

**Consortio de Investigación Económica y Social
Proyectos Medianos**

**Evaluación de Programas Sociales:
Importancia y Metodologías.
Estimación Econométrica para el Caso de PROJoven**

Informe Final

**Hugo Ñopo
Miguel Robles***

Lima, Noviembre del 2002

* Agradecemos la valiosa colaboración de Manuel Hernandez.

Índice

1.	Introducción.....	3
2.	Importancia del Monitoreo y la Evaluación de Impacto de Proyectos Sociales	5
3.	Experiencias de Evaluación de Proyectos Sociales en el Perú.....	7
4.	Evaluación de impacto: Aspectos conceptuales y metodológicos	15
4.1.	El problema de la evaluación de impacto	15
4.2.	El sesgo de selección.....	19
4.3.	Preselección de la Muestra de Usuarios y Grupo de Control a Evaluar. Los métodos de evaluación de impacto	21
4.4.	Las Diferencias en Características Observables y No Observables: “Ashenfelter’s Dip”	24
4.5.	Los estimadores de impacto	26
4.6.	Métodos de estimación para diseños no experimentales	33
4.7.	Descomposición de las Mediciones de Impacto: Efectos en las Características de los Beneficiarios Vs. Efectos en los Retornos a Dichas Características.	38
4.8.	Evaluación Social de los Programas. Analisis de Equilibrio: Parcial vs. General.....	39
5.	Aplicación de evaluación de impacto para PROJoven.....	41
5.1.	Antecedentes del programa PROJoven.....	41
5.2.	La información disponible.....	43
5.3.	Análisis comparado de los distintos estimadores de impacto obtenidos por medio de la técnica de emparejamiento.....	46
5.4.	Análisis de impacto por medio de técnicas de regresión.....	57
6.	Bibliografía	63

1. Introducción

A pesar que en la década de los noventa se hizo explícita en la gestión pública la importancia de los sistemas de monitoreo y evaluación de proyectos sociales, ellos son aún limitados y en muchos casos deficientes. A raíz de esto “el Estado peruano y los contribuyentes no tienen cómo saber la efectividad y eficiencia de los programas y proyectos que se desarrollan en el país para ayudar a los peruanos en situaciones difíciles”. Si se toma en cuenta que en el año 2000 cerca de unos US\$770 millones se destinaron a la ejecución de proyectos y programas sociales, no cabe duda que evaluar y cuantificar la efectividad de tales desembolsos es de vital relevancia para conocer cuán cerca o lejos se está de haberlos utilizado de la mejor manera posible. Ello permitirá identificar y aprender de aquellos programas con mayor rentabilidad social y enmendar rumbos en aquellos de dudosa efectividad. Así, dada la escasez de recursos, resulta necesario tener presente que la inversión y gasto social debe hacerse de manera no sólo eficaz sino también eficiente.

Por tanto, los esfuerzos por implementar sistemas de evaluación y monitoreo de programas sociales no sólo deben incrementarse, también debe asegurarse la calidad de los mismos. De lo contrario, poca o ninguna utilidad tendrán.

La presente investigación pretende contribuir al uso de rigurosas evaluaciones de impacto de programas sociales que mejor respondan a nuestra realidad y nuestras necesidades como país. Para ello se hace una revisión sobre los aspectos conceptuales y metodológicos de evaluación de impacto de programas sociales y se pretende explicar de manera sencilla sus alcances. El objetivo es acercar una vasta y muchas veces compleja literatura sobre el tema a quienes tienen entre sus manos el manejo, diseño, financiamiento, seguimiento y evaluación de programas sociales en el país.

Este Informe Final afina los conceptos vertidos en el Informe de Avance y concentra sus esfuerzos en el último capítulo. En él se muestran las estimaciones del impacto que tiene ProJoven sobre un conjunto de indicadores de interés, entre ellos los ingresos de los beneficiarios

del programa. Más allá de la importancia del impacto estimado para PROJoven interesa comparar los resultados que se encuentran con distintos estimadores de impacto y/o distintos criterios de emparejamiento.

2. **Importancia del Monitoreo y la Evaluación de Impacto de Proyectos Sociales**

Antes de argumentar sobre la importancia del monitoreo y evaluación de proyectos sociales debe aclararse qué implican estas actividades.

El monitoreo comprende el seguimiento continuo de cómo se va implementando un proyecto en comparación a cómo se acordó en sus inicios que debería implementarse. Así, el monitoreo permite conocer si se van cumpliendo los cronogramas acordados y si están llegando a los beneficiarios del proyecto aquellos elementos que se planearon, ya sean insumos, infraestructura o servicios.

La evaluación de proyectos en un sentido amplio también puede ser entendida como una actividad periódica, y ella permite dimensionar la relevancia del proyecto, su desempeño, eficiencia e impacto con relación a los objetivos del mismo. Pueden realizarse evaluaciones intermedias, antes de la finalización del proyecto, para conocer el progreso del mismo y la medida en que los objetivos propuestos se van alcanzando. Asimismo, una evaluación final debe informar sobre los efectos alcanzados por el proyecto y la sostenibilidad de los mismos.

Entonces, por un lado, el monitoreo guarda relación con la manera cómo se va implementando un proyecto en comparación a lo planeado; por otro lado, la evaluación tiene que ver con la medición de los logros, resultados y/o impactos que se derivan de la existencia del proyecto y que fueron trazados como objetivos a lograr.

La importancia del monitoreo radica en que es una herramienta útil para la gerencia efectiva de los proyectos, en tanto provee las alertas necesarias para saber si la implementación del proyecto diverge de lo planeado y detectar a tiempo potenciales problemas. Gracias al monitoreo se pueden corregir rumbos en la implementación, en el entendido que desviaciones de lo planeado ponen en duda el logro de los objetivos trazados. Una herramienta útil para las labores de monitoreo son los llamados planes operativos anuales o POA, en los cuales quedan expresados de manera explícita las distintas actividades y los plazos en que éstas deben realizarse durante la

implementación de los proyectos. No son pocos los proyectos sociales en los cuales el apoyo de la agencia donante de fondos está condicionado a los resultados que arroje el POA.

Por su parte, la importancia de la evaluación de un proyecto radica en el hecho de qué solo a través de ella se podrá conocer si el proyecto valió la pena o no y si el mismo debe continuar, modificarse o cancelarse. Por lo común se entiende que un proyecto vale la pena cuando logra los efectos que se esperaban de él, sin embargo un criterio más técnico debe buscar establecer la rentabilidad del mismo. Ello estará en función de la magnitud de los logros alcanzados, la permanencia de los mismos y los costos económicos que ha significado la implementación del mismo.

El presente documento se concentra en explicar las metodologías apropiadas para llevar adelante una correcta evaluación de impacto. Ello significa determinar cuáles son los logros atribuibles a la existencia de un programa social. Esta tarea no siempre es fácil dado que a la vez que un proyecto se ejecuta pueden estar ejecutándose otros proyectos o pueden existir cambios en el entorno económico o social, tal que los logros del programa que se evalúa pueden confundirse con los logros de esos otros programas o los efectos de un cambio en el entorno global. La discusión conceptual de tales metodologías será acompañada de una aplicación práctica para el caso del programa de capacitación juvenil, PROJoven.

Asimismo, debe decirse que la discusión aquí planteada se centra principalmente en cómo cuantificar los impactos de un programa social y no se ahonda en la determinación de la rentabilidad (o tasa interna de retorno) de los mismos o el valor presente neto que ellos generan. Estos conceptos implican la valoración inter-temporal de los impactos logrados, es decir la estimación del valor presente de los mismos, y su comparación con el correspondiente valor presente de los costos del programa. En general, existe una vasta literatura estándar sobre evaluación privada y social de proyectos para la estimación de esos indicadores. Para aplicaciones a proyectos sociales en el Perú se sugiere revisar el trabajo de Vásquez et al (2000).

3. Experiencias de Evaluación de Proyectos Sociales en el Perú¹.

En general, en el Perú son aún deficientes y escasos los sistemas de evaluación y monitoreo de programas sociales. Un reciente estudio efectuado por el Instituto Apoyo (2000) sobre “El Estado de los Sistemas de Monitoreo y Evaluación de los Programas Sociales Públicos en el Perú”, concluye que los sistemas vigentes de evaluación (y monitoreo) de los programas y/o proyectos sociales muestran serias deficiencias, a pesar de los esfuerzos de la gestión pública desde la segunda mitad de la década pasada por introducir y fortalecer tales sistemas, y del apoyo de la cooperación internacional. En la mayoría de los casos el Estado y los contribuyentes no tienen cómo conocer la efectividad y eficiencia de los proyectos que se desarrollan en el país.

Así, si bien 9 de los 20 proyectos de la muestra de proyectos analizados en el mencionado estudio realizaron alguna evaluación intermedia o *ex-post*, ésta habría sido de cobertura limitada, o sólo de tipo cualitativo, o sólo de una parte del proyecto. Asimismo, en pocas oportunidades se habrían realizado evaluaciones de impacto y sostenibilidad, al punto que en ninguno de los programas existe una línea de base y sólo tres, que representan cerca del 19% de la muestra en términos de los desembolsos efectuados en el 2000, utilizan grupos de control o comparación².

Por otro lado, en los cuatro estudios de caso analizados por Apoyo (2000), que incluyen el Programa de Capacitación Laboral Juvenil – PROJoven, el Proyecto de Salud y Nutrición Básica (PSNB), el Programa Nacional de Wawa Wasi (PNWW), y el Programa de Mejoramiento de la Calidad de la Educación Primaria (MECEP), se encuentra que el origen de los sistemas de monitoreo y evaluación está en los convenios firmados con los organismos internacionales. En segundo lugar, que no existe una marcada cultura de monitoreo y evaluación en las autoridades políticas. Se constata también que sólo en el caso de PROJoven hubo interés por llevar adelante una evaluación de resultados por parte de la alta dirección cuando inició sus operaciones; en el

¹ En la descripción de trabajos sobre evaluación de impacto realizados en el Perú aparecerán términos propios de la literatura sobre el tema. Cada uno de ellos será aclarado en los siguientes capítulos del presente documento.

² Al respecto, los autores también apuntan que en la nueva Ley de Inversiones, publicada en el 2000 por la Oficina de Inversiones, se establece que en todas las intervenciones públicas de carácter eventual las unidades ejecutoras deben llevar a cabo evaluaciones *ex-post*; sin embargo, es probable que en muchos casos estas unidades se encuentren desactivadas cuando culminen los proyectos, hecho que impediría realizar evaluaciones de sostenibilidad.

PSNB y el MECEP la alta dirección no manifestó decidido interés pero tampoco obstruyó; y en PNWW el interés de la alta dirección se dio recientemente. En tercer lugar, que los organismos financieros cumplen un doble rol al exigir evaluaciones, vinculadas al desembolso o la operación del proyecto, y al brindar apoyo técnico y flexibilidad para la contratación de consultores. Por último, que los proyectos de mayor envergadura, como el MECEP, tienen un menor desarrollo de sistemas de monitoreo y evaluación en comparación a los proyectos de menor envergadura, como PROJoven y el PSNB.

En este sentido, se resalta que PROJoven haya podido desarrollar un sistema de monitoreo y evaluación bastante completo, pero muy ajeno a la estructura tradicional que rige los proyectos sociales en el Perú. El caso de PROJoven destaca por ser el único programa que realiza evaluaciones de impacto internas y periódicas mediante una metodología cuasi-experimental, donde se identifica a un grupo de beneficiarios y a un grupo de control comparables. La selección del grupo de control se da a partir del método del *matching comparison*, sobre la base de variables como edad, sexo, nivel educativo, situación laboral previa y nivel socioeconómico de los individuos. En este caso se hace una evaluación de impacto en diversas variables laborales de los jóvenes beneficiarios del programa, como son la inserción laboral y la calidad de empleo. Para llevar adelante el proceso de evaluación PROJoven contempla la realización de encuestas de tipo socio-laboral, a grupos de beneficiarios y no beneficiarios del programa, al inicio del programa y a los 6, 12 y 18 meses posteriores al egreso del programa de los beneficiarios. En casi todas las convocatorias (nueve a la fecha), las acciones de PROJoven han incluido la selección de un grupo de control y la construcción de una línea de base.

En cuanto al PSNB, se señala que al interior del programa existe una oficina de Planificación y Monitoreo que se encarga de medir el avance de metas por línea de acción, mediante reportes cualitativos mensuales y reportes cuantitativos trimestrales. Para ello se utilizan indicadores de metas físicas por componente y actividad del PSNB, los cuales fueron diseñados al inicio del proyecto y que posteriormente fueron ajustados según los cambios introducidos en el mismo. Respecto a las evaluaciones externas, se ha realizado una de medio término y otra luego del cierre de sus actividades de campo, donde se elaboró un Plan de Evaluación en el que se especificaron

los diferentes estudios de evaluación a realizarse para el levantamiento y análisis de los indicadores seleccionados.

En el caso de PNWW, se indica que a pesar de ser un proyecto pequeño y de contar con reportes de avance, no se ha podido desarrollar un sistema de monitoreo y evaluación adecuado, dado que existen serios problemas en el levantamiento y procesamiento de la información. En la evaluación de medio término, efectuada por el Instituto de Información y Metodologías para el Desarrollo Organizacional, se encuentran considerables deficiencias en los sistemas de monitoreo y evaluación.

Finalmente, se precisa que en el MECEP existen dificultades para estandarizar instrumentos y lograr un sistema de monitoreo y evaluación integrado, debido a la extensión del programa y a que cada componente del MECEP tiene autonomía para establecer sus propios mecanismos de monitoreo y evaluación. La evaluación externa efectuada por el Instituto Apoyo muestra, entre otras cosas, que las acciones de monitoreo y evaluación no constituyen un sistema integral (a nivel agregado se hace un seguimiento de metas físicas y financieras más no en gran medida de procesos), que no existe suficiente coordinación entre los distintos componentes del proyecto (en términos del seguimiento de avances); y que en algunos componentes, no se dispone de suficientes instrumentos para la evaluación y no existe representatividad en las muestras seleccionadas.

De otro lado, en Vásquez (2000) “Monitoreo y Evaluación de Impacto de Programas Sociales: el caso de algunas Instituciones Públicas, Privadas y Fuentes Cooperantes”, se analiza el caso del Fondo Nacional de Compensación y Desarrollo (Foncodes), donde la evaluación de cada proyecto implica un análisis previo a la implementación del mismo que incluye un resumen de las características del proyecto y su viabilidad técnica, ambiental, social y económica; un seguimiento y monitoreo que supone una evaluación de la gestión realizada internamente y la elaboración de cronogramas específicos de evaluación de resultados; y una evaluación *ex-post* que responde en gran parte a las demandas de los organismos internacionales y que tiene como

meta comparar la situación de la población objetivo antes del proyecto y después de él a través de distintos indicadores³.

Sin embargo, entre los principales problemas detectados en la evaluación *ex-post* se encuentran la escasa representatividad estadística de los indicadores obtenidos así como la inexistencia de indicadores precisos; el reducido tamaño de la muestra y la complejidad de las encuestas de evaluación; la ausencia de técnicos idóneos y profesionales capacitados para evaluar y supervisar los proyectos; y la divergencia del análisis llevado a cabo.

Dada la escasa tradición sobre evaluación de impacto de los programas sociales en el país son muy pocos los trabajos sobre el tema que den cuenta de la metodología empleada para tal evaluación y los resultados encontrados. Sin embargo, entre los recientes trabajos de evaluación de impacto de programas sociales en el país, se puede destacar el “Costo Efectividad del Programa de Desayunos Escolares de Foncodes y el Programa de Alimentación Escolar del Pronaa”, efectuado por Ravina, Paulini y Cancho (2002); “Una Metodología de Evaluación de Impacto de los Programas de Educación y/o Capacitación en el Mercado Laboral: el Caso de PROJoven” de Galdo (2000); y “Una Medición del Impacto del Programa de Capacitación Laboral Juvenil PROJoven” de Ñopo, Robles y Saavedra (2002).

El primero de ellos, analiza el costo-efectividad de dos programas de ayuda nutricional implementados al interior de escuelas (el PDE de Foncodes y el PAE del Pronaa). Emplea una técnica analítica (Análisis Costo-Efectividad) que permite comparar los costos de un proyecto con los beneficios resultantes, no expresados en la misma unidad de medida. En este caso, los costos de cada programa incluyen tanto los costos en alimentos como los costos operativos, mientras que la efectividad es evaluada en términos del impacto del programa sobre la asistencia de los niños a la escuela primaria.

³ La evaluación *ex-post* también implica analizar cada una de las fases del ciclo de un proyecto (focalización y priorización, promoción, evaluación y aprobación y control y seguimiento) y proponer una serie de sugerencias según cada caso.

Para la estimación cuantitativa de dicho impacto, se utiliza la información de la Encuesta Nacional de Hogares (ENAH) correspondiente al segundo trimestre de 1999. En concreto, se aplica un modelo Probit binomial (para cada programa) donde la variable dependiente toma valor 1 si el niño asiste a la escuela y 0 si no lo hace; y se considera como variables explicativas a la variable de intervención (que toma valor 1 en caso el niño habite en un centro poblado donde se implemente el programa y 0 en caso contrario), así como algunas características de los niños, del hogar, de la vivienda y del contexto geográfico en el que habitan. De esta forma, se estima el efecto “puro” del programa sobre la asistencia escolar.

Con el propósito de atenuar el problema del “sesgo de selección” por “variables observables” que estaría presentando la variable de intervención, se encontró previamente para cada programa un grupo de comparación a través de la metodología del *propensity score*. A partir de un modelo Probit binomial, se calculó la probabilidad de cada niño de ser potencial beneficiario del proyecto (pertenecer a un centro poblado donde se implementa el proyecto); seguidamente se seleccionó a aquellos potenciales beneficiarios para llevar a cabo un proceso de emparejamiento entre dicho grupo y el de no beneficiarios, minimizando las diferencias en las probabilidades estimadas mediante un proceso de apareamiento con reemplazo. De este modo, se aseguró que a cada niño, potencial beneficiario del programa, se le asigne un niño de un centro poblado donde no se implementa el mismo.

Para dar solución al problema de “sesgo de selección” por variables “no observables”, se halló, a través del modelo Probit planteado para el cálculo del *propensity score*, una función monótona decreciente conocida como “ratio inverso de Mills” que luego se incorporó como un regresor adicional en la estimación del impacto de cada programa, es decir en la regresión de asistencia escolar⁴. De este estudio se concluye que el PDE de Foncodes ha tenido mejores resultados que el PAE del Pronaa (menores costos por unidad de efectividad).

⁴ Sin embargo, la metodología utilizada en este trabajo incurre en error al pretender combinar la metodología de emparejamiento o “matching comparisons” con los modelos econométricos de sesgo de selección que hacen uso del ratio inverso de Mills. En particular, al usarse una corrección por sesgo de selección del tipo ratio inverso de Mills se debió haber estimado una ecuación de asistencia escolar con sólo la muestra de “no beneficiarios” del programa, o

Por otro lado, el estudio de Galdo evalúa el impacto de la primera convocatoria del programa de capacitación laboral juvenil PROJoven sobre la inserción laboral remunerada, los ingresos y la calidad de empleo de los jóvenes participantes de este programa, comparándolos con un grupo de control para aislar los posibles efectos del entorno macroeconómico o del proceso de maduración de los propios individuos. La metodología de evaluación empleada se basa en el modelo no experimental del *matched comparisons*. Inicialmente se estableció una equivalencia entre una muestra de beneficiarios y un grupo de control mediante un estricto apareamiento en dos etapas: se identificó a los potenciales integrantes del grupo de control mediante un mapeo de las zonas de origen de los beneficiarios y luego se procedió a aparearlos uno a uno (*cell matching*) con los integrantes del grupo de beneficiarios, sobre la base de la edad, sexo, nivel educativo, estrato socioeconómico y situación laboral previa al inicio del programa de los individuos. Seis meses después de concluido el programa (o 12 meses después del inicio del mismo), se hizo una comparación de los cambios mostrados por los beneficiarios y controles en las variables bajo estudio. En algunos casos incluso se distinguió el efecto por género, edad y nivel educativo de los individuos.

La evaluación de impacto también incluye una estimación econométrica del diferencial de ingresos, antes y después del programa, entre el grupo de beneficiarios y controles. Para ello, se considera como variables independientes a una variable dicotómica que toma valor 1 para el grupo de beneficiarios y 0 para el grupo de control, y a un conjunto de variables de control post-programa (variables laborales y de capital humano). Para corregir el problema de “sesgo de selección” que pudiera estar presentando la variable dicotómica se utilizó, al igual que en el trabajo de Ravina et al. (2002), la técnica econométrica del “ratio inverso de Mills”: se estimó un modelo Probit binomial para identificar las variables que inciden mayormente en la participación de los jóvenes en PROJoven y con los residuos de esta relación se obtuvo el ratio en mención, que luego fue incorporado como regresor en una ecuación que estima el diferencial de ingresos después y antes de la participación en el programa, usando para dicha estimación la muestra tanto

alternativamente con sólo la muestra de los beneficiarios. La explicación posterior de estas metodologías permitirá aclarar el error incurrido.

de beneficiarios como de controles⁵. Se concluye en este trabajo que PROJoven logra un impacto estadísticamente positivo sobre la situación laboral de los jóvenes beneficiarios.

El trabajo de Ñopo, Robles y Saavedra, por su parte, analiza el impacto de la sexta convocatoria de PROJoven sobre la inserción laboral, los ingresos por hora y las horas de trabajo de los jóvenes beneficiarios. La evaluación utiliza información socio-laboral de una muestra de beneficiarios y un conjunto de individuos que no pasaron por el programa y que conforman el grupo de control. Esta información fue recogida antes de iniciados los cursos de capacitación (línea de base) y seis meses después de concluido el programa (primera medición de salida). El estimador que se emplea para medir el impacto es el de “diferencias en diferencias”, donde se compara la situación promedio antes-después de los beneficiarios con aquella de los controles.

Pese a que originalmente se contaba con un emparejamiento entre beneficiarios y controles efectuado por el propio equipo de PROJoven, para la cuantificación del impacto se consideró un nuevo emparejamiento dado que se encontraron algunas diferencias importantes entre los beneficiarios y controles en el periodo pre-programa. Se utilizó el método de la minimización de distancias euclidianas, considerando las siguientes variables: sexo, edad, nivel de educación, situación laboral, nivel de pobreza, ingresos por hora antes del programa, horas trabajadas a la semana antes del programa, y la propensión a participar en el programa. Esta última se obtuvo aplicando la metodología del *propensity score* (mediante un modelo Probit binomial se estimó la probabilidad para cada individuo de participar en PROJoven). El proceso de emparejamiento fue uno a uno sin reemplazo.

En la evaluación del impacto de PROJoven sobre la inserción laboral y el ingreso por hora, se incluye a su vez una estimación econométrica (para efectos comparativos). En el primer caso se analiza la probabilidad de encontrarse en situación de ocupado luego del programa utilizando un modelo Probit; en el segundo caso se estima un modelo de regresión donde la variable dependiente es el cambio porcentual de los ingresos entre el momento pre-programa y el

⁵ En este trabajo metodológicamente se incurre en similar error que el comentado en el trabajo de Ravina et al. (2002).

momento post-programa. En ambos modelos se incluye una variable dicotómica que distingue a los beneficiarios de los controles y a partir de la cual se captura el efecto aislado del programa.

De otro lado, el estudio también evalúa el impacto del programa sobre la segregación ocupacional por género y sobre la calidad del empleo. Asimismo, estima una tasa interna de retorno tomando en cuenta los costos (directos y de oportunidad) y beneficios que genera el programa y que pueden ser atribuidos directamente a PROJoven o a los jóvenes participantes del mismo. Esto implicó valorizar previamente los efectos hallados en la inserción laboral, el ingreso por hora y las horas trabajadas a la semana. Cabe señalar que en todos los aspectos evaluados se encuentra que PROJoven tiene efectos positivos.

4. Evaluación de impacto: Aspectos conceptuales y metodológicos⁶

4.1. El problema de la evaluación de impacto

Para entender cuál es el objetivo de la evaluación de impacto de un programa social se introducen primero tres conceptos que nos ayuden a entender el marco conceptual a plantearse. Un primer concepto es el de los “estados de la naturaleza”⁷. Este concepto hace referencia al hecho de que en el futuro pueden darse distintas realidades o estados posibles; por ejemplo, mañana puede llover o puede no llover. Ello implica que hoy podemos imaginarnos cómo será el mundo mañana bajo dos escenarios posibles, el primero sería en un “estado de lluvia”, el segundo bajo un “estado de no lluvia”. Sin duda que llegado mañana sólo uno de esos estados sucederá. Por supuesto, los “estados de la naturaleza” imaginables no tienen porque ser sólo dos, podríamos haber planteado por ejemplo cuatro: sin lluvia, con llovizna, con lluvia fuerte, con tormenta.

El concepto de “estados de la naturaleza” se usa típicamente en la evaluación de programas sociales para describir dos estados posibles: el “estado de tratamiento” y “el estado de no-tratamiento”. El primero hace referencia a una realidad en que un individuo cualquiera ha participado efectivamente de un programa, es decir si ha sido “tratado” por el programa. Por lo general decimos que este es el estado 1. Como puede adelantarse, el segundo estado hace referencia a una realidad en la que un individuo cualquiera no ha participado efectivamente de un programa, es decir si no ha sido “tratado” por el programa. Así, antes que se inicie un programa social podríamos imaginar la vida futura de cualquier individuo bajo dos escenarios posibles, si pasa o es tratado por el programa, si no pasa o no es tratado por el programa.

⁶ Una revisión rigurosa sobre aspectos conceptuales y econométricos en torno a evaluación de impacto se encuentra en Heckman, LaLonde y Smith (1998). El lector no familiarizado con técnicas estadísticas y econométricas complejas puede acudir a Baker (2000) y Ravallion(1999).

⁷ El concepto de “estados de la naturaleza” es ampliamente utilizado en teoría de Finanzas. Así, a través del “precio” de cada uno los “estados de la naturaleza” es posible valorar cualquier activo financiero.

Un segundo concepto útil es el de los potenciales “resultados”⁸ que puede experimentar un individuo cualquiera durante su vida. Así, los resultados que puede alcanzar un individuo cualquiera durante los próximos cinco años podrían ser, por ejemplo, los siguientes: estará empleado todos los años, sus ingresos serán US\$ 500 del primer al tercer año y US\$ 700 el cuarto y quinto año, estará casado hasta el cuarto año y en el quinto año permanecerá divorciado, aumentará 2 kilos cada año, en el tercer año se mudará a otro barrio, etc. En general, el “resultado” que presente un individuo puede estar referido a distintas variables o aspectos de su vida y durante varios momentos en el tiempo. En la evaluación de impacto interesará concentrarse en aquellos aspectos en los cuales se espera que el programa tenga alguna implicancia. Así, por ejemplo, si se evalúa un programa de desayunos escolares posiblemente definiremos los “resultados” en función del estado nutricional y del rendimiento académico. Normalmente, las evaluaciones de impacto se concentran en el “resultado” de unas pocas variables y para pocos períodos de tiempo. Típicamente se utiliza la letra Y para representar los resultados potenciales que puede alcanzar un individuo. Así, Y puede representar a una única variable o ser un vector que abarque varias variables, y puede referirse a los “resultados” de un período en particular o de varios períodos. Si Y representa a varias variables en varios períodos será una matriz.

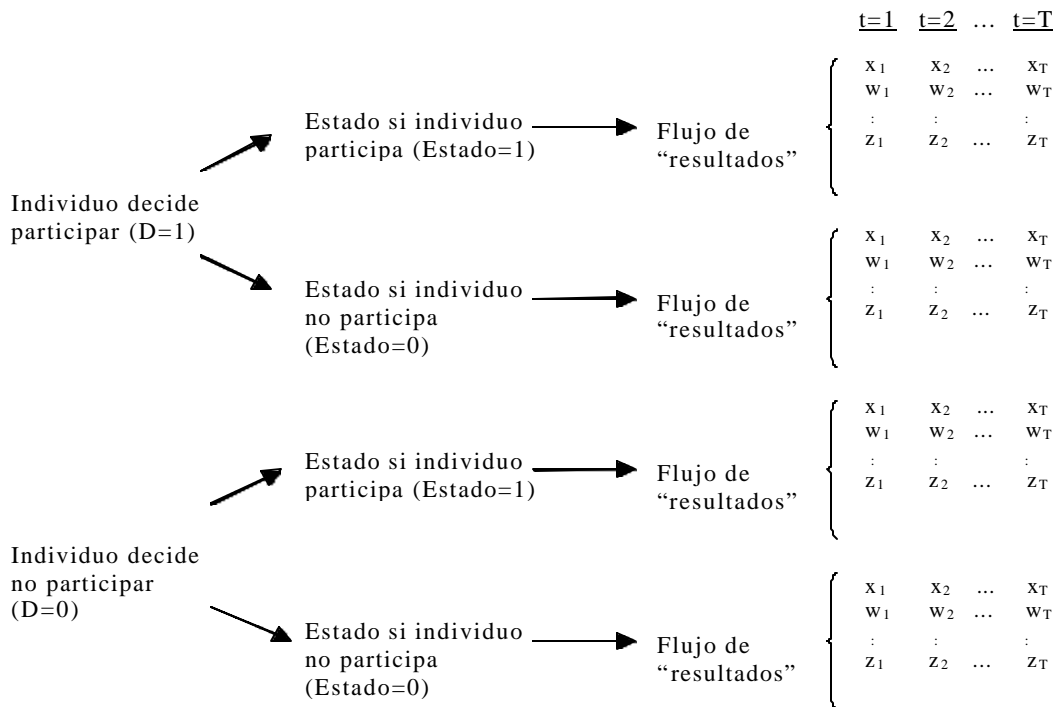
Un tercer concepto a revisar es el de la “decisión de participar” en un programa social. Vale la pena distinguir entre aquellos individuos que deciden participar en un programa de aquellos que deciden no hacerlo. La mayoría de los programas sociales suelen ser voluntarios en el sentido de que todos aquellos individuos que quieren participar (y cumplen con los requisitos para hacerlo) pueden potencialmente hacerlo. Un ejemplo sería la decisión de acudir a un comedor popular. En algunos otros programas la decisión de participar en el programa no está en manos de los participantes sino de terceros. Por ejemplo, en un programa de caminos rurales puede ser el gobierno o el ente ejecutivo del programa el que decida qué caminos serán rehabilitados o construidos y cuales no lo serán. En este caso puede pensarse en los pobladores aledaños a esos caminos como los participantes o beneficiarios del programa.

⁸ En inglés un término adecuado para este concepto es el de “outcomes”

Es importante distinguir también entre aquellos que deciden participar en un programa y aquellos que efectivamente lo hacen. En muchos programas se pueden tener más personas queriendo participar que vacantes disponibles, de tal manera que muchos deciden participar pero no se puede atender a todos, y por tanto sólo algunos efectivamente logran participar. En la literatura sobre evaluación de impacto es común utilizar la letra D para distinguir entre quienes deciden participar ($D=1$) y los que no lo hacen ($D=0$).

Luego de haberse revisado algunos conceptos será más fácil entender qué se entiende por evaluación de impacto. En la Figura 1 se aprecian los flujos de “resultados” que puede tener un individuo, para un conjunto de variables (x, w, \dots, z) entre el período 1 y el período T. Ese flujo de resultados dependerá de si el individuo decidió participar en el programa que se está evaluando y de si estamos considerando un “estado de tratamiento” o un “estado de no tratamiento”.

Figura 1
Flujos de resultados según la decisión de participación
y los estados de la naturaleza



Bajo este marco pueden evaluarse distintos impactos de un programa. En primer lugar, el impacto comúnmente analizado en la mayor parte de las evaluaciones corresponde a tratar de estimar cómo cambia el “flujo de resultados” de los individuos que decidieron participar en un programa, y que efectivamente lo hacen, cuando se compara su flujo de resultados en un “estado de tratamiento” con un “estado de no-tratamiento”. En términos más simples ello significa responder a la pregunta ¿qué ganan los participantes de un programa en comparación a no haber participado? Nótese que esta pregunta es muy distinta a ¿cuál es la situación de los participantes de un programa después del programa en comparación a antes del programa?⁹. Sin embargo, es claro que para aquellos individuos que deciden y efectivamente participan (los beneficiarios) de un programa social sólo podremos conocer su flujo de resultados en el “estado de tratamiento”, ya que ese es el “estado” que efectivamente se materializa para ellos. Preguntarse por el “flujo de resultados” de los beneficiarios en un “estado de no-tratamiento” es preguntarse por una situación hipotética que no se dará y que por tanto no podremos observar directamente (pero, bajo ciertos supuestos estadísticos, se puede *estimar*). Es equivalente a preguntarse por ejemplo cuál hubiese sido el número de asistentes a los desfiles de fiestas patrias en Lima si esos días hubiesen sido soleados en lugar de nublados, como efectivamente lo fueron.

Así, el problema central de la evaluación de impacto radica en tratar de averiguar cómo habría sido esa situación hipotética para los beneficiarios de un programa, es decir su “flujo de resultados”, si no hubiesen sido “tratados” por el programa. A esa situación hipotética se le conoce como “contra-factual”.

En segundo lugar, uno puede preguntarse por el impacto que tendría el programa entre aquellos que decidieron no participar en él. En este caso, la pregunta a responder es ¿qué hubiesen ganado los no participantes de un programa social si efectivamente hubiesen participado? Esta pregunta puede ser muy relevante cuando se quiere extrapolar el impacto que un programa ha tenido sobre los beneficiarios a grupos no beneficiarios del mismo. ¿Será cierto que el impacto de un programa social es el mismo sobre los individuos que deciden participar en él que sobre los

⁹ Debe diferenciarse entonces el análisis que toma en cuenta el “con o sin programa” de aquel que se concentra en el “antes y después del programa”.

individuos que deciden no participar en él? La mayor parte de las veces la respuesta es negativa. En este tipo de evaluación la situación que no podrá ser observada es aquella que corresponde al “flujo de resultados” en “estado de tratamiento” de los que deciden no participar en el programa. En este caso vamos a tener que hacer una aproximación de esa situación hipotética para conocer el impacto del programa sobre los no beneficiarios.

Formalmente, el impacto que un programa tenga sobre un individuo cualquiera corresponde a la mejora en el flujo de resultados que éste alcanza cuando lo trasladamos de un “estado de no-tratamiento” a un “estado de tratamiento”. Si Y_1 representa el flujo de resultados en el estado 1, es decir de “tratamiento”, y Y_0 representa el flujo de resultados en el estado 0 de “no-tratamiento” el impacto sobre un individuo cualquiera puede representarse por: $D = Y_1 - Y_0$

4.2. El sesgo de selección

Un concepto recurrente en la literatura sobre evaluación de impacto es el de “sesgo de selección”. Esta idea hace referencia al hecho que quienes deciden participar en un programa lo hacen porque poseen algunas características particulares en comparación a quienes deciden no participar. Supóngase que se inaugura un comedor popular en un distrito de la ciudad y que puede acudir todo aquel que quiera hacerlo, a condición de que resida en el mismo distrito. Lo que vamos a observar es que entre los adultos que acuden a ese comedor popular la proporción de mujeres es mayor que la de hombres, cuando se compara con la proporción de mujeres adultas al nivel de todo el distrito. Supóngase que la explicación a ese fenómeno radica en el hecho de que la mayor parte de los hombres del distrito trabajan lo suficientemente lejos de ese comedor popular como para que sea costo-eficiente asistir a él. En este caso, la población que decide “participar” en el comedor popular presenta un “sesgo” a ser mayoritariamente población que trabaja lo suficientemente cerca del lugar o que no trabaja. Ello es más común entre las mujeres.

Como regla general, puede pensarse que quienes deciden participar en un programa son aquellos para quienes luego de una evaluación costo-beneficio el programa representa su mejor opción,

quienes deciden no hacerlo es porque tienen mejores opciones de acción disponibles. La pregunta que surge es por qué para algunos individuos el programa puede ser una buena opción y para otros no. Ello se explica por el hecho de que para un individuo cualquiera su “flujo de resultados” en un determinado “estado de la naturaleza” depende de sus propias características o las de su entorno. Por tanto, el “impacto” del programa sobre ese individuo dependerá también de esas características. Un ejemplo de lo anterior sería pensar cómo será el impacto del programa de vaso de leche si éste se lleva a colegios privados. Sin duda que el impacto sería menor que el potencial impacto que se puede alcanzar en colegios públicos, ya que los ingresos familiares, acceso a agua, educación de los padres, etc. es mayor entre este último grupo de niños y por tanto es mucho más probable cuenten con un mejor nivel nutricional. El programa vaso de leche poco les aportará a estos niños, pero sí mucho a los niños de colegios públicos.

Una distinción adicional que vale la pena establecer es el “sesgo de selección” sobre variables observables del “sesgo de selección” sobre variables no observables. La primera se da cuando los que deciden participar en un programa lo hacen sobre la base de contar con características que pueden ser observadas por el evaluador o investigador. En el ejemplo del comedor popular sería muy sencillo preguntar a una muestra de personas del distrito si asisten o no al comedor popular y cuán lejos se encuentra su centro de trabajo. En este caso podremos observar que los asistentes al comedor se “seleccionan” de entre quienes trabajan cerca y los que no asisten se “seleccionan” de entre quienes trabajan lejos.

En otros casos no será posible observar la característica que produce el “sesgo de selección”. Por ejemplo, si los individuos que deciden participar en un programa son aquellos que cuentan una cualidad de liderazgo difícilmente el evaluador dispondrá de una variable que mida esa cualidad. Buena parte de la literatura sobre evaluación de impacto conviene en usar el término “sesgo de selección” para referirse únicamente al sesgo de selección en variables no observables.

4.3. Preselección de la Muestra de Usuarios y Grupo de Control a Evaluar. Los métodos de evaluación de impacto

Como se ha mencionado previamente el problema central de la evaluación de impacto es la construcción de un “contra-factual”. Ello implica poder reconstruir de una u otra manera esa situación hipotética que no podrá ser observada: ¿Cuál hubiese sido el flujo de resultados de los que participaron en un programa si no lo hubiesen hecho? ¿Cuál hubiese sido el flujo de resultados de los que no participaron en un programa si lo hubiesen hecho?

Si lo que se busca es estimar el impacto del programa sobre el grupo de beneficiarios habrá que tratar de responder a la primera de esas preguntas. Para tratar de reconstruir ese estado “contra-factual” lo que se hace es buscar a un grupo de individuos que no han participado del programa para que a partir de sus “flujo de resultados” podamos decir algo sobre cuáles hubiesen sido los “flujos de resultados” de los beneficiarios en el “estado de no tratamiento”, eso que es imposible observar. Ese grupo de individuos que nos permitirá reconstruir el estado “contra-factual” se conoce como “grupo de control”. Un objetivo que debe buscar el evaluador es que el “grupo de control” sea tal que le permita hacer una muy buena reconstrucción del “estado contra-factual”.

¿Cómo elegir el grupo de control? La respuesta a dicha pregunta es fundamental en los procesos de evaluación y a partir de ella se pueden clasificar los diseños de evaluación en experimentales o no experimentales.

Los métodos de evaluación experimentales

Estos métodos se basan en una selección aleatoria del “grupo de control”. Operativamente funcionan de la siguiente manera. Se convoca a concurso a todos aquellos que quieren participar en el programa en cuestión. Aquellos individuos que deciden participar y cumplen con los requisitos para hacerlo se inscriben¹⁰. Luego, se elige al azar a los individuos que podrán

¹⁰ La mayor parte de los programas sociales exigen que los participantes cumplan ciertos requisitos para poder participar. Por ejemplo, que se encuentren desempleados, que sean de bajos ingresos, que tengan hijos menores, que vivan en una determinada zona, que no cuenten con educación superior, etc.

participar y a los que no lo harán. Estos últimos (o una muestra de ellos) son los candidatos ideales para conformar el grupo de control.

¿Cuál es la ventaja de los métodos de evaluación experimentales? Ella radica en que al haber sido escogidos al azar tanto el grupo de beneficiarios como el grupo de control es de esperarse que ambos grupos posean características muy similares. Para entender por qué ello es así imagínese que del total de personas que viven en su distrito elegimos completamente al azar dos grupos de 500 personas, ¿usted esperaría que el porcentaje de mujeres sea muy parecido en ambos grupos? Claro que sí.

Lo interesante de los diseños de evaluación experimentales es que el grupo de beneficiarios y el grupo de control serán parecidos en cuanto a sus características observables y también en cuanto a sus características no observables. ¿Esto quiere decir que no existirá “sesgo de selección” en los diseños experimentales? No. Si comparamos a los individuos que decidieron participar (aunque no todos salieron sorteados para efectivamente participar) con los individuos que decidieron no participar es muy probable que encontremos características particulares en los primeros que les hizo tomar la decisión de inscribirse en el programa. Sin embargo, la ventaja de los diseños experimentales es que permite balancear el “sesgo de selección” entre el grupo de beneficiarios y de controles, es decir dicho sesgo puede existir pero será similar en ambos grupos.

¿Cuál es la desventaja de los diseños experimentales? Por lo general los encargados de la política pública rechazan este tipo de diseños experimentales porque sienten que pueden ser percibidos como injustos o políticamente inviables. Escoger por sorteo a quienes entran y a quienes quedan excluidos de un programa social puede generar desencanto entre éstos últimos y mermar la viabilidad política de su ejecución. Más aún si quienes fueron excluidos por sorteo tendrán que colaborar para hacer las veces de grupo de control. No es raro encontrar disputas entre los evaluadores, abogando por los diseños experimentales, y los responsables de la política pública argumentando en contra de ellos.

Una segunda desventaja de los diseños experimentales es que hacen muy difícil extrapolar los impactos encontrados hacia grupos de la población que decidió no participar en el programa. Esto se conoce como escasa “validez externa”. Ello por cuanto en el grupo de control no se cuenta con información de individuos que abiertamente decidieron no participar.

Los métodos de evaluación no experimentales

Los métodos de evaluación no experimentales, no utilizan un grupo de control seleccionado al azar de un conjunto de individuos que deciden participar del programa. Más bien, el grupo de control se forma con personas que quisieron participar en el programa pero que llegaron tarde o al último a la inscripción y por tanto no pudieron hacerlo; o bien con individuos que siendo elegibles para el programa por cualquier otro motivo no participaron del mismo.

El requisito central para formar un grupo de control es lograr que éste este conformado por individuos “muy parecidos” a los beneficiarios. En este contexto “muy parecido” significa que los individuos seleccionados para el grupo de control repliquen las características de los individuos beneficiarios del programa. ¿Qué características deben replicar? Aquellas que se utilicen como criterios de participación en el programa, por un lado, y también aquellas que inciden sobre los “flujos de resultados”, por otro. Así, si nos interesa evaluar el impacto de un programa sobre los ingresos laborales entonces tendrá que ponerse cuidado en que el grupo de control sea parecido al grupo de beneficiarios en: niveles de educación, experiencia laboral, sectores en los cuales trabajan, género, edades, educación de los padres, composición familiar, etc., es decir, todas aquellas variables que explican el mayor o menor ingreso que pueda tener un individuo. En buena medida al hacer que el grupo de control sea parecido al grupo de beneficiarios lo que se busca es eliminar cualquier “sesgo de selección” derivado de variables observables.

¿Cómo se puede conseguir a los miembros de un grupo de control en el caso de diseños experimentales? Una primera fuente de miembros del grupo de control puede ser aquel grupo de individuos que decidieron participar en el programa pero que llegaron tarde o al último y por

tanto no lograron una vacante en el mismo. Por lo general, estos individuos serán bastante similares a los beneficiarios. Sin embargo, cabe preguntarse ¿por qué llegaron al último? Podría ser porque son individuos con menor iniciativa. Esto representaría un “sesgo de selección” en una variable “no observable”.

Una segunda fuente para hallar a los miembros del grupo de control es recurrir a zonas donde el programa opera e identificar a individuos que sean elegibles para el programa (es decir, individuos que cumplen con los requisitos, si los hay) y que sean parecidos en variables observables a los beneficiarios del programa. Para reducir potenciales “sesgos de selección” en variables no observables puede darse preferencia a aquellos individuos que no tuvieron ninguna información sobre la existencia del programa a ser evaluado.

En tercer lugar, pueden utilizarse fuentes secundarias para hallar el grupo de control. Éstas pueden ser las encuestas de hogares u cualquier otra base de datos con un amplio número de individuos de donde se puede extraer a individuos con parecido a los beneficiarios. Sin embargo, esta alternativa debe ser utilizada con sumo cuidado: mientras que por un lado deben tomarse precauciones en asegurar cierta representatividad de la fuente, por el otro debe verificarse la compatibilidad de los instrumentos utilizados en el levantamiento de información (las encuestas deben contener preguntas similares) y en las técnicas de muestreo¹¹.

4.4. Las Diferencias en Características Observables y No Observables: “Ashenfelter’s Dip”

Toda técnica de emparejamiento va acompañada de un nivel de tolerancia en la mínima distancia a considerar (“caliper”). Según esto, si para cierto beneficiario sucede que el control que minimiza la distancia (Euclideana, Mahalanobis, Propensity o cualquier otra) aún se encuentra muy alejado, es decir, la distancia entre ambos es mayor al nivel de tolerancia, el emparejamiento

¹¹ Para mayor detalle acerca de los cuidados a tomar en cuenta para una correcta utilización de fuentes secundarias, vease Heckman *et al.* (1998).

no se realiza. Así, el beneficiario a evaluar deja de ser considerado en el análisis por falta de un adecuado control.

A medida que el nivel de tolerancia se hace más exigente, dos efectos interesantes (y correlacionados) son observados: por un lado la probabilidad de encontrar beneficiarios no emparejados aumenta, pero por el otro, el “grado de semejanza” entre la muestra de beneficiarios y controles también aumenta. Se conoce a esta coexistencia de efectos como el balance entre el emparejamiento incompleto y el imperfecto. El emparejamiento incompleto es no deseado porque eventualmente implica sesgos (o falta de representatividad) en la muestra de beneficiarios emparejados. El emparejamiento imperfecto es no deseado porque implica diferencias en características observables entre los beneficiarios y sus respectivos. La elección del nivel de tolerancia requiere de un adecuado balance entre estos dos efectos.

Un problema en el emparejamiento sobre el cual no se tiene control es el de la existencia de diferencias en características no-observables. Un supuesto modelístico en el que se basa la literatura de evaluación es que la minimización de las diferencias en características observables estará ligada a mínimas diferencias en características no-observables. Este supuesto, aunque razonablemente aceptable, es imposible de validar empíricamente. En la medida que se pueda trabajar con un vasto conjunto de características observables y bajos niveles de tolerancia en las diferencias en dichas características entre beneficiarios y controles, se podrá suponer que las diferencias en características no-observables están bajo control.

Particularmente, para programas de capacitación/inserción laboral se presenta un fenómeno interesante cuando se comparan las trayectorias de ingresos de beneficiarios y de individuos comparables pero no participantes en el programa. Lo que se observa son diferencias entre los ingresos percibidos en los períodos previos al inicio del programa pero que no responden a diferencias estructurales entre los dos grupos: el llamado “Ashenfelter’s Dip”¹². Según Ashenfelter, los beneficiarios de este tipo de programas deciden participar en ellos

¹² La literatura acuñó este término debido a que fué en Ashenfelter (1978) donde se documentó por primera vez este fenómeno.

inmediatamente después de haber sufrido un “shock negativo de ingresos” (un despido inesperado del centro de trabajo, una disminución de la jornada laboral, etc.). Así, si la variable “ingresos” no es tomada en cuenta como herramienta de control en el proceso de emparejamiento de beneficiarios y controles, será posible encontrar que el ingreso medio, durante períodos previos al programa, de los controles es mayor que el ingreso medio de los beneficiarios. Por otro lado, considerando únicamente la variable “ingresos al momento de inscribirse en el Programa” como herramienta para el emparejamiento, se corre el riesgo de emparejar a beneficiarios de ingresos medios o altos (que sufrieron un “shock” instantáneo y negativo en sus ingresos) con controles de ingresos estructuralmente bajos, generando de esta manera un sesgo hacia arriba (sobre-estimación) en la medición del efecto del Programa. La recomendación de Ashenfelter, y que nosotros suscribimos, es añadir una variable al emparejamiento: “historia de ingresos”. En el caso particular de ProJoven, o de programas similares como “A Trabajar”, el horizonte temporal podría remontarse a 6 o 12 meses previos al programa. Así, teniendo en cuenta la trayectoria temporal de los ingresos para beneficiarios y controles será posible una mejor identificación de aquellos efectos sobre los ingresos, u otras variables, que son realmente atribuibles al programa y aquellos atribuibles a choques aleatorios o a la existencia de otras medidas macroeconómicas que afecten la dinámica del mercado de trabajo.

4.5. Los estimadores de impacto

Como se ha mencionado previamente la evaluación de impacto busca conocer como cambia el “flujo de resultados” de un individuo cualquiera cuando éste pasa de un “estado de no-tratamiento” a un “estado de tratamiento”. Sin embargo, se ha mencionado también que será inviable conocer ese cambio con certeza. Ello, porque sólo uno de esos estados será el que se materializa, el otro no será nada más que un estado hipotético sobre el cual uno puede elucubrar “cómo hubiese sido”. Si hoy llovió, simplemente perdimos la oportunidad de observar el día sin lluvia. No obstante, esto no quiere decir que no sea posible lograr una buena idea de cómo hubiese sido el día sin lluvia.

Dada la imposibilidad de conocer el verdadero impacto de un programa para un individuo cualquier tenemos que recurrir a un “estimador” de ese efecto. Para ello necesitamos de un grupo de control. Por lo general, las evaluaciones de impacto se concentran en estimar el impacto del programa sobre los beneficiarios. Y los estimadores disponibles son buenos para aproximar el impacto medio del programa (el promedio de los impactos sobre un grupo de individuos, por ejemplo el grupo de beneficiarios) y no para aproximar el impacto sobre un individuo en particular.

En términos formales el objetivo de la evaluación de impacto de un programa es determinar el cambio esperado en Y (los flujos de resultados después de pasar por el programa) que evidencia un beneficiario ($D=1$) con determinadas características X , cuando pasa del estado 0 al estado 1: $E(Y_1 - Y_0 | X, D = 1)$. Dado que al grupo de beneficiarios sólo se le puede observar en el estado de “tratamiento” $E(Y_1 | X, D = 1)$, el principal problema consiste en estimar el estado contra-factual $E(Y_0 | X, D = 1)$. Los distintos métodos de evaluación suelen resolver este problema utilizando un grupo de comparación o control ($D=0$), con individuos que no han participado en el programa y que poseen similares características X . De hecho, al grupo de control sólo se le puede observar en el estado de “no tratamiento” $E(Y_0 | X, D = 0)$.

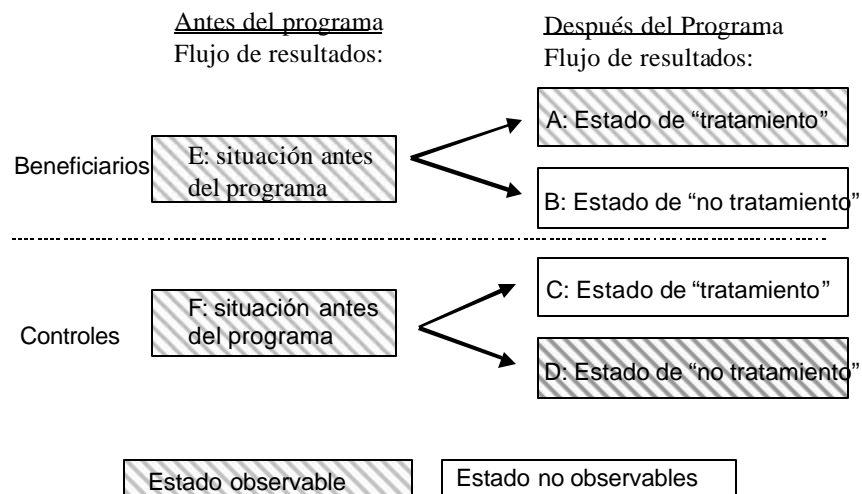
Existen tres estimadores ampliamente usados en la literatura sobre evaluación para medir el impacto de un programa: estimador antes-después, estimador de diferencia en diferencias (también conocido como doble diferencia) y estimador de corte transversal. Para un mejor entendimiento de estos tres estimadores usaremos la Figura 2. En ella se representa el “flujo de resultados” para el grupo de beneficiarios y el grupo de control en dos momentos del tiempo: antes del programa y después del programa (puede pensarse que ambos momentos abarcan uno o más períodos). Asimismo se representa el flujo de resultados en dos estados de la naturaleza (para el momento después del programa): estado de tratamiento y estado de no-tratamiento. La nomenclatura utilizada es la siguiente:

A: “flujo de resultados” de los beneficiarios después del programa en estado de tratamiento

B: “flujo de resultados” de los beneficiarios después del programa en estado de no-tratamiento

- C: “flujo de resultados” de los controles después del programa en estado de tratamiento
- D: “flujo de resultados” de los controles después del programa en estado de no-tratamiento
- E: “flujo de resultados” de los beneficiarios antes del programa
- F: “flujo de resultados” de los controles antes del programa

Figura 2
Estados posibles para beneficiarios y controles



De esta representación debe quedar claro que el impacto sobre el grupo de beneficiarios es igual a A-B. Si lo que se quiere medir es el impacto sobre los controles entonces queremos encontrar C-D. En lo que sigue nos concentramos en el impacto sobre el grupo de beneficiarios.

El estimador antes-después. Resulta de comparar la situación de los beneficiarios del programa antes del programa (E) con su situación después de haber pasado por el programa (A). En este caso se usa como grupo de control a los mismos beneficiarios en su situación anterior al programa, siendo el estimador de impacto el equivalente a A-E. El supuesto detrás de la utilización correcta de este estimador es que (E) logra aproximar bien la situación de los beneficiarios en caso no hubieran participado en el programa (B). El principal problema de este estimador radica en que típicamente, la simple comparación antes y después puede llevara a atribuir erróneamente al programa cambios que se hubieran dado en ese grupo de beneficiarios independientemente de su participación en el programa.

Una explicación formal: Este método supone que se cuenta con i) data longitudinal de los resultados (Y) mostrados por los participantes antes y después del programa, o ii) data de la misma población en distintos periodos en el tiempo donde al menos uno de estos periodos es anterior al programa. Sea Y_{1t} el resultado post-programa de una persona que participa en el mismo e $Y_{0t'}$ el resultado pre-programa de esta persona. Asumiendo que la participación en el programa sólo ocurre en el periodo k , donde $t > k > t'$, y que $Y_{0t'}$ es una buena aproximación del resultado post-programa de los beneficiarios en una situación de “no-tratamiento”, es decir

$$E(Y_{0t} - Y_{0t'} | D = 1) = 0 \quad (\text{S.1}),$$

el estimador antes-después viene dado por: $(\bar{Y}_{1t} - \bar{Y}_{0t'})_1$ donde el subíndice “1” indica que $D=1$ y “-”denota la media muestral. Nótese que a nivel individual, el impacto del programa puede expresarse de la siguiente manera: $Y_{1t} - Y_{0t} = (Y_{1t} - Y_{0t'}) + (Y_{0t'} - Y_{0t})$ donde el segundo término $(Y_{0t'} - Y_{0t})$ se conoce como el error de aproximación. Si este término tiende a promediar cero, entonces se puede estimar el impacto del programa a través de la diferencia entre la media del resultado post-programa y la media del resultado pre-programa de los participantes. Estas medias, naturalmente, pueden definirse para distintas características de los participantes (X).

Dado que se trabaja con medias y se asume que el promedio del error de aproximación tiende a cero, este estimador, como se mencionó líneas arriba, también se puede construir a partir de información sobre una misma población (no necesariamente conformada por los mismos individuos) en distintos momentos en el tiempo. En este sentido, la ventaja de utilizar este estimador es que sólo se requiere de información de los participantes antes y después del programa. Sin embargo, la principal desventaja es el supuesto de que el error de aproximación tiende a promediar cero. Bajo este supuesto, el resultado promedio de los participantes en el estado de “no tratamiento” es el mismo en t y t' , lo cual requeriría que no se hubiesen dado

cambios en el entorno económico o cambios en la posición en el ciclo de vida del cohorte de participantes durante ese lapso, algo difícil de imaginar¹³.

El estimador de diferencia de diferencias. Resulta de comparar las situaciones antes-después de los beneficiarios con aquella de los controles. Así, el impacto del programa se estima mediante la siguiente fórmula $(A-E) - (D-F)$. El supuesto detrás de este estimador es que el cambio en la situación de los controles $(D-F)$ entre el momento previo al programa y el momento posterior al programa es una buena aproximación del cambio que hubiesen experimentado los beneficiarios durante ese mismo período de no haber pasado por el programa $(B-E)$. Una ventaja que se atribuye a este estimador es que si existe un “sesgo de selección” en no observables y éste es constante en el tiempo, entonces la doble diferencia permite limpiarlo.

Una explicación formal: Este estimador supone que se cuenta con i) data longitudinal, o ii) data de los no participantes en el periodo t y t' . De esta manera, si el cambio promedio del resultado experimentado por los participantes en el estado de “no-tratamiento” es igual al de los no participantes, es decir, si la siguiente afirmación es válida:

$$E(Y_{0t} - Y_{0t'} | D = 1) = E(Y_{0t} - Y_{0t'} | D = 0) \quad (S.2)$$

entonces el estimador de diferencias en diferencias que viene dado por:

$$(\bar{Y}_{1t} - \bar{Y}_{0t'})_1 - (\bar{Y}_{0t} - \bar{Y}_{0t'})_0 \quad t > k > t'$$

es válido para estimar $E(\mathbf{D}_t | D = 1) = E(Y_{1t} - Y_{0t} | D = 1)$, donde $\mathbf{D}_t = Y_{1t} - Y_{0t}$, dado que

$$E[(\bar{Y}_{1t} - \bar{Y}_{0t'})_1 - (\bar{Y}_{0t} - \bar{Y}_{0t'})_0] = E(\mathbf{D}_t | D = 1) \quad ^{14}.$$

Adviértase que dado que no se puede observar el paso de un “estado de no-tratamiento” a un “estado de tratamiento” será imposible observar $(Y_{1t} - Y_{0t'})_1 - (Y_{0t} - Y_{0t'})_0$ para cualquier

¹³ Ashenfelter (1978) encuentra, por ejemplo, que es común que los participantes experimenten una caída (transitoria) en sus ingresos antes de inscribirse en un programa de capacitación, hecho que lleva a una sobre-estimación del impacto del programa.

¹⁴ Para demostrar esto basta advertir que: $(\bar{Y}_{1t} - \bar{Y}_{0t'})_1 = (\bar{Y}_{1t} - \bar{Y}_{0t})_1 + (\bar{Y}_{0t} - \bar{Y}_{0t'})_1$

individuo. Pero sí se puede obtener ya sea uno o el otro de estos términos dependiendo de si el individuo es un beneficiario o un control. Ello implica que se requieren supuestos adicionales para obtener una estimación de toda la distribución del impacto a través de este estimador¹⁵. Con el estimador de doble diferencias sólo podremos estimar la media de esa distribución. Debe tenerse en cuenta también que no es indispensable recurrir al mismo individuo en el periodo t y t' , sólo a individuos de una misma población.

La principal desventaja de emplear este estimador, es que no necesariamente la senda que siguen los resultados de los participantes entre el periodo t y t' en el estado de “no tratamiento” es igual a la de los no participantes, se viola el supuesto (S.2)¹⁶.

El estimador de corte transversal. Sólo toma en cuenta la situación de beneficiarios y controles después del programa. Directamente el impacto se estima a través de A-D. Aquí, el supuesto es que D es una buena representación de B. Esta metodología se utiliza en los casos en los que no hay información acerca del programa al inicio del mismo, es decir no se cuenta con una línea de línea de base, y sólo se puede recoger información luego del programa. Nótese que si las situaciones de los beneficiarios y controles son similares en el período anterior al programa, el estimador de diferencia en diferencias es equivalente al estimador de corte transversal.

Una explicación formal: El estimador de corte transversal compara la situación promedio de los beneficiarios y el grupo de control en el periodo t . Dado que no compara a las misma personas, no permite obtener la distribución del impacto, a menos que se efectúen supuestos adicionales. Este estimador asume que:

$$E(Y_{0t} | D = 1) = E(Y_{0t} | D = 0) \quad (S.3)$$

¹⁵ Se debe asumir que $(Y_{1t} - Y_{0t})_1$ es independiente de $(Y_{0t} - Y_{0t'})_1$ y que la distribución de $(Y_{1t} - Y_{0t})_1$ es igual a la distribución de $(Y_{0t} - Y_{0t'})_0$.

¹⁶ La caída transitoria en los ingresos de los participantes observada por Ashenfelter (1978) también es aplicable en este caso.

es decir, que en promedio aquellos que no participan en el programa muestran el mismo resultado de “no tratamiento” que aquellos que sí lo hacen. Bajo este supuesto, el estimador de corte transversal viene dado por: $(\bar{Y}_{1t})_1 - (\bar{Y}_{0t})_0$.

El estimador es válido bajo el supuesto (S.3) debido a que:

$E[(\bar{Y}_{1t})_1 - (\bar{Y}_{0t})_0] = E(\mathbf{D}_t | D = 1)$ ¹⁷. El supuesto (S.3) es válido si la participación en el programa no está relacionada con el resultado post-programa en el estado de “no-tratamiento”. Asimismo, si los cambios en el entorno económico o en la posición en el ciclo de vida de los individuos tienen el mismo efecto sobre los beneficiarios y el grupo de control, este estimador deja de ser vulnerable a los problemas que enfrenta el estimador antes-después.

Es importante señalar que tanto el estimador de diferencia en diferencias como el de corte transversal utilizan como grupo de control a individuos distintos a los beneficiarios. En general, el grupo de control está conformado por individuos que por alguna razón no participan del programa o que si quisieron entrar lo hicieron tarde, o es un grupo explícitamente identificado. Lo crucial es lograr que el grupo de control y el de beneficiarios sean los más similares posibles.

Cabe precisar también que los tres estimadores antes descritos se pueden definir condicionados a determinadas características observables X . Al condicionar en X o en variables “instrumentales” adicionales es más factible satisfacer versiones modificadas de los supuestos (S.1), (S.2) y (S.3), pero esto no está garantizado. Por ejemplo, si la distribución de las características X es distinta entre los participantes ($D=1$) y los no participantes ($D=0$), al condicionar en X se pueden eliminar las diferencias sistemáticas entre los resultados de los dos grupos. Sin embargo, si las diferencias entre los participantes y los no participantes se debe a características no observables, al condicionar en X se puede acentuar, en vez de eliminar, las diferencias entre ambos grupos en el estado de “no tratamiento”¹⁸.

¹⁷ Para demostrar ello se debe realizar la siguiente descomposición:

$$(\bar{Y}_{1t})_1 - (\bar{Y}_{0t})_0 = (\bar{Y}_{1t})_1 - (\bar{Y}_{0t})_1 + (\bar{Y}_{0t})_1 - (\bar{Y}_{0t})_0$$

¹⁸ En este sentido, el hecho que $|E(Y_0 | D = 1) - E(Y_0 | D = 0)| = M$ no significa que $|E(Y_0 | D = 1, X) - E(Y_0 | D = 0, X)| < M$. Para algunos valores de X , la brecha puede ser mayor.

4.6. Métodos de estimación para diseños no experimentales

Existen diversos métodos para estimar el impacto de un programa social cuyo diseño de evaluación corresponda a un diseño no experimental. En particular se revisarán dos de ellos: el método del “matched comparisons” o comparaciones emparejadas y el modelo de selección econométrico.

4.6.1 El método del “matched comparisons”.

Este método consiste en asignar para cada beneficiario un respectivo individuo control. Dicha asignación se realiza sobre la base de emparejar aquel beneficiario con aquel control que sean lo “más parecidos”. Una vez establecidas las parejas de beneficiarios y controles “emparejados” se podrá estimar el impacto promedio del programa como el promedio del impacto para cada pareja. La idea de encontrar un “match” o pareja para cada beneficiario apunta a reducir potenciales “sesgos de selección”, especialmente en variables observables. Existen diversas maneras de realizar el proceso de emparejamiento y de entender la noción del “más parecido” que serán explicadas luego.

Antes, debe entenderse cuáles son las potenciales fuentes de sesgo de un estimador de impacto y cómo el método de “matching” logra evitarlos. Es decir porque un estimador puede sub-estimar o sobre-estimar el verdadero impacto medio de un programa. Si se considera el estimador de corte transversal el sesgo de estimación viene dado por:

$$B = E(Y_{0t} | D = 1) - E(Y_{0t} | D = 0)$$

Así, en la medida que no se cumpla el supuesto que para el grupo de control (no beneficiarios) el valor esperado de sus resultados en “estado de no-tratamiento” sea igual al valor esperado de los resultados del grupo de beneficiarios en similar estado, nuestro estimador no apuntará correctamente hacia el verdadero efecto medio del programa sobre los beneficiarios.

Ese sesgo puede deberse a tres fuentes. En primer lugar a la ausencia de “soporte común”, ello quiere decir al hecho de que las características observables de beneficiarios y controles no están comprendidas en rangos similares. Así por ejemplo, si nuestro grupo de beneficiarios presenta edades que fluctúan entre los 15 y 30 años y en nuestro grupo de control esas van de los 21 a los 35 años, vamos a decir que no hay soporte común. El método de “matching” al buscar para cada beneficiario un control “muy parecido” logra evitar este problema. Formalmente ello implica condicionar el estimador de impacto a un set común de valores para las características X . Una consecuencia directa de la comprensión de esta fuente de sesgo es que cuando no exista un control suficientemente parecido para un beneficiario resulta mejor eliminar a ese beneficiario de la muestra y del proceso de evaluación. En nuestro ejemplo ello llevaría a descartar a los beneficiarios entre 15 y 20 años y entre los 31 y 35 años. Sin embargo la desventaja de ello es que nuestro estimador de impacto será sólo válido para aquel grupo de beneficiarios sólo entre los 21 y 30 años, no para el grupo completo.

Una segunda fuente de sesgo en el estimador de impacto proviene de potenciales “sesgos de selección” en la participación del programa debido a variables no observables. Para corregir este problema pueden combinarse técnicas de econométricas de corrección de sesgo de selección con el proceso de emparejamiento.

Una tercera fuente de sesgo en el estimador de impacto proviene del hecho de que la distribución de las características observables (X) para beneficiarios sea diferente de la distribución para los controles. En nuestro ejemplo, puede que hayamos logrado un soporte común pero aún así es posible que la proporción de individuos entre 21 y 25 años dentro del grupo de beneficiarios sea distinta de la proporción que muestra los controles. Para minimizar este sesgo se han propuesto algunos métodos de emparejamiento en el cuál se construyen controles ficticios como promedios ponderados del grupo de controles disponibles

A continuación se revisan algunos métodos particulares de “matching”.

Minimización de Distancia Euclídeana. Se asigna a cada beneficiario aquel control que tenga la mínima distancia euclídeana¹⁹ entre sus características y las del control. En este caso más parecido significa la menor distancia euclídeana. Operativamente, primero se estandarizan (restar el promedio y dividir por el desvío estándar) las variables que representan aquellas características en las cuales queremos que beneficiarios y controles sean parecidos; ello evita el problema de trabajar con variables que se encuentre en distintas unidades de medida. En segundo lugar se computa la distancia entre un beneficiario y cada uno de los controles disponibles. Se asigna a como control para ese beneficiario al control que arroje la menor distancia. Se repite el mismo procedimiento para cada uno de los beneficiarios.

Minimización de Distancia Mahalanobis. Una variante a la medición de distancias euclídeanas consiste en utilizar como métrica de distancia la propuesta por Mahalanobis, que es similar al de la distancia euclídeana pero además de corregir por las varianzas, lo hace también por la covarianzas entre las variables. Es superior al método de las distancias euclídeanas en un doble sentido: en primer lugar no hace falta tomar en cuenta el problema de las distintas unidades de medida porque la corrección es hecha automáticamente y, por otro lado, corrige por la posible correlación entre las variables consideradas. Ahora bien, como sucede muy a menudo, dichas ventajas traen un costo, en este caso computacional: es necesario computar la matriz de varianzas-covarianzas de las variables consideradas en el emparejamiento y luego realizar transformaciones lineales sobre las variables originales²⁰.

El método del “propensity score”. El “propensity score” es una medida que trata de establecer el grado en que un individuo cualquiera hubiese tendido a participar en el programa que se está evaluando. Así, se utiliza un modelo de regresión del tipo “probit” o “logit” para estimar la propensión a participar en el programa utilizando la muestra disponible de beneficiarios y controles. Como variable dependiente se usa una variable dicotómica que toma valor 1 si se trata de un beneficiario y valor 0 si se trata de un control y como variables del lado derecho de la

¹⁹ La distancia euclídeana es la manera tradicional de estimar la distancia entre dos puntos en un espacio determinado:

$$d = \sqrt{(x_A - x_B)^2 + (y_A - y_B)^2}$$

²⁰ $|I| = (X_i - X_j)' \mathbf{\hat{\alpha}}^{-1} (X_i - X_j)$

regresión se utilizan todas aquellas variables en las que se desea similitud entre beneficiarios y controles y aquellas que puedan afectar la decisión de participar o no en el programa. Una vez estimada la regresión se computa para cada individuo la “propensión” predicha por el modelo, es decir el llamado “propensity score”. Luego se asigna como pareja de un beneficiario a aquel control que posea el “propensity score” más cercano al de ese beneficiario.

Una variante en los métodos de emparejamiento consiste en construir controles ficticios. Así, ya sea que se haya usado el método de distancias euclidianas o “propensity score” se sugiere utilizar como control de un beneficiario el promedio de los 5 o 10 controles más cercanos a él. Así, se puede obtener un promedio para la(s) variable(s) resultado de los controles más cercanos. Otra modificación es construir la pareja de un beneficiario a partir de promedios ponderados de todos los controles disponibles. En este caso se dará mayor ponderación al control más cercano y la menor ponderación al control más lejano. El valor de las ponderaciones se puede determinar a partir de lograr estimaciones de la distribución de las distancias entre un beneficiario dado y todos los controles.

Una vez hecho el proceso de emparejamiento o “matching”, es decir una vez que se cuente ya con una pareja (verdadera o ficticia) para cada beneficiario, se puede estimar el estimador de impacto de corte transversal o el de doble diferencias.

4.6.2 *El modelo de Estimación Econométrica.*

Considérese el siguiente modelo para representar el impacto de participar en un programa²¹:

$$Y_{it} = c_t X_i + b_t P_{i0} + u_{it}, \quad t > 0$$

$$P_{i0} = a_0 Z_i + e_{i0}$$

Donde Y_{it} es la variable “resultado” de interés para el individuo “ i ” sobre la cual se desea medir el impacto del programa (por ejemplo ingresos), medida en el momento “ t ” (post-programa). X_i y

²¹ Friedlander, Greenberg and Robins (1997). “Evaluating Government Training Programs for the Economically Disadvantaged”. *Journal of Economic Literature*. Vol XXXV.

Z_i representan un conjunto de variables (algunas pueden ser similares) exógenos y características personales de los individuos. P_{io} representa una variable binaria, la cual toma valor cero para los individuos que no participan del programa y valor uno para los beneficiarios del mismo; u_{it} y e_{it} constituyen términos de error. Bajo esta formulación el efecto del programa puede ser medido a través de la estimación del parámetro b_t .

Sin embargo, como se ha mencionado previamente, en el caso de estudios de evaluación no-experimentales debe tenerse en cuenta el “problema de selección” del grupo de beneficiarios. Así, en el caso que $E(P_{io}, u_{it}) = 0$ será posible estimar el parámetro de interés b_t mediante una regresión de Y en X y P . En caso contrario esta última estimación arrojará un estimador sesgado.

La correlación entre P_{io} y u_{it} puede deberse a que $E(Z_i, u_{it}) \neq 0$ pero $E(u_{it}, e_{it}) = 0$, ó a que $E(Z_i, u_{it}) = 0$ pero $E(u_{it}, e_{it}) \neq 0$. En el primer caso se estaría en presencia de un “sesgo de selección” debido a variables observables. Para resolver este problema se podría incluir Z_i como regresores adicionales en la primera ecuación o alternativamente acudir al método de “matching” para el eliminar este tipo de sesgo. En el segundo caso el sesgo de selección proviene por variables no observables. Una manera de abordar este problema es a través del uso del “ratio inverso de Mills”²². Ello supone correr una regresión primero para explicar la participación en el programa, teniendo cuidado de incluir entre los regresores alguna variable que explique la participación pero que no explique la variable “resultado” que se está evaluando. Luego se puede construir el ratio “inverso de Mills” para cada individuo, que es un indicador que aporta información sobre las características no observables de los individuos. Posteriormente, se puede estimar una ecuación sólo para la muestra de no beneficiarios para predecir la variable “resultado” (por ejemplo, ingresos) empleando el ratio inverso de mills como un regresor adicional. Finalmente, para cada beneficiario se puede estimar el impacto como la resta del resultado efectivamente alcanzado por él menos una predicción de su resultado utilizando los parámetros estimados mediante el último modelo de regresión.

²² Ver Maddala (1999), capítulo 9

4.7. Descomposición de las Mediciones de Impacto: Efectos en las Características de los Beneficiarios Vs. Efectos en los Retornos a Dichas Características.

Habiendo estimado el impacto del programa por cualquiera de los enfoques previamente mencionados, es posible desagregar el mismo en dos componentes: uno atribuible a cambios en las características individuales de los beneficiarios y el otro a cambios en los retornos que estas características tienen. Así por ejemplo, siguiendo con el caso ProJoven, uno de los principales cambios en las características observables de los beneficiarios se observaría en la capacitación ocupacional que ellos obtienen al participar del programa, pero a su vez, esta capacitación ocupacional adicional podría generar una mejora en los retornos a la experiencia que ellos tengan.

Para esto, se hace necesario estimar por separado “ecuaciones de cambio” en la variable de impacto tanto para los beneficiarios como para sus respectivos controles. Esto es, se necesita estimar un modelo econométrico en el que la variable explicada sea el cambio en la variable resultado del individuo, explicada en función de una serie de características individuales relevantes.

$$\mathbf{DY}_{it}^1 = c_t^1 X_i^1 + u_{it}^1,$$
$$\mathbf{DY}_{it}^0 = c_t^0 X_i^0 + u_{it}^0,$$

De esta manera obtenemos un estimador de “retornos”, c , para los beneficiarios y otro para los controles, así como estimadores de las características medias de ambos grupos. Bajo el supuesto habitual de media cero de los errores, las ecuaciones se convierten en

$$\overline{\mathbf{DY}_{it}^1} = \hat{c}_t^1 \overline{X_i^1},$$
$$\overline{\mathbf{DY}_{it}^0} = \hat{c}_t^0 \overline{X_i^0}.$$

Entonces, el estimador de diferencias en diferencias puede ser descompuesto utilizando la conocida técnica de Blinder-Oaxaca,

$$\overline{\mathbf{DY}}_{it}^1 - \overline{\mathbf{DY}}_{it}^0 = (\hat{c}_t^1 \overline{X}_{it}^1 - \hat{c}_t^0 \overline{X}_{it}^0),$$

$$\overline{\mathbf{DY}}_{it}^1 - \overline{\mathbf{DY}}_{it}^0 = \hat{c}_t^1 (\overline{X}_{it}^1 - \overline{X}_{it}^0) + (\hat{c}_t^1 - \hat{c}_t^0) \overline{X}_{it}^0.$$

Así, el estimador de diferencias en diferencias para el impacto del programa es desagregado en dos componentes aditivos, uno explicando cambios en las características individuales de los participantes, x , y el otro cambios en los retornos a dichas características, c .

4.8. Evaluación Social de los Programas. Analisis de Equilibrio: Parcial vs. General.

La metodología de evaluación propuesta en los párrafos anteriores puede ser considerada como una de “análisis de equilibrio parcial”. Ello se debe al hecho de que se concentra en la evaluación de los beneficiarios y los correspondientes controles (o conterfactuales) sin tomar en consideración la manera en que ellos interactúan con el entorno económico. Por ejemplo, como se verá más adelante en el caso de ProJoven, se evalúan los cambios en ingresos, inserción laboral y segregación ocupacional de los beneficiarios y controles sin tomar en consideración que ProJoven podría generar los incentivos necesarios para alterar la demanda por mano de obra en los segmentos del mercado donde opera. ¿Cómo afecta ese cambio a los no participantes en el programa? ¿Cómo afecta a la economía en general? Típicamente, se presume que estos efectos son mínimos y por lo tanto se deja de lado la cuantificación de los mismos.

En general, todo programa social tiene un efecto directo sobre los beneficiarios del mismo, pero también tendrá efectos indirectos sobre otros agentes de los mercados donde tales beneficiarios operan, generando así reacciones que afectarán a su vez las relaciones económicas de los mismos en el futuro. Analizar los efectos de un programa social en un contexto en el que se presume que existen efectos indirectos no despreciables requiere un enfoque de equilibrio general, para el cual es necesario afinar y complementar las herramientas econométricas presentadas aquí.

La descomposición de los efectos del programa, en términos de cambios en las características individuales y cambios en los retornos a estas características, presentada en la sub-sección anterior, podría ser considerada como un primer paso hacia una estimación de equilibrio general, en el sentido que los cambios a los retornos a las características implican una interacción en el mercado laboral. Sin embargo, un verdadero enfoque de equilibrio general requiere de un nivel de sofisticación aun mayor en la modelación de esos cambios en los retornos, pues en realidad corresponden a cambios en las interacciones del mercado. En el ejemplo de ProJoven que se maneja en este trabajo, éstos corresponden a los cambios en los patrones de demanda por mano de obra por parte de los empleadores que eventualmente pudieran generarse por efecto del programa.

Finalmente, puede argumentarse sin embargo que, dadas las dimensiones típicas de este tipo de programas sociales, sus efectos sobre el mercado serán muy pequeños en comparación con los efectos directos sobre los beneficiarios. La validez de un argumento de este tipo aboga en favor de la utilización de análisis de equilibrios parciales en la evaluación de impacto. La complejidad adicional que trae el análisis de equilibrio general no aportaría nuevas luces en la medición de los efectos de los programas.

5. Aplicación de evaluación de impacto para PROJoven

5.1. Antecedentes del programa PROJoven

El Programa de Capacitación Laboral Juvenil (PROJoven) se creó a mediados de 1996 por iniciativa del Ministerio de Trabajo y Promoción Social como respuesta a la precaria situación laboral de los jóvenes de menores ingresos en el país. Así, se establecen como objetivos fundamentales de PROJoven mejorar las oportunidades de empleo de jóvenes de escasos recursos y dinamizar el sistema de capacitación en el país, mejorando su calidad técnica y pertinencia. En términos prácticos, se espera aumentar la tasa de inserción en el mercado laboral de los beneficiarios del programa y sus niveles de ingreso. Asimismo, se espera que los beneficiarios empleen un menor tiempo en la búsqueda de empleo y que haya una menor segregación ocupacional.

A inicios de 1999, PROJoven obtuvo el primer puesto en un Concurso de Proyectos de Capacitación Laboral convocado por el Fondo de Capacitación para el Empleo (FONDOEMPLEO). Gracias a ello, este programa alcanzó el financiamiento necesario para ampliar sus actividades en departamentos del interior del país.

El programa consiste básicamente en financiar la capacitación teórica y práctica de un grupo de beneficiarios pertenecientes a una determinada población objetivo. En este sentido, PROJoven convoca a entidades de capacitación (ECAPs) para que organicen e implementen cursos diseñados en función de las necesidades del sector productivo y de las características de los beneficiarios. Previamente las ECAPs pasan por un proceso de selección que toma en cuenta su experiencia, su capacidad de gestión y los recursos humanos con los que cuenta. Así, las entidades seleccionadas pasan a formar parte del registro de entidades de capacitación (RECAP) y se encargan de presentar propuestas de cursos en las licitaciones que el programa convoca. Los cursos seleccionados son aquellos cuyo ratio precio-calidad técnica resultan los más adecuados para el programa. Cabe destacar que para asegurar la pertinencia de los cursos capacitación, se

deben presentar adicionalmente cartas de compromiso de empresas mediante las cuales se asegure las prácticas laborales de los beneficiarios.

La formación técnica de los beneficiarios, desarrollada directamente en las ECAPs, tiene una duración promedio de 3 meses y es financiada en su integridad por PROJoven. Incluye además una subvención mensual por gastos de movilidad, refrigerio y un seguro médico. Luego de la fase teórica, los beneficiarios realizan prácticas laborales en empresas bajo la modalidad contractual de convenio de formación laboral juvenil. Esta fase práctica tiene una duración mínima de tres meses, durante la cual las propias empresas pagan una subvención a los beneficiarios.

La población objetivo de este programa está compuesta por jóvenes de ambos sexos, entre 16 y 25 años de edad, con bajos niveles de escolaridad, pertenecientes a hogares pobres o en extrema pobreza, y que se encuentran desempleados, subempleados o en actividad forzosa. A través de un elaborado proceso de acreditación se asegura el cumplimiento de estos requisitos en el grupo de beneficiarios del programa. El programa, por su parte, se encarga de asignar a los beneficiarios entre los cursos disponibles mediante un mecanismo que combina información de los cursos de capacitación disponibles y las preferencias de los beneficiarios.

Desde el inicio de su ejecución en 1997, PROJoven ha realizado a la fecha nueve convocatorias (actualmente se encuentra ejecutando la novena), alcanzando a beneficiar a un poco más de 30 mil jóvenes de bajos ingresos en diversas ciudades del país. Asimismo, son más de 300 las entidades de capacitación, entre públicas y privadas, que han participado en el programa.

Cabe señalar además que PROJoven es el primer programa social que ha incorporado, desde su diseño inicial, un proceso de evaluación sistemático y riguroso. Ello ha significado la recolección de información para grupos de beneficiarios y de control a partir de la cual se han venido realizando algunas evaluaciones para las diversas convocatorias.

Para el ejercicio de evaluación aquí desarrollado se utiliza la información de los beneficiarios y grupo de control de la sexta convocatoria del programa, ejecutada entre fines de 1999 e inicios del 2000.

5.2. La información disponible

Se cuenta con información de jóvenes que han participado en la sexta convocatoria del programa, “beneficiarios”, y jóvenes que no lo han hecho, que serán llamados “controles”. Dicha información ha sido recolectada tanto al inicio del programa, la cual constituye una línea de base, como después de seis meses de terminado el programa. Esta última información se conoce como la primera medición de salida. A los 12 meses y 18 meses de culminado el programa correspondiente a la sexta convocatoria se realizarán la segunda y tercera medición de salida²³.

Sobre la Línea de Base. Dado el diseño del programa, la evaluación utilizó una metodología no experimental, por lo que la línea de base se estableció con una muestra de beneficiarios del programa y un grupo de individuos seleccionados en base a un conjunto de criterios y que servirían como grupo de control. Se colectó información en marzo y abril del 2000, justo antes del inicio de los cursos de la sexta convocatoria de PROJoven. La información recogida en la construcción de la línea de base, tanto para los beneficiarios como el grupo de control, comprende principalmente datos personales y familiares del individuo y preguntas sobre su situación laboral.

Los entrevistados son una muestra de 1,034 beneficiarios²⁴ del total de beneficiarios matriculados en la sexta convocatoria, 3,586 jóvenes de quienes se recogió información a través de una encuesta sociolaboral. Se utilizó un muestreo aleatorio estratificado. Así, el universo analizado

²³ En el caso de la segunda medición, ya se ha realizado el trabajo de campo pero la información aún no se encuentra disponible en versión electrónica.

²⁴ De estos 1043 jóvenes beneficiarios 949 son titulares y 94 pertenecen a un grupo de reserva. La idea de contar con un grupo de reserva en la línea de base surge por la presunción de que en mediciones posteriores existirán algunos beneficiarios que, por razones diversas, no podrán ser entrevistados nuevamente. En ese caso se reemplazará ese beneficiario por uno del grupo de reserva.

fue estratificado por ciudades, género y dos rangos de edades (de 16 a 20 años y de 21 a 25 años de edad).

**Universo y muestra de beneficiarios en línea de base
(Distribución por ciudades)**

Ciudad	Universo Beneficiarios	Muestra de beneficiarios	
		Grupo Principal	Grupo Reserva
Lima	2,586	346	35
Trujillo	345	185	19
Arequipa	373	193	20
Chiclayo	154	111	11
Cuzco	128	114	9
Total	3,586	949	94

Fuente: PROJoven

Asimismo, en la elaboración de la línea de base se obtuvo información para los individuos que constituirían el grupo de control, a los cuales se les aplicó una encuesta socio-laboral similar a la utilizada para los beneficiarios. Los jóvenes para el grupo de control se eligieron de los lugares de origen de la muestra de beneficiarios establecida previamente. Una vez conocidos los lugares de procedencia de los beneficiarios se realizó una encuesta de POTENCIALES integrantes del grupo de control, es decir, individuos “elegibles” para PROJoven pero que no participan en él. Para ello se ubicó la vivienda del beneficiario y a partir de allí se "peinaba" la cuadra o manzana y se ubicaban a los posibles integrantes a través de 5 variables de decisión (sexo, edad, situación laboral, educación y puntajes de pobreza) las cuales debían de coincidir ANTES de iniciar la participación en el programa. Si más de un potencial control de la cuadra coincidía con el beneficiario en estas variables de decisión se procedía a seleccionar aleatoriamente al respectivo control. Si no se ubicaba a un control en la misma cuadra la búsqueda se hacía en las cuadras cercanas, teniendo como límite inicial el mismo distrito, pasando en casos extremos a otro distrito con similar puntaje de pobreza. Todos los controles provienen de zonas en donde existen beneficiarios.

En total, se dispone de un total de 1,561 controles seleccionados bajo los procedimientos antes descritos. Existen más controles que beneficiarios porque originalmente se consideraron tres potenciales controles para cada beneficiario.

Encuesta de salida: primera medición

Después de seis meses de terminados los cursos y prácticas correspondientes a la sexta convocatoria del programa entre los meses de mayo y junio del 2001 se realizó la primera medición. Se re-entrevistaron a los beneficiarios y controles previamente encuestados en la línea de base utilizando el mismo cuestionario sobre información sociolaboral. El total de beneficiarios y controles para los cuales existe esta información se resume en la siguiente tabla.

**Distribución de individuos con información disponible
en Línea de Base y Primera Medición
Projoven - 6ta Convocatoria**

Ciudad	Beneficiarios		Grupo de control	
	Número	Porcentaje	Número	Porcentaje
Arequipa	206	20.2	306	19.6
Chiclayo	122	12.0	195	12.5
Lima	372	36.5	561	35.9
Cuzco	116	11.4	193	12.4
Trujillo	202	19.8	306	19.6
Total	1018	100.0	1561	100.0

Fuente: PROJoven

En total se dispone de 1,018 beneficiarios y 1,561 controles a partir de los cuales se puede estimar el impacto del programa PROJoven en diversas variables de interés.

En la siguiente tabla se muestra la distribución de las principales características de los beneficiarios y controles disponibles.

**Características de las Muestras de Beneficiarios y Controles
PROJoven - 6ta Convocatoria**

	Beneficiarios	Controles
Número	1018	1561
Distribución según ciudad (%)		
Arequipa	20.2	19.6
Chiclayo	12.0	12.5
Lima	36.5	35.9
Cuzco	11.4	12.4
Trujillo	19.8	19.6
Participación femenina (%)	52.1	51.7
Distribución según edad (%)		
15 – 17 años	24.2	20.3
18 – 19 años	28.9	30.1
20 – 21 años	21.4	22.4
22 – 23 años	16.8	18.6
24 – 25 años	8.7	8.6
Distribución según nivel educativo (%)		
Sin educación	0.2	0.0
Primaria incompleta	0.5	0.6
Primaria completa	2.5	2.6
Secundaria incompleta	16.9	18.1
Secundaria completa	80.0	78.7
Distribución según situación laboral (%)		
Inactivo	16.1	18.6
Desempleado	18.8	19.1
Ocupado	65.1	62.3
Nivel de pobreza		
Promedio	16.69	16.73
Desviación Estándar	4.13	3.46

Fuente: PROJoven

5.3. Análisis comparado de los distintos estimadores de impacto obtenidos por medio de la técnica de emparejamiento.

En esta sección se analiza el desempeño empírico de los estimadores propuestos previamente. Se hace un análisis de impacto para diferentes variables de interés para el programa: ingresos mensuales, ingresos por hora, inserción laboral y segregación ocupacional según género. Para esto, se comparan los resultados obtenidos a partir de cuatro diferentes emparejamientos

practicados sobre la muestra original de beneficiarios y controles. Estos cuatro emparejamientos responden a variaciones en las especificaciones econométricas de las técnicas que la literatura provee. A continuación una breve descripción de los mismos.

En primer lugar, se procedió al computo de los “propensity scores” por medio de un modelo dicotómico (probit) utilizando la información de todas las observaciones de las muestras de controles y beneficiarios. Para esto, la participación en la sexta convocatoria de ProJoven se modeló como una variable aleatoria que toma el valor 1 para los beneficiarios y 0 para los controles. En el lado derecho de la ecuación se consideró una serie de características individuales, familiares y del hogar: género, edad, nivel educativo, situación laboral, una dummy para controlar si el individuo tiene hijos o no, una dummy para controlar por la existencia de ocupación secundaria, dummies con información de los niveles educativos de los padres, una dummy para controlar si el individuo asistió previamente a algún curso técnico superior, una variable construida por ProJoven para medir la situación de pobreza del hogar (basada en necesidades básicas insatisfechas) y finalmente se consideró el ingreso per capita del hogar (excluyendo al individuo en cuestión).

Habiendo estimado estos “propensity scores” se procedió a explorar los distintos criterios de emparejamiento, emparejando a nivel de ciudad con la finalidad de evitar la posibilidad de emparejar individuos que enfrentan mercados laborales distintos (en ciudades distintas).

El primer criterio de emparejamiento seleccionó, sin reemplazo, un control para cada beneficiario, minimizando las diferencias de propensity scores entre beneficiarios y controles.²⁵ Un segundo criterio de emparejamiento, muy similar al primero, consideró la selección de los controles con reemplazo. De esta manera se abrió la posibilidad de incrementar el tamaño de la muestra al permitir que un control sea pareja de más de un beneficiario.

²⁵ En aquellas situaciones donde la mínima diferencia de propensity scores entre un beneficiario y los controles fuese mayor que cierto nivel de tolerancia (caliper), el beneficiario en cuestión ha sido descartado de la muestra.

En estos dos primeros criterios de emparejamiento se utiliza una ligera variante a lo que la literatura tradicional recomienda (el cual funcionó como mecanismo de control adicional): además de buscar que la diferencia en “propensity scores” entre beneficiarios y controles emparejados fuera muy pequeña, establecimos criterios adicionales de emparejamiento en algunas de las características individuales. Estos criterios adicionales fueron: que un beneficiario y su respectivo control difieran a lo más en un año de edad, que sean del mismo sexo, que tengan el mismo nivel educativo, que se encuentren en la misma situación laboral y que se encuentren en el mismo quintil para la variable de situación de pobreza construida por ProJoven. Así, se consiguió una muestra de 842 pares beneficiario-control con el primer criterio de emparejamiento y 893 pares con el segundo.

Para un tercer criterio de emparejamiento se prescindió del mecanismo de control adicional descrito en el párrafo anterior y, haciendo un re-muestreo con reemplazo de los controles se consiguieron 884 pares beneficiario-control.

En un cuarto criterio de emparejamiento se dio mayor énfasis en las características mismas que en los propensity scores. Así, el nuevo criterio consistió en emparejar beneficiarios y controles que sean semejantes en cada una de las variables que fueron consideradas en el cálculo de los propensity scores. Esto es lo que en la literatura se conoce como “emparejamiento en características” y es presentado como una alternativa al emparejamiento en propensity scores. Es una alternativa superior a propensity scores en el sentido que consigue emparejar individuos que sean exactamente iguales en todas las características (y no solo en una combinación lineal de ellas); pero por otro lado resulta una alternativa inferior porque consigue un número de parejas sustancialmente menor (lo que a su vez implica mayores errores estándares del estimador medio). Bajo este criterio de emparejamiento en características conseguimos 350 pares beneficiario-control.

A continuación, se presentan en la siguiente tabla un resumen con los criterios de cada emparejamiento y el número de parejas resultante de cada uno.

ProJoven 6ta Convocatoria
Comparación de Criterios de Emparejamiento

	Emparejamiento			
	1	2	3	4
Emparejamiento por "Propensity Scores"	Si	Si	Si	No
Emparejamiento por Características	Si	Si	No	Si
Criterio de Emparejamiento Predominante	Propensity	Propensity	Propensity	Características
Muestreo con Reemplazo	No	Si	Si	Si
Número de Parejas	842	893	884	350

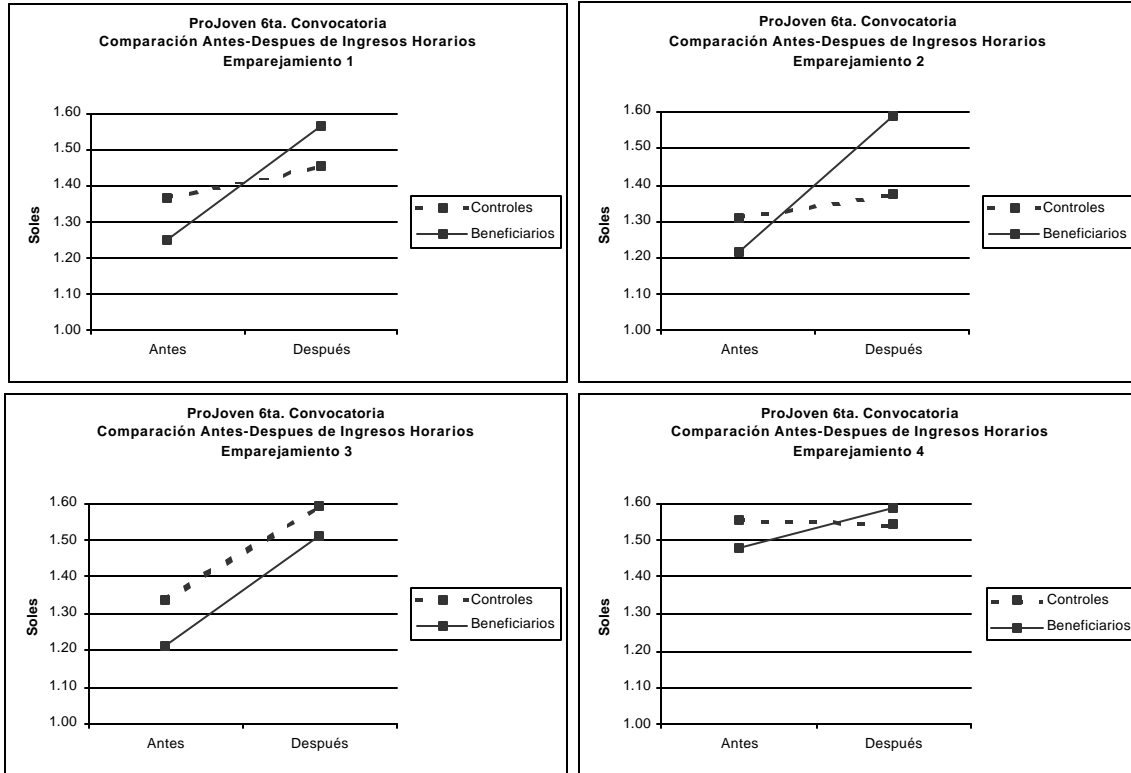
En lo que sigue, se describen los resultados que cada uno de estos emparejamientos arroja. Los resultados serán mostrados para los cuatro emparejamientos de manera paralela. Esto facilitará la comparación de resultados y también, de alguna manera, servirá para lograr una idea (informal) acerca de la robustez de los parámetros estimados. Con el ánimo de presentar un análisis de confiabilidad de los estimadores, también se reportan los p-values²⁶ obtenidos a partir de bootstrap²⁷.

a) Impacto sobre los ingresos por hora

De acuerdo a la literatura, la primera variable a analizar (o la más importante) es el efecto del programa sobre los ingresos por hora. Para ello se han computado, en primer lugar, los estimadores de la media de ingresos por hora para los beneficiarios y controles, antes y después de la aplicación de la sexta convocatoria de ProJoven. Los resultados para los cuatro emparejamientos considerados en este análisis empírico se muestran en el siguiente panel.

²⁶ Un p-value es el nivel de confianza para el cual no se puede rechazar la hipótesis nula que establece “el estimador es significativamente positivo”.

²⁷ Bootstrap es una de las técnicas de remuestreo que han cobrado popularidad gracias al uso intensivo de herramientas computacionales. En este caso, repetimos 10,000 veces un proceso de selección aleatoria de muestras de beneficiarios, con sus respectivos controles. Véase Horowitz(2000).



Con estos estimadores de tendencia central es posible observar que, intuitivamente en primer lugar, el crecimiento de los ingresos por hora de los beneficiarios es mayor al crecimiento de los ingresos por hora de los controles en todos los criterios de emparejamiento. Un siguiente paso es la estimación precisa de los valores de los estimadores para los tres tipos contemplados en el desarrollo teórico del presente proyecto. La siguiente tabla resume esos resultados, computados como porcentajes de los ingresos promedio de los beneficiarios antes de participar en el programa. Los números más pequeños debajo de cada estimador corresponden a los p-valores de la hipótesis nula que dice: “el efecto es significativamente positivo”.

ProJoven 6ta Convocatoria
Comparación de Estimadores de Impacto Sobre Ingresos Por Hora

	Emparejamiento 1	Emparejamiento 2	Emparejamiento 3	Emparejamiento 4
Estimador Diferencia de Diferencias	18.6% 0.9919	25.7% 0.9994	3.5% 0.6233	7.9% 0.6901
Estimador Antes-Después	25.4% 0.9996	30.7% 1.0000	24.9% 0.9953	6.9% 0.7313
Estimador de Corte Transversal	7.8% 0.8989	15.9% 0.9885	-5.2% 0.1627	2.8% 0.6105

El estimador de diferencia en diferencias asume que los “choques aleatorios” que pudieran haber sufrido los beneficiarios en el periodo comprendido entre las dos mediciones, y que hubiesen podido afectar sus ingresos por hora, son los mismos que han sufrido los controles. De acuerdo a esta medida, el impacto de la sexta convocatoria de ProJoven puede ser tan pequeño como del orden del 3.5% como tan grande como 25.7% para los tres primeros emparejamientos, pero del orden del 7.9% para el último, que es más restrictivo.

Los estimadores antes-después, que ignoran la posibilidad de observar cambios en los ingresos de los controles, arrojan resultados sustancialmente mayores para los tres primeros emparejamientos (en el rango de 25%-31%), pero en el cuarto emparejamiento reporta una cifra muy similar a la que reporta el estimador de diferencia de diferencias (6.9%).

Los estimadores de corte transversal, que asumen un emparejamiento perfecto antes del programa, son los que arrojan los resultados más bajos en términos de impacto de ProJoven. Esto sucede porque se ignora una regularidad empírica muy importante encontrada en los datos de ProJoven y que es común en este tipo de programas: antes de participar en el programa, los beneficiarios de ProJoven muestran ingresos por hora sustancialmente por debajo de los ingresos por hora de los controles, pese al riguroso control en características observables hecho a nivel del individuo, del hogar y de la ubicación geográfica. Este problema, que la literatura conoce como el “Ashenfelter’s Dip”, implica sesgos en la estimación de los impactos reales de los programas. Un tratamiento adecuado para remover estos sesgos requiere de un levantamiento

adicional de información acerca de la historia de ingresos de los individuos (tanto beneficiarios como controles).

b) Impacto sobre la inserción laboral

Considerando que la población objetivo de ProJoven forma parte de la cohorte con más alta rotación laboral en el país, que la elasticidad-precio de su oferta laboral es también la más alta de la población, y por último, que “mejorar la inserción laboral de los participantes” es uno de los objetivos explícitos de ProJoven; es importante analizar el impacto de ProJoven en la inserción laboral.

Para esto, se muestran también los tres estimadores propuestos previamente, considerando los cuatro emparejamientos realizados.

ProJoven 6ta Convocatoria
Comparación de Estimadores de Impacto Sobre la Inserción Laboral

	Emparejamiento 1	Emparejamiento 2	Emparejamiento 3	Emparejamiento 4
Estimador Diferencia de Diferencias	6.1% 0.9975	4.7% 0.9866	4.3% 0.9430	6.9% 0.9713
Estimador Antes-Después	3.3% 0.9446	1.9% 0.8234	3.3% 0.9423	21.1% 1.0000
Estimador de Corte Transversal	6.1% 0.9975	4.7% 0.9866	7.0% 0.9990	6.9% 0.9713

Los estimadores de corte transversal reportan valores iguales a los estimadores de diferencias en diferencias en tres de los emparejamientos. Esto responde a la construcción de los mismos: una de las variables consideradas en los emparejamientos 1,2 y 4 ha sido precisamente la “situación laboral del individuo”. Según estos, el impacto de ProJoven en la inserción laboral de los participantes se encuentra en el rango 4%-7%. Es decir, la población participante en ProJoven consiguió una tasa de inserción laboral que está entre los 4 y 7 puntos porcentuales por encima de la que hubiesen conseguido en caso de no participar en el programa.

Los estimadores antes-después reportan valores substancialmente por debajo de los otros dos (diferencia de diferencias y corte transversal) en tres de los cuatro emparejamientos. El resultado que obtenemos para el cuarto emparejamiento con este estimador está considerablemente alejado de los demás y creemos que se debe a una aleatoriedad estadística del emparejamiento o al carácter estricto del criterio de emparejamiento que origina ciertos sesgos.²⁸

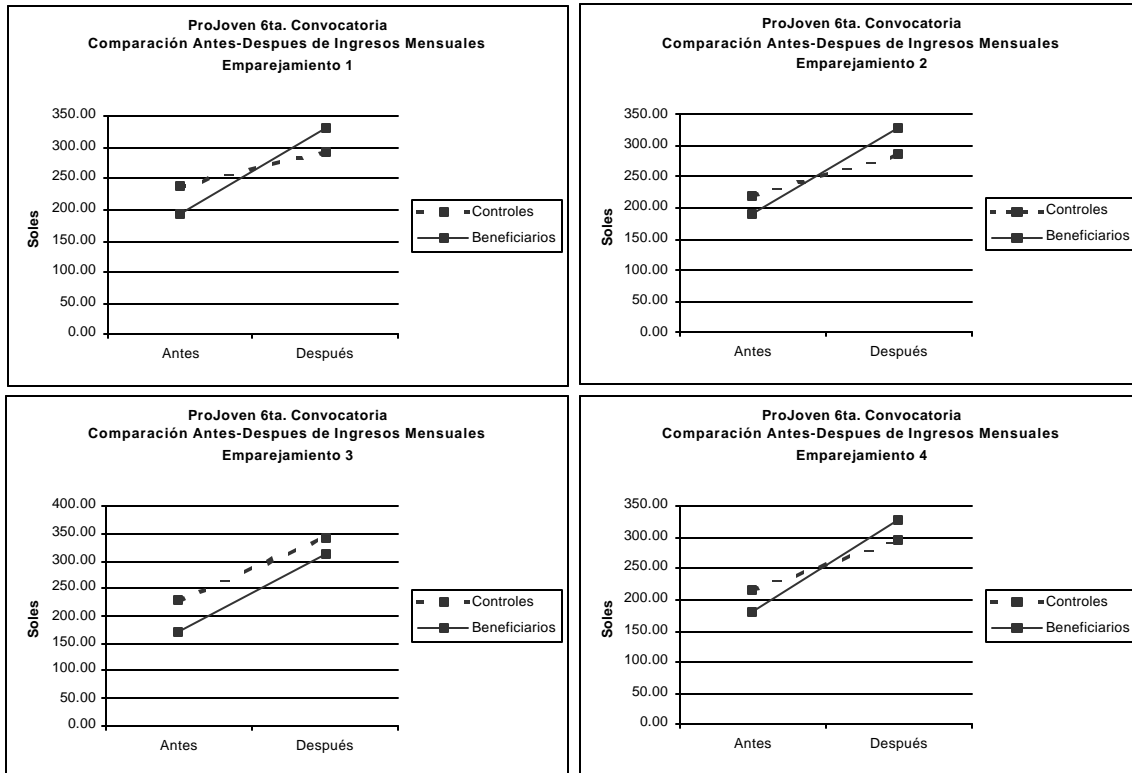
En cualquier caso, el “efecto ProJoven” es positivo y significativo.

c) Impacto sobre los ingresos mensuales

Un análisis de impacto sobre los ingresos por hora ignora efectos de inserción laboral “al margen intensivo”, es decir, ignora la posibilidad de observar a los participantes del programa trabajando más horas al mes en comparación con el grupo de control, gracias a su participación en el programa. Es necesario complementar el análisis de impacto con una mirada a los indicadores de ingreso mensuales.

Tal como presentamos los resultados de impacto sobre los ingresos por hora, aquí presentaremos en primer lugar un panel con cuatro gráficos representando los cambios en ingresos mensuales tanto para beneficiarios como para controles en los cuatro emparejamientos que consideramos en esta sección empírica del informe. Después presentaremos unas tablas con los valores precisos de los estimadores obtenidos para el impacto de ProJoven en esta variable y sus correspondientes p-values.

²⁸ Vease Rosenbaum y Rubin (1985)



Se puede observar que tanto los ingresos mensuales de los beneficiarios como los de los controles crecieron entre las dos mediciones (una antes y una después de la aplicación de la sexta convocatoria de ProJoven). En tres de los cuatro emparejamientos se puede observar que este crecimiento es más marcado entre los beneficiarios que entre los controles (la líneas rectas que conectan los puntos que representan los ingresos medios antes y después, para beneficiarios y controles, se intersecan). Esto da una idea del impacto positivo de ProJoven. La cuantificación precisa de este impacto, por medio de los tres estimadores que se están utilizando en esta sección así como sus correspondientes p-values, son mostrados a continuación.

ProJoven 6ta Convocatoria
Comparación de Estimadores de Impacto Sobre Ingresos Mensuales

	Emparejamiento 1	Emparejamiento 2	Emparejamiento 3	Emparejamiento 4
Estimador Diferencia de Diferencias	42.9% 1.0000	37.7% 1.0000	14.0% 0.7822	37.0% 0.9712
Estimador Antes-Después	73.0% 1.0000	73.1% 1.0000	80.6% 1.0000	80.4% 1.0000
Estimador de Corte Transversal	12.2% 0.9679	14.1% 0.9840	-9.1% 0.1627	11.2% 0.8605

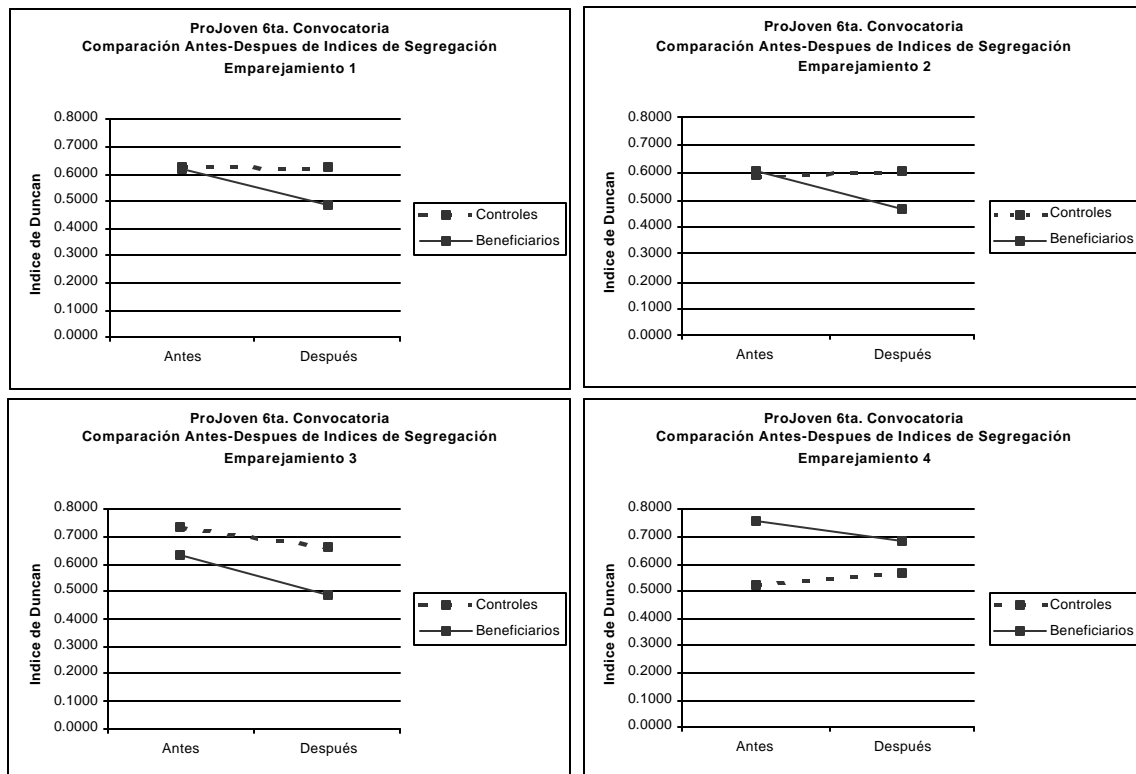
Los estimadores de impacto sobre ingresos mensuales son considerablemente mayores a los estimadores obtenidos para el impacto en los ingresos por hora. Esto denota la existencia de un impacto positivo de ProJoven en el margen intensivo de la oferta de trabajo de los beneficiarios. Es decir, ellos trabajan más horas por mes, en comparación al grupo de control. Solo en uno de los casos (estimador de corte transversal en el emparejamiento 3), reportamos un impacto negativo del programa; aunque como puede apreciarse, la significancia de tal valor negativo no es lo suficientemente alta.

d) Impacto sobre la segregación ocupacional

El impacto de ProJoven sobre la segregación ocupacional por género es medido a través de los Índices de Segregación de Duncan. De acuerdo a esta medida, una economía completamente segregada reportará un valor de 1 para el índice de segregación, mientras que una economía en la que hombres y mujeres están equitativamente distribuidos sobre las diferentes ocupaciones reportará un valor de 0 para tal índice. Es decir, a menor valor del índice, menor segregación ocupacional (lo cual puede ser entendido como mayores oportunidades de acceso de las mujeres a las ocupaciones típicamente masculinas).

Se espera que, por efecto de su participación en ProJoven, la población beneficiaria reporte índices de segregación menores a los que reportaba antes de la aplicación del programa. Es decir, se espera que el efecto de ProJoven, sea una tendencia a la baja para tal índice de Duncan. A continuación mostramos los paneles con los cuatro gráficos que muestran la evolución de los

índices de Duncan para la muestra de beneficiarios y controles, antes y después de la sexta convocatoria, para los cuatro emparejamientos considerados en esta sección.



Antes de la aplicación de la sexta convocatoria, los niveles de segregación ocupacional para beneficiarios y controles en los primeros tres emparejamientos se encuentran alrededor de 0.6, cifra bastante cercana a los niveles de segregación ocupacional que reporta el Índice de Duncan para la cohorte 18-25 a nivel nacional. Los niveles de segregación ocupacional que se observan en los beneficiarios y controles del emparejamiento 4 son muy disímiles entre si y no guardan semejanza con las cifras nacionales. Esto da nuevas luces acerca del sesgo potencial que pudiera existir en la muestra obtenida por el emparejamiento 4.

En cualquier caso, es fácil observar un efecto de ProJoven en el sentido esperado: hacia la disminución de la segregación. La medición precisa de tales efectos, con las mismas consideraciones de emparejamientos, estimadores y p-values se presenta a continuación

ProJoven 6ta Convocatoria
Comparación de Estimadores de Impacto Sobre la Segregación Ocupacional

	Emparejamiento 1	Emparejamiento 2	Emparejamiento 3	Emparejamiento 4
Estimador Diferencia de Diferencias	-0.1176 0.9413	-0.1498 0.9429	-0.0701 0.7773	-0.1214 0.7966
Estimador Antes-Después	-0.1243 0.9830	-0.1369 0.9802	-0.1409 0.9584	-0.0738 0.7357
Estimador de Corte Transversal	-0.1310 0.9803	-0.1341 0.9560	-0.1721 0.9891	0.1155 0.2260

El efecto de ProJoven resulta positivo solo en uno de los casos (usando el estimador de corte transversal con el cuarto emparejamiento). Los resultados que reportan los dos primeros emparejamientos son razonablemente estables y alrededor de los trece puntos. Eso se interpreta de la siguiente manera: el efecto de ProJoven sobre la segregación ocupacional por género de los participantes equivale a cambiar de ocupación al 13% de las mujeres (hombres), cambiándoles de ocupaciones típicamente femeninas (masculinas) a ocupaciones típicamente masculinas (femeninas).

5.4. Análisis de impacto por medio de técnicas de regresión lineal.

En esta sub-sección mostraremos los resultados obtenidos a través de las reconocidas técnicas de regresión lineal. La evaluación de impacto, en este caso se concentrará en dos variables: ingresos por hora e inserción laboral²⁹.

²⁹ El análisis de impacto sobre los ingresos mensuales por medio de técnicas de regression lineal requiere de una serie de ajustes al modelo que van mas alla de los alcances y el objetivo de este estudio. Tampoco es posible un análisis de impacto sobre la segregación ocupacional por medio de técnicas de regresión lineal (debido a la no-linealidad del Índice de Segregación de Duncan).

a) Impacto sobre Ingresos por Hora.

El análisis de impacto sobre los ingresos por hora requiere en este caso de la estimación del siguiente modelo

$$Y_{it} = b_{it}P_i + c_{it}X_i + a_t \mathbf{1}_{it} + u_{it},$$
$$1_{\{Y_{it}>0\}} = f(\mathbf{a}_{it}Z_i + e_{it}), \quad t = 0,1$$

La primera ecuación corresponde a una típica ecuación de ingresos, donde la variable explicada Y_{it} representa el ingreso por hora del individuo i en el momento t , la variable P_i es una dummy que toma el valor 1 si el individuo i participa en ProJoven y 0 en caso contrario, X_i es un vector de características socio-demográficas y laborales del individuo i , $\mathbf{1}_{it}$ es el ratio inverso de Mills estimado para el individuo i en el instante t (es el elemento de corrección por sesgo de selección en el ecuación de ingresos).³⁰ La segunda ecuación corresponde a un modelo dicotómico (en este caso, probit) en el que la variable explicada $1_{\{Y_{it}>0\}}$ toma el valor 1 si el individuo i posee empleo remunerado en el instante t y toma el valor 0 en caso contrario, el vector Z_i representa una serie de características socio-demográficas y de experiencia laboral del individuo i . Los elementos u_{it} y e_{it} denotan los componentes de error en las estimaciones.

Esto es, estimaremos una ecuación de ingresos con la adecuada corrección por sesgo de selección de aquellos individuos cuyas observaciones participan en la ecuación de ingresos: personas ocupadas reportando ingresos positivos. Esta estimación es realizada tanto para $t=0$ (antes de la realización de la Sexta Convocatoria) como para $t=1$ (después). La técnica a seguir es la estándar de dos etapas propuesta por Heckman (1979).³¹ Los resultados se muestran a continuación.

³⁰ La corrección por sesgo de selección se hace necesaria porque la estimación de la ecuación de ingresos se hace seleccionando una sub-muestra de individuos con empleo remunerado (y esta sub-muestra no es seleccionada aleatoriamente de la muestra total disponible para la evaluación).

³¹ El lector interesado en los detalles de esta técnica de estimación puede consultar Heckman (1979).

ProJoven 6ta Convocatoria
Estimación de la Ecuación de Ingresos Antes y Después del Programa
(Modelo Heckman de 2 etapas)

	Antes		Después	
	Coefficiente	Error Est.	Coefficiente	Error Est.
Ecuación de Ingresos				
Dummy de Participación en ProJoven (1=Participa)	-0.3130	0.0358	0.0195	0.0363
Sexo (1=Hombre)	0.1724	0.0375	0.1483	0.0371
Dummy de Instrucción Técnica (1=Si)	0.0553	0.0419	0.0811	0.0442
Dummy de Educación Secundaria Incompleta (1=Si)	0.1813	0.1028	0.1200	0.1076
Dummy de Educación Secundaria Completa (1=Si)	0.1261	0.1001	0.0872	0.1035
Dummy de Ubicación Geográfica (1=Lima, 0=Resto del Pais)	0.2021	0.0374	0.2564	0.0367
Trabaja en Empresa Pequeña (6-20 trabajadores)	0.2015	0.0489	0.1870	0.0472
Trabaja en Empresa Mediana (21-99 trabajadores)	0.2833	0.0687	0.2762	0.0689
Trabaja en Empresa Grande (más de 99 trabajadores)	0.3462	0.0605	0.3483	0.0577
Constante	-0.0226	0.1102	-0.1157	0.1100
Ecuación de Inserción Laboral				
Experiencia Potencial del Individuo	0.1236	0.0106	0.0687	0.0102
Logaritmo del Ingreso per Capita de los demás Miembros de su Hogar	-0.1196	0.0241	-0.2781	0.0261
Dummy de Mujeres con Hijo (1=Si)	-0.4141	0.0735	-0.6126	0.0686
Constante	0.4778	0.1209	1.4375	0.1323
Ratio Inverso de Mills (Corrección por Sesgo de Selección)	-0.3296	0.1002	-0.2158	0.0780
Total de Observaciones	2579		2579	
Observaciones en la Ec. De Ingresos	1552		1624	

El parametro de interés en nuestra evaluación de impacto es el coeficiente de la dummy de participación en ProJoven. Precisamente, nos interesa el cambio en este coeficiente entre las ecuaciones de ingresos estimadas después y antes del programa. Como bien puede observarse, este coeficiente reporta una evolución positiva, de -0.3130 a 0.0195. Esto significa que la participación en ProJoven esta asociada con un incremento de 33.26% en los ingresos por hora de los participantes. Esto, después de haber controlado por una serie de variables que explican los ingresos de los individuos.

Esta medida ofrece solo información acerca del impacto promedio del programa. Es necesario complementar esta tendencia central con una medida de dispersión. Para esto, hemos realizado un ejercicio de remuestreo bootstrap en el que hemos estimado el mismo modelo de dos etapas en cada iteración.³² Concentrando nuestra atención en la diferencia de los “efectos de ProJoven” en la ecuación de ingresos después y antes del programa, hemos construido una distribución empírica para tal diferencia. Los resultados se muestran a continuación.

³² El ejercicio bootstrap consiste en seleccionar una nueva sub-muestra aleatoria (con reemplazo) de la muestra original en cada iteración.

ProJoven 6ta Convocatoria
Resultados del Impacto de Projoven Sobre los Ingresos por Hora
(10000 Repeticiones Bootstrap)

	Coeficiente ProJoven	Sesgo	Error Estándar
Antes	-0.3130	0.0003	0.0372
Después	0.0195	0.0004	0.0371
Diferencia	0.3326	0.0001	0.0495

De esta manera, el estimador de 33.26% para el impacto de ProJoven obtenido por medio de esta técnica, lleva asociado un error estándar de 4.95% y un sesgo de 0.01%. Un intervalo de confianza del 95% para este estimador va de 23.65% a 42.80%.

b) Impacto Sobre Inserción Laboral.

El análisis de impacto de Projoven en la inserción laboral de los participantes requiere la estimación del siguiente modelo de participación laboral

$$1_{\{Y_{it}>0\}} = f(\mathbf{a}_i Z_i + e_{it}), \quad t = 0,1$$

En este modelo dicotómico (probit) la variable explicada $1_{\{Y_{it}>0\}}$ toma el valor 1 si el individuo i posee empleo remunerado en el instante t y toma el valor 0 en caso contrario, el vector Z_i representa una serie de características socio-demográficas y de experiencia laboral del individuo i , el elemento e_{it} denota al componente de error en la estimación.³³ Para poder aprovechar el poder informativo de los estimadores obtenidos a partir del modelo probit es necesaria la estimación de los efectos marginales de los estimadores. Esto se hace tomando las derivadas parciales de la función de distribución con que se asume que el error se encuentra distribuido (que en nuestro caso es normal); estas derivadas parciales se toman con respecto de cada uno de

³³ El vector de características Z_i de este modelo dicotómico no necesariamente coincide con el vector de características Z_i utilizado en la estimación del modelo dicotómico que sirve de base para la estimación de la ecuación de ingresos. La selección de las características que conforman estos vectores responde a razones modelísticas y econométricas, cuya discusión dejamos de lado por el momento.

los elementos del vector de características Z. Los resultados de la estimación del modelo probit original y de los efectos marginales se muestran a continuación.

ProJoven 6ta Convocatoria				
Estimación de la Ecuación de Participación Antes y Después del Programa				
(Modelo Probit)				
	Antes		Después	
	Coefficiente	Error Est.	Coefficiente	Error Est.
Ecuación de Participación				
Dummy de Participación en ProJoven (1=Participa)	0.0808	0.0539	0.0932	0.0549
Experiencia Potencial del Individuo	0.0970	0.0109	0.0511	0.0105
Sexo (1=Hombre)	0.4043	0.0585	0.1918	0.0614
Dummy de Ubicación Geográfica (1=Lima, 0=Resto del Pais)	-0.1562	0.0547	0.0055	0.0556
Dummy de Instrucción Técnica (1=Si)	0.1043	0.0627	-0.2817	0.0644
Dummy de Mujeres con Hijo (1=Si)	-0.1353	0.0828	-0.5391	0.0807
Logaritmo del Ingreso per Capita de los demás Miembros de su Hogar	-0.0874	0.0244	-0.2763	0.0280
Número de Miembros del Hogar	0.0206	0.0110	-0.0245	0.0112
Constante	0.1527	0.1378	1.6108	0.1651
Efectos Marginales Asociados				
Dummy de Participación en ProJoven (1=Participa)	0.0300	0.0199	0.0335	0.0197
Experiencia Potencial del Individuo	0.0361	0.0040	0.0185	0.0038
Sexo (1=Hombre)	0.1492	0.0212	0.0691	0.0220
Dummy de Ubicación Geográfica (1=Lima, 0=Resto del Pais)	-0.0586	0.0206	0.0020	0.0201
Dummy de Instrucción Técnica (1=Si)	0.0384	0.0228	-0.1048	0.0245
Dummy de Mujeres con Hijo (1=Si)	-0.0512	0.0317	-0.2045	0.0313
Logaritmo del Ingreso per Capita de los demás Miembros de su Hogar	-0.0325	0.0091	-0.0998	0.0098
Número de Miembros del Hogar	0.0077	0.0041	-0.0088	0.0040
Observaciones	2579		2579	

Y tal como se procedió para la estimación del impacto en ingresos, el impacto de ProJoven en la inserción laboral será medido a partir del cambio en el coeficiente de la dummy de participación en ProJoven después y antes de la aplicación de la Sexta Convocatoria (en sus efectos marginales). De manera similar a como procedimos en la estimación del impacto en ingresos, la estimación de una medida de dispersión para el estimador obtenido por regresiones fue hecha a partir de re-muestreo (bootstrap). Obteniéndose los siguientes que se muestran en la siguiente tabla. Ahí podemos observar que la participación en ProJoven esta asociada con una mayor probabilidad de encontrar empleo remunerado, después de haber controlado por una serie de variables que explican la inserción laboral de los individuos. Pero este efecto positivo reporta también errores estándares considerables lo que implica que este impacto positivo sea no significativo.

ProJoven 6ta Convocatoria
 Resultados del Impacto de Projoven Sobre la Inserción Laboral
 (10000 Repeticiones Bootstrap)

	Coeficiente ProJoven	Sesgo	Error Estándar
Antes	0.0300	0.0000	0.0198
Después	0.0335	-0.0003	0.0197
Diferencia	0.0035	-0.0002	0.0258

Es importante notar que los efectos significativamente positivos que encontramos para el impacto de ProJoven en los ingresos de los participantes, no se encuentran en la inserción laboral de los mismos. ProJoven ayuda a mejorar los ingresos de los participantes que consiguen empleo remunerado, pero no necesariamente ayuda mucho a los participantes en la consecución de tal empleo remunerado.

6. Bibliografía

Abadie, Alberto y Guido Imbens (2002). "Simple and Bias-Corrected Matching Estimators for Average Treatment Effects". No publicado.

Aedo, Cristián y Sergio Núñez (2001). "The Impact of Training Policies in Latin America and the Caribbean: The case of Programa Joven". ILADES/Georgetown University. Graduate Program in Economics.

Ashenfelter, Orley (1978). "Estimating the Effect of Training Programs on Earnings". The Review of Economics and Statistics. Vol 6, Issue 1, pp 47-57.

Ashenfelter, Orley y David Card (1984). "Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs". Working Paper #174. Industrial Relations Section. Princeton University.

Baker, Judy; (2000). "Evaluating the Impact of Development Projects on Poverty". World Bank, Washington D.C.

Bravo, David y Dante Contreras (2000). "The Impact of Financial Incentives to Training Providers: The Case of Chile Joven". Inter-American Seminar on Economics. Micro-Data Research in Latin America. National Bureau of Economics Research.

Deutsch, Ruthanne, Andrew Morrison, Claudia Piras & Hugo Ñopo (2001). "Working Within Confines: Occupational Segregation by Gender for three Latin American Countries". Inter-American Development Bank.

Friedlander, Daniel, David Green & Philip Robins (1997). "Evaluating Government Training Programs for the Economically Disadvantaged". Journal of Economic Literature, Vol XXXV pp 1809-1855.

Galdo, Jose Carlos. (2000) “Una Metodología de Evaluación de Impacto de los Programas de Educación y/o Capacitación en el Mercado Laboral: el Caso de ProJoven”.En Impacto de la Inversión Social en el Perú, Ed. Enrique Vásquez, Lima CIUP-IDRC.

Heckman, James J, 1979. 'Sample Selection Bias as a Specification Error," *Econometrica*, Vol. 47 (1) pp. 153-61.

Heckman, James, Hidehiko Ichimura & Petra Todd (1997). “Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program”. *Review of Economics Studies*, 65, 261-294.

Heckman, James y Jeffrey Smith (1997). “The Sensitivity of Experimental Impact Estimates: Evidence from the National JTPA Study”. Working Paper 6105. National Bureau of Economic Research. <http://www.nber.org/papers/w6105>

Heckman, James y Jeffrey Smith (1999). “The Pre-Program Earnings Dip and The Determinants of Participation in a Social Program: Implications for Simple Program Evaluation Strategies”. Working Paper 6983. National Bureau of Economic Research. <http://www.nber.org/papers/w6983>

Heckman, James, Robert LaLonde & Jeffrey Smith (1998). “The Economics and Econometrics of Active Labor Market Programs” in *Handbook of Labor Economics*, Volume III, Orley Ashenfelter and David Card, Eds.

Heckman, James, Jeffrey Smith & Nancy Clements (1999). “Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts”. *Review of Economic Studies*, 64, 487-535.

Horowitz, Joel (2000). “The Bootstrap”. En *Handbook of Econometrics*, Vol. 5, J.J. Heckman and E. Leamer, eds., Elsevier Science B.V., 2001, pp. 3159-3228.

Jacinto, Claudia (2000). “Contextos y Actores Sociales en la Evaluación de los Programas de Capacitación de Jóvenes” en “Los jóvenes y el Trabajo. La Educación frente a la Exclusión Social”, Universidad Iberoamericana, México.

Maddala, G.S. (1999), "Limited dependent and qualitative variables in econometrics", Cambridge University Press, New York.

Mitnik, F. “La experiencia Projoven: apoyo a la inserción de los jóvenes en el mercado laboral del Perú”. En: Boletín Cinterfor/OIT N° 139-140, abr-set 1997. 250p.

Ñopo, Hugo, Miguel Robles y Jaime Saavedra (2002). “Una medición del impacto del Programa de capacitación Laboral Juvenil PROJoven”. GRADE, Documento de Trabajo # 36.

Oaxaca, Ronald (1973). “Male-Female Wage Differentials in Urban Labor Markets", International Economic Review, October 1973.

Oaxaca, Ronald y Michael R. Ransom (1994). “On Discrimination and the Decomposition of Wage Differentials", Journal of Econometrics, March 1994.

Preston, J.A. (1999). “Occupational gender segregation: Trends and explanations”. The Quarterly Review of Economics and Finance, 1999. No. 39, pp. 611-624.

Ravallion, Martin; (1999). "The Mystery of the Vanishing Benefits: Ms. Speedy Analyst's Introduction to Evaluation." Policy Research Working Paper 2153, World Bank

Ravina, Renato, Paulini Javier y Cancho César (2002). “Costo efectividad del programa de desayunos escolares de FONCODES y el programa de alimentación escolar del PRONAA” Informe Final (sujeto a revisiones y comentarios). Documento no publicado.

Rosenbaum, Paul y Donald Rubin (1985). “The Bias Due to Incomplete Matching”. Biometrics, Vol. 41, No. 1. (Mar., 1985), pp. 103-116.

Saavedra, Jaime y Eduardo Maruyama (1999) “Los Retornos a la Educación y a la Experiencia en el Perú: 1985-1997”. Lima: GRADE

Saavedra, Jaime y Juan Chacaltana (2001). “Exclusión y Oportunidad. Jóvenes urbanos y su inserción en el mercado de trabajo y en el mercado de capacitación”. Lima: GRADE.

Santiago Consultores Asociados (1999). "Evaluación Ex-Post Chile Joven Fase II". Informe Final Preliminar. Documento presentado en Seminario internacional: Modelos de Evaluación para Programas de Capacitación de Jóvenes: 17 y 18 de julio 2,000, Medellín, Colombia, OIT-CINTERFOR.

Vásquez E., Aramburú C., Figueroa C. y C. Parodi (2000). "Gerencia social. Diseño, monitoreo y evaluación de proyectos sociales". Lima, Centro de Investigación de la Universidad del Pacífico.