



Metodologías

DE EVALUACIÓN DE IMPACTO

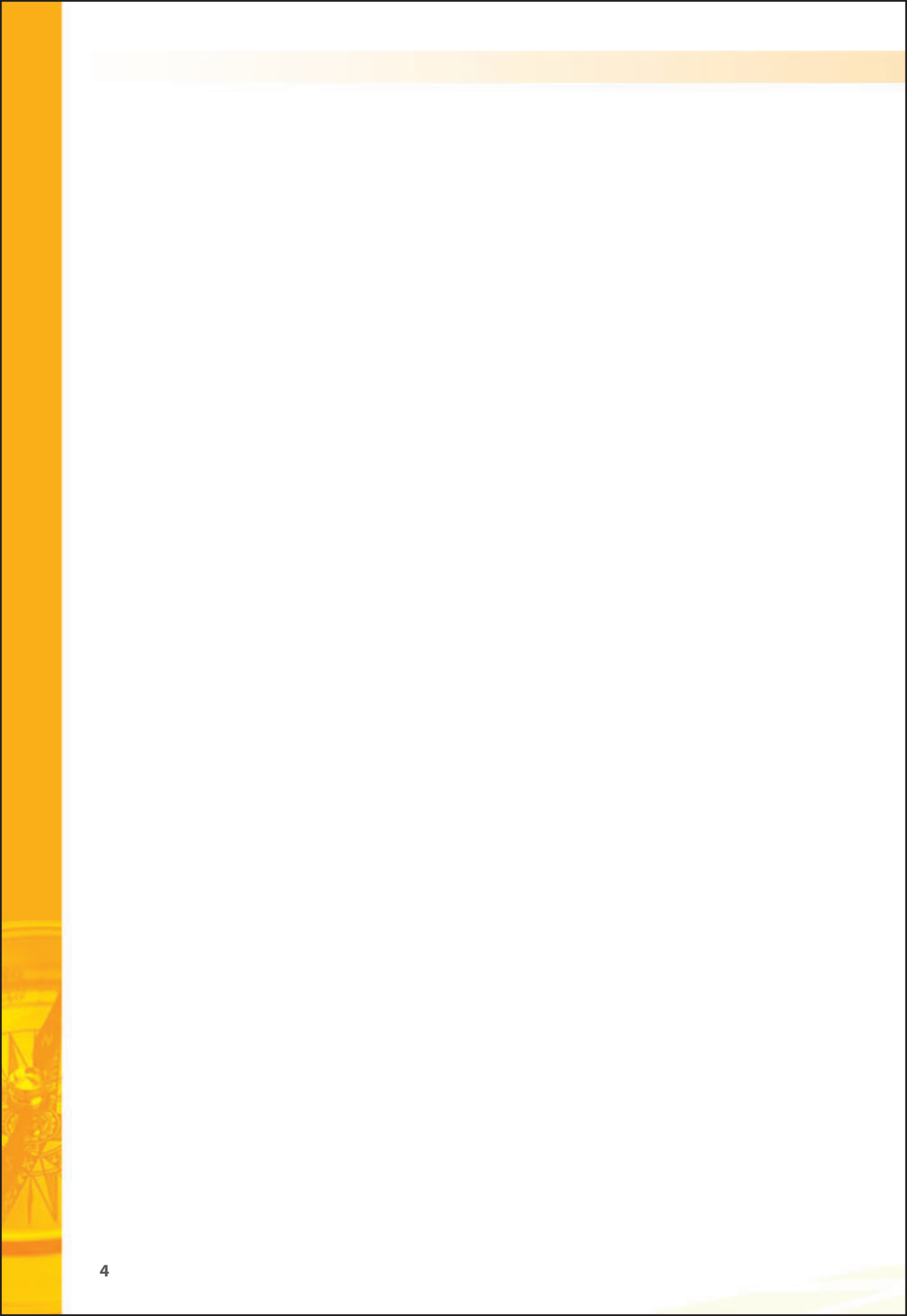
Secretaría Nacional de Planificación y Desarrollo
Subsecretaría de Seguimiento y Evaluación

INDICE

RESUMEN EJECUTIVO	5
I. Introducción	6
II. Formulación de los problemas en la evaluación	8
III. ¿Cómo podemos estimar los parámetros de impacto?	9
IV. Métodos para construir el contrafáctico	10
V. Elementos para tomar en cuenta en una evaluación de impacto	27
VI. Conclusiones	28
Bibliografía	29

Elaborado por la Subsecretaría de Seguimiento y Evaluación de la Secretaría Nacional de Planificación y Desarrollo, sobre la base del informe del consultor Hessel Oosterbeek¹.

1. Ph.D. en Economía, Profesor de Economía de la Educación, Universidad de Amsterdam, Blasiusstraat 34/2, 1091 CS Amsterdam, +31 (0)64254231.



RESUMEN EJECUTIVO

Las instituciones públicas y, en especial, la Subsecretaría de Seguimiento y Evaluación de SENPLADES requiere contar con herramientas que permitan realizar el seguimiento y la evaluación de programas y proyectos ejecutados por las diferentes instituciones que forman parte del Estado, y determinar el cambio en el buen vivir de la población que genera la inversión pública.

El impacto de los proyectos o programas puede ser determinado a través de métodos exclusivamente cuantitativos (experimentales y cuasi - experimentales) que exigen construir un grupo de control, tarea ardua y compleja. La comparación estadística entre el grupo de control y el grupo beneficiario de las acciones del proyecto o programa permite identificar, cuantificar y evaluar la conveniencia y eficacia de la inversión pública. Sin embargo, al momento de utilizar esta metodología, es importante considerar sus limitaciones, como el hecho de no tomar en cuenta el contexto en el que se realizan las intervenciones, los problemas técnicos y éticos en la selección del grupo de control y el elevado costo requerido para la aplicación.

En este contexto se presenta el siguiente documento que busca proporcionar una guía para realizar evaluaciones de impacto. En el capítulo 1 se hace una introducción referente a la metodología de evaluación de impacto. En el capítulo 2, se plantea la formulación de los problemas en la evaluación de impacto, de manera formal. El siguiente capítulo responde a la pregunta: “¿Cómo podemos estimar los parámetros para medir el impacto?” En el capítulo 4, se describen los dos grupos de métodos que permiten diseñar el contrafáctico: métodos experimentales y cuasi - experimentales. El capítulo 5 brinda una descripción de los elementos necesarios que se deben tomar en cuenta para realizar una evaluación de impacto. Finalmente, se presentan las conclusiones.

I. Introducción

La evaluación de proyectos, programas y/o políticas públicas se aplica con la finalidad de valorar la utilidad y los beneficios generados por la intervención pública mediante la aplicación de una serie de metodologías, que son aplicadas con objetivos distintos y permiten establecer conclusiones que buscan mejorar el desempeño de las instituciones; mejorar la eficacia y eficiencia de las intervenciones; que haya transparencia en la asignación y uso de recursos; y determinar el impacto en el buen vivir de la población.

La evaluación de impacto es una de estas metodologías. Tiene por debilidad el alto costo, debido al levantamiento de la línea base y, como principal cuestionamiento, el aspecto ético, por el hecho de requerir un grupo de control, conformado por personas de similares características a las atendidas por el proyecto, pero que no recibirán sus beneficios. Además, llega a establecer el impacto final de la intervención pero no hace un análisis de los factores que permitieron conseguir tales impactos.

La principal fortaleza de la metodología es, en cambio, la robustez con la que identifica los impactos finales de una intervención en las variables de interés, que constituyen el fin del proyecto, programa o política. A continuación, se describe la metodología.

La evaluación de impacto mide los cambios del buen vivir generados por una intervención en un grupo de beneficiarios; esto implica que se evalúa el grado al que se logran cumplir los objetivos propuestos en la intervención. Los impactos se expresan a través de las variables de resultado de la intervención. Es decir, evalúa los resultados de un programa específico sobre el grupo beneficiario, con relación a un grupo de personas de similares características que no reciben el beneficio (contrafactual).

Cualquier análisis y evaluación de programas que no utiliza grupos de comparación (contrafácticos), se ha comprobado que no es satisfactorio.

En términos generales, se requiere cumplir con dos objetivos cruciales al momento de evaluar un programa: el primero es lograr aislar el efecto de la intervención, y el segundo es crear un contrafactual que responde a la pregunta: ¿Qué hubiera sucedido si no se intervenía? Por lo tanto, presenta la situación sin intervención del programa. De esta manera, se logra comparar entre el grupo que recibe el tratamiento con el contrafáctico o grupo de control. Existen dos momentos en los cuales se puede evaluar:

* **Ex – ante:** se trata de simular la implementación:
“¿Qué pasaría si...?”

* **Ex – post:** se evalúa una vez ocurrida la intervención.

La evaluación ex – post no significa que deba empezar después de que el programa termine, o una vez iniciado. Las evaluaciones ex - post que entregan mejores resultados son aquellas que fueron diseñadas ex – ante, por lo general, simultáneamente con el programa.

Para evaluar el impacto se requieren datos de uno o más indicadores de resultado. El indicador seleccionado depende de lo que se quiera medir y de los objetivos que se haya propuesto alcanzar a través de la intervención.

Por ejemplo, en el caso de un programa que entregue transferencias monetarias a las familias más pobres con la condición de que estas lleven a sus niños y niñas a las escuelas y a centros de salud para controles periódicos (inversiones en recursos humanos), por ejemplo, el programa PROGRESA en México los indicadores relevantes serían el nivel de pobreza de esos hogares, la escolaridad y el estado de salud de los niños y niñas de los hogares beneficiados (indicadores que interpretan el nivel de pobreza).

Así mismo, se requiere algún método que permita inferir el grupo de comparación o contrafactual. Esto es intrínsecamente no observable, ya que es físicamente imposible observar a una misma persona en dos estados de la naturaleza (por un lado, participando en el programa, y la misma persona sin participar en el programa).

Las evaluaciones ex ante se llevan a cabo de dos maneras: aritméticamente o estructuralmente.

Aritméticas: A base de los datos recogidos en la línea de base, se podrían simular los cambios esperados en las variables de interés, así como los efectos en la población beneficiaria. Cabe resaltar que, este método no incluye modelos de comportamiento y no es muy confiable, pero nos da una idea de los posibles efectos de la futura intervención.

Estructurales: Permiten analizar los cambios en los comportamientos de los hogares, ya que incluyen modelos y ecuaciones simultáneas que simulan dichos comportamientos.

II. Formulación de los problemas en la evaluación²

El problema esencial, como se mencionó anteriormente, es que no es posible observar a la misma persona en dos estados de la naturaleza al mismo tiempo.

La descripción formal del problema es como sigue: se recolecta la información de un indicador de resultado Y_i para cada unidad i en una muestra de tamaño n . Por ejemplo: Y_i puede ser el ingreso del hogar i normalizado por una línea de pobreza específica de cada hogar (que refleje las diferencias en los precios que enfrentan en cada diferente localidad y en el tamaño y composición de cada hogar). Algunas de las unidades de la muestra reciben el programa y otras no; una variable dicotómica toma el valor $D_i=1$ para las unidades que reciben el programa y $D_i=0$ para las que no lo reciben. Se define la variable Y_iT si la unidad i recibe el programa (T se usa por "tratamiento") y Y_iC (C es por "contrafactual") si el hogar no recibió el programa. La ganancia individual del programa está dada por $G_i = Y_iT + Y_iC$.

También se recolecta información de un vector de covariables para los resultados ("variables de control"), X_i , que incluye la unidad como un elemento. El método más común para controlar por covariables asume que el resultado es lineal en el control de los parámetros, dando lo siguiente:

$$Y_iT = X_i \beta T + \mu_i T \quad (i=1, \dots, n) \quad (0.1)$$

$$Y_iC = X_i \beta C + \mu_i C \quad (i=1, \dots, n) \quad (0.2)$$

Se asume que los términos de error satisfacen $E(\mu_i | X_i) = 0$. Existen dos parámetros que se utilizan ampliamente que son "efecto promedio de tratamiento" (average treatment effect (ATE)), $E(G_i)$, y "efecto promedio de tratamiento en los tratados" (average treatment effect on the treated ($ATET$)), $E(G_i | D_i = 1)$. El ATE condicional es:

$$E(G_i | X_i) = X_i (\beta T - \beta C) \quad (1)$$

Mientras que el $ATET$ condicionado es:

$$E(G_i | X_i, D_i = 1) = X_i (\beta T - \beta C) + E(\mu_i T - \mu_i C | X_i, D_i = 1) \quad (2)$$

Como está claramente reconocido en la literatura, el problema esencial es que al estimar (1) y (2), (0.1) y (0.2) no son estimables, ya que no es posible saber los resultados de un participante en el contrafactual (Y_iC cuando $D_i=1$) ni en el resultado contrafactual bajo tratamiento (Y_iT cuando $D_i=0$). Para tratar de resolver este problema, se supone que se estima (0.1) en una sub-muestra para la cual $D_i=1$ mientras que (0.2) se estima en el resto de la muestra. El modelo estimable es entonces:

$$Y_iT = X_i \beta T + \mu_i T \quad \text{si } D_i=1 \quad (3.1)$$

$$Y_iC = X_i \beta C + \mu_i C \quad \text{si } D_i=0 \quad (3.2)$$

2. La mayor parte del texto fue tomado de Martin Ravallion "Evaluating Anti-Poverty Programs"

Se puede estimar una regresión simple (“switching”) para los resultados observables medidos en una muestra colectiva.

$$Y_i = D_i Y_i^T + (1 - D_i) Y_i^C = X_i \beta_C + X_i (\beta_T - \beta_C) D_i + \epsilon_i \quad (i=1, \dots, n) \quad (4)$$

en donde el término de error es:

$$\epsilon_i = D_i (\mu_i^T - \mu_i^C) + \mu_i^C \quad (5)$$

Los impactos entonces van a estar reflejados en los coeficientes de D_i en la ecuación (4). Existe un caso especial que es muy utilizado en la práctica; consiste en la especificación del “efecto común”, en el que todos, excepto los interceptos y los vectores de los parámetros β_C y β_T , se asumen que no varían con el tratamiento y por lo tanto se refieren a las ecuaciones (0.1) y (0.2). (Generalmente, este supuesto es impuesto sin ninguna justificación obvia, más allá del hecho de que uno puede inmediatamente ver el impacto promedio del resultado de una regresión estándar). Entonces la ecuación (4) termina siendo una regresión de resultados de participación y de las variables de control:

$$Y_i = (\beta_T - \beta_C) D_i + X_i \beta_C + \epsilon_i \quad (6)$$

en donde $\beta_T - \beta_C$ son los interceptos de las ecuaciones (0.1) y (0.2).

III. ¿Cómo estimar los parámetros para medir el impacto?

El método de mínimos cuadrados ordinarios (**MCO**) que se aplica a las ecuaciones (3.1) y (3.2), da estimadores consistentes de los parámetros de impacto en la ecuación (1) si no existe sesgo de selección en X , es decir que $E(\mu_i^T - \mu_i^C | X_i, D_i=1) = 0$; o equivalentemente, que los resultados de la media condicional no dependan del tratamiento, $E(Y_i^C | X_i, D_i=1) = E(Y_i^C | X_i, D_i=0)$. Entonces se dice que la ubicación del programa –la asignación a las unidades de (3.1) y (3.2)– es exógena. Un requerimiento quizás más estricto es que los resultados son independientes del tratamiento condicional en X . A esto se denomina “exogeneidad” en la literatura. En este caso, el término de error que se define en la ecuación (5) desaparece en las expectativas dados los regresores, asegurándose de que los **MCO** den estimadores consistentes bajo las condiciones estándares. También **ATE** y **ATET** se vuelven idénticos. El sesgo de una intervención viene dado por la siguiente diferencia: $E(Y_i^C | X_i, D_i=1) - E(Y_i^C | X_i, D_i=0)$, tal que la diferencia de la media de los resultados entre los participantes y no participantes esté dada por la siguiente identidad:

$$E(Y_i^T | X_i, D_i=1) - E(Y_i^C | X_i, D_i=0) = ATET + SESGO$$

donde $ATET = E(Y_i^T | X_i, D_i=1) - E(Y_i^C | X_i, D_i=1)$

Una manera para asegurar exogeneidad es aleatorizar la entrega del beneficio, este método se denomina evaluación experimental. En contraste, en una evaluación no-experimental (también llamada “estudio de observación” o “cuasi-experimental”) el programa es colocado con un propósito o fin (o sea, no aleatorio). Esto ocurre comúnmente en estudios de pobreza, ya que se lo coloca con relación a las características observables, por ejemplo, el número de dependientes en un hogar en áreas pobres. Además, puede existir una selección por parte de los participantes, como el hecho que algunos deciden no participar a pesar de que son elegibles, de manera que el proceso de tipo auto-orientado caracteriza a los programas antipobreza.

IV. Métodos para construir el contrafáctico

Existen dos tipos de diseños de evaluaciones de impacto:

1. Métodos Experimentales;
2. Métodos Cuasi – experimentales.

4.1 Métodos Experimentales

Uno de los métodos experimentales es la aleatorización por doble diferencia o “diferencia en diferencias”, que se considera el más exacto y más seguro.

4.1.1 Métodos de Doble Diferencia o “diferencias en diferencias”

En primer lugar, se debe aleatoriamente sortear quién recibe el tratamiento y quién no, de una forma balanceada. En segunda instancia, se debe levantar la línea de base previa a la intervención. En el tiempo 0 (inicial), es decir, el momento previo a la intervención, se esperaría que las características de los beneficiarios (grupo de tratamiento) y los no beneficiarios (grupo de control) sean iguales, dicho de otra manera que: $YT0 = YC0$. En tercer lugar viene la intervención, se debe asegurar que ésta ocurra DESPUÉS de la toma de la línea de base. Finalmente, se espera un tiempo, que dependerá de la lógica del programa, y se realiza la siguiente toma de datos (línea de seguimiento). En esta segunda toma se espera que exista ya una diferencia entre $YT1 \neq YC1$. Si existiera una diferencia, se dice que existe un impacto en el grupo de tratamiento debido a la intervención.

Existen dos supuestos básicos para el estimador de doble diferencia:

- i) El sesgo por selección es separable y no varía en el tiempo, de tal manera que desaparece al tomar las diferencias a través del tiempo.
- ii) Los resultados en el período inicial no son contaminados por las expectativas que se generan por la futura implementación del programa.

Bajo estos supuestos, el estimador de doble diferencia se lo puede conceptualizar como el impacto de la diferencia simple del segundo período menos la diferencia simple de la línea de base³.

La diferencia en diferencia requiere de una comparación entre el grupo de tratamiento y control antes del programa y una segunda diferencia entre el grupo de tratamiento y control después del programa, de tal manera que:

$$\begin{array}{cc} Y_{T1} & Y_{C1} \\ \bar{\quad} & \bar{\quad} \\ Y_{T0} & Y_{C0} \\ \hline Y_T & - & Y_C & \text{Impacto irreatable} \end{array}$$

- $YT1 - YT0 = YT$
- $YC1 - YC0 = YC$
- $YT - YC = \text{Efecto}$

El diseño de diferencias en diferencias es muy útil para solucionar posibles sesgos en un experimento social donde existe algún tipo de cumplimiento selectivo u otra distorsión al momento de realizar la asignación aleatoria. Esta metodología es la más robusta, exacta, confiable y segura ya que no presenta sesgo.

3. Existe un ejemplo interesante realizado en Indonesia por Dulfo (2001) que estima el impacto en la escolaridad y en el ingreso al construir escuelas.

Por otra parte, se suele considerar un problema ético el beneficiar a un grupo con una intervención y a otro no, ya que, muchas veces, ésta afecta de forma determinante la calidad de vida, inclusive la misma vida de las personas, como es el caso de los programas que ayudan a la disminución o erradicación de la desnutrición infantil. Sin embargo, en algunos casos se puede acumular los beneficios, y recibirlos juntos en momentos posteriores a la realización de la evaluación de impacto; esto es factible con programas que implican transferencias de dinero y no con programas como el mencionado anteriormente. Además, la metodología descrita anteriormente puede también traer problemas políticos. Este tipo de análisis en muchos casos resulta muy costoso y, además, pueden darse casos de contaminación por la posibilidad de movilización.

4.1.2 Método de diferencias en órdenes más altos: seguimiento de ex – participantes⁴

En muchos casos no es posible conseguir información antes de la intervención o información de línea de base. En varias ocasiones, por requerimientos urgentes, no es posible retrasar la operación hasta lograr obtener una toma en la línea de base. A pesar de estos problemas, aún sin levantar la línea de base se puede identificar el impacto, observando los resultados de los participantes en la ausencia del programa después de la intervención, en lugar de hacerlo antes. Una vez que se realiza la diferencia en diferencias, se debe considerar que, en este caso, el programa ya está operativo en el período 1. El alcance de la metodología de identificación surge del hecho de que algunos participantes en el período 1 se salen del programa, por lo tanto, existe un estimador que es la diferencia entre las dobles diferencias entre los que se quedan y los que se retiran.

$$DDD=[E(G2|D2=1,D1=1)-E(G2|D2=0,d1=1)]-[E(G1|D2=1,D1=1)-E(G1|D2=0,D1=1)]$$

La primera parte es la ganancia neta de continuar participando en el programa, que está dada por la diferencia entre la ganancia de participar en el periodo 2 y la ganancia de quienes dejaron el programa. Si únicamente interesan las ganancias marginales de la participación, se debe tomar en cuenta la primera parte de la ecuación; la selección para entrar al programa en el inicio no es un problema. Se debe considerar que existen algunas ganancias para los que dejan el programa, ya que $E(G2|D2=0, D1=1) \neq 0$. Se puede pensar, por ejemplo, que quizás los participantes aprendieron algo durante ese tiempo. La segunda parte del término se refiere al sesgo por selección que surge del efecto de las ganancias en el período 1 y de la participación en el período 2.

La triple diferencia identifica las ganancias de los participantes en el periodo 2, o sea $E(G2|D2=1,D1=1)$, si cumple con las siguientes condiciones:

- i) No hay sesgo de selección en cuanto a quién deja el programa.
- ii) No hay ganancias en los no participantes.

Un tercer cuestionario permite realizar una prueba de que estas dos condiciones se mantengan.

4.1.3 Evaluación de programas “antipobreza”

Existen dos tipos de problemas que confunden los esfuerzos para identificar un impacto. El primero es el sesgo por selección, que puede surgir de variables observables o de variables no observables. El segundo es la existencia de efectos que se desbordan hacia otras variables (efectos “spillover”), que confunden los efectos que pueden atribuirse a los impactos del programa a sus participantes directos.

4. El texto se tomó principalmente del documento de Martin Ravallion: “Evaluating Anti-Poverty Programs”. Development Research Group, World Bank. WPS3625, Junio 2005.

4.1.4 Selección de observables

Las ecuaciones (3) y (4) tratan con la selección de observables de una manera especial, en el sentido de que los controles entran en una forma lineal con respecto a los parámetros. Este supuesto ad hoc es difícilmente justificable por otro argumento, excepto conveniencia computacional. En evaluaciones cuasi-experimentales de programas antipobreza, es difícil asegurar que las variables observables estén balanceadas entre los dos sets de observaciones.

Para entender este problema, se puede suponer que la colocación está determinada por “*proxy-mean test*”, lo cual es comúnmente usado en países subdesarrollados para definir a quién dar los beneficios del programa antipobreza. Esto asigna una calificación a todos los potenciales participantes en función de variables o características observables como, por ejemplo, el caso de Ecuador, que utiliza el índice Selben para definir a quien beneficiar. En ciertos casos, se determina a los beneficiarios del programa solamente si se encuentran por debajo o por encima de estos índices o niveles críticos, determinados según el presupuesto disponible. Con un 100% de aceptación, no existen intervalos de la calificación para que se puedan observar tanto participantes como no participantes en una muestra de cualquier tamaño. Este ejemplo se conoce como “fracaso en apoyo común” en la literatura de evaluación.

Como se mencionó anteriormente, el problema es esencialmente sencillo: ¿cómo se infiere el contrafactual para los participantes con relación a no participantes, quienes no tienen las mismas características en común, como está sintetizado en el puntaje del *test* de medias *proxy*? Si se quiere inferir un impacto promedio para aquellos que reciben el programa, entonces se debe dar mucha importancia a la validez del diseño de cualquier grupo de comparación. Afortunadamente, en la práctica, existe siempre un grado de áreas grises en la aplicación de cualquier *test proxy* y existe cobertura incompleta de quienes incluso pasan esta prueba. Además, no se necesita conocer el impacto de todo el grupo de tratamiento. Por ejemplo, puede existir una política que desee incrementar el presupuesto variando el puntaje del índice que determina quién recibe o no el programa, en este caso se pueden enfocar los impactos alrededor de este puntaje, esto se llama hacer un diseño de discontinuidad.

Se tendría que trincar la muestra de no participantes para asegurar la validez del grupo de comparación, mas allá de la ineficiencia al recolectar la información innecesaria, esto en general no es una preocupación. Más preocupante es que una submuestra no aleatoria de participantes tendría que ser descartada por falta de suficiencia similar de los comparativos. Esto señala un intercambio entre dos fuentes de sesgo. Por un lado, se tiene la necesidad de asegurar comparabilidad en términos de características iniciales. Y, por otro lado, se crea un posible sesgo muestral por inferencias sobre el impacto, hasta el punto en que se tienen que descartar unidades de tratamiento para obtener comparabilidad.

4.1.5 Resolver el sesgo por selección

Una preocupación común en evaluaciones cuasi-experimentales es si el proceso de selección para la evaluación de un programa se captura adecuadamente por las variables de control X . Esta preocupación no se puede separar del problema de la colocación no aleatoria condicionada a observables. No se puede juzgar si la endogeneidad de la colocación es un supuesto plausible sin primero establecer si se realizó adecuadamente la heterogeneidad de las observables.

Por ejemplo, pensemos en el caso en el que se decida identificar los determinantes de participación en el mercado laboral con relación al sueldo de mujeres que trabajan en el mercado laboral. Lo que se quiere es interpretar o llegar a conclusiones sobre los determinantes de los salarios de todas las mujeres. La diferencia entre trabajadores y no trabajadores determina si existe sesgo por selección. Si se supone que se escoge una submuestra aleatoria de la población, y la submuestra tiene atributos similares a los del grupo de los que no trabajan, no existe ninguna razón para suponer que hay un sesgo por selección al examinar únicamente las mujeres que están trabajando. Eso es porque al realizar una selección aleatoria, las características (tanto observables como no observables) de la muestra de trabajadoras capturan un poco de todas las características de la población.

Suponiendo que no hubo una selección aleatoria y, en consecuencia, hay un potencial riesgo de que sean distintas las características de las trabajadoras y no trabajadoras, el sesgo por selección de la muestra ocurre cuando algún componente de la decisión de trabajar es relevante al proceso de determinación de salarios. En otras palabras, cuando alguno de los determinantes de la decisión de trabajar esta influyendo también en el salario. Cuando la relación entre la decisión de trabajar y el sueldo es puramente a través de variables observables, se puede controlar esto al incluir las variables apropiadas en la ecuación del salario. Por lo tanto, no surgirá sesgo por selección en el caso de que existan solo diferencias observables.

Si en lugar de pensar en variables observables, se consideran variables no observables que afecten la decisión de trabajar, estas variables están correlacionadas con las características no observables que afectan el salario y, por lo tanto, crearán una relación entre la decisión de trabajar y el proceso que determina los salarios. Entonces, controlar solo por variables observables no es suficiente. Si las características no observables están correlacionadas con las observables, al no incluir un estimado de estas variables no observables se genera un sesgo que surge por la selección⁵.

Como se dijo anteriormente, al igual que muchos programas públicos, la participación en intervenciones directas que combaten la pobreza casi nunca es aleatoria. Este es un problema si existen algunas variables que conjuntamente influyen en los resultados, y la colocación de los programas no es observable para el evaluador. En este caso, no se puede atribuir al programa las diferencias observables de los resultados medidos entre las unidades que reciben el programa y las que no (condicionadas por las variables de control X). Las diferencias en medias condicionadas que se observan en los datos podrían ser el resultado del hecho de que los participantes del programa fueron seleccionados a propósito por un proceso que no es posible observar del todo. Cuando la aceptación al programa está en las manos del individuo, tiene que existir una presunción razonable de que la selección del programa depende de las ganancias que este le brinde.

En términos de la formulación clásica del problema de la evaluación mencionada, se puede suponer que los participantes tienen atributos latentes que rinden resultados más altos que los no participantes (para una de las variables X). Entonces, el término de error en la ecuación (3.1) estaría centrado en el lado derecho, relacionado con los no participantes (3.2). El término de error en la ecuación (4) no puede desaparecer en expectativa y el **MCO** nos dará estimadores sesgados e inconsistentes. Se debe recordar que el sesgo en la media del estimador de impacto condicional es la diferencia en las medias contrafactuales, $E(Y_i T | X_i, D_i = 1) = E(Y_i C | X_i, D_i = 0)$. Nuevamente se debe enfatizar que el grado de preocupación de esta forma de sesgo de selección en la práctica no puede separarse de la pregunta anterior sobre cuán bien se controla la heterogeneidad observable.

Existen ejemplos que se enfocan a sesgos en estimadores de impacto no experimentales en ciertos

5. Para un mayor detalle de la sección de sesgo por selección propuesta por James Heckman (1979) referirse al paper de Francis Vella: "Estimating Models with simple Selection Bias: A Survey". The Journal of Human Resources, Vol.33, No.1 (1998 de donde se tomó parte del texto además de Ravallion, Martin en su paper, "Evaluating Anti-Poverty Programs"

casos específicos. En algunos estudios se han encontrado grandes sesgos en métodos cuasi - experimentales cuando se comparan a métodos de evaluación aleatorizados. No obstante, no se pueden rechazar los métodos cuasi - experimentales en algunas aplicaciones de ciertos estudios; lo importante es mejorar los datos y la metodología que se necesitan para conocer si los programas sirven y cumplen sus objetivos. Sin embargo, algunos de los métodos cuasi-experimentales más comunes –como comparaciones de corte transversal de participantes y no participantes o comparaciones reflexivas de los resultados de los participantes a lo largo del tiempo– pueden dar resultados severamente sesgados cuando la información disponible da controles inadecuados de heterogeneidad.

Existen varias formas para corregir este problema, sin embargo, las metodologías más comunes son dos. La primera consiste en el método de máxima verosimilitud por el supuesto de la distribución de los errores. La segunda se caracteriza por un procedimiento en dos etapas que elimina el supuesto de que los errores no son cero. El uso de la metodología de máxima verosimilitud es poco común, por lo que únicamente se define la estimación en dos etapas en particular de un modelo parametrizado.

El modelo de sesgo por selección tiene la siguiente forma:

- i) $Y_i^* = X_i' \beta + \epsilon_i \quad i=1, \dots, N$
- ii) $d_i^* = Z_i' \gamma + v_i \quad i=1, \dots, N$
- iii) $d_i = 1$ si $d_i^* > 0$ y $d_i = 0$ si no lo es
- iv) $Y_i = Y_i^* \cdot d_i$

donde Y_i^* es la variable latente endógena con su parte observable Y_i ; d_i^* es la variable latente con la función d_i . Las ecuaciones iii) y iv) reflejan si la variable dependiente principal es observable y si las relaciones entre d_i y d_i^* y Y_i y Y_i^* respectivamente, X_i y Z_i , son vectores exógenos de variables de control.

La ecuación de interés sobre la submuestra n que corresponden a $d_i = 1$

$$v) \quad Y_i = X_i' \beta + \epsilon_i \quad i=1, \dots, N$$

Se debe recordar que la estimación de **MCO** da estimaciones sesgadas de β porque $E[\epsilon_i | Z_i, d_i = 1] \neq 0$. La estrategia entonces propuesta por Heckman (1976, 1979) es tratar de corregir esto mediante la inclusión de un término de corrección que toma en cuenta $E[\epsilon_i | Z_i, d_i = 1]$. Para emplear este método, se debe tomar la expectativa condicional de la ecuación

$$v) \quad E[Y_i | Z_i, d_i = 1] = X_i' \beta + E[\epsilon_i | Z_i, d_i = 1]; \quad i=1, \dots, N$$

Usando el supuesto i) y la fórmula de la expectativa condicional de una variable aleatoria truncada, se ve que:

$$E[\epsilon_i | Z_i, d_i = 1] = \frac{\sigma \epsilon v}{\sigma^2} \left\{ \frac{\varphi(Z_i' \gamma)}{\Phi(Z_i' \gamma)} \right\}$$

en donde $\Phi()$ y $\varphi()$ son las funciones de distribución acumulada y la densidad de probabilidad de la distribución normal estandarizada. El término entre corchetes se conoce como la razón inversa de Mills (el inverso de Mills ratio). Se utiliza el modelo *probit* para estimar los parámetros desconocidos γ y σ . En la primera etapa se estima γ sobre el total de las N observaciones por el método de máximo verosimilitud (con *probit*) y luego se construye el estimado de la razón inversa de Mills. Se estiman consistentemente los parámetros utilizando el **MCO** de las n observaciones

que reportan valores para Y_i e incluyen la razón inversa de Mills, en este caso λ_i , como un regresor adicional en la ecuación v):

$$vi) Y_i = X_i' \beta + \mu \lambda_i + \eta_i \quad i=1 \dots N$$

Con **MCO** se obtienen estimadores consistentes de β y de μ , esto se conoce como el estimador de la fusión de control.

Es importante incluir variables adicionales en el vector Z en la primera etapa para identificar los estimadores de la segunda etapa.

4.1.6 Desbordamientos (Spillovers) en no participantes⁶

Eliminar el sesgo por selección no asegura que los impactos se pueden identificar. La formulación clásica del problema de evaluación asume que solamente el tratamiento en la unidad i puede afectar los resultados en esa unidad. Luego se asume que se puede observar al grupo de no participantes (control) y que estos no han sido afectados en ningún sentido por el programa en cuestión. Esto se puede hacer una vez que el programa esté operando (generando un diseño de diferencia simple) o se puede hacer para los participantes antes del programa (dando una comparación reflexiva). Cuando se hacen ambas cosas, se está utilizando un método de doble diferencia. Bajo esas condiciones, se infieren impactos al comparar a aquellos que han dejado el programa con quienes se han mantenido. Sin embargo, en todos estos casos se asume que se puede observar el estado de los no participantes de una manera no contaminada por el programa.

Este supuesto puede ser problemático al realizar evaluaciones de impacto de programas anti – pobreza. Por ejemplo, se evalúa un programa donde el Gobierno se compromete a dar trabajo a quien desee trabajar, por un sueldo estipulado, de manera universal, es decir, se entrega el beneficio a todo aquel que lo desee. Todos los participantes en el programa, obtienen como beneficio el sueldo, el mismo que se considera bajo en un contexto específico. La universalidad de acceso significa que bajo este esquema se puede brindar un seguro efectivo para combatir el riesgo. El que una persona requiera un trabajo cuya remuneración es baja, podría pensarse automáticamente como un programa que está dirigido a las personas más pobres.

Entonces, este tipo de programas dirigidos a ciertos grupos poblacionales tendrían bien definidos quiénes son los participantes y quiénes no. En un primer momento, puede parecer apropiado el recolectar la información de los dos grupos y comparar los indicadores de resultado de estos como una manera de identificar el impacto (posiblemente después de corregir cualquier heterogeneidad observada). Sin embargo, el diseño de evaluación clásica puede dar un resultado severamente sesgado. Las ganancias de dicho programa se esparcen en el mercado laboral privado e ignoran el efecto de desbordamiento (*spillover*) que esto tiene y por lo tanto subestiman los beneficios. Ciertamente, si la garantía de empleo es efectiva, entonces el esquema establecerá un umbral más bajo para toda la distribución de salarios, ya que ningún trabajador aceptará un empleo que no sea el del programa a un sueldo menor que el del programa ya que siempre puede unirse a este. Incluso si se escoge un grupo de comparación perfecto de no participantes, se puede concluir que el esquema no tiene impacto, ya que los salarios son los mismos para participantes y no participantes, y esto no capturaría las ventajas del programa en lo absoluto.

Los efectos de *spillovers* pueden surgir también por intervención del gobierno. Generalmente, no es muy claro si los recursos que se transfieren a los participantes realmente financian al proyecto

6. La mayor parte del texto fue tomado de Martin Ravallion "Evaluating Anti-Poverty Programs" en donde se encuentra una descripción en mayor detalle sobre spillovers en no participantes.

identificado. Hasta cierto grado, toda la ayuda externa puede ser intercambiable. Se puede verificar mediante supervisión que el subproyecto propuesto fue en efecto completado, pero no se puede descartar la posibilidad de que no lo fue. Participantes y líderes locales naturalmente tratarían de llevar a cabo la mejor opción de desarrollo que consideren, a pesar de que sea algo que pensaban hacer de todas maneras con los recursos disponibles. Entonces existe otro gasto que en realidad esta siendo financiado por esta ayuda. De igual manera, no hay que descartar la posibilidad de que localidades que no participen, sean beneficiadas por una reasignación de gasto público realizada por autoridades locales, de tal manera que disminuyan el impacto de participar en el programa.

Los efectos de los desbordamientos (*spillovers*) pueden subestimar los impactos reales del programa. Sin embargo, el sesgo también puede orientarse en la otra dirección; por ejemplo, si el Gobierno desea demostrar el impacto de un proyecto antipobreza y decide añadir mayores recursos externos enfocados en localidades donde está el proyecto. La diferencia de resultados de los participantes y del grupo de control estaría sobredimensionando el impacto de la ayuda externa.

4.1.7 Información necesaria⁷

Está claro que la información inadecuada e imperfecta yace en el centro del problema de la evaluación. Cuando se realiza una evaluación de impacto, es importante conocer la mayor cantidad de detalles tanto administrativos como institucionales sobre el programa; información que proviene de la misma administración del programa. Para evaluaciones no experimentales, esta información es clave para diseñar un cuestionario que recolecte la información correcta para controlar el proceso de selección. El conocimiento del contexto del programa y las características del diseño pueden ayudar para manejar la selección en cuanto a las características no observables. La elaboración de evaluaciones cuasi experimentales para programas antipobreza pueden demandar grandes requerimientos de información. Las fuentes de información utilizadas pueden incluir tanto entrevistas formales como informales o no estructuradas con los participantes del programa, al igual que información cuantitativa de muestras representativas. Sin embargo, es extremadamente difícil realizar preguntas contrafactuales en entrevistas o grupos focales; por ejemplo, tratar de preguntar a alguien que se encuentra participando en un programa social lo siguiente: “¿qué haría si no hubiese recibido ayuda del programa?” Hablar con participantes del programa puede ser muy valioso, pero es poco probable que provean información creíble por sí solo. Se requiere además información de indicadores de resultado y variables explicativas relevantes.

La información de resultados y sus determinantes, incluyendo la participación en el programa proviene de las encuestas. La unidad de observación puede ser individual, de hogar, de un área geográfica o de alguna institución (escuela, clínica, etc.) dependiendo del tipo de programa. La información de encuestas puede ser complementada con información del programa (tal como la base de datos proveniente del monitoreo del programa) o con información de lugar (bases de datos de análisis geográficos). Una preocupación que puede surgir es la *comparabilidad* o consistencia de diferentes fuentes de información, particularmente las utilizadas para la observación a los participantes y no participantes. Diferencias en el diseño de los instrumentos para los cuestionarios pueden implicar diferencias significativas en las medidas de resultados, de tal manera que sesguen el análisis del verdadero impacto⁸.

7. La mayor parte del texto fue tomado de Martin Ravallio “Evaluating Anti-Poverty Programs”.

8. Para ver otras experiencias, referirse a análisis realizados por P. Shultz (2004) sobre el programa PROGRESA de México, que daba transferencias monetarias a familias pobres. Las condiciones que debían cumplir eran: tener a hijos en escuelas, asistir a centros de salud y tomar suplementos nutricionales. Este estudio realizó una colocación aleatoria del programa para obtener un grupo de control o contrafáctico. Otro ejemplo es el estudio de Newman et al. (2002) de un programa del Banco Mundial para Bolivia.

4.1.8 Comparaciones de diferencias simples con asignación aleatoria

Un experimento social busca aleatorizar la colocación del programa, de tal manera que las unidades (en un conjunto definido) tengan la misma oportunidad ex – ante de recibir el programa. Todas las características (observables y no observables) antes de la intervención son idénticamente distribuidas entre las unidades que reciben el programa y las que no lo reciben. Por implicación, las diferencias en las medias de los resultados ex – post entre los dos grupos son atribuibles al programa. La aleatorización garantiza que no existe sesgo por selección al estimar (3.1) y (3.2) o equivalentemente la ecuación del término de error (4) es ortogonal a los regresores. Los no participantes son entonces un grupo válido de control para identificar el contrafactual, y ATE es estimado consistentemente (no paramétricamente) por las diferencias entre las medias muestrales de Y_iT y Y_iC (incluyendo las submuestras con los valores dados de X_i).

Sin embargo, a pesar de que los diseños aleatorios serían ideales para identificar el impacto, pueden surgir varios problemas en la práctica como mencionamos en la sección de métodos experimentales: políticos, éticos, etc.

4.2 Métodos cuasi - experimentales⁹

Esta metodología intenta replicar el método experimental. Existen varios tipos que se detallan a continuación.

4.2.1 Método de Grupos Emparejados (*Propensity score matching*)

Es una metodología que corrige por sesgo de selección al realizar las estimaciones. Aplica una probabilidad predicha a un grupo de miembros (tratamiento vs. control) basándose en predictores observables, en general obtenidos por una regresión logística para crear un grupo contrafáctico. Trata de emparejar entre los que tienen probabilidades casi iguales, o sea, entre los más parecidos. Este método selecciona el grupo de comparación con relación a las probabilidades predichas de participación (llamadas “*propensity scores*”). Generan un efecto de tratamiento promedio o **ATE**; sin embargo, tiene un problema, y es que únicamente emparejan tomando en cuenta las características observables, pero pueden existir características no observables que sesguen los resultados. Para comprender la metodología se debe entender los criterios de asignación del programa. Como se mencionó anteriormente, a los participantes y a los no participantes se les empareja con base en sus características similares (*propensity score*), $P(Z_i) = E(D_i|Z_i)$ ($0 < P(Z_i) < 1$), donde Z_i es el vector de las variables de control antes de exponerse al programa, esto incluye valores antes del programa del indicador de resultado. Si las D_i 's son independientes a lo largo de todas las i , y si los resultados son independientes de la participación dado Z_i , entonces los resultados son también independientes de la participación dado $P(Z_i)$. De esta manera el emparejamiento elimina el sesgo por selección.

El método de grupos emparejados (*propensity score matching*) difiere de otros métodos, en especial del método experimental social en su forma más pura, en que los grupos emparejados (*propensity score*) podría enfrentar problemas de sesgo de selección en el momento que se postula la existencia de una variable latente que conjuntamente influye sobre la colocación y los resultados, de tal manera que invalida los supuestos realizados por el *propensity score*.

No se puede asumir que al eliminar el sesgo por selección basado en observables, también se reduce el sesgo agregado, eso solo se dará en el caso de que las dos fuentes de sesgo – que se asocian con las observables y que surgen por factores no observables – van en la misma dirección, lo que no puede ser refutado a priori. Una comparación natural entre el **MCO** y el *propensity score*,

9. La mayor parte del texto fue tomado de Martin Ravallion “Evaluating Anti-Poverty Programs”.

se da cuando en este método se introducen variables dicotómicas que definen la participación en el programa, lo que permite que las covariables observables entren como controles lineales. El **MCO** requiere esencialmente el mismo supuesto de independencia condicional (exogeneidad) que el método de grupos emparejados (*propensity score matching*), pero además impone supuestos de la forma funcional arbitraria de los efectos del tratamiento y variables de control. En cambio, el método de grupos emparejados no requiere un modelo paramétrico que junte los resultados de la participación del programa. De tal manera que el método de grupos emparejados permita la estimación de impactos sin los supuestos arbitrarios de la forma funcional y la distribución del término de error. Además difieren con respecto a la muestra.

El Método de Grupos Emparejados (**MGE**) da mayor atención a las submuestras emparejadas, las comparaciones que no están emparejadas se eliminan, por lo tanto algunos participantes pueden ser excluidos porque tienen un puntaje fuera del rango de los que se encuentran en la muestra de participantes. Otra diferencia tiene que ver con la elección de variables de control **X**. En el método de regresión estándar se mira a los predictores para los resultados, y se da preferencia a las variables que se pueden considerar exógenas al resultado. En el **MGE**, en cambio, se mira a las covariables de participación, posiblemente se incluye variables no significativas o variables que no arrojen predicciones efectivas de los resultados. Ciertamente, resultados analíticos y simulaciones indican que las variables con poca habilidad de predicción de los resultados pueden ayudar a reducir el sesgo al estimar efectos causales cuando se utiliza el método de grupos emparejados (*propensity score matching*).

En resumen, el método de grupos emparejados (**MGE**) es una importante herramienta que permite analizar el impacto de un programa mediante la generación de un contrafactual adecuado (bajo el supuesto de que no existen variables no observables que influyeran en la participación). Lo más importante en éste método es entender el mecanismo de asignación –los determinantes observables de la colocación del programa. Utilizar este método es más convincente que suponer que la asignación realizada en un programa específico responde a los supuestos de un experimento natural y que no existe sesgo por selección

4.2.2 Diseños discontinuos

La idea, en este caso es inferir los impactos que surgen de las diferencias entre un grupo de tratamiento y control, pero de las unidades que están cerca del punto de corte (que se utilizó para definir quién es elegible y quien no). Para comprender este método de mejor manera, es posible pensar que **Mi** es el puntaje que la unidad *i* recibe como una prueba de medias *proxy* y **m** es el punto de corte para la elegibilidad, de tal manera que **Di=1** para **Mi ≤ m** y **Di=0** en el caso contrario. El impacto entonces se define como:

$$E(Y_i|Mi = m-\epsilon) - E(Y_iC|Mi = m+\epsilon)$$

para un valor arbitrario muy pequeño de $\epsilon > 0$.

En la práctica, existe siempre un grado de área gris inevitable en la aplicación de pruebas de elegibilidad. Entonces, en lugar de asumir que existe un estricto control y cumplimiento, se puede pensar en la probabilidad de la participación del programa $P(Mi) = E(Di|Mi)$, lo que es una función creciente de **Mi** con una discontinuidad en **m**. La idea esencial se refiere a que los impactos están medidos por la diferencia en los resultados de la media en el vecindario de **m**. El supuesto clave de identificación es que no hay discontinuidad en los resultados contrafactuales en **m**. Se debe considerar que existe una variación entre la estrictez de cada programa, esto no descarta la posibilidad de que la continuidad sea un supuesto posible. Un cuestionario en la línea de base

puede ser de gran ayuda para limpiar las diferencias en los resultados antes de la intervención en cualquiera de los lados de la discontinuidad, de tal manera que se estaría combinando el diseño de discontinuidad con el método de dobles diferencias.

Comparado con el diseño del Método de Grupos Emparejados, el de discontinuidad da el impacto para una muestra escogida de participantes, mientras que el método de grupos emparejados trata de precisar el impacto del grupo total de tratamiento. Sin embargo, el problema que muchas veces se genera por el criterio de elegibilidad puede significar que este método también esté acotado a una submuestra seleccionada; la pregunta entonces sería si ésta es una submuestra interesante. El truncar la muestra del grupo de tratamiento bajo este método (*PSM*), generalmente tiende a excluir a aquellos que tengan la probabilidad más alta de participar (para lo que es más difícil encontrar una pareja de no participantes), mientras que el diseño de discontinuidad solo incluyó a aquellos que tienen la probabilidad más baja. A pesar que los diseños de discontinuidad son identificados como no paramétricos, en muchos libros de economía se usan otros métodos paramétricos en los que la discontinuidad en el criterio de elegibilidad se usa como una variable instrumental para la colocación del programa.

4.2.3 Comparaciones de lista de espera

En este caso, se utiliza como grupo de comparación a aquellos que han aplicado (es decir, son elegibles) pero aún no han recibido el programa, por lo tanto, la comparación de lista de espera es el grupo de control que se forma por un retraso en el programa. En muchos casos, este grupo que se queda a la espera es elegido por una aleatorización, sin embargo, existen estudios que lo utilizan en un diseño cuasi-experimental. Este método en cierta manera trata con el problema de heterogeneidad latente en los estimadores de diferencias simples, al igual que el método de grupos emparejados; los participantes exitosos, en general, suelen tener las mismas características no observables, ya sea que hayan recibido el programa o todavía no.

El supuesto principal es que el tiempo en que se da el tratamiento es aleatorio. En la práctica se debe anticipar el sesgo potencial que puede surgir del tratamiento selectivo entre los participantes o comportamientos que resulten en los participantes que están a la espera de recibir el tratamiento. Cuando es factible, las comparaciones de lista de espera dan un estimador de diferencias simples robustas, considerando la heterogeneidad latente. Sin embargo, se deben realizar pruebas que permitan verificar que no existe sesgo por selección en las observables. Muchas veces se utiliza el método de grupos emparejados (*PSM*) para limpiar la heterogeneidad observable en las comparaciones de lista de espera.

4.2.4 Variables Instrumentales¹⁰

Si se trata de explicar el logaritmo de un salario de un individuo Y_i con una cantidad de variables de características personales $X1_i$, y se incluye el número de horas que una persona i trabaja $X2_i$, con un modelo lineal:

$$Y_i = X1_i \beta_1 + X2_i \beta_2 + \epsilon_i \quad (7)$$

este modelo no tiene interpretación a no ser que se incluya supuestos al término de error ϵ . De otra manera se podría simplemente dejar que β_1 y β_2 tomen valores arbitrarios y suponer que el término de error de la ecuación (7) se mantiene para todas las observaciones. El método más común para interpretar esto describe las expectativas condicionadas o la mejor aproximación lineal de Y_i dado $X1_i$ y $X2_i$. Esto requiere que se impongan los siguientes supuestos:

10. La mayor parte del texto fue tomado de Verbeek, Marno. A guide to Modern Econometrics. John Wiley & Sons, Ltd. 2000 en donde además se detalla en mayor profundidad el concepto de endogeneidad y el método generalizado de momentos.

$$E\{\epsilon_i X_{1i}\} = 0 \quad (8.1)$$

$$E\{\epsilon_i X_{2i}\} = 0 \quad (8.2)$$

Estas son las condiciones necesarias para la consistencia de los estimadores **MCO**. Al momento en que se relaja cualquiera de estas condiciones, el modelo ya no corresponde a la expectativa condicional de Y_i dado X_{1i} y X_{2i} .

En la ecuación del logaritmo del ingreso, ϵ_i incluye todos los factores no observables que afectan al salario de un individuo, tales como: "habilidad innata" o "inteligencia". Se dice que el número de horas que trabaja una persona dependen parcialmente de estas características no observables. Si éste es el caso, los estimadores **MCO** están consistentemente estimando el valor esperado condicional del salario de una persona, entre otras cosas, el número de horas que trabajó el individuo, pero no están estimando consistentemente el efecto causal de las horas de trabajo. Esto quiere decir que, el estimador de **MCO** β_2 va a reflejar la diferencia en el salario esperado de dos personas arbitrarias con las mismas características de X_{1i} observables, pero que trabajan X_2 y X_2+1 horas, respectivamente. Esto de ninguna manera mide la diferencia entre el sueldo de una persona arbitraria (por alguna razón exógena) que decide incrementar sus horas de trabajo de X_2 a X_2+1 . La razón es que en la primera interpretación, los factores no observables que afectan el salario de un individuo no son constantes en las dos personas, y en la segunda interpretación, los no observables se mantienen constantes. Puesto de otra manera, cuando se interpreta un modelo como una esperanza condicionada, la condición "todo lo demás constante (Ceteris Paribus)" únicamente se refiere a las variables incluidas, mientras que en la interpretación causal también incluye las variables omitidas, es decir, las no observables en el término de error.

Los coeficientes del modelo de regresión se interpretan como efectos causales. En estos casos tiene sentido discutir sobre la validez de las condiciones (8.1) y (8.2). Si $E\{\epsilon_i X_{2i}\} \neq 0$. Se dice que X_{2i} es endógena (con respecto al efecto causal de β_2). Para ecuaciones del salario de microeconomía, usualmente se argumenta que muchas variables explicativas son potencialmente endógenas, incluyendo variables como el nivel de educación, estado civil, salud, industria. Por ejemplo, se observa que en promedio los salarios son un 10% más que para personas que están casadas. Esto no quiere decir y no refleja un efecto causal de estar casado, sino una consecuencia de la diferencia de características no observables de gente que está casada y gente que no.

Si no se impone la condición que $E\{\epsilon_i X_{2i}\} = 0$, el método **MCO** produce estimadores sesgados e inconsistentes para los parámetros del modelo. La solución requiere un método de estimación alternativo. Para derivar un estimador consistente, es necesario asegurarse de que el modelo está estadísticamente identificado. Esto significa que se necesita imponer supuestos adicionales; de otra manera, el modelo no está identificado y ningún estimador es necesariamente inconsistente. Para poder entender esto, hay que fijarse en las ecuaciones (8.1) y (8.2). Estas condiciones se llaman condiciones de momento o condiciones en términos de expectativas (momentos) que se acogen en el modelo; y deben ser suficientes para identificar los parámetros desconocidos en el modelo. Esto quiere decir que los parámetros K de los β_1 y β_2 tiene que ser tales que las siguientes igualdades se mantengan:

$$E(Y_i - X_{1i} \beta_1 - X_{2i} \beta_2) * X_{1i} = 0 \quad (9.1)$$

$$E(Y_i - X_{1i} \beta_1 - X_{2i} \beta_2) * X_{2i} = 0 \quad (9.2)$$

Cuando se estima el modelo por MCO, se imponen estas condiciones en el estimador, por medio de los correspondientes momentos muestrales. Esto es, el estimador $b = (b_1, b_2)$ por $\beta = (\beta_1, \beta_2)$

surgen de las siguientes ecuaciones:

$$1/N * \Sigma (Y_i - X_{1i} b_1 - X_{2i} b_2) * X_{1i} = 0 \quad (10.1)$$

$$1/N * \Sigma (Y_i - X_{1i} b_1 - X_{2i} b_2) * X_{2i} = 0 \quad (10.2)$$

De hecho, estas son las condiciones de primer orden para la minimización del criterio de mínimos cuadrados. El número de condiciones es exactamente igual al número de parámetros desconocidos, tal que b_1 y b_2 se puedan resolver de las ecuaciones (10.1) y (10.2). Sin embargo, apenas se viola la ecuación (8.2), la condición (10.2) desaparece y a partir de ese momento ya no se puede resolver el b_1 y b_2 . Esto significa que el β_1 y β_2 ya no se pueden identificar.

Entonces para identificar β_1 y β_2 en un caso más general, se necesita al menos una condición de momento adicional. Tal condición adicional se deriva usualmente de la disponibilidad de un instrumento, o una variable instrumental. Una variable instrumental z_{2i} , es una variable que se asume no tiene correlación con el término de error en el modelo, pero tiene una alta correlación con la variable endógena X_{2i} . Si se logra encontrar dicho instrumento, la condición (9.2) se puede reemplazar por:

$$E((Y_i - X_{1i} \beta_1 - X_{2i} \beta_2) * z_{2i}) = 0 \quad (11)$$

Dado que esta condición de momento no es una combinación de otras variables (z_{2i} no puede ser una combinación lineal de X_{1i} 's), la introducción del instrumento es suficiente para identificar los K parámetros β_1 y β_2 . El estimador de la variable instrumental B_{IV} , se puede resolver de la siguiente manera:

$$1/N * \Sigma (Y_i - X_{1i} \beta_{1IV} - X_{2i} \beta_{2IV}) * X_{1i} = 0 \quad (12.1)$$

$$1/N * \Sigma (Y_i - X_{1i} \beta_{1IV} - X_{2i} \beta_{2IV}) * z_{2i} = 0 \quad (12.2)$$

La solución se puede determinar analíticamente y al hacerlo se obtiene la siguiente expresión para el estimador IV :

$$B_{IV} = (\Sigma z_i X_i)^{-1} \Sigma z_i Y_i \quad (13)$$

donde $X_i = (X_{1i}, X_{2i})$ y $z_i = (X_{1i}, z_{2i})$. Claramente, si $z_{2i} = X_{2i}$, esta expresión se reduce al estimado MCO.

Bajo los supuestos (9.1) y (11) y algunas condiciones regulares, el estimador de las variables instrumentales es consistente y asintóticamente normal. Lo más importante de estas condiciones regulares es que la matriz $K'XK$

$$Plim 1/N \Sigma z_i X_i' \Sigma z_i X_i$$

es finita e invertible. Una condición necesaria para esto es que el instrumento z_{2i} esté correlacionado con X_{2i} y no sea una combinación lineal de los elementos en X_{1i} . La covarianza asintótica de B_{IV} depende de los supuestos que se hacen de la distribución de ϵ_i . En el caso estándar donde ϵ_i es $IID(0, \sigma^2)$, independientemente de z_i , se puede mostrar que:

$$\sqrt{N} (B_{IV} - \beta) \rightarrow N(0, \sigma^2 (\Sigma z_i X_i' \Sigma z_i X_i)^{-1}) \quad (14)$$

donde la matriz simétrica $K'XK$:

$$\Sigma z_i X_i' \Sigma z_i X_i$$

se asume que es invertible, y $\Sigma_{zx} = \Sigma_{xz}$. La no singularidad de Σ_{zz} requiere que no exista multicolinealidad entre los elementos de K en el vector z_i . En muestras finitas se puede estimar la matriz de covarianzas de BIV por:

$$V\{BIV\} = \sigma^2 ((\Sigma_{xizi})(\Sigma_{zizi})^{-1}(\Sigma_{zixi}))^{-1} \quad (15)$$

Donde σ^2 es un estimador consistente de σ^2 basado en la sumatoria de errores al cuadrado, por ejemplo,

$$\sigma^2 = 1/N \Sigma(Y_i - X_i BIV)^2 \quad (16)$$

Como en el caso de mínimos cuadrados, es posible ajustar por los grados de libertad y dividirlo por $N - K$ en lugar de N .

El problema que puede surgir es que la mayoría de veces no es obvio determinar cuál de las variables puede utilizarse apropiadamente como instrumento. En el ejemplo mencionado se necesita una variable que esté correlacionada con las horas de trabajo X_2i , pero no correlacionada con los factores no observables de "habilidad innata" incluidos en el término de error. Se puede argumentar que las variables relacionadas a la composición de la familia pueden servir como variables instrumentales.

El único caso en el que las condiciones de momento pueden ser parcialmente comprobables es cuando existen más condiciones de las requeridas para la identificación. En este caso, se pueden comprobar las restricciones sobre-identificativas, sin necesariamente ser capaces de especificar cuál de las condiciones de momento corresponde a estas restricciones.

Se puede generalizar el modelo que se vio en el caso de permitir un número arbitrario de instrumentos. Considérese el siguiente modelo:

$$Y_i = X_i \beta + \epsilon_i \quad (17)$$

donde X_i es una dimensión de K . El estimador **MCO** se basa en las condiciones de momento K .

$$E(\epsilon_i X_i) = E(Y_i - X_i \beta) X_i = 0$$

En general, se asume que existen R instrumentos disponibles para el vector z_i , los cuales pueden superponerse con X_i . Las condiciones de momento relevantes están dadas por las siguientes R restricciones:

$$E(\epsilon_i z_i) = E(Y_i - X_i \beta) z_i = 0 \quad (18)$$

Si $R=K$, se estaría en la situación anterior y el estimador de las variables instrumentales se puede resolver con las condiciones de momento de la muestra,

$$1/N * \Sigma(Y_i - X_i BIV) z_i = 0$$

y se obtiene:

$$BIV = ((\Sigma z_i X_i)^{-1} \Sigma z_i Y_i)$$

Si el modelo se escribe en forma matricial:

$$Y = X\beta + \varepsilon$$

y la matriz Z es una matriz $N \times R$ de los valores de los instrumentos, este estimador de las variables instrumentales puede ser escrito también de la siguiente manera:

$$\beta_{IV} = (Z'X)^{-1}Z'Y \quad (19)$$

Si $R > K$, existen más instrumentos que regresores. En este caso no es posible resolver el estimado β reemplazando en la ecuación (18) con su contraparte de la muestra. No es posible resolver el sistema ya que existen más ecuaciones que incógnitas. En lugar de quitar instrumentos (y perder eficiencia) en general se escoge β de tal manera que los R momentos de la muestra cumpla con lo siguiente:

$$\frac{1}{N} \sum (Y_i - X_i \beta) z_i$$

Que estén lo más cerca posible de cero, se logra minimizando la siguiente ecuación cuadrática:

$$QN(\beta) = [1/N \sum (Y_i - X_i \beta) z_i]' W N [\sum (Y_i - X_i \beta) z_i] \quad (20)$$

Donde WN es una matriz $R \times R$ positiva definida y simétrica. Esta es una matriz que otorga diferentes ponderaciones a cada momento de la muestra. En general puede depender de un tamaño muestral N , porque por sí solo puede ser un estimado. Para las propiedades asintóticas del estimador de resultado de β , la probabilidad del límite de WN , denotado por $W = \text{plim } WN$, es importante. Esta matriz W debe ser simétrica y positiva definida. Usando notación matricial por conveniencia se puede reescribir la ecuación (20) como:

$$QN(\beta) = [1/N \sum Z'(Y - XB)]' W N [1/N \sum Z'(Y - XB)] \quad (21)$$

diferenciando con respecto a β , dado por las condiciones de primer orden

$$-2X'ZWNZ'Y + 2X'ZWNZ'X\beta_{IV} = 0$$

lo que implica:

$$X'ZWNZ'Y = X'ZWNZ'X\beta_{IV} \quad (22)$$

este es un sistema con K ecuaciones y K incógnitas de β_{IV} , donde $X'Z$ tiene dimensión $K \times R$ y Z' es $R \times 1$. Dado que la matriz $X'Z$ tiene rango K , la solución para la ecuación (21) es:

$$\beta_{IV} = (X'ZWNZ'X)^{-1} X'ZWNZ'Y \quad (23)$$

lo que en general, depende de los ponderados de la matriz WN .

Si $R = K$, la matriz $X'Z$ es cuadrada (y por los supuestos) invertible. Esto puede escribirse de la siguiente manera:

$$\begin{aligned} \beta_{IV} &= (Z'X)^{-1} W N^{-1} (X'Z)^{-1} X'ZWNZ'Y \\ &= (Z'X)^{-1} Z'Y \end{aligned}$$

En lo que corresponde a la ecuación (19), la matriz de ponderación es irrelevante. En esta situación el número de condiciones de momento es exactamente igual al número de parámetros que están siendo estimados. Se puede pensar en una situación en la que β se "identifica exactamente" porque se tiene justo la información necesaria (o sea las condiciones de momento necesarias) para estimar β . Una consecuencia inmediata de todo esto es que el mínimo de la ecuación (21) es cero, lo que implica que a todos los momentos de la muestra puede darse el valor de cero al escoger β apropiadamente. Eso es, $QN(\beta|V)$ es igual a cero. En este caso, $\beta|V$ no depende de WN y el mismo estimador se obtiene independientemente de la matriz de ponderación que se escoja.

Si $R < K$, el número de parámetros para estimar excede el número de condiciones de momento. En este caso β está "subidentificado" (no identificado) porque no existe suficiente información (eso es, condiciones de momento) para estimar de una única forma a β . Técnicamente, esto quiere decir que el inverso de la ecuación (23) no existe y que un número infinito de soluciones satisfacen las condiciones de primer orden de la ecuación (22). A no ser que se piense en una condición de momento adicional, este problema de identificación es fatal en el sentido de que no existe un estimador consistente para β . Cualquier estimador es necesariamente inconsistente.

Si $R > K$, entonces el número de condiciones de momento excede el número de parámetros para ser estimados, y entonces β está "sobre-identificado" porque hay más información de la que es necesaria para conseguir un estimador consistente para β . En este caso se tiene un rango de estimadores para β que corresponden a las posibles alternativas para la matriz de ponderación WN . Siempre que la matriz de ponderación es (asintóticamente) positiva definida, los estimadores que resultan para β son todos consistentes. La idea detrás del resultado consistente viene de que se está minimizando una función cuadrática en un conjunto de momentos de la muestra que convergen asintóticamente a los correspondientes momentos de la población, mientras que los momentos de la población son iguales a cero para los verdaderos valores de los parámetros. Este es el principio básico detrás del método de momentos.

Distintas matrices de ponderación WN dan diferentes estimadores consistentes con distintas matrices de covarianza asintóticas. Esto permite escoger la matriz de ponderación óptima que da un estimador de las variables instrumentales eficiente. Se puede demostrar que la matriz de ponderación óptima es proporcional al inverso de la matriz de covarianza de los momentos de la muestra. Intuitivamente, esto significa que los momentos de la muestra con una varianza pequeña, dan información precisa sobre los parámetros en β , reciben más peso en la estimación que los momentos de la muestra con varianza mayor. Esencialmente, esta es la misma idea que la del enfoque de los mínimos cuadrados ponderados, a pesar de que los pesos ahora reflejan distintos momentos de la muestra en lugar de diferentes observaciones.

Es claro que la matriz de covarianzas de los momentos de la muestra es igual a:

$$1/N \sum \epsilon_i z_i z_i'$$

y dependen de los supuestos que se hacen tanto en ϵ_i como en z_i ; si se asume que ϵ_i es $IID(0, \sigma^2)$ e independiente de z_i , la matriz de varianza asintótica de los momentos de la muestra está dado por:

$$\sigma^2 \Sigma_{zz} = \sigma^2 \text{plim } 1/N \sum z_i z_i'$$

consecuentemente, una matriz de ponderaciones óptima se obtiene de:

$$WN^{opt} = (1/N \sum z_i z_i')^{-1} = (1/NZ'Z)^{-1}$$

Y el estimador **IV** resultante es:

$$BIV = (X'Z(Z'Z)^{-1} Z'X)^{-1} X'Z(Z'Z)^{-1} Z'Y$$

Esta expresión es la que se encuentra en la mayoría de libros. El estimador es comúnmente llamado Estimador Generalizado de Variables Instrumentales (**GIVE**). Es también conocido como **2SLS** (mínimos cuadrados en dos etapas). Si ϵ_i es heteroscedástico o muestra autocorrelación, la matriz óptima de ponderación se puede ajustar respectivamente.

Comparado con otros métodos, al usar variables instrumentales no se requiere que se asuma exogeneidad. En lugar de eso, se asume exogeneidad en la variable Z_i , que Z_i tiene relación con la colocación y que Z_i no es un elemento del vector de controles X , lo cual es comúnmente conocido como "exclusión restricción". Si estos supuestos se mantienen, entonces se puede decir que Z_i identifica el impacto del programa.

4.2.5 Fuentes de Variables Instrumentales¹¹

4.2.5.1 Diseños parcialmente aleatorios

Es muy común en experimentos sociales, que muchas de las unidades que han sido seleccionadas aleatoriamente para el programa, no participen. En este caso se puede utilizar como variable instrumental a la asignación aleatoria. La "exclusión restricción" en este caso se cumple ya que se asignó aleatoriamente quién debería participar en el programa, entonces solo afecta los resultados por medio de su participación actual.

El estimador de variables instrumentales es ahora una dicotómica para la asignación del programa (uno si fue asignado y cero para control). Esta variable, si bien está correlacionada con la variable de tratamiento real, no tiene ninguna correlación con el error por lo tanto es un buen instrumento.

Esto muchas veces ocurre en casos reales por lo cual es muy importante corregirlo de esta manera.

4.2.5.2 Fuentes no experimentales de variables instrumentales

A pesar de que la existencia de asignaciones aleatorias permite utilizar variables instrumentales muy fácilmente cuando se trata con cumplimientos de endogeneidad, en muchos casos no es posible. Otras fuentes para variables instrumentales son: geografía, política y discontinuidades creadas por el diseño del programa.

El programa de colocación geográfica ha servido como identificación en muchos estudios que utilizan el estimador de variables instrumentales. Por ejemplo, en el estudio de Atanasio, Vera y Hernández (2004), que utiliza la distancia del hogar al centro de comunidad más cercana como variable instrumental¹²; estaba altamente correlacionada con la variable endógena y no tenía

11. La mayor parte del texto fue tomado de Ravallion Martin "Evaluating Anti-Poverty Programs". World Bank Policy Research Working paper 3625, Junio 2005.

12. Atanasio, Vera y Hernández (2004), en un estudio realizado en Colombia para medir el impacto de un programa nutricional en los niños menores de 5 años.

nada que ver con el hecho de recibir o no los alimentos. Se podría pensar que la distancia puede ser endógena por la decisión del lugar donde vivir del hogar o de la comunidad. Sin embargo, en el cuestionario, ninguno de los encuestados atribuye su decisión de mudarse con la ubicación del centro comunitario, por lo tanto consideran el instrumento como exógeno.

Las características políticas de ciertas áreas geográficas también han sido fuente de variables instrumentales. Entender la economía política de la colocación de un programa también puede ayudar a identificar impactos. Un ejemplo de esto es el estudio de Besley y Case (2000) que utiliza la presencia de mujeres en los parlamentos estatales en los Estados Unidos como una variable instrumental del seguro de compensación de los trabajadores al tratar de estimar los impactos de la compensación en los salarios y el desempleo. En su trabajo argumentan que las abogadas mujeres implementan leyes que favorecen la compensación de trabajadores, pero que esto es poco probable que tenga un efecto independiente en el mercado laboral.

Otra posibilidad para encontrar variables instrumentales es el diseño discontinuo del programa, ya que el ALTE es el vecindario donde está la línea de corte de la elegibilidad del programa. Por ejemplo, en el estudio realizado por Angrist y Lavy (1999) quienes miran el impacto de calificaciones escolares del tamaño de la clase en Israel, utilizan el hecho de que una profesora extra se asigna cuando la clase aumenta sobre 40 estudiantes.

Estos ejemplos demuestran la justificación de un estimador de variable instrumental que surge de información externa a los confines del análisis cuantitativo. Esas fuentes pueden ser argumentos teóricos, sentido común, o argumentos empíricos que se basan en distintos tipos de información, que muchas veces se centran en entender cómo funciona y opera el programa en la práctica.

V. Elementos para tomar en cuenta en una evaluación de impacto¹³

Hasta ahora, el principal enfoque en la literatura ha sido tratar de ver si el diseño de la evaluación permite obtener estimadores de impacto confiables en un contexto específico. Sin embargo, Deaton llama la atención sobre otros temas importantes con respecto a lo que se puede aprender de una evaluación, para aplicar los resultados en otros ambientes e implementar políticas en el futuro.

Muchas veces se trata de inferir desde una evaluación de impacto si es posible ampliar el programa. Este tipo de preocupaciones de “validez externa” incluye los estudios experimentales y no experimentales. Un mismo programa puede funcionar y tener impactos importantes en una comunidad, pero no tener el mismo impacto en otra, por lo tanto, es complicado realizar generalizaciones en cuanto a ampliar las coberturas de un programa y replicar los resultados. Esto quiere decir que el contexto institucional es sumamente importante para el impacto. Entonces, las preocupaciones sobre la validez externa de las evaluaciones de impacto surgen cuando se requiere que ciertas instituciones estén presentes para facilitar los experimentos.

Otra preocupación respecto a la validez externa es que, a pesar de que supuestos de equilibrios parciales pueden ser apropiados para un programa piloto, esto puede dejar de serlo cuando se amplía la cobertura del ámbito nacional, y los efectos de equilibrio general se vuelven importantes, muchas veces, a esto se conoce en la literatura económica como efectos feedback o macro. Para entenderlo mejor se debe considerar que, por ejemplo, un estimador de impacto de la escolaridad dado un subsidio de educación basado en una prueba aleatoria puede ser engañoso en el momento de ampliar la cobertura, dado que se alterará la estructura de los retornos de la escolaridad. Consecuentemente, un análisis de equilibrio parcial puede sobreestimar el impacto del subsidio de educación una vez que los salarios relativos se ajusten. De esta manera, los efectos de una prueba pueden diferir sustancialmente de los efectos una vez ampliada la cobertura del programa.

Los factores de contexto son cruciales para el desempeño del programa al existir el riesgo de sobreestimar el impacto ya que en ciertos contextos todo parece funcionar perfectamente y en otros, en cambio, nada funciona. Por lo tanto, es muy importante para el programa adaptarse apropiadamente al contexto institucional y socioeconómico en el cual debe operar. Muchas veces, la diferencia que se encuentra en el impacto de la prueba piloto comparada con el impacto que se genera una vez ampliada la cobertura, es que la composición socio económica cambia según la escala. En este sentido, es importante entender qué determina el impacto. Se necesita suplementar las herramientas de la evaluación con otras fuentes de información que pueden aclarar los procesos que influyen en los resultados medidos. Un método es repetir la evaluación en distintos contextos, considerando las características de una localidad. El programa puede tener mayor o menor efecto, por ejemplo, en comunidades con menor desigualdad de género, o con menor tasa de criminalidad, etc.

Otra alternativa es tratar de explorar más a fondo por qué el programa tiene mayores o menores impactos en un contexto específico, como una manera de inferir si será útil implementarlo en distintos contextos. En este caso, se pueden utilizar indicadores intermedios.

En lugar de emplear únicamente los resultados finales de la evaluación de impacto, se pueden mirar con mayor detalle los diferentes efectos en un amplio número de indicadores de comportamiento.

13. La mayor parte del texto fue tomado de Deaton Angus "Instrument of development: Randomization in the topics, and the search for the elusive keys to economic development".

Por ejemplo, saber cuál es el comportamiento de la gente una vez que reciben la transferencia de dinero. Una cosa sería únicamente mirar el efecto que esto tiene en la escolaridad y en la salud, sin embargo, es importante analizar el cambio en el comportamiento de las personas, como por ejemplo: ¿qué tipos de inversión realizaron y por qué? Las respuestas de comportamiento en cuanto a ahorro y consumo pueden proveer información para entender qué determina el impacto.

Además, se pueden obtener otras fuentes de información externa para la evaluación, por ejemplo, fuentes que provienen de la investigación cualitativa. La idea es realizar una prueba de los supuestos que se hacen en la intervención, lo que comúnmente se llama “theory-based evaluation”. Ver: ¿quién participó; se capturaron ciertas partes de la comunidad? En la práctica, no siempre se puede descomponer de esta manera el programa para realizar pruebas de cada nexo de la cadena.

Los efectos de desbordamientos (*spillover*) contienen información más profunda para entender los efectos y cómo un programa opera. Los efectos indirectos ocasionados por el programa en los no participantes son muy comunes.

VI. Conclusiones

Es importante tener en cuenta que no existe una manera única que permita evaluar el impacto de un proyecto, programa o política pública. Cada circunstancia y requerimiento de información necesita de un análisis específico, de la utilización de una herramienta determinada que permita generar información para tomar decisiones que corrijan el desempeño u orientaciones que no conduzcan al logro de los objetivos propuestos

En la práctica, se pueden encontrar estrategias de identificación experimental que permiten aislar el efecto endógeno. La mayoría de veces se combinan varias metodologías para obtener resultados más robustos y con mayor credibilidad. Buenas evaluaciones requieren que el evaluador esté informado, desde el inicio, del funcionamiento y características operativas del programa, que muchas veces son una fuente crucial de información. Por esta razón, se debe considerar que la evaluación no es el último paso por realizar una vez finalizadas las actividades o acciones, si no que se integra desde el primer momento en que estas se conciben y diseñan.

BIBLIOGRAFÍA

- Angrist J, Lavy, 1999. "Using Maimonides' Rule to Estimate the Effect of Class Size on Scholastic Achievement". *Quarterly Journal of Economics*, 114: 533-75.
- Atanasio, Vera, et al, 2004. "The short-term Impact of a Conditional Cash Subsidy on Child Health and Nutrition in Colombia". [en línea]. London: The Institute For Fiscal Studies.
- Besley, Timothy; Case Anne, 2000. "Unnatural Experiments: Estimating the Incidence of Endogenous Policies". *Economic Journal*, 110, F672 – F694.
- Deaton, Angus, 2009. "Instruments of development: Randomization in the topics, and the search for the elusive keys to economic development". Princeton University; National Bureau of Economic Research. NBER Working Paper No. w14690
- Dulfo, Esther, 2001. "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment". *THE AMERICAN ECONOMIC REVIEW*. Indonesia.
- Newman, John, et al, 2002. "Pobreza y Desigualdad en Municipios de Bolivia". Banco Mundial, Bolivia.
- Ravallion Martin, 2005. "Evaluating Anti-Poverty Programs". World Bank Policy Research Working paper 3625.
- Shultz, Paul, 2004. "School subsidies for the poor: evaluating the Mexican Progresa poverty program". *Journal of Development Economics*, Elsevier, vol. 74(1), pages 199-250, June
- Vella Francis, 1998. "Estimating Models with simple Selection Bias: A Survey". *The Journal of Human Resources*, Vol.33, No.1.
- Verbeek, Marno, 2000. "A guide to Modern Econometrics". John Wiley & Sons, Ltd.

Metodologías de evaluación de impacto

Secretaría Nacional de Planificación y Desarrollo
Subsecretaría de Seguimiento y Evaluación