

RE-EVALUANDO PROJOVEN: PROPENSITY SCORE MATCHING Y UNA EVALUACIÓN PARAMÉTRICA

Cybele Burga*

Versión Final: 5 de Febrero de 2003

*Este documento se realizó bajo el auspicio del CEDEP.

Abstract

El presente estudio tiene como objetivo evaluar el impacto del programa de capacitación laboral juvenil PROJoven sobre la variable resultado "horas trabajadas a la semana" e "ingresos laborales". La principal técnica de estimación de dicho impacto se basará en el matching propensity score, la cual se describe detalladamente en la sección metodológica. Esta técnica tiene la ventaja de lograr estimaciones del efecto de tratamiento más precisas, dado que se compara los resultados promedio de dos grupos que son similares en términos de un vector de características. Adicionalmente, el propensity score permite reducir el problema de la multidimensionalidad ya que finalmente el tratamiento se calcula de forma condicional sólo a un escalar (la probabilidad de participación en el programa), en vez de a un vector de características. La hipótesis detrás de toda implementación de programas de capacitación es que luego de pasar por el tratamiento (programa) los individuos que lo recibieron (beneficiarios) tienen mejores resultados en las variables mencionadas en comparación con los individuos que no lo hicieron (grupo de control). Esta hipótesis se valida para el caso de los ingresos laborales mensuales, sin embargo, cuando se estima el efecto sobre los ingresos por hora estos resultados son nulos. La hipótesis también se valida si se comparan las horas de trabajo a la semana para cada uno de los grupos. Esto quiere decir que el programa también tiene efectos positivos en cuanto al subempleo por horas trabajadas ya que los beneficiarios terminan trabajando más horas en promedio que los controles.

1 Introducción

Uno de los segmentos demográficos más perjudicados en los últimos años, en lo que se refiere a inserción laboral ha sido el de los individuos cuyas edades están entre los 14 y 25 años de edad. Estos jóvenes no sólo enfrentan altas tasas de desempleo, sino también las más altas tasas de subempleo por ingresos.

Las explicaciones a este fenómeno son varias: la explosión demográfica de los 70's y el aumento de la tasa de actividad; el deterioro de la calidad de la educación, así como el incremento de jóvenes en las ciudades debido a las migraciones desde la Sierra del Perú hacia las ciudades capitales, principalmente Lima (capital del Perú). Según el Ministerio de Trabajo y Promoción Social -MTPS (1998), la migración ha determinado que la tasa de crecimiento de la población de Lima sea significativamente mayor a la del total del país. Todo esto genera que los jóvenes de menos recursos no tengan una adecuada formación que les permita tener adecuadas competencias para el trabajo y por lo tanto ser competitivos en el mercado laboral, sino que terminan siendo desplazados y marginados en la sociedad.

En el Cuadro 3 se muestra que la tasa de desempleo de los jóvenes es de 15% en Lima Metropolitana y 12% en el Resto Urbano, mientras que la de población adulta bordea en 6%. Otro aspecto preocupante es que el desempleo de los jóvenes es mayor entre los que han tenido acceso a algún tipo de capacitación. Esto puede deberse a la baja calidad de los servicios de capacitación, que de alguna manera puede manifestarse en la poca credibilidad que tiene para los empresarios la capacitación como elemento que permitiría incrementar la productividad de los trabajadores. Esta poca credibilidad se plasma en los bajos niveles de capacitación de los jóvenes en las empresas, 1.6% (ver Cuadro 1).

Debido a esto, y a otros factores más, surgió a mediados de 1996 el programa de Capacitación Laboral Juvenil PROJoven como iniciativa del Ministerio de Trabajo y Promoción Social (MTPS). El problema que buscaba enfrentar el PROJoven era el de las limitadas oportunidades de empleo que tenían los jóvenes, debido a: los altos costos de la información, las limitadas redes sociales a los que éstos tienen acceso, la deficiente calidad y escasa pertinencia de la capacitación ocupacional a la que tienen acceso los jóvenes pobres, educación técnica y ocupacional con equipos y materiales obsoletos y a la selección y diseño de cursos de capacitación desvinculados de los requerimientos del sector empresarial y del mercado de trabajo. Esto se plasma en la creciente discordancia ocupacional que existe entre los jóvenes peruanos, lo que termina generando un problema de "sobreeducación. Esta se produce si un individuo tiene más educación que la requerida para desempeñar adecuadamente un trabajo, por lo que sus habilidades no estarían siendo plenamente utilizadas (Burga y Moreno, 2000).

El presente documento presenta una medición del impacto del programa de capacitación laboral juvenil PROJoven. En particular, en el estudio se analizarán los resultados que provienen de la primera y segunda medición de la Segunda Convocatoria del programa, llevada a cabo en 1998.

En este estudio se ha utilizado la información socio-laboral que provee el programa para un grupo de individuos que conformaron los grupos de beneficiarios y de control.

La información puede dividirse en dos rubros: (1) la información antes que el individuo pase por el programa, es decir, antes de recibir la capacitación y de realizar las pasantías (línea de base); (2) y la información después que el individuo pasó por el programa (ésta puede ser a los 6, 12 o 18 meses después de culminada esta fase lectiva). Para ello se aplican encuestas de seguimiento utilizando los mismos instrumentos y además se aplican al mismo tiempo para ambos grupos y de esta forma garantizar la equivalencia durante el experimento.

El número total de beneficiarios elegidos por el programa, para la segunda convocatoria, fue de 1807 individuos. Sobre este grupo, y aplicando la técnica del muestreo aleatorio estratificado¹, se obtuvo una muestra de 327 beneficiarios y 327 controles, con los cuales se realizará la evaluación de impacto del programa². Entre la línea de base, primera y segunda medición hay una pérdida de muestra del orden de 12%. Como se explicará en detalle en secciones posteriores, a pesar que este muestreo pretenda realizar estimaciones de los parámetros poblacionales para los estratos de la población, el efecto de tratamiento (impacto del programa) sólo se calculará para el grupo de los tratados y no será extrapolable para el resto de la población.

En este sentido, el objetivo del estudio será estimar el impacto sobre algunos indicadores laborales como resultado de haber pasado por el programa. Para calcular este impacto se comparará la situación con tratamiento (haber pasado por el programa) versus la situación sin tratamiento (no haberlo hecho) en variables como ingreso laboral y horas trabajadas por semana del individuo (número de veces que estuvo ocupado entre enero y julio de 1999).

El aspecto central en los métodos de evaluación radica en la construcción de los contrafactuales (grupo de comparación). Para este estudio, la construcción de los contrafactuales o controles ha sido realizada a través de distintos métodos de matching. Esta técnica permite seleccionar factores observables suficientes como para que dos individuos con los mismos valores para estos factores no presenten diferencias sistemáticas en sus reacciones al programa. De esta forma, el impacto puede ser calculado al comparar individuos similares en estas características con la única diferencia que uno participó y el otro no. Lo que se estimará será el impacto sobre individuos de ciertas características entre todas, que fueron asignados al programa. Este efecto tratamiento, comúnmente se ha llamado "efecto de tratamiento sobre los tratados".

En nuestro proceso de matching se buscó que los beneficiarios y controles sean comparables en variables demográficas y laborales, las cuales detallaremos en la siguiente sección. Las distintas técnicas utilizadas para hacer el matching también se detallarán posteriormente.

En este trabajo se presentan las estimaciones del impacto del programa sobre variables laborales para los jóvenes beneficiarios. El estimador de matching indica que el impacto promedio del programa sobre los tratados en la variable ingreso laboral men-

¹Una muestra aleatoria estratificada es obtenida mediante la separación de los elementos de la población en grupos (estratos) que no presenten traslapes, y la selección posterior de una muestra ir-restricta aleatoria simple en cada estrato.

²En el caso de la segunda convocatoria, la evaluación de impacto sólo se hizo para la ciudad de Lima.

sual es estrictamente positivo (156 soles a un año después de la fase lectiva), aunque resulta muy cercano a cero para los ingresos laborales por hora. Esta diferencia se puede explicar por medio del impacto del programa sobre las horas trabajadas por los beneficiarios (el cual es positivo en 8 horas en promedio, para la misma medición).

El documento está organizado de la siguiente manera: en la primera sección, a modo de introducción se hace una breve descripción del programa, luego en la segunda, se presenta el marco teórico discutiendo la metodología usada para la evaluación de impacto y finalmente, en la última sección se presentan los resultados de las estimaciones.

2 El Programa de Capacitación Laboral Juvenil (PRO-Joven)

El Programa de Capacitación Laboral Juvenil PROJoven capacita a jóvenes en oficios para ocupaciones específicas a nivel básico o de semicalificación. Esta capacitación comprende dos fases: formación técnica y formación laboral. La primera tiene una duración de tres meses y está a cargo de las ECAP's que se vinculan previamente con empresas del sector productivo y de servicios para diseñar cursos que respondan a las actuales necesidades de capacitación (esto sin embargo no siempre sucede en la práctica.) La segunda fase es la de las pasantías y es remunerada en las empresas de régimen privado laboral. Idealmente es en estas empresas donde los jóvenes adquieren experiencia de trabajo. La duración mínima de esta fase es de tres meses.

El diseño del programa permite una nueva participación del estado, enfatizando los roles de promoción, financiamiento, contratación y supervisión de los servicios de capacitación antes que la prestación directa.

2.1 Objetivos del Programa

El PROJoven tiene dos objetivos fundamentales: (1) mejorar las oportunidades de empleo de los jóvenes (productividad y empleabilidad), especialmente para aquéllos de escasos recursos, y, (2) dinamizar el sistema de capacitación en el país, mejorando la calidad técnica del diseño e implementación de cursos de capacitación ocupacional. Por lo tanto, contribuiría a elevar el nivel de eficiencia y eficacia del mercado de capacitación laboral, promoviendo la competencia y una mejor interacción entre las entidades de capacitación (ECAP's), las necesidades reales del sector empresarial y el mercado de trabajo.

Inicialmente, sólo participaban en el programa jóvenes que vivían en la ciudad de Lima, sin embargo, en la actualidad, participan las principales ciudades del país en sus zonas urbano y urbano periféricas³.

2.2 Usuarios potenciales del programa

Los potenciales usuarios del programa son los jóvenes que pertenecen a hogares pobres, cuyas edades fluctúan entre los 16 y 24 años de edad, con escasas o nula experiencia laboral y que estén desempleados o subempleados o inactivos, con o sin educación secundaria completa (como máximo nivel educativo). Asimismo, si es que han recibido o reciben capacitación, ésta no puede ser mayor a 300 horas. Para determinar la situación de pobreza de los jóvenes se tomó en cuenta lo siguiente:

1. Dimensión familiar de la pobreza
2. La pobreza estructural y la coyuntural. La pobreza estructural se define en función al método de las necesidades básicas insatisfechas, siendo algunas de

³Las actuales áreas de influencia del Perú son: Lima, Callao, Arequipa, Trujillo, Chiclayo, Cusco, Piura, Huancayo, Chimbote e Iquitos.

las variables empleadas: hacinamiento, materiales de vivienda y acceso del hogar a servicios públicos básicos. Por otro lado, la pobreza coyuntural se mide con el indicador de tasa de dependencia económica. (ver Cuadro 5 - Criterios de acreditación)

2.3 Selección de las ECAP's

El proceso de selección de las entidades de capacitación (públicas y privadas) consiste en pasar por un proceso de pre calificación, en donde se evalúan aspectos como la experiencia de capacitación previa, la capacidad administrativa, la capacidad de gestión y los recursos humanos con los que cuenta. Una vez que las entidades han sido seleccionadas pasan a formar parte del Registro de Entidades de Capacitación (RECAP).

Únicamente las entidades registradas en el RECAP están autorizadas a presentar propuestas de cursos en las licitaciones que el programa convoca. Mediante este proceso de convocatoria a concurso público, PROJoven selecciona y contrata los cursos de capacitación para ocupaciones específicas de semicalificación cuya demanda está comprobada en el mercado.⁴

Desde agosto de 1996 hasta enero del 2002, se han adjudicado 1516 cursos a 379 ECAP's que benefician a 30,283 jóvenes de escasos recursos económicos a nivel nacional. En cada una de las convocatorias, la oferta de cursos por parte de las ECAP's ha tenido una tendencia creciente. En las nueve convocatorias realizadas, el programa ha realizado una evaluación de 3,511 cursos.

Una de las singularidades del PROJoven (y que la distingue por ejemplo de su similar chileno, Chile Joven) es que permite a los jóvenes adquirir una experiencia concreta de trabajo (al exigir de forma obligatoria a las ECAP's pasantías pagadas -por las empresas- para los jóvenes por las empresas) y de esta forma mejorar sus posibilidades de inserción en el mercado laboral. Sin embargo, cabe mencionar que las prácticas dependen en gran medida de los planes de producción y ventas, así como de la evolución del ciclo económico. Estas serían algunas de las razones por las cuales menos del 100% de los beneficiarios terminan realizando prácticas laborales en las empresas con las cuales las ECAP's tenían una carta de compromiso de trabajo. (ver Cuadro 4).

Bajo este esquema se busca que la capacitación sea el instrumento que brinde no solo conocimientos teóricos, sino actitudes, hábitos, destrezas y habilidades, que permitan el desarrollo de competencias básicas para el trabajo.

2.4 Cursos

Los cursos tienen como objetivo lograr que los jóvenes beneficiarios del programa adquieran habilidades y competencias que les permitan desempeñarse en ocupaciones específicas que requieran semi calificación o calificación inicial. El desarrollo de los cursos implica dos fases: la primera, denominada "Formación Técnica", se desarrolla en

⁴Una vez que las entidades pasaron el proceso de evaluación, pasan a concursar mediante la aplicación de un algoritmo matemático, aquellas ofertas cuyo ratio precio-calidad técnica resulten las más adecuadas para el programa.

las entidades de capacitación y otra de "Práctica Laboral", realizada en las empresas con las cuales se tiene convenio.

La fase de Formación Técnica tiene una duración promedio de 3 meses (300 horas como máximo), e involucra el desarrollo de habilidades técnicas específicas, y promueven actitudes y valores indispensables para el trabajo . Durante esta fase, la capacitación es gratuita y el Programa otorga a los beneficiarios una subvención mensual para gastos de movilidad y refrigerio y un seguro médico. En el caso de las mujeres con hijos, se brinda un subsidio adicional por cada niño menor de cinco años, a fin de promover su participación en el programa. Luego de la fase lectiva se realizan prácticas laborales en empresas sujetas al régimen laboral privado. PRO Joven exige a las entidades que suscriban "Cartas de Intención" con empresas a fin de asegurar que los cursos tendrán una fase práctica y que los jóvenes se capacitarán en actividades efectivamente demandadas por el sector productivo . Las empresas pagan la subvención a los alumnos durante la fase práctica y coordinan con las entidades de capacitación el contenido de los cursos que se van a dictar, logrando una mayor pertinencia. Estas prácticas tienen una duración mínima de 3 meses e implican la suscripción de un Convenio de Formación Laboral Juvenil (CFLJ)

Finalmente, actualmente el PROJoven cuenta con financiamiento del gobierno alemán (KFW) en un esquema de reconversión de deuda externa por inversión social por DM 12 millones (primer tramo) y 4.52 millones (segundo tramo) de marcos alemanes (además de las contrapartidas nacionales proporcionadas por el MTPS). También se cuenta con donaciones de US\$ 1.5 millones del Fondo Nacional de Capacitación Laboral y Promoción del Empleo (FONDOEMPLEO), UNFPA, donaciones PNUD y préstamo del BID.

3 Metodología

En esta sección, se plantea el problema de la evaluación de programas sociales en general y dentro de este contexto, describimos y justificamos las estrategias de estimación que se utilizan en el presente documento para la evaluación de PROJoven.

3.1 El Problema la Evaluación de Programas Sociales

En términos generales, para evaluar un programa es importante evaluar sus efectos (“efecto de tratamiento”) sobre algún resultado de interés (ej: ingresos) experimentado por los individuos de una población determinada (ej: beneficiarios de PROJoven).

Para definir formalmente el problema usaremos la notación estándar de la literatura de evaluación de programas sociales⁵:

- Y_{1i} : el resultado del individuo i si éste fue expuesto al tratamiento.
- Y_{0i} : el resultado del individuo i si éste no fue expuesto al tratamiento.
- $D_i \in \{0, 1\}$: indicador del tratamiento recibido por el individuo i ; 1 si recibió el tratamiento y 0 si no fue así.
- X : un conjunto de características del individuo fuera del tratamiento.

Consecuentemente, el efecto causal del tratamiento (programa) sobre el individuo i , en términos de la variable de resultados Y , sería $\Delta_i = Y_{1i} - Y_{0i}$. No obstante, en la realidad sólo podemos observar $Y_i = D_i Y_{1i} + (1 - D_i) Y_{0i}$ para cada uno de los individuos: es imposible observar Y_{1i} e Y_{0i} para el mismo individuo i . Entonces, el problema fundamental de la evaluación de programas sociales consiste en que es imposible *observar* el efecto de tratamiento individual.

Dada esta restricción, mucha literatura se ha concentrado en la estimación de impactos promedio. El parámetro que más atención ha recibido en este sentido es el efecto promedio de tratamiento sobre los tratados (*average treatment effect on the treated*). Este se define como:

$$\Delta(X) = E(\Delta | X, D = 1) = E(Y_1 | X, D = 1) - E(Y_0 | X, D = 1) \quad (1)$$

Donde el símbolo “ $|$ ” significa “*condicional en*”. Este efecto, como su nombre lo indica, nos da el impacto que en promedio tiene el programa sobre la variable de resultados para los individuos que participan en el programa, condicional a un vector de características individuales. Su utilidad reside en que permite evaluar el beneficio bruto de un programa existente, y constituye por tanto una parte importante de la evaluación costo - beneficio del mismo.

⁵En particular, en el presente documento seguimos la notación de Heckman, Ichimura y Todd, “Matching as an Econometric Evaluation Estimator”, *Review of Economic Studies*, vol. 65(2), Abril 1998.

Si bien los datos que normalmente se poseen para realizar la evaluación de un programa permiten estimar $E(Y_1 | X, D = 1)$, las principales dificultades estriban en la estimación de $E(Y_0 | X, D = 1)$ ⁶.

3.2 Experimentos Sociales y Matching

Los métodos basados en la creación de grupos de comparación usan datos de individuos no participantes para estimar esta esperanza, para lo cual se asume que, condicional en el vector de características X , los resultados de los no participantes se aproximan a lo que los participantes hubieran obtenido de no haber participado en el programa. Es decir, de manera formal, se asume que $\{Y_{1i}, Y_{0i} \perp D_i\} | X_i$. Este supuesto⁷, conocido como *selección en observables* (Rubin, 1974, 1977), implica:

$$E(Y_0 | X, D = 0) = E(Y_0 | X, D = 1) \quad (2)$$

Lo cual permite identificar $\Delta(X)$. Este supuesto puede garantizarse por definición por medio de la realización de un experimento social⁸, donde la admisión al programa se realice por medio de determinadas variables de certificación (X 's) y luego la asignación al tratamiento sea aleatoria.

Sin embargo, aunque no se haya realizado un experimento de esta naturaleza, bajo este supuesto podemos generar grupos de comparación que sean similares al grupo de control de un experimento, lo cual nos garantizaría que se cumpla la implicancia de la última ecuación. Este es el principal objetivo de los métodos de *matching* (calce o pareo): generar un grupo de control con el cual se pueda comparar el efecto sobre los tratados. La creación de un grupo de control adecuado facilita la identificación de $\Delta(X)$, en la medida que $E(Y_0 | X, D = 0)$ sí puede ser estimado usando datos observables en la realidad. No obstante lo anterior, la realización de esta estimación posee los típicos problemas econométricos de selección de modelos: hay que escoger adecuadamente las variables X 's y las formas funcionales necesarias para estimar las ecuaciones que relacionan Y_1 , Y_0 y X , para nombrar las dos principales dificultades.

El evitar estos problemas es la importante ventaja de los experimentos sociales aleatorios en comparación con otros métodos de evaluación no experimentales. El gran inconveniente de la aleatorización consiste en su costo, a veces no sólo alto económicamente sino también políticamente.

3.3 Propensity Score

Uno de los principales problemas de los métodos tradicionales de matching consiste en el condicionamiento por las variables X 's. Para la construcción del grupo de control debemos encontrar individuos no tratados que sean similares a individuos tratados,

⁶Como podemos ver, teóricamente este problema es equivalente al de observaciones perdidas (*“missing data”*): no se posee las observaciones de Y_{0i} para los participantes en el programa.

⁷En realidad, para la presente evaluación de PROJoven necesitamos sólo $\{Y_{0i} \perp D_i\} | X_i$, lo cual nos permite también obtener como implicancia la ecuación siguiente.

⁸Bajo ciertas condiciones. Ver Heckman, Ichimura, Smith y Todd: “Characterizing Selection Bias Using Experimental Data”, *Econometrica* (66), Setiembre, 1998.

en términos de las variables X 's. Es decir, el vector de características individuales X debe estar cercano entre estos individuos, bajo cierta métrica. Si este vector X está compuesto por muchas variables, resulta difícil definir en la práctica el grado de cercanía entre dos valores de X .

Esta dificultad en la práctica hace muy compleja la formación de grupos de control adecuados, perjudicando cualquier posible evaluación de programas sociales. En uno de los trabajos más influyentes en esta línea, Lalonde (1986) mostró que existían grandes diferencias entre los resultados de estimadores no experimentales estándar, en comparación con los resultados arrojados por un experimento aleatorio, en la tarea de estimar los efectos de un programa de entrenamiento laboral en Estados Unidos.

Las importantes diferencias entre estos resultados han sido atribuidas a la deficiente conformación de los grupos de control en el estudio de Lalonde. Dehejia y Wahba (1999) mostraron que existían grandes diferencias entre los grupos de control y los individuos sujetos a tratamiento en términos de las variables de características individuales. La principal causa de esta deficiencia yacía en el hecho de que Lalonde conformó los grupos de control restringiendo en algunas características básicas a los controles, para que se asemejaran a los individuos del grupo de tratamiento. Sin embargo, este procedimiento resultó a todas luces insuficiente en la medida que los controles resultaban muy distintos de los tratados en variables clave como edad, nivel educativo, situación marital e ingresos antes del tratamiento (ver Tabla 1 en Dehejia y Wahba, 1999).

Para reducir la complejidad del condicionamiento por las variables X 's en dicho trabajo se propuso el uso del *propensity score*, o probabilidad de haber sido sujeto al tratamiento, para reflejar cualquier diferencia en las variables definidas antes del tratamiento X (*pre-treatment variables*). En este sentido, resulta muy importante el siguiente teorema planteado por Rosebaum y Rubin (1983):

(RR 1983): *Sea $p(X_i)$ la probabilidad de que el individuo i haya sido sujeto de tratamiento, definida como $p(X_i) \equiv \text{Prob}(D_i = 1 \mid X_i) = E(D_i \mid X_i)$, donde $0 < p(X_i) < 1$. Entonces:*

$$\{Y_{1i}, Y_{0i} \perp D_i\} \mid X_i \Rightarrow \{Y_{1i}, Y_{0i} \perp D_i\} \mid p(X_i) \quad (3)$$

Este teorema establece, en otras palabras, que obtenemos los mismos resultados si en vez condicionar las distribuciones de las variables de resultados por el vector de características X , controlamos únicamente por la probabilidad de haber sido tratados dado el valor de este vector, valor que es sólo un escalar: $p(X)$, conocido como *propensity score*. Esta tarea resulta a todas luces mucho más sencilla y por tanto más factible de implementar en la práctica.

En este sentido, si podemos razonablemente suponer la selección en determinadas variables observables X , entonces podemos estimar el efecto de tratamiento sobre los tratados de la siguiente forma:

$$\Delta(X) = E(Y_1 - Y_0 \mid p(X), D = 1) = E(Y_1 \mid p(X), D = 1) - E(Y_0 \mid p(X), D = 0) \quad (4)$$

3.3.1 Estimación del Propensity Score

El *propensity score*, que no es más que la probabilidad de haber pasado por el programa, $p(X)$, se puede estimar para cada uno de los individuos de la muestra de tratados y potenciales controles. El objetivo es obtener una medida de la comparabilidad de los individuos del grupo de control (individuos “no tratados”) con respecto a aquellos del grupo de individuos sujetos a tratamiento; en términos de tener valores de *propensity score* más cercanos.

La estimación del propensity score consiste simplemente en la estimación un modelo de elección discreta para modelar la variable de participación en el programa, condicional a un vector de características individuales que pueden haber influido en dicha posibilidad. En este sentido, el propensity score puede estimarse con un modelo PROBIT o LOGIT, utilizando como variables explicativas diferentes características individuales que pueden determinar la posibilidad de que el individuo haya pasado por PROJoven.

3.3.2 Matching No Paramétrico sobre el Propensity Score

Conocido (habiendo estimado) el propensity score, es posible identificar (estimar) el efecto del tratamiento sobre las personas participantes en el programa, tal como se puede apreciar en la ecuación (4).

En particular, una estimación no paramétrica puede realizarse del siguiente modo: para cada observación i en la muestra de tratados, se conforma un promedio ponderado de observaciones de la muestra de comparación. Luego, se estima el efecto de tratamiento sobre i , tal como se aprecia a continuación:

$$Y_{1i} - \sum_{j \in \{D=0\}} W_{N_0 N_1}(i, j) Y_{0j} \quad (5)$$

donde $\{D = 0\}$ es el conjunto de índices para los no tratados, N_0 es el número de observaciones en el grupo de comparación, N_1 es el número de observaciones en el grupo de tratamiento y $\sum_{j \in \{D=0\}} W_{N_0 N_1}(i, j) = 1$ para todo i . Esta ponderación se construye sobre la base de un supuesto de distribución paramétrica local, que es la base de la estimación no paramétrica.

En el caso del estimador Kernel, el efecto de tratamiento promedio estimado para todo el programa está dado por:

$$\frac{1}{N_1} \sum_{i \in \{D=1\}} \left[Y_{1i}(X_i) - \sum_{j \in \{D=0\}} W_{N_0 N_1}(i, j) Y_{0j} \right] \quad (6)$$

En esta ecuación, el término $\frac{1}{N_0} \sum_{j \in \{D=0\}} W_{N_0 N_1}(i, j) Y_{0j}$ es un estimador de $E(Y_{0i} | p(X_i), D_i = 0)$. Es respecto de este término que difieren los diversos estimadores de matching no paramétrico sobre el *propensity score* que se describen en la subsiguiente sección.

3.3.3 Matching Paramétrico sobre el Propensity Score

Los estimadores de regresión paramétricos más frecuentemente utilizados en el trabajo empírico imponen supuestos de distribución comunes a todo el rango de datos y supuestos de linealidad.

La discusión que sigue se basa en Heckman, Ichimura, Smith y Todd (1997). Uno de los principales modelos de estimación paramétrica es el de Barnow, Cain y Goldberger (1980) que asume que Y_0 está relacionado linealmente con las variables en el vector X y un componente no observable U_0 , de modo que tenemos:

$$E(Y_0 | X, D = 0) = X\beta + E(U_0 | X, D = 0) \quad (7)$$

con $E(U_0 | X, D = 0) = E(U_0 | X)$ lineal en X , de modo que es posible estimar $E(Y_0 | X, D = 1)$ usando data de individuos no participantes: $E(Y_0 | X, D = 0)$ (recordar que es posible condicionar simplemente por $p(X)$). En general, no se imponen condiciones de soporte común (usar individuos participantes y no participantes que tengan similares *propensity scores*) para las estimaciones según este método, con lo cual la distribución de las X pueden ser muy diferentes en las muestras de tratados y no tratados. Las comparaciones se realizan imponiendo linealidad y extrapolando los resultados (proyección lineal).

La principal diferencia con los estimadores no paramétricos es que éstos últimos realizan supuestos paramétricos sobre la distribución de Y de manera local, es decir, para sólo intervalos predefinidos (anchos de banda o *bandwidths* alrededor de las observaciones de individuos tratados). Contrariamente, los estimadores paramétricos asumen una única distribución para todo el rango de datos.

Una de las principales ventajas de la estimación paramétrica es su simplicidad, y que no requiere de muchas observaciones para que sus resultados sean robustos a diferentes especificaciones (sobre todo cuando se trabaja con tamaños de muestra pequeños). No obstante, cabe resaltar, implican de todas maneras supuestos algo más restrictivos sobre los procesos generadores de los datos. Entre los estimadores paramétricos más comunes, podemos encontrar al de diferencia en diferencias.

El estimador de diferencia en diferencias consiste en el cálculo de la diferencia en el cambio promedio de la variable en evaluación, desde antes a después del programa o tratamiento. Estas estimaciones, no obstante, pueden presentar serias distorsiones debido a que, por un lado, los resultados no controlan por variables como características individuales (personales y laborales), y por otro, a que los grupos de beneficiarios y controles presentan heterogeneidades al interior de categorías laborales (ver estadísticas descriptivas), siendo éstas relevantes para la acreditación en el programa. Dadas estas observaciones, en el presente trabajo se intenta lidiar con estos problemas a través de diferentes técnicas de estimación, paramétricas y no paramétricas.

Con respecto al tema de la simplicidad, los estimadores paramétricos son de gran ayuda en las estimaciones posteriores, en la medida que ayudan a resolver problemas respecto a la naturaleza de los datos sin demasiadas complicaciones. En particular, es el caso en el presente trabajo que la variable de evaluación del proyecto se trata de una variable censurada, como la variable ingresos laborales. Esta variable está cen-

surada en cero, donde colapsa la distribución debido a los individuos que se encuentran desempleados o inactivos, y que por lo tanto perciben ingresos laborales iguales a cero.

En este sentido, una especificación TOBIT permite incorporar este problema a las estimaciones e intentar corregirlo. En particular, se puede plantear una relación lineal de las características del individuo (Z_i) con su resultado sobre el nivel de ingresos luego del programa como la siguiente:

$$Y_i = Z_i\beta + U_i \quad (8)$$

donde la variable de resultado (“ingresos laborales” por ejemplo) depende de las características de cada individuo, incluyendo la probabilidad de haber pasado por el programa (que nos ayuda a controlar por el vector de características X que puede haber determinado la participación del individuo), y se encuentra censurada en cero debido a los individuos que se encuentran desempleados. Corrigiendo por este problema inherente a este tipo de datos podemos obtener estimaciones razonables del efecto de tratamiento sobre los ingresos laborales ex-post de cada individuo.

3.3.4 Estimadores No Paramétricos Utilizados

Estimadores “Uno a Uno”

El estimador de matching tradicional es el del vecino más cercano (*nearest neighbour*) según el cual cada individuo del grupo de tratamiento se para con un individuo del grupo de control, de modo que sea el más cercano en términos de $p(X)$. Es decir, se escoge el individuo no tratado j para ser el contrafactual del individuo i de modo que, formalmente, el grupo de control del individuo i , $C^o(p_i)$, con *propensity score* p_i es sólo un individuo j que cumple:

$$C^o(p_i) = \{j : |p_i - p_j| = \min_{k \in \{D=0\}} \{|p_i - p_k|\}\} \quad (9)$$

de modo que $W_{N_0N_1}(i, k) = 1$ si $k = j$ y es igual a cero de otro modo. Este estimador utiliza a sólo un individuo del grupo de control para comparar con cada individuo del grupo de tratamiento. Este estimador puede refinarse si es que se especifica que el individuo de control no puede tener un *propensity score* demasiado alejado de aquel del individuo del grupo de tratamiento. Por ejemplo se puede especificar que nunca la distancia entre ambos debe ser mayor a δ de modo que el grupo de control sería:

$$C^o(p_i) = \{j : \delta > |p_i - p_j| = \min_{k \in \{D=0\}} \{|p_i - p_k|\}\} \quad (10)$$

de modo que las ponderaciones serían $W_{N_0N_1}(i, k) = 1$ si $k = j$ y si $\delta > |p_i - p_j|$ y cero de otro modo. Este estimador se conoce como el *Caliper Matching*. En este caso, si no existe individuo del grupo de control que sea comparable en estos términos con el individuo del grupo de tratamiento, entonces se elimina a este último de la estimación del efecto del tratamiento.

Estimadores "Ponderados Suavizados"

Usando los estimadores anteriores podemos perder a varios individuos del grupo de tratamiento si es que no existen individuos en el grupo de control que sean "parecidos" en los términos definidos previamente. Esto puede llevar a desperdiciar información de individuos del grupo de control. Asimismo, según estos métodos confiamos en un único individuo del grupo de control como contrafactual para cada uno de los integrantes del grupo de tratamiento. Esto último nos indica que podríamos mejorar la eficiencia de nuestras estimaciones si pudiésemos incluir a otros individuos del grupo de control que sean también "similares" a los del grupo de tratamiento, donde como siempre la similitud está dada por la cercanía de sus valores de *propensity score*. Para este fin usamos en las estimaciones presentadas en este trabajo otro tipo de estimadores que utilizan una mayor cantidad de individuos del grupo de control, ponderándolos para compararlos con los del grupo de tratados. Estos estimadores se conocen como "Estimadores de Matching Ponderados Suavizados" (*"smoothed weighted matching estimators"*).

Uno de los estimadores de este tipo más utilizados es el estimador Kernel. Este puede utilizar a todos los individuos del grupo de control⁹ ponderándolos a la hora de compararlos con los individuos del grupo de tratamiento con los siguientes pesos $W_{N_0N_1}(i, j)$ (ver ecuación 5):

$$W_{N_0N_1}(i, j) = \frac{K\left(\frac{p(X_i) - p(X_j)}{h_n}\right)}{\sum_{k \in \{D=0\}}^{N_0} K\left(\frac{p(X_i) - p(X_k)}{h_n}\right)} \quad (11)$$

siendo $K(\cdot)$ la función kernel y h_n el ancho de banda, o parámetro de suavización. Dado que si utilizamos toda la muestra de potenciales controles es probable que las distribuciones de características individuales X sean muy distintas con respecto a las del grupo de control, y que a pesar de ello estos individuos tengan ponderaciones positivas a la hora de compararse con individuos tratados, muchas veces es útil restringir el grupo de control a un subconjunto menor para tener comparaciones más precisas. En este sentido, en general se suele imponer la condición de *soporte común*, que significa que sólo se usarán a individuos del grupo de control que no tengan valores de *propensity score* por debajo del valor mínimo para los tratados, y se descartarán también a aquellos individuos tratados con valores de *propensity score* por encima del máximo de aquellos del grupo de control. Es decir, se toman las regiones en las cuales las distribuciones de *propensity score* para cada uno de los grupos de control y tratamiento se traslapan. Por las razones presentadas líneas arriba, en el presente trabajo se utilizan tanto este último estimador, como aquellos descritos anteriormente como el de "vecino más cercano" y el "*Caliper Matching*".

⁹De modo que $C^o(p_i) = \{D = 0\}$.

4 Estimación

Para cuantificar el impacto del programa es necesario definir las variables sobre las cuales se estimará el mencionado impacto. Dadas las características y objetivos del programa, una de las variables más importantes sobre la cual se estimará este impacto será el ingreso laboral. Se espera que el ingreso del grupo de tratamiento (beneficiarios) luego de haber pasado por el programa sean superiores a los del grupo de control. Adicionalmente, debido a que el total de horas trabajadas puede afectar las medidas de ingresos en sus distintas frecuencias también es bueno calcular el impacto sobre los ingresos por hora; asimismo, de manera separada, también es interesante estimar el impacto sobre el total de horas trabajadas a la semana.

4.1 Estadísticas Descriptivas

Brevemente describiremos las características principales de la línea de base, primera y segunda medición, para los beneficiarios y controles. Para construir la línea de base del grupo de beneficiarios se realizó un muestreo aleatorio y estratificado (por género y edad). Los beneficiarios entrevistados son una muestra de 327, de un total de 1807 beneficiarios que terminaron el programa. La encuesta (sociolaboral) a este grupo se aplicó en las ECAP's en donde estaban siendo capacitados.

Como mencionamos anteriormente, los 327 controles o potenciales beneficiarios se encontraron en las zonas que se identificaron con mayor concentración de muestra de beneficiarios (conglomerados mapeados - puerta por puerta o entorno de vivienda) a los cuales se les aplicaba la misma encuesta.

La información de la segunda medición tanto para el grupo de beneficiarios como de control se obtuvo a partir de la encuesta que se les aplicó 12 meses después de finalizada la fase lectiva. La información adicional que permite obtener esta encuesta está relacionada a los indicadores laborales y de historial laboral (entre los meses de enero de 1999 y julio del mismo año). Asimismo, permite identificar en ambos grupos el acceso a capacitación (y su duración) recibida durante ese período. Esta última variable es importante ya que permite controlar los resultados evitando que el impacto sobre los ingresos se sobrestime o subestime.

En los Cuadros 7 y 8 del anexo se presentan las principales características del grupo de control elaborado por PROJoven y el grupo de beneficiarios del programa (antes de pasar por el tratamiento). En términos de las variables de género, edad y escolaridad, ambas muestras tienen valores promedios muy cercanos, lo cual indica que el grupo de control puede ser un buen instrumento para obtener el contrafactual del grupo de beneficiarios. Las mayores diferencias se presentan en las variables de estado civil (variable dummy soltero) y en la de carga familiar (variable dummy si el individuo tiene hijos). Con respecto a los indicadores laborales, observamos que los ingresos promedios (tanto mensual como por hora) de ambos grupos son muy similares.¹⁰

Las similitudes entre los grupos de control y beneficiarios en la línea de base se repiten en variables de situación laboral. En el Cuadro 9 del anexo se presentan las

¹⁰En este cálculo sólo se consideran a los individuos ocupados en el mercado laboral.

estadísticas de ocupación e inactividad para ambos grupos, las cuales reflejan las mencionadas similitudes. Según estos resultados, el porcentaje de empleados de 56% para los individuos del grupo de control y de un 60% para los del grupo de beneficiarios, y porcentajes de desocupados de 23% y 21% del mismo modo¹¹. Sin embargo, al mirar al interior de ambos grupos observamos que existen diferencias en lo que concierne al desagregado de las categorías laborales en la línea de base, por lo que el grupo de control y el de tratamiento no serían exactamente iguales. Como podemos observar, en la línea de base, el porcentaje de controles que se encontraban trabajado como trabajadores familiares no remunerados era de 12%, mientras que en el grupo de beneficiarios 19%.

En la siguiente sección haremos una primera aproximación del efecto de tratamiento utilizando comparaciones simples en los resultados de los grupos de control y beneficiarios.

4.2 Estimación de Diferencia en Diferencias

El primer acercamiento a la estimación del efecto del tratamiento se realiza a través del estimador de diferencia en diferencias, tradicionalmente utilizado en la literatura de evaluación de impacto, lo que servirá para fines referenciales comparativos.

Este efecto se puede estimar a través del coeficiente de la variable muda que indica el tratamiento en la siguiente regresión:

$$Y_j - Y_0 = \alpha + \beta D + \epsilon \quad (12)$$

Donde Y es la variable de resultado (ingresos u horas trabajadas) y D es el indicador de tratamiento. El subíndice j denota el número de la medición (primera, segunda o tercera) y 0 la línea de base.

Los resultados de este estimador sobre las variables de ingresos laborales mensuales, ingreso laboral por hora y horas trabajadas se presentan en los Cuadros 11 y 12. En el primero, se presenta los resultados considerando la muestra censurada por el lado izquierdo (cero), puesto existe un porcentaje de individuos que no reciben ingresos laborales. En el segundo, se presentan los resultados para los datos truncados, es decir, considerando sólo aquellos individuos que perciben ingresos. Estas dos formas de analizar los resultados buscan corregir los potenciales sesgos por el censuramiento de las variables de resultados. La presencia de estos sesgos podría tener efectos distorsionadores¹² sobre las estimaciones de impacto subsecuentes.

Los resultados son consistentes con lo encontrado en algunas evaluaciones previas de PROJoven¹³. El impacto promedio sobre la variable ingresos es positivo, aunque no son siempre estadísticamente significativos. En particular, si bien el efecto sobre el ingreso promedio mensual es positivo y significativo, los efectos positivos sobre horas trabajadas e ingreso laboral por hora no lo son.

¹¹Por cuestiones de comparación, en el cuadro 9 hemos incorporado la información de la segunda medición.

¹²Los efectos pueden sobreestimar o subestimar las estimaciones de impacto del programa.

¹³Galdo, José 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, CIUP-IRDC.

Con respecto a los efectos sobre el ingreso promedio mensual cabe mencionar que el mayor impacto se observa a los 12 meses de finalizada la fase lectiva (es decir, en la segunda medición). Estos resultados serían consistentes con algunas hipótesis sobre la capacitación laboral en actividades de calificación inicial, la cuales sostienen que por un lado las habilidades de estos individuos se deprecian rápidamente y por otro, este segmento demográfico se enfrenta a una alta rotación laboral (Ver Cuadros 11 y 12).

4.3 Estimación del Propensity Score

Como se ha descrito en la sección metodológica, la técnica que aplicamos en este trabajo requiere la estimación de la probabilidad de haber sido sujeto al tratamiento, condicional en variables de atributos individuales y familiares observables. En particular, las variables que encontramos relevantes para realizar el emparejamiento (*matching*) con los individuos del grupo de beneficiarios fueron las siguientes: edad, género, nivel educativo (dummy por secundaria completa), estado civil (dummy si el individuo es soltero), ingreso laboral por hora (antes del programa), horas trabajadas a la semana, categoría ocupacional (dummies por trabajador familiar no remunerado, trabajador por cuenta propia, obrero del sector privado), educación de los padres, tamaño de la empresa en que trabaja (dummy que representa empresas con menos de 5 trabajadores), carga familiar (dummy si el individuo tiene hijos), entre otras (ver Anexo).

En el Cuadro 13 del Anexo se presentan los resultados de la estimación del *propensity score* para los individuos de la muestra completa. El bajo poder explicativo del modelo especificado (pseudo R^2 de 7,0%) radica en el hecho de que el grupo de control ha sido creado de manera bastante minuciosa: las diferencias entre los grupos de control y el de beneficiarios deben ser explicadas mayormente por variables no observables. En el caso de un experimento social, con asignación perfectamente aleatoria al tratamiento, ninguna variable podría explicar la asignación al mismo por definición. En este sentido, esto es una medida indirecta del grado de efectividad del grupo de control creado por PROJoven.

Cabe resaltar que aunque el modelo no sirva para explicar en gran medida las diferencias en la distribución de los ingresos, nos servirá para definir los conjuntos de comparación¹⁴ para cada individuo del grupo de beneficiarios y la estimación no paramétrica del efecto de tratamiento sobre estos. Las ventajas de la estimación no paramétrica sobre los métodos de regresión tradicionales han sido delineadas en la sección metodológica.

Otro punto necesario de aclarar es la inclusión de variables con relativo bajo poder explicativo en la estimación del *propensity score*. La idea principal en este sentido, es la de lograr el mayor grado de ajuste posible: se busca la mejor predicción de probabilidades para los individuos de la muestra. En los Gráficos 5 y 6 del anexo, se presentan estas proyecciones para los individuos de los grupos de control y tratamiento. Específicamente, en el gráfico 3 se presenta el histograma de estas proyecciones, mientras que en el 4 se muestran las densidades estimadas a partir de estas. Uno de los temas

¹⁴Específicamente, el grado de comparabilidad de cada individuo del grupo de control respecto de cada individuo del grupo de tratamiento.

más importantes para la precisión de los estimaciones del efecto de tratamiento por medio del matching radica en la comparabilidad de ambos grupos. En estos gráficos se muestra un alto grado de coincidencia entre las distribuciones de probabilidades de ambos grupos lo cual es evidencia en favor de este supuesto. En el caso negativo (extremo) de que ambas distribuciones no se traslapen lo suficiente, sería imposible realizar estimaciones precisas del efecto de tratamiento. Debido a que las estimaciones realizadas con la ENAHO dieron resultados poco satisfactorios en este sentido, en este trabajo sólo se presentan resultados utilizando el grupo de comparación propuesto por el propio programa PROJoven.

El paso siguiente es la estimación del efecto de tratamiento, lo cual se aborda en la siguiente sección.

4.4 Estimación del Efecto de Tratamiento sobre los Tratados

Las estimaciones de impacto del programa se focalizan en las variables de ingresos laborales del participante, ingreso familiar, horas trabajadas y probabilidad de estar empleado. En las siguientes secciones se presentan los resultados según las técnicas descritas en la sección metodológica.

4.4.1 Técnica no Paramétrica: Propensity Score

En los Cuadros 14 y 15 del anexo se presentan los resultados de la estimación censurada y truncada de los efectos de tratamiento para las variables mencionadas. En cuanto a los de la muestra censurada, el Cuadro 14 muestra el impacto positivo del programa sobre la variable ingreso, tanto mensual como por hora. Sin embargo, al analizar la significancia estadística de estos resultados comprobamos que, mientras la variable ingreso mensual es significativa en todas las mediciones, el ingreso por hora no lo es en ninguna. Asimismo, la variable de horas trabajadas refleja también un efecto positivo y significativo para los individuos participantes. En este sentido, podría pensarse que el programa habría tenido efectos sobre la empleabilidad y en horas trabajadas, los cuales habrían sido los causantes del mayor ingreso laboral de los participantes. Los resultados sobre la muestra truncada presentados en el mismo cuadro, reflejan una situación similar.

Observamos que el mayor impacto sobre ingresos se presenta en la segunda medición, lo cual puede deberse a un aumento de las horas trabajadas, mientras que el efecto en la tercera medición es positivo y significativo, lo cual puede explicarse por un incremento tanto en las horas trabajadas como en el ingreso. Con respecto a los resultados de la primera medición muestran un impacto significativo sobre el ingreso mensual mas no en el ingreso por hora y esto se debería a que tanto las horas trabajadas como el ingreso mensual se incrementaron para ambos grupos.

El Cuadro 15 presentan los mismos resultados pero para una muestra en la que no se incluyen individuos con cursos de capacitación adicional, recibidos luego de la finalización de la fase lectiva de PROJoven. Estas últimas estimaciones proveen evidencia de la importancia del programa sobre el efecto de tratamiento, fuera del hecho de que algunos participantes hayan tenido acceso a cursos de capacitación adicional.

Este resultado indica claramente la importancia del programa sobre las variables de resultados.

Las metodologías empleadas en esta sección, sin embargo, no consideran las posibles distorsiones generadas por la naturaleza censurada de los datos (la distribución de los ingresos tiene un punto de masa en el valor cero). Esto se intenta corregir con las metodologías utilizadas en la siguiente sección.

4.4.2 Técnica Paramétrica: Tobit

En los Cuadros 16 y 17 se presentan los resultados de la estimación Tobit planteada en la sección metodológica. El Cuadro 16 presenta la estimación de la probabilidad de que el individuo se encuentre remunerado en la etapa posterior al programa. Esta estimación se realizó con el objetivo de intentar corregir por el sesgo que podría generarse por el hecho de que no todos los individuos se encuentran remunerados. Aunque conjuntamente se rechaza la hipótesis de que las variables explicativas no son significativas, el poder explicativo de este modelo parece muy reducido, como se puede inferir de su bajo ajuste (bajo R^2).

El Cuadro 17 presenta los resultados de la regresión final del modelo Tobit. En esta ecuación se relacionan diferentes variables explicativas con el nivel de la variable de impacto del programa. Los resultados encontrados son diferentes a los anteriores. En la presente estimación, la variable dummy tratado no es significativa, lo cual implicaría que el programa no tiene un efecto significativo sobre la variable de resultado: el ingreso laboral por hora de los individuos.

Este resultado presenta algunas dificultades. En primer lugar, el modelo que intenta estimar la probabilidad de estar remunerado resulta poco significativo (bajo coeficiente de determinación), por lo que no se estaría identificando claramente la probabilidad de pertenecer a la muestra (censurada) de cada individuo. En segundo lugar, la misma estimación del modelo Tobit presenta un bajo poder explicativo. Esto es un indicador asimismo de que el modelo en su conjunto no sería el más adecuado para estimar el efecto del programa. Consecuentemente, en el presente trabajo consideramos más precisas las estimaciones presentadas anteriormente (propensity score matching), por lo que no se toma en cuenta en las conclusiones la poca significancia del indicador de tratamiento sobre el nivel de ingresos en esta regresión. Asimismo, estos resultados dan evidencia a favor de la hipótesis de ausencia de sesgo de selección en la muestra considerada.

5 Conclusiones

El presente estudio estimó los impactos del programa de capacitación laboral juvenil PROJoven, sobre los participantes en el mismo. Para este objetivo, se recurrieron a diferentes metodologías estadísticas que mejoran la comparabilidad entre los grupos de participantes y no participantes. Estas metodologías permiten, por ende, mejorar la estimación del efecto antes mencionado.

De las diversas metodologías empleadas, las más robustas fueron las técnicas no paramétricas de propensity score matching; contrariamente, las técnicas paramétricas, con sus supuestos aparentemente muy restrictivos (hipótesis de sesgo de selección, linealidad, etc.), tuvieron un pobre desempeño.

Es así que consideramos como el principal resultado el proveniente de la técnica de propensity score matching, cuyos resultados fueron además consistentes con los obtenidos por medio de estimadores de diferencia en diferencias, tradicionalmente utilizados en la evaluación del programa. En cuanto a la variable ingresos laborales mensuales, el hecho que los jóvenes sean sujetos de tratamiento (participen en PROJoven) incrementaría sus ingresos en 156 soles (para el caso de la muestra censurada), y en 144 soles (para el caso de la muestra truncada) con respecto al grupo de no participantes¹⁵. Estos resultados son robustos aún si consideramos únicamente individuos que no han recibido capacitación posterior al programa. Para el caso de la muestra censurada, también se obtiene un aumento en el número de horas trabajadas a la semana para los participantes en comparación con los no participantes. Esto no se repite, sin embargo, para la muestra truncada.

Consecuentemente, podemos concluir que, independientemente de la técnica utilizada para la evaluación de impacto del programa, los efectos parecen ser significativos y robustos. Sin embargo, creemos recomendable y necesario complementar la evaluación de impacto del programa con una evaluación financiera, que permita identificar el costo económico y social del programa, para una evaluación costo-beneficio completa. Asimismo, creemos que evaluaciones económicas de este tipo deben ser complementadas con evaluaciones cualitativas, como las denominadas “evaluaciones de proceso” para las siguientes convocatorias, de tal forma que sea posible identificar la eficiencia de cada una de las áreas que conforman el programa¹⁶.

Finalmente, dados los resultados de nuestra evaluación, sugerimos la implementación de evaluaciones permanentes del programa que permitan discernir la factibilidad de institucionalizarlo (de ser necesario) dentro de la política pública que realiza el Ministerio de Trabajo.

¹⁵Esto si consideramos los efectos después de 12 meses de finalizada la fase lectiva del programa

¹⁶Por ejemplo, debe ser importante verificar la eficiencia del componente de intermediación laboral del programa. Esto, en el sentido de que existen actualmente otras alternativas de intermediación laboral que pueden ser menos costosas: bolsas virtuales, bolsas privadas, sistema nacional de empleo -SIL-, etc.

Referencias

Bravo David y Contreras Dante (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 Economic Research.

Burga Cybele y Moreno Martín (2001), "¿Existe Subempleo Profesional en el Perú?. Publicaciones Breves No.13, CIES.

Dehejia R, Wahba S, (1998), "Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs", Working paper series, 6586.

Heckman J, Lalonde R, Smith J (1998) "The Economics and Econometrics of Active Labor Market Programs, Handbook of Labor Economics, vol III. (cap 1- 7)

Heckman J, Ichimura H., Todd Petra, (1997) "Matching As An Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme", Review of Economic Studies, 64,605-654

Heckman J, Ichimura H., Todd Petra, (1998) "Matching As An Econometric Evaluation Estimator, Review of Economic Studies, 65,261-294

Heckman J, Ichimura H., Smith J, Todd Petra, (1997) "Characterizing Selection Bias Using Experimental Data".

Heckman J, Vytkačil Edward (2000) "Causal Parameters, Structural Equations, Treatment Effects and Randomized Evaluations of Social Programs", Universidad de Chicago y American Bar Foundation. Version a Revisar.

Informes Internos del Equipo de PROJoven, Marzo del 2002.

Boletines de Economía Laboral (BEL), Junio 1998, Marzo 1999. (MTPS)

Lalonde Robert J. (1986) "Evaluating the Econometrics Evaluations of Training Programs with Experimental Data". The American Economic Review , vol 76, issue 4, 604-620.

Rosenbaum P.R, Rubin D.B.(1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects", Biometrika,70,1,41-55

ANEXO 1

Tabla 1: Descripción de Variables

Nombre de Variables	Descripción
Edad	Años de edad cumplidos por el beneficiario y el control
Sexo	Variable dummy por género
Secundaria de padre	Variable dummy equivalente a 1 si el padre terminó la enseñanza secundaria, 0 en caso contrario
Secundaria de madre	Variable dummy equivalente 1 si la madre terminó la enseñanza secundaria, 0 en caso contrario
Variabes relacionadas a la ocupación	Variabes dummy : Obrero del sector privado (=1 y 0 en caso contrario) Trabajador familiar no remunerado (=1 y 0 en caso contrario) Trabajador ambulante(=1 y 0 en caso contrario) Trabajado independiente(=1 y 0 en caso contrario)
Estado Civil	Variable dummy según el estado civil: soltero=1 y 0 en caso contrario
Categoría ocupacional	Variable dummy por participación en el mercado laboral equivalente a 1 si es inactivo, 0 en caso contrario
Secundaria completa	Variable dummy =1 si el individuo terminó la enseñanza secundaria, 0 en caso contrario
Tamaño de empresa	Variable que indica el tamaño de las empresas donde estaban trabajando los individuos que estaban ocupados en la línea base. Empresas con numero de hasta 5 trabajadores son consideradas pequeñas. Esta es una variable dummy =1 por empresa pequeña, 0 en caso contrario
Hijos	Variable dummy para indicar si el individuo tiene carga familiar
Duración del desempleo	Variable discreta que indica el tiempo que lleva desempleado un individuo que se encuentra desempleado medido en semanas, en la línea base
Ingreso laboral por hora	Variable continua que indica el ingreso laboral por hora de los individuos en la línea base
Horas	Variable que indica las horas trabajadas a la semana de los individuos en la línea base

ANEXO 2

Cuadro 1

Algunas características básicas de la población joven: Perú Total

	Rangos de edad			%
	14 - 19 años	20- 24 años	Jovenes	
Total	2,198,112	1,567,333	3,765,445	
<i>Nivel Educativo</i>				
Sin Nivel	8,474	6,969	15,443	0.4
Primaria incompleta	90,557	39,559	130,116	3.2
Primaria completa	131,812	58,844	190,656	4.7
Secundaria incompleta	1274422	223,720	1,498,142	37.0
Secundaria completa	561,862	640,368	1,202,230	29.7
Sup. No. Uni. Incompleta	59,352	191,564	250,916	6.2
Sup. No. Uni. Completa	5,828	121,840	127,668	3.1
Sup. Uni. Incompleta	65,805	244,097	309,902	7.6
Sup. Uni. Completa		40,372	40,372	1.0
<i>Acceso a Capacitacion</i>				
No	1752089	916,813	2,668,902	65.8
Si	446,023	650,520	1,096,543	27.1
Total	2198112	1854956	4,053,068	100.0
<i>Tipo de capacitacion *</i>				
CEO	153,721	228,193	381,914	31.3
CFS	26,303	67,721	94,024	7.7
IST	133,109	189,972	323,081	26.5
Universidad	22,396	58,395	80,791	6.6
CSVIT	40,729	15,617	56,346	4.6
Empresa	2,517	16,921	19,438	1.6
Centro de instruccion tecnica	5,447	13,517	18,964	1.6
Por Correspondencia	621	589	1,210	0.1
Otro	61,180	59,595	120,775	9.9
Total	446,023	773,296	1,219,319	100.0

Notas

* El total de personas que contestaron "tipo de capacitacion" fueron los que dijeron que se habian cap

Fuente:

Encuesta Nacional de Hogares del INEI (ENAHO), Tercer Trimestre del 2000

Cuadro 2

Indicadores laborales de la población joven y adulta: Perú Total

Indicadores Laborales	Total					Jóvenes		
	Jóvenes		Total	Adultos		Acceso a capacitación		
	[14-19]	[20-24]		de 25 a mas	Total	No	Si	Total
Lima Metropolitana								
Ocupado	272,730	411,840	684,570	2580220	3264790	438,037	246,533	684,570
No subempleado por ingresos	93,451	275,569	369,020	1878430	2247450	221,723	147,297	369,020
Subempleado por ingresos	179,279	136,271	315,550	701,790	1017340	216,314	99,236	315,550
No subempleado por horas	162,884	301,055	463,939	1868473	2332412	302,888	161,051	463,939
Subempleado por horas	109,846	110,785	220,631	711,747	932,378	135,149	85,482	220,631
Desocupado	55,457	69,044	124,501	153,493	277,994	84,767	39,734	124,501
Cesante	39,243	65,155	104,398	151,470	255,868	70,243	34,155	104,398
Anterior	16,214	3,889	20,103	2,023	22,126	14,524	5,579	20,103
Inactivos	606,724	249,439	856,163	1187833	2043996	621,271	234,892	856,163
Total	934,911	730,323	1,665,234	3,921,546	5,586,780	1,144,075	521,159	1,665,234
Resto Urbano								
Ocupado	436,333	479,869	916,202	2947383	3863585	651,986	264,216	916,202
No subempleado por ingresos	55,330	162,621	217,951	1556021	1773972	145,083	72,868	217,951
Subempleado por ingresos	381,003	317,248	698,251	1391362	2089613	506,903	191,348	698,251
No subempleado por horas	214,048	293,167	507,215	1896447	2403662	361,697	145,518	507,215
Subempleado por horas	222,285	186,702	408,987	1050936	1459923	290,289	118,698	408,987
Desocupado	53,132	71,099	124,231	164,267	288,498	73,206	51,025	124,231
Cesante	42,903	65,380	108,283	159,585	267,868	62,021	46,262	108,283
Anterior	10,229	5,719	15,948	4,682	20,630	11,185	4,763	15,948
Inactivos	773,736	286,042	1,059,778	1144106	2203884	799,635	260,143	1059778
Total	1,263,201	837,010	2,100,211	4,255,756	6,355,967	1,524,827	575,384	2,100,211

Fuente: Encuesta Nacional de Hogares del INEI (ENAHO), Tercer Trimestre del 2000

Nota:

1. Tasa de Desempleo: TD= PEA desempleada/PEA
2. Tasa de subempleo: TS= PEA subempleada (por horas o ingresos)/Ocupados
3. Tasa de inactividad: TI= Total de inactivos/PET

Elaboración propia

Cuadro 3

Indicadores laborales de la población joven y adulta: Perú Total

Indicadores Laborales	Total					Jóvenes		
	Jóvenes		Total	Adultos de 25 a mas	Total	Acceso a capacitación		
	[14-19]	[20-24]				No	Si	Total
Lima Metropolitana								
Tasa de actividad	35.1	65.8	48.6	69.7	63.4	45.7	54.9	48.6
Tasa de subempleo por ingresos	65.7	33.1	46.1	27.2	31.2	49.4	40.3	46.1
Tasa de subempleo por horas	40.3	26.9	32.2	27.6	28.6	30.9	34.7	32.2
Tasa de desempleo	16.9	14.4	15.4	5.6	7.8	16.2	13.9	15.4
TD de cesantes	12.0	13.5	12.9	5.5	7.2	13.4	11.9	12.9
TD de aspirantes	4.9	0.8	2.5	0.1	0.6	2.8	1.9	2.5
Tasa de inactividad	64.9	34.2	51.4	30.3	36.6	54.3	45.1	51.4
Resto Urbano								
Tasa de actividad	38.7	65.8	49.5	73.1	65.3	47.6	54.8	49.5
Tasa de subempleo por ingresos	87.3	66.1	76.2	47.2	54.1	77.7	72.4	76.2
Tasa de subempleo por horas	50.9	38.9	44.6	35.7	37.8	44.5	44.9	44.6
Tasa de desempleo	10.9	12.9	11.9	5.3	6.9	10.1	16.2	11.9
TD de cesantes	8.8	11.9	10.4	5.1	6.5	8.6	14.7	10.4
TD de aspirantes	2.1	1.0	1.5	0.2	0.5	1.5	1.5	1.5
Tasa de inactividad	61.3	34.2	50.5	26.9	34.7	52.4	45.2	50.5

Fuente: Encuesta Nacional de Hogares del INEI (ENAHO), Tercer Trimestre del 2000

Nota:

1. Tasa de Desempleo: TD= PEA desempleada/PEA
2. Tasa de subempleo: TS= PEA subempleada (por horas o ingresos)/Ocupados
3. Tasa de inactividad: TI= Total de inactivos/PET

Elaboración propia

Cuadro 4

**Resultados obtenidos por el PROJoven por convocatoria y según etapas de formación
Lima Metropolitana, Arequipa, Trujillo, Chiclayo, Cusco, Piura, Huancayo y Chimbote
(1996-2001)**

ETAPAS	Convocatorias								Total	%
	1a	2a	3a	4a	5a	6a	7a	8a		
	1996	1997	1998	1998	1999	2000	2001	2001		
Formación técnica	1505	1807	2243	2671	3075	3651	4179	5157	24288	100.0%
Alumnos egresados	1450	1729	2146	2527	2945	3481	4052	5010	23360	96.2%
Alumnos que abandonaron el curso	55	78	97	144	130	170	127	147	928	3.8%
Formación Práctico-Laboral	1450	1729	2146	2527	2945	3481	4052	5010	23360	100.0%
Alumnos con prácticas	1201	1443	1762	2056	2267	2768	3106	3880	18483	79.4%
Prácticas concluidas	861	909	1247	1385	1593	1968	2295	3097	13355	72.3%
Prácticas por períodos menores a tres n	340	534	515	671	674	760	811	783	5088	27.5%
Inserción laboral autónoma	25	28	40	50	77	95	102	220	637	2.7%
Prácticas pendientes	155	145	230	275	382	409	485	590	2671	11.5%
Salida del programa antes de las prácticas	69	113	114	166	219	249	224	320	1474	6.3%

Fuente: Ministerio de Trabajo y Promoción Social (MTPS)

PROJoven

*Cifras Preliminares - Marzo 2002

Cuadro 5

Criterios de Acreditación:

A. Primer criterio: Condición de Pobreza

Criterios Primarios

ITEM's	Categoría	Ptje.	Categoría	Ptje.	Categoría	Ptje.	Categoría	Ptje.
PISO	Tierra	4	Madera (Entablado)	3	Cemento	2	Losetas, vinílicos o parquet	0
TECHO	Materiales residuales	4	Esteras o caña	3	Calamina, triplay o eternit	2	Concreto armado o cemento	0
PAREDES	Esteras o cartón	4	Adobe o quincha o madera	3	Ladrillo sin tarrajear	2	Ladrillo tarrajeados	0
AGUA	Cursos Naturales. Camión Cisterna	4	Pozo o Pilón	3	Red Pública Fuera de la Vivienda pero dentro del Edificio		1	
					Red Pública Dentro de la Vivienda		0	
BAÑO	No tiene servicio	4	Pozo ciego, silo o letrina	3	Red pública fuera de la vivienda y compartida		2	
					Red pública fuera de la vivienda y exclusivo		1	
					Red pública dentro de la vivienda		0	
HACINAMIENTO	De 5 a más personas	4	De 4 a 5 personas	3	De 3 a 4 personas	2	Menos de 3 personas	0
NIVEL DE INSTRUCCIÓN DEL JH	Primaria incomp. o analfabeto	4	Secundaria incomp. o completa	2	Superior incompleta	1	Superior completa	0
TASA DE DEPENDENCIA	De 3 a más	4	Entre 2 y 3	2	Menor a 2	0		

Criterios Secundarios

1. Al menos 1/3 de los integrantes del hogar hace uso de algún programa de asistencia social.
2. El hogar cocina con combustible diferente a gas y no tiene teléfono.
3. Al menos 1 niño(a) de 8 - 12 años miembro del hogar, no asiste al colegio
4. El jefe del hogar se encuentra desempleado hace tres meses.

B. Segundo y Tercer Criterio de Acreditación

NIVEL DE INSTRUCCIÓN

- | | |
|---|----------------------|
| 1. Asiste a un colegio de tipo regular o no escolarizado de lunes a viernes | NO ACREDITADO |
| 2. Ha estudiado en centros pre-universitarios, institutos superiores o universidades . | NO ACREDITADO |
| 3. Ha estudiado cursos técnicos durante más de 300 horas. | NO ACREDITADO |
| 4. No asiste al colegio regular ni ha seguido cursos técnicos durante más de 300 horas. | ACREDITADO |

CONDICION LABORAL

- | | |
|---|----------------------|
| 1. No está desempleado ni subempleado. | NO ACREDITADO |
| 2. Está desempleado (<i>se cansó de buscar trabajo o está buscando trabajo</i>). | ACREDITADO |
| 3. Está subempleado por ingresos (<i>gana menos de un salario mínimo vital mensual</i>) | ACREDITADO |

Fuente:

Fuentes Internas de PROJoven.

Cuadro 6
Número de observaciones en cada base de datos

	Línea de Base	6 meses después de FP	12 meses después de FP	18 meses después de FP
Controles	317	303	282	277
Usuarios	317	321	299	282

Luego del Merge

Controles	272
Usuarios	288

Cuadro 7
Estadísticas Descriptivas
(Características personales)

	Hombre (%)	Edad (Años)	Escolaridad (Años)	Soltero (%)	Hijos (tiene) (%)
Grupo de Beneficiarios	40.6	20.5	10.7	91.0	15.6
Grupo de Control	40.8	20.3	10.7	74.6	32.0

Nota:

Muestra final de 560 individuos, 272 controles y 288 beneficiarios.

Cuadro 8
Estadísticas Descriptivas
(Características laborales)

	Ingreso Laboral Mensual	Ingreso Laboral por hora	Horas traba- jadas a la semana
Grupo de Beneficiarios	273.0	2.0	40.3
Grupo de Control	228.6	1.9	37.8

Nota:

Muestra final de 560 individuos, 272 controles y 288 beneficiarios.

Cuadro 9
Inserción Laboral como porcentaje del total de jóvenes según categoría laboral
2da Convocatoria

	Grupo de Control				Grupo de Beneficiarios			
	Línea de Base	%	Segunda Medición	%	Línea de Base	%	Segunda Medición	%
Trabajadores F	119	43.8	136	50.6	123	42.7	180	61.9
TFNR	32	11.8	6	2.2	50	17.4	9	3.1
Desocupados	62	22.8	102	37.9	61	21.2	77	26.5
Inactivos	59	21.7	25	9.3	54	18.8	25	8.6
Total	272	100.0	269	100.0	288	100.0	291	100.0

Nota:

1/ Estos resultados corrigen el hecho que los individuos contestaron la pregunta de categoría laboral cuando no se encontraban trabajando. Ese porcentaje de individuos fue asignado a la condición de desempleados.

Cuadro 10
Categorías Ocupacionales

	Grupo de Control				Grupo de Beneficiarios			
	Línea de Base	%	Segunda Medición	%	Línea de Base	%	Segunda Medición	%
TPCP	24	15.9	35	22.7	33	19.1	22	11.6
OAPR	71	47.0	79	51.3	73	42.2	114	60.3
EAPR	9	6.0	13	8.4	6	3.5	26	13.8
OAPU	1	0.7	3	1.9	1	0.6	2	1.1
EAPU	0	0.0	0	0.0	1	0.6	3	1.6
TFNR	32	21.2	18	11.7	50	28.9	9	4.8
TH	14	9.3	6	3.9	8	4.6	13	6.9
otro	0	0.0	0	0.0	1	0.6	0	0.0

Nota:

Resultados de la segunda convocatoria.

Cuadro 11
Ingreso Laboral Promedio y Horas Trabajadas
(Línea de Base, Primera, Segunda y Tercera Medición)

	No. de obs	Variables de interés					
		Ing. Lab. Mens.		Ing. Lab. por hora		Horas trabajadas	
<i>Línea de Base</i>							
Controles		273.06		1.93		40.28	
Beneficiarios		228.61		1.81		37.83	
<i>Primera Medición</i>							
Controles		430.35		2.32		52.57	
Beneficiarios		509.71		2.34		53.25	
<i>Segunda Medición</i>							
Controles		258.50		1.54		33.49	
Beneficiarios		419.88		2.43		40.74	
<i>Tercera Medición</i>							
Controles		367.97		1.91		52.78	
Beneficiarios		425.29		1.88		52.91	
Estimador de D-D 1/		Promedio	Int. Conf	Promedio	Int. Conf	Promedio	Int. Conf
PM - Línea de Base	560	99.21	49.1 220.1	0.18	-0.20 1.21	4.29	-3.37 12.45
SM - Línea de Base	560	163.37	90.7 284.6	0.81	-0.25 1.10	7.99	-6.45 12.06
TM - Línea de Base	560	75.52	25.7 204.2	0.11	-0.51 1.03	3.76	-4.60 11.59

Notas:

* A soles del febrero del 2002

1/ Muestra censurada, es decir, para los individuos que no reportaron ingresos el valor imputado es "0".

Cuadro 12
Ingreso Laboral Promedio y Horas Trabajadas
(Línea de Base, Primera, Segunda y Tercera Medición)

	No. de obs	Variables de interés								
		Ing. Lab. Mens.			Ing. Lab. por hora			Horas trabajadas		
<i>Línea de Base</i>										
Controles		288.29			2.04			40.28		
Beneficiarios		242.64			1.92			37.83		
<i>Primera Medición</i>										
Controles		439.08			2.36			52.57		
Beneficiarios		515.35			2.39			53.25		
<i>Segunda Medición</i>										
Controles		411.25			2.48			53.97		
Beneficiarios		588.28			3.41			56.47		
<i>Tercera Medición</i>										
Controles		367.97			1.94			52.78		
Beneficiarios		425.29			1.92			52.91		
<hr/>										
Estimador de D-D	1/	Promedio	Int. Conf		Promedio	Int. Conf		Promedio	Int. Conf	
PM - Línea de Base	198	134.62	49.1	220.1	0.50	-0.20	1.21	4.54	-3.37	12.45
SM - Línea de Base	212	187.65	90.7	284.6	0.43	-0.25	1.10	2.80	-6.45	12.06
TM - Línea de Base	200	114.93	25.7	204.2	0.26	-0.51	1.03	3.50	-4.60	11.59

Notas:

* A soles del febrero del 2002

1/ Muestra truncada, solo se considera los casos en que ambos periodos (LB-PM, LB-SM, o LB-TM) el ingreso es positivo o las horas trabajadas son positivas.

Cuadro 13**Estimación de la probabilidad de haber sido beneficiario de PROJoven (propensity score)***(PROBIT: Variable dependiente: Tratado - Beneficiario de PROJoven)*

Vars. independientes	Coef.	Err. Est.	z	Prob.
Constante	-2.317	0.592	-3.910	0.000
Edad	0.090	0.026	3.430	0.001
Hombre	-0.175	0.121	-1.450	0.148
Padre Secundaria 1/	-0.053	0.127	-0.420	0.676
Madre Secundaria 2/	0.020	0.140	0.140	0.886
Soltero	0.777	0.206	3.780	0.000
Hijos 3/	-0.419	0.181	-2.310	0.021
Ingreso por hora	-0.012	0.036	-0.350	0.727
Horas trabajadas	-0.002	0.002	-0.730	0.465
TFNR	-0.284	0.347	-0.820	0.413
Independiente	0.281	0.209	1.340	0.179
Ambulante	0.476	0.263	1.810	0.070
Pequeña Empresa	0.125	0.147	0.860	0.392

Número de obs.	560
Log Likelihood	-360.824
LR Chi-C(14)	54.220
P-Value	0.000
Pseudo R2	0.070

Cuadro 14
Efecto del tratamiento sobre las variables de ingreso y horas trabajadas
(Matching tipo Kernel)

Tipo de Muestra	VARIABLES	Número de repeticiones Bootstraping	Efecto del tratamiento	Intervalo al 95% de Confianza	
Censurada					
	INGRE1	200	94.00	50.91	151.84
	INGRE2	200	156.83	91.68	195.26
	INGRE3	200	90.05	43.76	137.54
	INGHOR1	200	0.27	-0.38	0.66
	INGHOR2	200	0.70	-0.12	2.19
	INGHOR3	200	0.27	-0.05	0.48
	Horas1	200	4.05	0.21	8.69
	Horas2	200	8.15	0.20	14.18
	Horas3	200	5.86	0.91	10.71
Truncada					
	INGRE1	200	106.17	48.86	223.18
	INGRE2	200	144.58	83.23	248.31
	INGRE3	200	88.56	33.98	186.78
	INGHOR1	200	0.24	-0.40	0.75
	INGHOR2	200	0.06	-1.47	0.85
	INGHOR3	200	0.08	-0.31	0.53
	Horas1	200	3.45	-3.29	9.91
	Horas2	200	1.42	-13.19	13.59
	Horas3	200	2.22	-3.28	9.97

Notas:

- 1/ INGRE1: Ingreso laboral mensual luego de 6 meses de finalizada la fase lectiva
- 2/ INGRE2: Ingreso laboral mensual luego de 12 meses de finalizada la fase lectiva
- 3/ INGRE3: Ingreso laboral mensual luego de 18 meses de finalizada la fase lectiva
- 4/ INGHOR1: Ingreso laboral por hora luego de 6 meses de finalizada la fase lectiva
- 5/ INGHOR2: Ingreso laboral por hora luego de 12 meses de finalizada la fase lectiva
- 6/ INGHOR3: Ingreso laboral por hora luego de 18 meses de finalizada la fase lectiva
- 7/ Horas1, horas2, horas3: Horas trabajadas a la semana.

Cuadro 15
Efecto del tratamiento sobre las variables de ingreso y horas trabajadas
para los individuos que no recibieron capacitación adicional
(Matching tipo Kernel)

Tipo de Muestra	Variables	Número de repeticiones Bootstrapping	Efecto del tratamiento	Intervalo al 95% de Confianza	
Censurada					
	INGRE2	200	211.09	51.78	326.67
	INGRE3	200	87.64	0.79	149.39
	INGHOR2	200	0.18	-1.16	0.93
	INGHOR3	200	0.12	-0.25	0.49
	Horas2	200	3.50	-8.95	16.33
	Horas3	200	2.31	-4.91	10.84
Truncada					
	INGRE2	200	162.06	106.73	347.16
	INGRE3	200	87.64	11.34	168.40
	INGHOR2	200	-0.03	-1.09	0.95
	INGHOR3	200	0.12	-0.16	0.48
	Horas2	200	2.55	-7.63	15.96
	Horas3	200	2.31	-4.53	10.31

Notas:

1/ INGRE2: Ingreso laboral mensual luego de 12 meses de finalizada la fase lectiva

2/ INGRE3: Ingreso laboral mensual luego de 18 meses de finalizada la fase lectiva

3/ INGHOR2: Ingreso laboral por hora luego de 12 meses de finalizada la fase lectiva

4/ INGHOR3: Ingreso laboral por hora luego de 18 meses de finalizada la fase lectiva

5/ Horas1, horas2, horas3: Horas trabajadas a la semana.

Cuadro 16
Estimación de la probabilidad de haber sido trabajador remunerado
(PROBIT: Variable dependiente: Remunerado - no Remunerado)

Vars. independientes	Coef.	Err. Est.	z	Prob.
Constante	-1.502	0.604	-2.490	0.013
Tratado	0.171	0.116	1.480	0.138
Hombre	0.235	0.123	1.910	0.056
Edad	0.036	0.026	1.390	0.165
Hijos	0.133	0.185	0.720	0.473
Soltero	0.320	0.207	1.550	0.122
Duración del Desempleo	-0.008	0.006	-1.260	0.208
Tamaño de familia	0.003	0.027	0.130	0.899
Ingreso Familiar	0.000	0.000	1.520	0.129
Capacitación	-0.134	0.129	-1.040	0.299
Secundaria Completa	0.187	0.177	1.060	0.290
Ocupado en Línea de Base	0.487	0.114	4.270	0.000

Número de obs.	560.000
Log Likelihood	-352.487
LR CHI2(12)	44.560
P-Value	0.000
Pseudo R2	0.06

Notas:

* Estas variables corresponden al segundo periodo, en este caso a 12 meses de finalizada la fase lectiva.

** Esta estimación pretende corregir por el hecho de que los datos de ingresos se encuentran censurados en cero.

También pretende corregir el sesgo de selección ocasionado por el hecho de que solo algunos jóvenes son los que se logran insertar de manera remunerada, a 12 meses de culminada su fase lectiva.

Cuadro 17**Estimación TOBIT para calcular el impacto en el ingreso horario de haber pasado por el PROJoven**

Vars. independientes	Coef.	Err. Est.	z	Prob.
Constante	-7.568	3.060	-2.470	0.014
Tratado	1.003	1.036	0.970	0.333
Propensity Score	3.365	3.446	0.980	0.329
Inv. De Ratio de Mills	-4.544	2.626	-1.730	0.084
Pequeña Empresa *	5.177	1.070	4.840	0.000
Obrero Privado *	7.121	1.122	6.350	0.000
Horas trabajadas a la semana *	0.033	0.017	1.970	0.050

Número de obs.	560
Log Likelihood	-1346.4
LR CHI2(12)	120.4
P-Value	0.00
Pseudo R2	0.043

Notas:

* Estas variables corresponden al segundo periodo, en este caso a 12 meses de finalizada la fase lectiva.

Gráfico 1

% de Ocupados en cada Medición

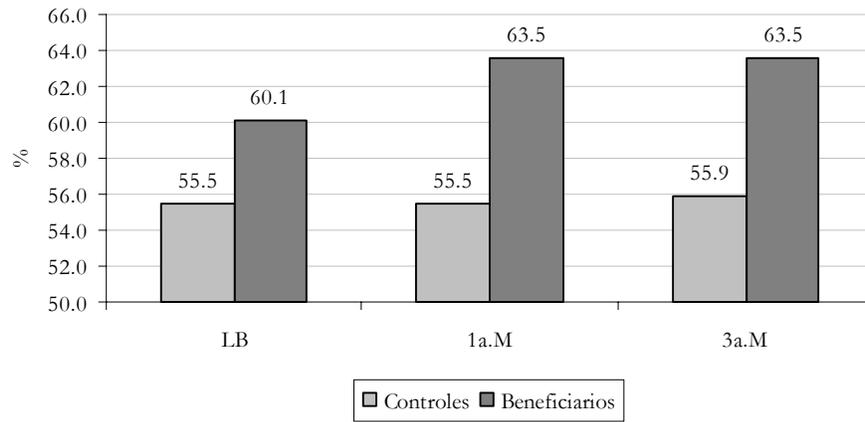


Gráfico 2

% de Trabajadores Remunerados en cada Medición

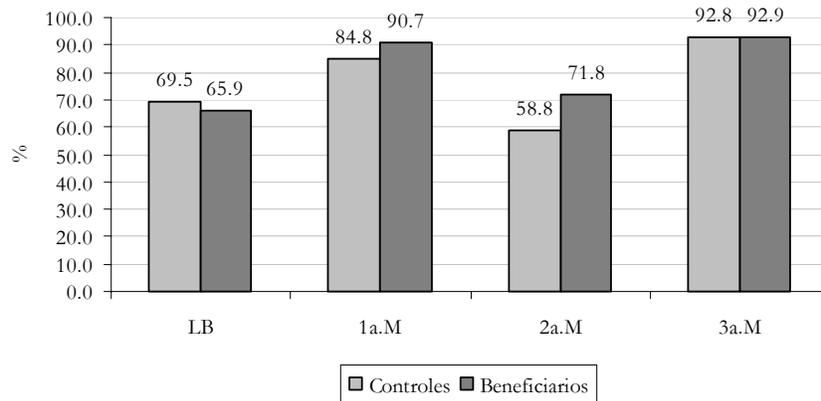


Gráfico 3

**Beneficiarios y Controles bajo Contrato:
(LB versus Segunda Medición- 18 meses)**

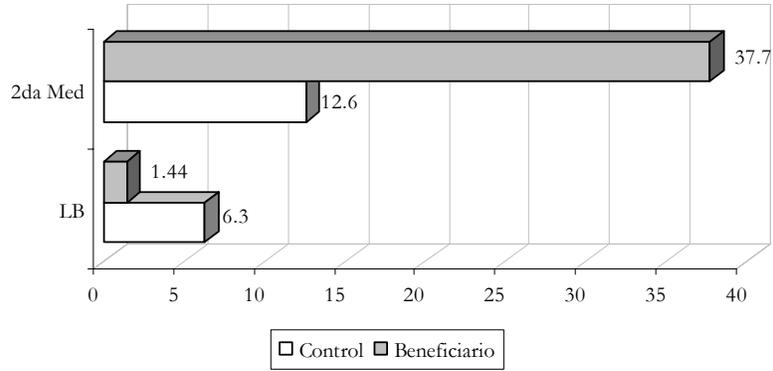
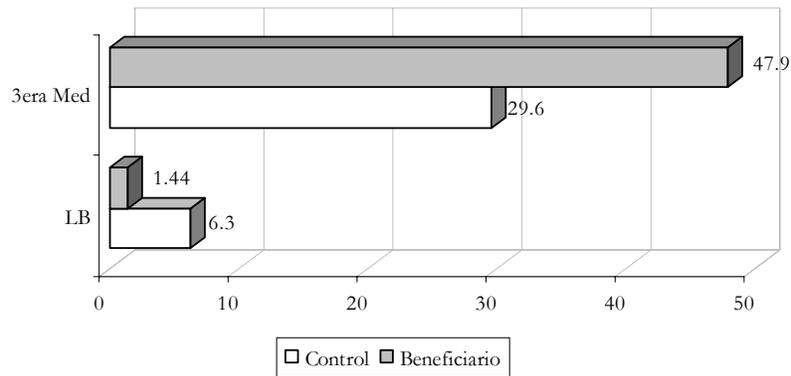


Gráfico 4

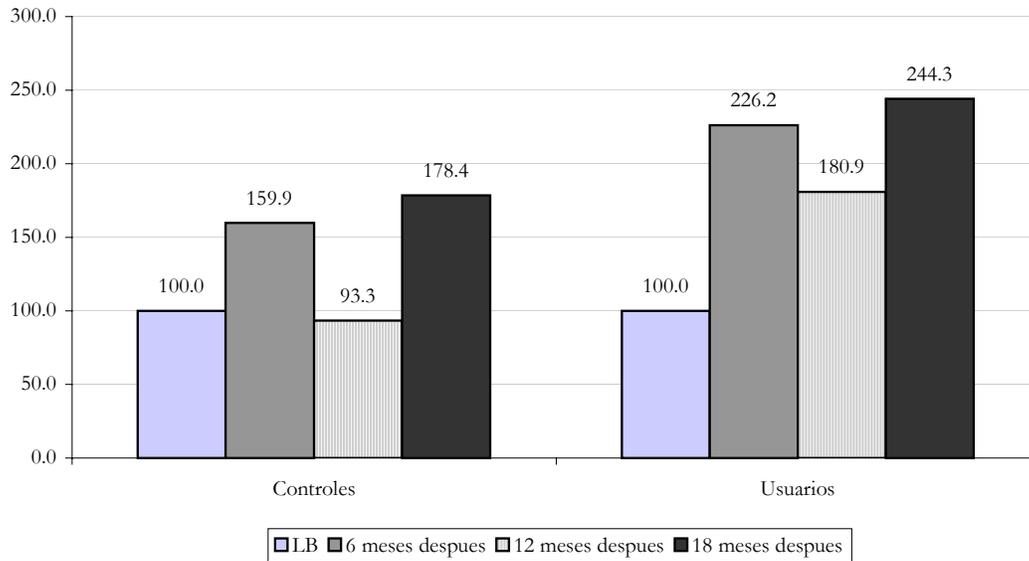
**Beneficiarios y Controles bajo Contrato:
(LB versus Tercera Medición-18 meses)**



Nota: El contrato puede ser: Contrato Indefinido, definido, prueba, CFIJ

Gráfico 5

**Evolución de los ingresos reales mensuales
(línea de base=100)**



Nota:

Los ingresos están expresados en soles constantes del 2000

Línea de Base: t=0

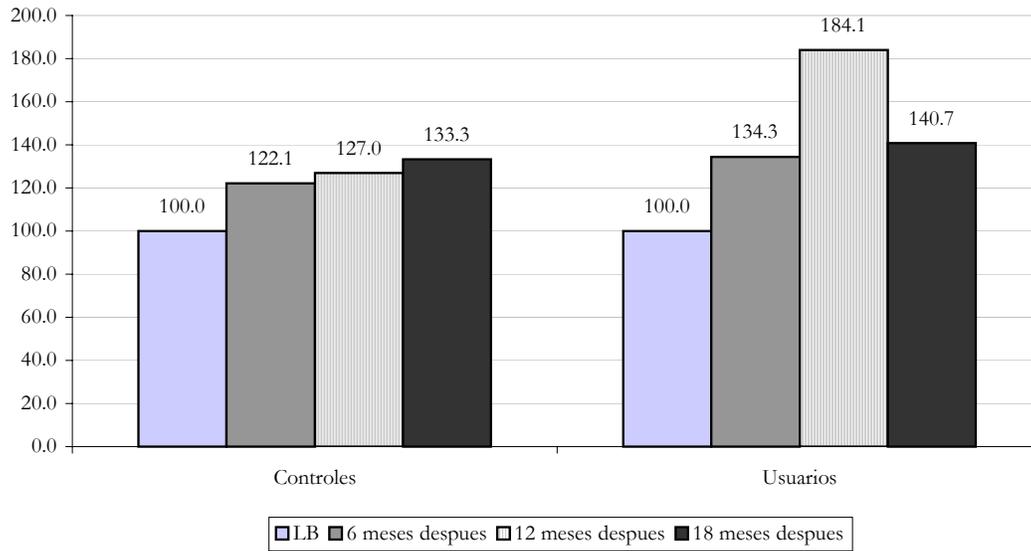
Primera Medición: t= 6 meses después de finalizada la fase práctica

Segunda Medición: t=12 meses después de finalizada la fase práctica

Tercera Medición: t=18 meses después de finalizada la fase práctica

Gráfico 6

**Evolución de los ingresos reales por hora
(línea de base=100)**



Nota:

Los ingresos están expresados en soles constantes del 2000

Línea de Base: t=0

Primera Medición: t= 6 meses después de finalizada la fase práctica

Segunda Medición: t=12 meses después de finalizada la fase práctica

Tercera Medición: t=18 meses después de finalizada la fase práctica

Figura 3:

Histograma de Propensity Scores Estimados

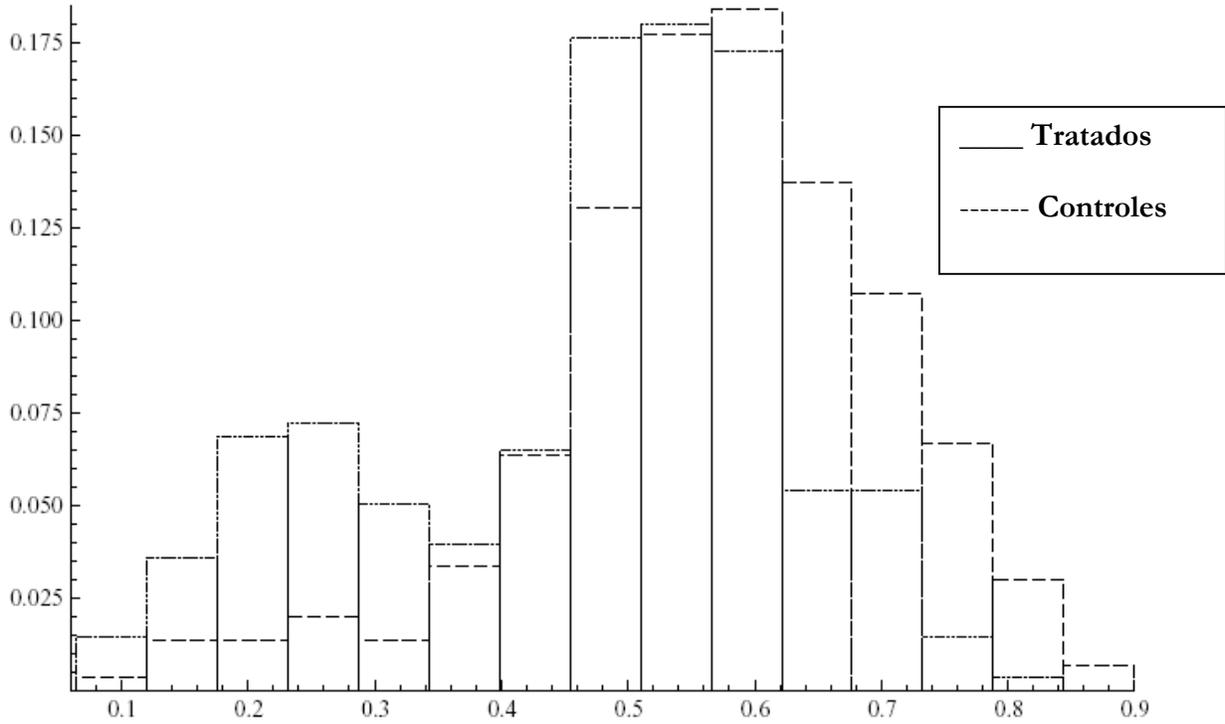


Figura 4:

Densidades Estimadas sobre los Propensity Scores

