individuos, lo que es bastante irreal.
- Luego, se podría pensar que el efecto es diferente entre individuos, y antes de la capacitación, ni los individuos ni los responsables del programa, saben nada acerca de quienes van a ganar más o menos del programa (las varianzas de los impactos tienen poca importancia).
- Finalmente, el supuesto más real es que el efecto es diferente para los participantes, y ellos y los responsables tendrían información antes de comenzar la capacitación sobre quienes van a resultar más o menos favorecidos del programa (seguramente en función de sus características personales, observadas y no observadas las que impactarán en el resultado del mismo).²⁴ Esto último tiene implicancias muy importantes para las políticas pues excluir o incluir grupos de personas con determinadas características provocará que los programas sean más o menos exitosos.

3 Características más importantes a estudiar en la evaluación de programas

Podrían citarse muchas características poblacionales o parámetros de interés para el estudio de la evaluación. Las mismas serían las respuestas a las preguntas usuales sobre la evaluación de los programas y dependerán de los tres supuestos citados al final del numeral anterior. En la definición de las características se supondrá que la participación en un programa de capacitación es voluntaria y no compulsiva, lo que se corresponde con la mayoría de los programas implementados.

El primer parámetro de interés sería el efecto del programa en sus participantes (*"treatment on the treated"* o *"training on the trained"*, TT). Este efecto sería el primero a evaluar, ya que es crucial: si el resultado es menor que el costo, el programa debe eliminarse. Este efecto no se puede estimar para cada individuo, y aun si se lograra

²⁴ Por ejemplo, en los programas de capacitación laboral para jóvenes se puede pensar que los jóvenes de bajos ingresos pero con cierto nivel de educación, o con cierta experiencia pueden aprovechar más los programas que el resto de los jóvenes, en términos de obtener mejores ingresos a posteriori.

hacerlo, no sería útil. Lo que se estima es un valor medio para la población bajo el tratamiento. En términos formales se puede definir como:

$$TT = E(Y_{1i} - Y_{0i} / X, P_i = 1)$$

donde el operador E es la esperanza del efecto Δ_i condicionada a que el individuo participe. En términos simples representa el promedio de los resultados de la capacitación para los participantes.

El segundo parámetro de interés es el efecto medio del tratamiento ("*average treatment effect*", ATE). Representa el impacto medio del tratamiento en *toda* la población elegible y no sólo los que voluntariamente quieren participar.

Si en vez de eliminar o reducir un programa se quiere expandir a todas las personas elegibles y hacerlo obligatorio para todas las personas que cumplen determinadas condiciones, el ATE se vuelve una herramienta sumamente útil. Formalmente se define como:

$$ATE = E(Y_{1i} - Y_{0i} / X)$$

Se observa que es un promedio de los resultados en toda la población. Heckman critica esta medida, ya que afirma que no sería útil considerar toda la población donde se incluyen personas que no deberían ser beneficiarias, como las personas con alto nivel de ingresos familiares o amas de casa. No obstante estas críticas no son tan relevantes pues se puede restringir el promedio ATE para que considere únicamente segmentos de población que se juzgue importante incluir.

En tercer término, Imbens y Angrist (1994) definen el efecto marginal del impacto ("*local average treatment effect*", LATE). Este indicador mide el impacto medio del programa en las personas cuya participación va a cambiar en función de cambios en el mismo.

Por ejemplo, supóngase que se desea decidir la reducción del programa en un determinado porcentaje, implantando una matrícula o restringiendo la entrada de beneficiarios de los programas de capacitación, utilizando el criterio de darle la capacitación al primero que llega. Interesa saber el impacto del programa en ese porcentaje de personas que ahora quedarían fuera del mismo a raíz del cambio.

Puede ocurrir que el impacto medio de ese grupo marginal que queda afuera no excede el costo, mientras que en el resto sí lo hace. Si los que se benefician más son los más entusiastas por participar (y por tanto están dispuestos a pagar alguna matrícula o a llegar antes que otros), entonces aquellos con más ganancias posteriores van a participar con mayor probabilidad.

Si se consideran los tres supuestos descritos al final del anterior numeral se puede afirmar que, con el supuesto primero, $TT = LATE = ATE$, ya que los efectos son iguales para toda la población, lo que emerge como un supuesto muy simplificador. Por otra parte, en los otros dos supuestos, la *heterogeneidad* de la población es la regla y en eso se basan los análisis cuasi experimentales.

Existen adicionalmente otros parámetros de interés para el evaluador como por ejemplo:

- la varianza de los impactos Δ_i entre participantes:

$$V(Y_{1i} - Y_{0i} / X, P_i = 1)$$

- proporción de participantes que se benefician del tratamiento

$$P(Y_{1i} > Y_{0i} / X, P_i = 1)$$

Este indicador ayuda a estimar el efecto “*cream skinning*”, es decir el sesgo que se encuentra en ocasiones a raíz de que los centros de capacitación elijan los mejores candidatos para el programa para mostrar mejores resultados.

- cuantiles de los resultados Δ_i , es decir: el menor Δ_i tal que $F(Y_{1i} - Y_{0i} / X, P_i = 1) \geq k$ (donde k es la cuantila correspondiente de la distribución de los resultados).²⁵ Esto interesa cuando se quiere cuantificar el efecto distributivo de la intervención.
- efecto de la capacitación en las personas más pobres de la distribución, definidas como aquellas que tienen un ingreso inicial $Y_{0i} = y_0$, es decir

$$F(Y_{1i} - Y_{0i} / X, P_i = 1, Y_{0i} = y_0)$$

4 El sesgo de selección

El concepto de sesgo de selección es de fundamental importancia en el análisis de los métodos empíricos de evaluación de los programas. Es de capital importancia en el caso de los métodos cuasi o no experimentales, ya que es en ellos donde se presenta especialmente este problema. Los diseños experimentales surgen precisamente, entre otras causas, para controlarlo con el fin de mejorar las estimaciones que explican los resultados de un programa.

Las personas que no participan en un programa son diferentes a las que sí lo hacen, en lo a que a generación de resultados se refiere. Los

²⁵ La función $F(\cdot)$ es la función de distribución acumulada de probabilidad. Se estima como la suma acumulada de las frecuencias muestrales.

resultados, por ejemplo, pueden ser la generación de mayores salarios o haber conseguido empleo luego de la intervención o plan de capacitación para el trabajo. A la hora de tratar el problema fundamental de la evaluación, los resultados medios de los participantes, de no haber efectivamente participado (resultados *contrafácticos*, es decir no son los reales) *no serán iguales a los que no participaron* aun suponiendo que se tiene las mismas características individuales (X) para ambos grupos. ¿Por qué? Por ejemplo, los participantes pueden ser más motivados, más entusiastas, más ambiciosos, con mayor gusto por el trabajo, que los hubieran llevado de todas maneras a generar ingresos más altos aun sin haber participado, que los que no son participantes. De no controlar este aspecto se tiende a sobrevaluar los efectos del programa, ya que los ingresos mayores de estas personas participantes son mayores, no sólo por haber participado en el programa sino por otros aspectos *no observados*. Los factores antes citados como ejemplos no se observan ni miden y por tanto pueden implicar un sesgo importante en la explicación de los ingresos. En particular es posible que existan problemas en la estimación de los modelos que intentan explicar los Y_{ji} , es decir, en las ecuaciones [1] y [2]. La propia decisión de participación en el programa [3] se verá influida por este sesgo y por tanto es fundamental su tratamiento en los diseños cuasi experimentales.

De lo anterior se deduce que la distribución de los resultados (antes y después del programa de capacitación) que los participantes hubieran tenido de no haber participado en el programa, no es igual a la distribución de los resultados de los no participantes. En general, suponiendo que $F(\cdot)$ es la función de distribución de probabilidad de los resultados, se tiene:

$$F(Y_{0i}/X, P_i=1) \neq F(Y_{0i}/X, P_i=0) \text{ y además } F(Y_{1i}/X, P_i=1) \neq F(Y_{1i}/X, P_i=0)$$

$$E(Y_{1i}-Y_{0i}/X, P_i=1) \neq E(Y_{1i}/X, P_i=1) - E(Y_{0i}/X, P_i=0)$$

El análisis más en profundidad de este sesgo se realizará al estudiar los métodos cuasi o no experimentales. A los efectos de su tratamiento econométrico, véase Diez de Medina (1992).

5 Tipología de los métodos formales para la evaluación

Aceptar que la implementación de políticas activas en el mercado laboral implica necesariamente la evaluación de su impacto y por ello la elección del método más apto para cada situación, parece ser la tendencia más extendida a nivel de la práctica, desde la década de los ochenta en adelante. La proliferación de nuevas metodologías tendientes a perfeccionar las formas de evaluar los programas de empleo y capacitación, han marcado un crecimiento vertiginoso a partir

de esa década, principalmente al influjo de trabajos de J. Heckman *et al.* (1979, 1986, 1996, 1999) y otros autores como J. Smith (2000), R. LaLonde (1986), G. Imbens y J. Angrist (1994), entre otros muchos. Los métodos que surgen se agregan a los más tradicionales ya utilizados desde tiempo atrás, como el clásico análisis costo-beneficio o costo-eficiencia, los estudios que buscan medir los efectos agregados y los estudios de impacto.

Los métodos de evaluación pueden clasificarse en:

- a) métodos experimentales
- b) métodos cuasi experimentales
- c) análisis de impacto agregado
- d) análisis costo-beneficio

Cada uno muestra ventajas e inconvenientes y han sido utilizados en mayor o menor medida para estudiar el impacto de los programas de capacitación.

Brevemente se pueden definir de la siguiente forma:

Métodos experimentales

El método experimental, eminentemente *estadístico*, implica tomar personas que se postulan voluntariamente y son elegibles para un programa o tratamiento y asignarlas aleatoriamente a dos grupos: uno que efectivamente va a ser tratado o capacitado ("*grupo de tratamiento*") y otro que no recibirá el mismo ("*grupo de control*"). La principal característica, por tanto, es que el grupo de control no se forma por personas que no participan naturalmente, ya sea porque desconocen el programa o porque no quieren hacerlo, sino por personas que desean y postulan al programa pero que el experimento las excluye para formar su grupo control. Esto, se asegura, evita –por propia construcción, si ésta se hace correctamente– el sesgo de selección, ya que las características observadas y no observadas en el grupo de control y el de tratamiento, se distribuyen igual. La principal característica del método es su intencionalidad: no se espera crear estadísticamente un grupo de control a posteriori del programa, sino que se lo genera especialmente, bajo condiciones controladas. Este método se estudia más adelante.

Métodos cuasi experimentales

La metodología cuasi experimental (o no experimental) parte de la definición de varios parámetros de interés o características ya citadas y busca conformar un grupo de control utilizando técnicas econométricas con datos relevados en el programa y externos al mismo. Acepta que la experimentación social es de difícil ejecución y

puede tener serios inconvenientes éticos y de otra índole que serán luego estudiados, y avanza en utilizar técnicas actualizadas para contar con un grupo de control adecuado, principalmente para evitar el sesgo de selección. La diferencia más importante es que introduce *modelos* que implican la utilización de supuestos sobre *comportamientos humanos*, sociales y del mercado laboral que aprovecha los avances de la economía laboral y microeconomía, y no se basa estrictamente en un diseño experimental estadístico. No obstante esta metodología cuenta con inconvenientes que serán analizados posteriormente.

Métodos de impacto agregado

Son métodos que implican vincular los efectos microeconómicos y sociales de los programas con el comportamiento macro de la economía en su conjunto. Se acepta que los efectos de un programa se propagan en toda la población (participante y no) y se elaboran metodologías tendientes a estimar el impacto. Por ejemplo, los modelos de equilibrio general computables o las matrices de contabilidad social son ejemplos de ensayos de impacto agregado. Estos métodos no serán desarrollados aquí (véase Baker, 2000).

Análisis costo-beneficio

Es un enfoque más tradicional de evaluación de proyectos, que se basa en el concepto de costo de oportunidad de las diferentes alternativas al programa específico y la estimación de costos directos e indirectos, así como en la estimación de tasas sociales y privadas de retorno como resultado de igualar los flujos de costos y beneficios temporalmente. Generalmente se compara esta tasa con la tasa de interés vigente en la economía y de ahí surge la conveniencia o no de ejecutar un programa.

En las partes siguientes se analizarán los métodos experimentales y cuasi experimentales.

6 Métodos experimentales

Desde una óptima estadística, tomando prestado metodologías utilizadas principalmente por ciencias físicas y biológicas, los métodos experimentales fueron utilizados frecuentemente en ciencias sociales como la psicología o la educación, desde principios de la década de 1930. Es a partir de 1967 con la implementación del programa New Jersey Income Maintenance, en Estados Unidos que se comienza a utilizar con fines de evaluación de programas antipobreza o generadores de empleo. Posteriormente el Departamento del Trabajo de Estados Unidos comienza su utilización frecuente, pero es la evalua-

ción del programa de capacitación Job Training Partnership Act (JTPA) también en ese país, el que implementa la metodología experimental en gran escala para la evaluación de programas de entrenamiento y capacitación. Esta modalidad, pronto es introducida en programas de política activa en materia laboral de Canadá y Europa.

El método experimental para la evaluación difiere de los no experimentales principalmente por la utilización de la *asignación aleatoria* de las personas a diferentes tratamientos. En general se asigna a un tratamiento y a un grupo de control pero puede utilizarse también para evaluar múltiples tratamientos. Es un experimento controlado que se realiza en diferentes momentos. Por ejemplo, en el caso de un programa de capacitación, una vez que se presentaron y aceptaron los futuros participantes, se sorteó aleatoriamente un conjunto que no va a recibir el tratamiento y se estudian sus características y resultados a posteriori, al tiempo que el resto se considera grupo de tratamiento. A veces el objetivo de los métodos experimentales es el de estimar los impactos diferenciales entre diferentes tratamientos.

Se pueden distinguir dos tipos de experimentos: aquellos que buscan estimar la respuesta de una población a diferentes impulsos o intervenciones y, los más frecuentes, los que buscan estimar el impacto de uno o varios tratamientos. En el primer tipo se prueba una cantidad de combinaciones de ingresos y tasas de imposición y se observa cómo responde la población en su participación laboral en función de diferentes combinaciones; y en el segundo se considera el impacto de una intervención, como por ejemplo, la capacitación de mano de obra y el efecto final, luego de pasar por una “caja negra”, en los ingresos de la población beneficiaria. El segundo tipo es el más frecuente en la praxis de evaluaciones. Explícitamente, no se busca estimar un modelo de comportamiento que busque explicar las causas por las que se dan las causalidades como se dan. Por ello los resultados no se pueden generalizar a otros programas muy fácilmente.

Como ya se afirmó anteriormente, si las personas se asignan aleatoriamente a los dos grupos, luego de que ellas ya habían aplicado y probaron ser elegidas por el programa para participar en él, el tratamiento se distribuye independientemente de otros efectos causales del resultado final del tratamiento, eliminando el problema del sesgo por omisión de variables. Por ejemplo, en un programa de capacitación laboral, la experiencia laboral, su educación previa y los ingresos de la familia influirán luego en los resultados, en términos de ingresos del programa, pero en la medida que el programa se asigne aleatoriamente, la distribución de la experiencia laboral, educación previa e ingresos familiares es la misma en el grupo que se capacitó y en el grupo control que quedó sin capacitar. En resumen, la participación en el programa y las variables como la experiencia laboral y otras, están incorrelacionadas, por lo que *no existirá el sesgo de selección*.

La estimación del impacto en la evaluación experimental

Dentro del contexto de un experimento controlado con asignación aleatoria, para evaluar el programa se desea explicar el resultado experimental del mismo: por ejemplo, el ingreso salarial o el estatus laboral posterior, entre otros resultados posibles, en función de la participación en el programa del individuo.

Además de la participación, el resultado dependerá de variables como la experiencia laboral, la edad, la educación previa, los ingresos familiares, el sexo, el grupo étnico, entre muchos otros factores posibles de explicarlo.

Sean:

Y_i = Ingreso salarial o resultado del programa

P_i = Participación en el programa, ya definida como una variable dicotómica (1= participa en el programa, 0 no participa)

X_{ji} = Variables explicativas adicionales no correlacionadas con la participación, con $j = 1, 2, \dots, k$; $i = 1, 2, \dots, n$, donde k es el número de variables y n el número de participantes,

entonces se puede explicitar un modelo explicativo que explique los resultados del programa de la siguiente manera:

$$[4] \quad Y_i = \beta_0 + \beta_1 \cdot P_i + \delta_1 \cdot X_{1i} + \delta_2 \cdot X_{2i} + \dots + \delta_k \cdot X_{ki} + u_i.$$

La estimación del efecto del programa se realizará por Mínimos Cuadrados Ordinarios (MCO), ya que se supone que las variables no observadas recogidas en u_i no tienen correlación con la participación en el programa condicional a las variables X_{ki} .²⁶

El efecto causal del resultado (Y) de la participación en la capacitación (P) es la diferencia:

$$E(Y_{1i}/X, P_i=1) - E(Y_{0i}/X, P_i=0)$$

Teniendo en cuenta el modelo [4], la expresión anterior se puede expresar:

$$E(\beta_0 + \beta_1 + \delta_1 \cdot X_{1i} + \delta_2 \cdot X_{2i} + \dots + \beta_k \cdot X_{ki} + u_i) - E(\beta_0 + \delta_1 \cdot X_{1i} + \delta_2 \cdot X_{2i} + \dots + \delta_k \cdot X_{ki} + u_i) = \beta_1$$

lo que equivale al parámetro β_1 , cuyo estimador se denomina “*efecto del tratamiento*” o también el “*estimador de diferencias*”, lo que significa que es la estimación del efecto en el resultado de la presencia del tratamiento en relación al grupo de control. Lo anterior es válido en la medida de que $E(u_i/X, P_i)=0$.

²⁶ Los estimadores por MCO cuando no se cumple la incorrelación entre las variables explicativas y el término residual son inconsistentes, es decir, no importa cuán grande sea el tamaño muestral, las estimaciones muestrales no van a converger al verdadero parámetro, lo que invalida totalmente la inferencia que se haga con ellos. De ahí la importancia de lograr la incorrelación.

En el caso de asignación aleatoria, como es el caso de la evaluación experimental, la participación no debería estar nunca correlacionada ni con las X ni con el término de error donde están las variables no observadas, lo que asegura que no existe sesgo de selección ni inconsistencia de los estimadores.

El estimador anterior es muy útil en la evaluación de los efectos del programa, ya que es un estimador eficiente en presencia de asignación aleatoria y, más importante aún, permite probar si la aleatorización ha sido bien hecha. Si ésta no hubiese sido correctamente realizada, la misma estaría relacionada con las variables explicativas X, lo que provocaría estimadores inconsistentes. Por ello si se realiza la estimación de [4] y también de un modelo más simple: $Y_i = \beta_0 + \beta_1 \cdot P_i + u_i$ que excluye las variables X y las estimaciones de $\hat{\alpha}_1$ son muy diferentes entre sí, se estaría en el caso de una mala asignación aleatoria.

Cuando se realiza experimentación social se puede contar en general con datos de panel, es decir, los mismos individuos se observan en dos –o más– momentos del tiempo, o quizá con datos de corte transversal²⁷ repetidos para los no participantes. Cuando así sucede se puede utilizar otro estimador denominado “*diferencias en diferencias*”, que es el cambio promedio del resultado en el grupo de tratamiento o capacitación en el curso del experimento, menos el cambio promedio del resultado del grupo control a lo largo del mismo lapso. A esta diferencia se la puede enriquecer con variables explicativas como en el estimador de diferencias visto anteriormente.

Si se supone que el cambio promedio del ingreso sin participación (Y_0), de t_1 a t_2 , es igual para los participantes que para los no participantes:

$$E(Y_{0t_2} - Y_{0t_1} / P = 1) = E(Y_{0t_2} - Y_{0t_1} / P = 0)$$

el estimador de “*diferencias en diferencias*” del impacto será:

$$\hat{\beta}_1 = (Y_{1t_2} - Y_{0t_2})_{P=1} - (Y_{0t_2} - Y_{0t_1})_{P=0}$$

Este mismo estimador puede ser deducido planteando un modelo alternativo al [4] que en lugar de utilizar el nivel se utilice la diferencia de los Y, obteniendo:

$$[5] \quad \Delta Y_i = \beta_0 + \beta_1 \cdot P_i + \delta_1 \cdot X_{1i} + \delta_2 \cdot X_{2i} + \dots + \delta_k \cdot X_{ki} + u_i$$

Si se aplica el método de MCO a [5] se obtendrá el estimador anterior de “*diferencias en diferencias*” recién definido.

²⁷ Un conjunto de datos de corte transversal (“*cross section*”) implica tener una muestra de individuos en un momento dado del tiempo. Una encuesta de hogares sería un corte transversal mientras que un estudio de panel sería longitudinal en el sentido de que serían observaciones de los mismos individuos en distintos momentos del tiempo.

El estimador anterior tiene una virtud muy interesante: la de eliminar las diferencias existentes en los resultados *antes del tratamiento*. En el caso de la capacitación, elimina las diferencias en los ingresos de las personas antes de la misma. Por ejemplo, si el resultado (ingreso por trabajo) promedio del grupo del tratamiento es 80 antes del experimento, y el del grupo control es 40 –también al inicio del tratamiento–, y al final el ingreso promedio es de 160 para el grupo del tratamiento y de 60 para el de control, el resultado de comparar los dos grupos en el momento que termina el tratamiento daría $160-60=100$. Sin embargo, el resultado anterior sobreestima el efecto de la capacitación, ya que, antes de la misma, existían diferencias en los ingresos promedio de ambos grupos. El estimador “*diferencias en diferencias*” sería: $(160-60)-(80-40)=60$, cifra que considera las diferencias iniciales y evita sobreestimar los resultados de un programa.

En la ecuación [5] se puede considerar que ΔY_i es el cambio en los salarios durante la duración de la capacitación. Si en las variables X, por ejemplo, se incluye la educación previa, implicaría que la regresión permite que los individuos con más educación tiendan a tener mayores cambios en los ingresos salariales, independientemente de estar en un grupo de tratamiento o de control.

Cuando el tipo de procedimiento del tratamiento (capacitación) no se cumple cabalmente, sino que hay un *cumplimiento parcial*, las variables explicativas del modelo [4] podrían estar correlacionadas con las características no observadas contenidas en el término de error u_i por lo que el método de estimación por MCO ya no sería correcto sino que se debería utilizar el método de Variables Instrumentales (IV).²⁸

En planes de capacitación laboral, por ejemplo, cuando se asigna aleatoriamente un individuo al grupo de tratamiento y luego éste no cumple con la asistencia o los requisitos mínimos para considerarlo como participante, el modelo explicativo no sería el correcto, sino que se debería utilizar la asignación aleatoria original como instrumento de la verdadera participación (parcial) en el experimento. Esta técnica econométrica consiste en encontrar una variable altamente correlacionada con las variables X y P pero que no esté correlacionada con el término de error. La asignación original cumple estos requisitos por la propia forma en que fue realizada, lo que asegura que sea un buen instrumento.

Se parte de considerar que, si el experimento en realidad fue aleatoriamente recibido, la variable de participación P_i estaría incorrelacionada con las características individuales del participante. Por ello, si se realiza un modelo explicativo del resultado P_i utilizando las variables explicativas X_{ji} y resulta que los coeficientes asociados a éstas son significativamente nulos, entonces se puede concluir que hubo recepción aleatoria de los participantes. Para ello se puede es-

²⁸ El método consiste en encontrar una o varias variables altamente correlacionadas con las X pero que no esté correlacionadas con el término de error. El método no implica la sustitución de las X sino la utilización de un modelo bietápico que se expone en Greene (2002).

pecificar modelos especiales (*logit* o *probit*) cuyas características principales son las de tener como variable a explicar una variable dicotómica o binaria. Lo que se estaría explicando sería estrictamente la probabilidad de participar en un programa de capacitación de un individuo i en función de variables explicativas X_{ij} .

Existen otros estimadores alternativos como el de “antes-después” y el estimador de “*cross-section*” que no se tratarán aquí (véase Heckman *et al.*, 1999).

A continuación se presentan las principales ventajas e inconvenientes del método experimental.

Ventajas

- Fácil implementación.
- El evaluador puede crear las condiciones que debe cumplir el experimento para evaluar el tratamiento específico, antes de ser implementado, mediante la utilización de la asignación aleatoria, lo que asegura un vínculo estrecho entre tratamiento y resultado.
- Los defensores de la experimentación afirman que la causalidad entre tratamiento y resultado es más identificable en la experimentación que en los métodos cuasi experimentales.
- Es más fácil de explicar a los que implementan y aprueban la ejecución de políticas, puesto que hay que explicar únicamente cómo fue hecha la asignación aleatoria.
- Los resultados de un experimento son menos discutidos puesto que no dependen de las diferentes metodologías y estrategias de estimación econométrica.
- Permiten evaluar las evaluaciones no experimentales.
- La estimación de los impactos se realiza por métodos de estimación más fáciles de implementar y no tan debatibles desde el punto de vista econométrico como en los modelos no experimentales. Esto, bajo el supuesto de que la asignación aleatoria se implementa bien y se evitan los sesgos discutidos más abajo.
- No se presenta el sesgo de selección, ya que el análisis controla los efectos de las características observables y no observables por la asignación aleatoria.

Inconvenientes

- Muy altos costos.
- Existen limitaciones éticas a la experimentación social. Como el tratamiento generalmente es visto como una política beneficiosa para los individuos, hacer una convocatoria, una selección y luego dejar aleatoriamente de lado un conjunto para que sea un grupo control puede conllevar muchas críticas desde el punto de vista ético. Considérese, por ejemplo, un plan de capacitación para

desempleados de larga duración. Aparte, es dudoso que el grupo control resultante sea fácilmente estudiado a posteriori puesto que generalmente desencadena resistencias.

- Necesidad de largos plazos para evaluar el impacto de los experimentos.
- Fracaso de la aleatorización (“*randomisation bias*”). Esta es una falla decisiva para invalidar el método. Si se realiza una asignación que presente un sesgo deliberado o involuntario el método no deberá ser utilizado para la evaluación, ya que se podría dar una correlación entre la participación y ciertas características de los individuos. Si, por ejemplo, el aviso de que unos participantes van a ser grupo de control puede llevarlos a cambiar su comportamiento.
- Fallas en seguir las reglas del tratamiento. Por ejemplo, si los participantes no asisten a la capacitación o lo hacen en forma esporádica, se presenta un *cumplimiento parcial*. También se presenta una imperfección en el método cuando los individuos del grupo control reciben la capacitación de alguna otra forma (*sesgo de sustitución*). En definitiva, para que un diseño experimental sea válido, los protocolos previstos para ambos grupos deben cumplirse. De lo contrario se está en presencia de un experimento inválido.
- Abandono del tratamiento (“*attrition*”). Se trata de personas que dejan la capacitación en algún momento de la misma. Si las causas son totalmente exógenas al experimento no habría problemas mayores de sesgo, aunque sí lo habría de existir correlación entre abandono y características de la capacitación. Por ejemplo, si los más capaces son tentados con mejores salarios fuera de la ciudad y sistemáticamente abandonan la capacitación, entonces quedarán los menos capaces y por tanto el tratamiento estará correlacionado con variables omitidas inherentes a la persona. El tratamiento del abandono y su influencia en la evaluación se trata en Heckman, Smith y Taber (1998).
- Efectos del experimento. Como se trata de seres humanos, el mero hecho de ser parte de un experimento y ser consciente de ello, puede alterar el propio comportamiento. Para mitigarlo, en ciencias físicas o médicas se realizan experimentos ciegos, es decir, los participantes no saben si están en el grupo de control o no. No obstante en programas laborales o de capacitación ello no es posible, ni por parte del participante ni por parte del propio ejecutor o administrador del programa, quienes saben que están bajo observación evaluadora y por tanto pueden alterar su comportamiento. Este efecto a veces se denomina “*efecto Hawthorne*”, a raíz de un experimento de productividad laboral de la Compañía General Electric en su planta de Hawthorne, Estados Unidos, en los años 1920 y 1930.
- Muestras chicas. Como los experimentos son usualmente muy caros se tiende a contar con muestras chicas, lo que lleva a estimaciones más imprecisas de los parámetros de interés.

- Muestra no representativa. Cuando los participantes se eligen de una forma sesgada, como por ejemplo, pertenecientes a una etnia en particular o la elección se hace en base a voluntarios, la inferencia posterior será inválida.
- Programa no representativo. Si el experimento se hace como programa demostrativo que es reducido y luego se implementa en gran escala, la evaluación en base al experimento quizá no sea válida para evaluar el grande. Sería peor aun si las condiciones del experimento difieren de las condiciones prevalecientes en la ejecución del programa en gran escala.
- Efectos de *equilibrio general*. La puesta en marcha de un programa puede tener efectos, tanto en el grupo tratado, como en el de control, así como en el resto de la población en su conjunto. Puede incurrirse en “*efectos de desplazamiento*”, por ejemplo, en el sentido de que programas de capacitación en gran escala pueden sustituir programas financiados por los empleadores. También, personas con desempleo de larga duración que mejoren en sus métodos de búsqueda, por un programa que los ayude a ello, pueden sustituir a personas con desempleo más esporádico, aumentando la duración para éstos que fueron desplazados.

7 Métodos cuasi experimentales

Los métodos cuasi experimentales son aquellos que se utilizan más frecuentemente, ya que existe poca práctica de la experimentación social, especialmente debido a los altos costos en tiempo y en recurso que supone esta última. Por otra parte, las bondades que se citan en la experimentación se ven relativizadas en la medida de que muchas veces la realización práctica de la misma no es lo pura que debería ser para concluir fehacientemente sobre los impactos de los programas (Heckman y Smith, 1995). Es así que con mayor frecuencia se trata de observar datos existentes que seguramente su concepción original no fue con el objetivo de evaluar programas específicos, combinados con datos que surgen de registros administrativos de la propia administración del programa. A esto se agregan los estudios que abordan las relaciones macro y micro que da la teoría social y económica para explicar comportamientos, ya sea en el tiempo o geográficamente. Estos avances, que han sido muy grandes a nivel de ciencias sociales, no son incorporados en los estudios experimentales, los que se sustentan fundamentalmente en el diseño de experimentos estadísticos.

El conjunto de datos disponibles, junto al conocimiento de la teoría social y económica, tiñen a los métodos cuasi experimentales de mayor relevancia, pero también los hace depender crucialmente de la disponibilidad de datos confiables, con diseños muestrales realizados

en forma científicamente correcta, así como de la aplicación de métodos econométricos especiales que fundamentalmente tratan de lidiar con el sesgo de selección ya estudiado, y también con otros aspectos de la estimación de los modelos que han motivado toda una vertiente de análisis, cuyo exponente mayor es J. Heckman.

Como el quid del problema en los métodos cuasi experimentales es el sesgo de selección, no es de extrañar que la mayoría del tema se centre en cómo conformar grupos de control adecuados que no son diseñados ad hoc, sino que hay que conformarlos analíticamente. También se centrará en cómo controlar las características no observables de los individuos que hacen que pueda existir el sesgo a la hora de explicar los resultados de los programas como en [4]. Para ello, en la práctica se cuenta con diferentes tipos de conjuntos de datos:

- cortes transversales o encuestas en diferentes momentos del tiempo
- paneles de individuos a lo largo de uno o varios puntos temporales
- registros administrativos del programa (datos sobre participantes).

En la explicación de los resultados (salarios o estatus laboral antes y después de la capacitación, por ejemplo) y en la explicación de la decisión de participar en un programa por parte del individuo i , ambos aspectos ya explicitados en las ecuaciones [1] a [3], se deberá considerar especialmente que la regla de decisión por participar (ecuación [3]) y las ecuaciones explicativas del ingreso (ecuaciones [1] y [2]) sean adecuadamente distinguidas del punto de vista de la estimación y que se puedan obtener estimadores que cumplan buenas propiedades estadísticas.

Si se utilizan cortes transversales repetidos o paneles, el problema no es tan grave, ya que ambos diseños implican que los factores observables y no observables que influyen los resultados son separables aditivamente y hacen algunos supuestos sobre las propiedades de la serie de tiempo de los no observables (errores o residuos del modelo). No obstante, si bien las variables explicativas de la decisión de participar pueden coincidir en su casi totalidad con las variables explicativas de los resultados, es conveniente tener variables que expliquen la decisión pero no los resultados, ya que ello ayudará a distinguir ambas relaciones.

Siguiendo a Heckman y Smith (1996) se puede formalizar lo anterior partiendo de que el individuo i decide participar en un programa de capacitación en el momento q . Así, sea Y_{oit} el ingreso del individuo i en el momento t en ausencia de capacitación (o sea cuando $P_i=0$), que depende de las características X_{it} .

Para cada momento posterior al momento de la capacitación (q), o sea, $t > q$, cada individuo i tiene una dupla de resultados (Y_{0it}, Y_{1it}), de resultados sin y con capacitación, que se explican por un conjunto de variables personales y familiares (X_{it}).²⁹

$$Y_{0it} = X_{it}\beta_{0t} + u_{0it}$$

$$Y_{1it} = X_{it}\beta_{1t} + u_{1it}$$

En las relaciones anteriores u_{0t} y u_{1t} representan todas las características no observadas de los individuos, que son los factores desencadenantes del sesgo de selección. Generalmente se supone que $\beta_{0t} = \beta_{1t}$ a excepción de las ordenadas en el origen y además, que $u_{0it} = u_{1it}$.

Las dos ecuaciones anteriores se pueden sintetizar en una, al considerarse que el resultado Y_{it} , en general, se puede expresar como un promedio ponderado entre participantes y no participantes del programa:

$$[6] \quad Y_{it} = (1-P_i) \cdot Y_{0it} + P_i \cdot Y_{1it} = P_i \cdot \alpha + X_{it}\beta_t + u_{it} \quad \text{para } t > q$$

Si no se desea hacer el supuesto de que la respuesta de X no varía para los individuos capacitados y los no capacitados para no ser tan restrictivos, se puede suponer: $\beta_{0t} \neq \beta_{1t}$ por lo que:

$$[7] \quad Y_{it} = P_i (\beta_{1t} - \beta_{0t}) X_{it} + X_{it}\beta_{0t} + u_{it}$$

Finalmente, si se supone que $\beta_{0t} \neq \beta_{1t}$ y, además, que $u_{0it} \neq u_{1it}$, entonces se está en el caso más general (para $t > q$):

$$[8] \quad Y_{it} = P_i (\beta_{1t} - \beta_{0t}) X_{it} + X_{it}\beta_{0t} + u_{0it} + P_i (u_{1it} - u_{0it})$$

Esta última relación es la menos restrictiva e implica que el impacto de un programa depende de variables observadas y no observadas. Las características de los resultados “*treatment on the treated*” (TT) y el “*average treatment effect*” (ATE) están dadas por las siguientes expresiones, respectivamente:

$$TT = E(Y_{1it} - Y_{0it} / X_{it}, P_i = 1) = (\beta_{1t} - \beta_{0t}) X_{it} + E(u_{1it} - u_{0it} / X_{it}, P_i = 1)$$

$$ATE = E(Y_{1it} - Y_{0it} / X_{it}) = (\beta_{1t} - \beta_{0t}) X_{it} + E(u_{1it} - u_{0it} / X_{it})$$

Ambas características coincidirán cuando:

$$E(u_{1it} / X_{it}, P_i = 1) = E(u_{0it} / X_{it}, P_i = 1) = E(u_{0it} / X_{it}) = E(u_{1it} / X_{it}),$$

y eso ocurre únicamente cuando los errores son iguales.

²⁹ Es obvio que no se puede obtener esta dupla para $t < q$ ya que no existiría posibilidad de tener el resultado Y_1 puesto que no hubo capacitación hasta ese momento.

Sin embargo, ahora la relación entre la decisión de participar P_i y las características no observables u_0 y u_1 es sumamente importante, puesto que da origen al sesgo de selección. Éste surge de la correlación entre P_i y las u_{ji} . Por ejemplo, sistemáticos valores altos de u_{ji} en ausencia del tratamiento quizá muestra mayor probabilidad de participar en el programa y por tanto mayor Y a posteriori (dentro de las u_{ji} se podría incluir variables imposibles de observar como la ambición, el entusiasmo, la motivación). Esta correlación implica que la estimación de efectos como en [8] realizada por MCO u otros métodos clásicos darían estimadores inconsistentes. De ahí es que surge la utilización del *método de Variables Instrumentales* (IV). Por ejemplo, podrían constituir buenos instrumentos variables que se comportaran en forma muy similar a como se hubiesen comportado de haber sido asignadas aleatoriamente.

Siguiendo el análisis de Heckman y MaCurdy (1986), se puede afirmar que la decisión de participar en capacitación depende de un índice I de beneficios netos para el que toma la decisión, que incluye los costos de oportunidad, es decir, los ingresos no generados por asistir al programa, entre otros factores. El índice depende de un conjunto de variables Z observadas (características personales, familiares, laborales, etc.) y de variables no captadas incluidas en v . Si se supone una relación lineal:

$$I_i = I(Z_i, v_i) = Z_i \cdot \gamma + V_i$$

Los individuos participarán en los programas de capacitación en la medida que el índice I sea positivo, es decir, la probabilidad de participación se puede expresar como:

$$[9] \quad \Pr(P_i=1 / Z_i) = \Pr(I_i > 0) = F(V_i < Z_i \cdot \gamma)$$

Se ha demostrado que si se utilizan cortes transversales sucesivos para estimar el efecto de un programa es necesario tener una variable en el vector Z_i que no esté incluida en las X_i con el fin de no tener que hacer supuestos más restrictivos acerca de los residuos (restricción de exclusión para la identificación del sistema formado por [8] y [9]).

La estimación del impacto utilizando métodos cuasi experimentales

Los modelos descriptos ilustran la forma teórica de resolver el problema de realizar una evaluación de los programas de capacitación. Si se supone que la participación en la capacitación se puede considerar como si fuera aleatoria totalmente sin ninguna influencia con variables no observables, o si se puede identificar un instrumento para

evitar la correlación entre variables explicativas y características no observables, entonces el estudio práctico del impacto se realizaría en forma similar a lo estudiado para el método experimental.

Lo usual es contar con cortes transversales en varios momentos del tiempo, donde cada uno de las muestras contiene diferentes individuos. Se parte de la premisa de que ambos cortes son realizados en forma probabilística, es decir, son observaciones sorteadas en base a un diseño muestral basado en técnicas de probabilidad. Por ello, el corte transversal anterior al tratamiento se usará como sustituto del grupo de control y el posterior al tratamiento, será el grupo de tratamiento.

Una vez definido el grupo de control –ahora es una aproximación, al contrario que en el caso experimental–, la estimación de los parámetros de interés ya introducidos se hace en forma similar al método experimental, utilizando el estimador de “*diferencias en diferencias*”, por ejemplo, y los métodos de estimación basados en variables instrumentales. Como se vio anteriormente los primeros comparan el grupo de tratamiento y el de control antes y después de la intervención, mientras que el que utiliza *Variables Instrumentales* identifica variables que tengan relación con la participación pero que no influyan los resultados condicionados a esa participación.

Uno de los métodos crecientemente más utilizados en este contexto son los métodos de “*apareamiento*” o de “*matching*” que son aquellos que buscan conformar un grupo de control que sea, en todas sus características personales, muy similar al grupo de tratamiento, y por ello controla por el efecto antes discutido y evalúa el impacto en términos de diferencia en el resultado de ambos grupos. En otras palabras, para cada participante se busca un individuo que tenga las mismas características personales X pero que no haya participado, es decir, que pertenezca al grupo de control. Dentro de estos métodos, el llamado “*propensity score matching*” (PSM) aparece como el más desarrollado y recientemente fue aplicado para la evaluación de programas de empleo y pobreza en Argentina (Jalan y Ravallion, 2001). Este método estadístico fue propuesto originalmente por Rosenbaum y Rubin (1983).

El método de apareamiento, en general, conforma el grupo de control en base a un conjunto de variables explicativas X de los individuos. Este método requiere un enorme conjunto de datos que puedan dar lugar a tener un grupo de control con las mismas características que el grupo tratado. Por este aspecto, el PSM vincula el grupo de control y el de tratamiento, a través de la probabilidad de participar de los individuos $\Pr(X_i)$, en función de variables explicativas definiendo el *propensity score* como:

$$e(X_i) = \Pr(P_i=1 / X_i)$$

con la virtud de que ahora se trata de un número (escalar) y no de un vector de variables.

Cuando hay asignación aleatoria como en los métodos experimentales, la evaluación permite hacer una comparación entre grupo control y de tratamiento que hace que la distribución de probabilidad de las variables explicativas X sea la misma en uno y en otro caso.

Cuando no hay experimentación, las X pueden no tener necesariamente la misma distribución de probabilidad. Rosenbaum y Rubin definen lo que llaman “*balancing score*” como una función $b(X_i)$ tal que la distribución condicional de X_i dada $b(X_i)$ es la misma para ambos grupos, lo que es decir que, condicionado a que se cumpla $b(X_i)$, X_i y P_i son independientes. La forma que estos autores proponen como más idónea para esta función, es aquella llamada PSM.

Parten de definir un tratamiento “*fuertemente ignorable*” si, condicionado a las X_i , los resultados (Y_{0i} , Y_{1i}) y P_i son independientes y $0 < \Pr(P_i=1 / X_i) < 1$. Se prueba que, bajo esta condición, la diferencia entre el resultado del grupo del tratamiento y el del grupo de control es un estimador insesgado de Δ_i . Es decir:

$$E_X[E(Y_{1i}/X_i, P_i=1) - E(Y_{0i}/X_i, P_i=0)] = E(Y_{1i} - Y_{0i})$$

En definitiva esto significa que, dadas ciertas características X_i de los individuos, o dada la probabilidad de participar $P(X_i)$ la participación en el programa es independiente de los resultados en el estado de no participación. Ello implica que todas las variables que afectan la participación y los resultados del grupo de control deban incluirse en el apareamiento.

En la práctica, la PSM debe ser estimada mediante un modelo *logit* o *probit*. El apareamiento se realiza en base a la cercanía de los PSM estimados para cada individuo del grupo de tratamiento y del grupo de control. La idea es minimizar las distancias entre ellos para poder luego estimar las diferencias en los resultados. Hay varias formas de definir estadísticamente esas distancias pero ello no se tratará aquí (véase Rosenbaum y Rubin, 1983; Wooldridge, 2002).

Baker (2000) resume muy bien los pasos a seguir para utilizar la PSM:

- 1) Utilizar una muestra representativa de participantes y de no participantes que pudieran haber sido elegidos para el programa (las muestras deberían ser de la encuesta más similar y comparable).
- 2) Estimar un modelo *logit* o *probit* para explicar la probabilidad de participar utilizando ambas submuestras de participantes y no participantes elegibles.

- 3) Para cada individuo estimar la probabilidad de participar en función de sus características y del modelo estimado en 2). Ella se llama PSM.
- 4) Algunos de los no participantes pueden haber sido excluidos al principio porque tenían una PSM que estuviera fuera del rango de variación de la PSM en el grupo tratado. El rango de PSM estimado para el grupo de tratamiento debe corresponder lo más posible al de la muestra de no participantes.
- 5) Para cada individuo del grupo de tratamiento se deberá encontrar una observación en el grupo de no participantes que tenga la PSM lo más cercana posible. Generalmente se utiliza la diferencia absoluta entre las PSM (*“el vecino más cercano”*).
- 6) Se calcula el valor promedio del resultado para un conjunto, por ejemplo de cinco, de observaciones que tengan los 5 PSM más similares o cercanos. La diferencia entre la media y el valor observado para las observaciones del grupo de tratamiento es el estimador de la ganancia debido al programa para esa observación.
- 7) Finalmente se calcula la media de las ganancias calculadas según 6) y se obtiene la ganancia promedio del programa.

Ventajas

- Menor costo que los diseños experimentales, ya que generalmente las muestras proceden de otras fuentes diferentes a la de la administración del programa.
- Puede dar respuesta a todos los parámetros de interés citados como necesarios de conocer a la hora de hacer una evaluación. En particular puede dar los indicadores de distribución del impacto ya citados.
- Implica la utilización de modelos de comportamiento económico y social en la modelización, lo que deja de lado el enfoque únicamente estadístico prevaleciente en los diseños experimentales (considérese la modelización de la probabilidad de participar [8], por ejemplo). En particular, se trata de incorporar los avances en la teoría de las decisiones y del comportamiento macroeconómico del individuo en la estimación de los modelos econométricos subyacentes.
- No hay limitaciones éticas como en la experimentación social.
- No se incurre en el sesgo de aleatorización, ni hay efectos del tratamiento pues los individuos no saben si están en el grupo de tratamiento o control ya que son datos que se observan y no se inducen.
- Generalmente las muestras que se utilizan son grandes.

Inconvenientes

- Si no se logra conformar un grupo de tratamiento que se comporte como si hubiera sido asignado aleatoriamente, el método fracasaría, puesto que toda evaluación de impacto tendría estimaciones inconsistentes del punto de vista estadístico. Existen muchas maneras de estimar grupos de control adecuados que no se han estudiado en el capítulo pero todas descansan en técnicas econométricas que consideran directa o indirectamente la manera cómo los individuos hacen sus decisiones y alcanzan resultados.
- Se necesita una elaboración econométrica muy afinada para evitar dar resultados diferentes utilizando el mismo conjunto de datos. El desarrollo de estos instrumentos ha sido muy importante en la última década pero la proliferación de métodos han vuelto confuso contar con un resultado fiable, ya que los resultados diferentes pueden dar lugar a que la utilización de los métodos se haga en función de querer mostrar mejores o peores resultados de un programa.
- El conjunto de datos disponible, no forzosamente llena los requisitos que el evaluador necesita que cumpla. Como se necesita que los individuos se comporten como si fueran participantes o grupo de control, a veces los datos no se adecuan a los perfiles de uno u otro grupo.
- Se necesita ineludiblemente abordar con rigurosidad la estimación del sesgo de selección mediante diferentes estrategias que pueden dar diversidad de resultados.
- Es posible que exista “*sesgo de contaminación*”, en el sentido de que en los datos como grupo de control puede haber individuos que hayan pasado por capacitación similar a la del programa a evaluar, por ejemplo, pero que no se explicita porqué el conjunto de datos no contiene esa información.
- No se consideran efectos de equilibrio general, como tampoco lo hacían los diseños experimentales. Ellos se deberán estimar de alguna otra forma que considere con mayor rigurosidad la interrelación entre los niveles micro y macroeconómicos.

8. Conclusión

Este capítulo pretende únicamente ser una introducción a los métodos experimentales y cuasi experimentales en la evaluación de los programas de capacitación. Se observa que, mientras los métodos experimentales son deseables teóricamente, son de muy costosa y difícil conformación en forma totalmente compatible con los objetivos de evaluar los resultados. Su enfoque principalmente estadístico los hacen atractivos pero despojados de muchos aspectos que hacen a la inclusión de los comportamientos humanos en la metodología. Por otra parte es de destacar la poca utilización en la evaluación de programas en América Latina y el Caribe, principalmente debido a sus altos costos y la resistencia de los administradores de los mismos basándose en razones éticas y políticas.

La utilización de métodos cuasi experimentales es la práctica más extendida, aunque el nivel de rigurosidad y exactitud metodológica varía en forma muy importante. La necesidad de contar con bases de datos adecuadas y comparables muchas veces limita la validez de los resultados estimados y pone en tela de juicio la credibilidad de la metodología, ya que es fácilmente debatible en función de la opción econométrica elegida para su modelización. La falta de bases de datos estadísticamente confiables diseñados en forma coordinada con los programas de capacitación, muchas veces debido a la debilidad de los sistemas oficiales de estadística o la poca coordinación con otras áreas del Estado, hace difícil la utilización de estos métodos en forma precisa. El prever el levantamiento de información sobre participantes y no participantes en un programa debería ser, desde el inicio del mismo, parte integrante de la presupuestación de cualquier programa de capacitación. Ello aseguraría una evaluación de medio término y final compatible metodológicamente con los objetivos del programa. Muchas veces los administradores y ejecutores ven a la evaluación con resistencia y no se plantea la necesidad de contar con información de base. Los diseños previos de los programas deben introducir explícitamente la conformación de bases de datos desde el propio inicio del programa.

El avance metodológico en el sentido de mejorar los métodos cuasi experimentales es muy importante a nivel académico y de la práctica de programas en países desarrollados. Es indudable su importancia con el fin de poder corregir a tiempo o eliminar programas que no alcancen, o lo hagan parcialmente, los objetivos fijados. La necesidad de generar información de base para implementar con rigurosidad las diferentes metodologías se debe considerar al inicio de los programas pues ello es lo que asegurará una evaluación compatible con los objetivos trazados en los mismos.

La aceptación de la necesidad de evaluar los programas con técnicas científicamente robustas es creciente a nivel de la ejecución de las políticas de empleo. De los métodos revisados se puede inferir que la necesidad de contrastar los resultados obtenidos de un conjunto de participantes con un grupo de control, es vital para poder llegar a una conclusión sobre la eficiencia de las intervenciones, pero también para monitorear el cumplimiento de las metas a lo largo de la ejecución de los programas y no únicamente al final. La transparencia o “*accountability*” de los mismos se logra con evaluaciones científicas y robustas desde el punto de vista técnico.

9. Bibliografía

- AEDO, C.; NUÑEZ, S. *The impact of training policies in Latin America and the Caribbean: the case of Programa Joven*. Santiago de Chile: ILADES/Georgetown University, 2001.
- BAKER, J. *Evaluating the impact of development projects on poverty*. Washington: World Bank, 2000.
- BURLESS, G. The case for randomized field trials in economic and policy research. *Journal of Economic Perspectives*. v. 9, 1995. p. 63-84.
- DAR, A.; TZANNATOS, Z. *Active labor market programs: a review of the evidence from evaluation*. Washington: World Bank, 1996.
- DIEZ DE MEDINA, R. El sesgo de selección en la actividad de jóvenes y mujeres. *Suma*. v. 7 n.13,1992. p. 69-85.
- FAY, R. Enhancing the effectiveness of active labour market policies: evidence from Programme Evaluations in OECD countries. *Labour market and social policy*. París: OCDE, 1996.
- GALASSO, E.; RAVALLION, M.; SALVIA, A. *Assisting the transition from workfare to work: a randomized experiment in Argentina*. Washington: World Bank, 2001. (Mimeo)
- GREENE, W.H. *Econometric analysis*. 5a. ed. Londres: Prentice-Hall, 2002.
- HECKMAN, J. Sample selection bias as a specification error. *Econometría*. n. 47, 1979. p. 153-161.
- . Instrumental variables: a study of implicit behavioural assumptions in one widely used estimator. *Journal of Human Resources*. v. 32, 1997. p. 441-461.
- HECKMAN, J.; ICHIMURA, H.; TODD, P. Matching as an econometric evaluation estimator: evidence from evaluating a job training program. *Review of Economic Studies*. v. 64, 1997. p. 605-654.
- HECKMAN, J.; LALONDE, R.; SMITH, J. The economics and econometrics of active labor market programs. En: Ashenfelter, O.; Card, D. (Eds.) *Handbook of labor economics*. v. IIIA. New York: Elsevier Science, 1999.

- HECKMAN, J.; MACURDY, T. Labor econometrics. En: Griliches, Z.; Intriligator, M. (Eds.) *Handbook of econometrics*. Amsterdam, 1986. v.3.
- HECKMAN, J.; SMITH, J. Experimental and nonexperimental evaluation. En: Schmid, G. et al. (Eds.) *International handbook of labour market policy and evaluation*. New York: Elsevier Science, 1996.
- . Assessing the case for social experiments. *Journal of Economic Perspectives*. v. 9, n. 2, 1995. p. 85-110.
- HECKMAN, J.; SMITH, J.; CLEMENTS, N. Making the most out of Programme Evaluations and Social Experiments: accounting for heterogeneity in programme impacts. *Review of Economic Studies*. v. 64, 1997. p. 487-535.
- HECKMAN, J.; SMITH, J.; TABER, C. Accounting for dropouts in evaluations of social programs. *The Review of Economics and Statistics*. v. 80, 1998. p. 1-14.
- IMBENS, G.; ANGRIST, J. Identification and estimation of local average treatment effects. *Econometrica*. v. 62, 1994. p. 467-476.
- JALAN, J.; RAVALLION, M. *Income gains to the poor from workfare: estimates for Argentina's Trabajar Program World Bank*. 2001 (Mimeo.)
- . Estimating the benefit incidence of an antipoverty program by propensity score matching. *Journal of Business and Economic Statistics*. ASA. 2001.
- LALONDE, R. Evaluating the econometric evaluations of training programs. *American Economic Review*. v. 76, 1986. p. 604-620.
- PIERRE, G. *A framework for active labour market policy evaluation*. Ginebra: OIT, 1999. Employment and training papers, 49.
- ROSENBAUM, P.; RUBIN, D. The central role of the propensity score in observational studies for causal effects. *Biometrika*. v. 70, 1983. p. 41-55.
- SMITH, J. *Evaluating active labor market policies: lessons from North America*. Ontario: University of Western Ontario. Departamento de Economía; NBER, 2000. (Borrador)
- . A critical survey of empirical methods for evaluating active labor market policies. *Swiss Journal for Economics and Statistics*. v. 136, 2000. p. 1-22.
- STOCK, J.; WATSON M. *Introduction to econometrics*. Addison Wesley, 2003.
- WOOLDRIDGE, J. *Econometric analysis of cross section and panel data*. The MIT Press, 2002.

