

Educación Superior y Movilidad Social en Chile*

Josefa Aguirre, Juan Matta

14 de junio de 2023

Abstract

En este trabajo estudiamos la contribución de la educación superior (ES) a la movilidad social en Chile. Usando datos administrativos de educación e ingresos laborales, y aprovechando las discontinuidades en la admisión a distintos programas, estimamos los retornos salariales de las distintos programas para estudiantes de primera y segunda generación en la ES. Encontramos que los retornos a la ES son mayores para estudiantes pertenecientes a la primera generación en sus familias en estudiar en la educación superior. De este modo, la ES contribuye a reducir las brechas de ingreso por origen social. La brecha en retornos se explica principalmente por brechas en retornos al interior de los distintos programas de estudio, aunque también ayuda el hecho de que estudiantes de primera generación tienden a ingresar a programas más rentables que estudiantes de segunda generación con puntajes similares.

JEL Codes:

*This version: 14 de junio de 2023.

textitAuthors' information: Josefa Aguirre - Assistant Professor of Economics, Pontificia Universidad Católica de Chile, e-mail: josefa.aguirre@uc.cl; Juan Matta - Facultad de Administración y Economía, Universidad Diego Portales, e-mail: juan.matta@udp.cl

1. Introducción

La educación superior suele considerarse un vehículo relevante de movilidad social. En años recientes, el acceso a la educación superior en Chile ha experimentado un aumento sin precedentes, en parte como resultado de políticas de financiamiento tales como el Fondo Solidario de Crédito Universitario, el Crédito con Aval del Estado y la gratuidad universitaria. En solo tres décadas la matrícula universitaria se multiplicó por un factor de seis, permitiendo la incorporación de grupos históricamente excluidos de la educación superior. Aquello ha resultado en que más del 50% de los estudiantes que ingresan a la universidad cada año pertenezcan a la primera generación de sus familias en acceder a la educación superior.

Un acceso más extendido a la educación superior, sin embargo, no es garantía de movilidad social. En primer lugar, acceder a la educación superior sería insuficiente si las nuevas generaciones ingresaran a programas menos rentables que sus pares más aventajados. En efecto, investigaciones recientes han revelado que los retornos económicos a la educación superior varían considerablemente dependiendo del área de estudios y del nivel de selectividad de las instituciones (Hastings et al., 2013; Rodríguez et al., 2016; Zimmerman, 2014). Según Kirkeboen et al. (2016), por ejemplo, la elección del área de estudios puede ser incluso más importante que la decisión de ir a la universidad. En este sentido, el potencial de la educación superior como vehículo de movilidad social dependerá en buena medida de que estudiantes de origen social desaventajado accedan a programas de estudio que ofrezcan altos retornos económicos a sus egresados.

El origen social podría determinar en parte las posibilidades de acceso a programas de estudio rentables. Por una parte, estudiantes desaventajados suelen arrastrar déficits de capital humano que limitan los programas a su disposición debido a la existencia de selección por desempeño académico. Por otra parte, es posible que los programas de estudio más rentables sean a su vez más costosos y por lo tanto menos asequibles para estudiantes provenientes de familias de menores ingresos. Esto último puede suceder a pesar de la creciente disponibilidad de oportunidades de financiamiento a través de becas y créditos, lo que ha

sido presentado como una paradoja en el caso de Estados Unidos (Hoxby and Avery, 2012; Hoxby and Turner, 2013, 2015).

Más allá del programa de ingreso, la contribución de la educación superior a la movilidad social podría verse limitada si los retornos salariales a programas específicas fueran menores para estudiantes desaventajados. El estudio de Zimmerman (2019), por ejemplo, muestra que el ingreso a ciertas programas de elite en Chile aumenta las posibilidades de acceder al 1% de mayores ingresos y a posiciones de liderazgo para hombres que estudiaron en colegios privados, pero no para mujeres o para quienes estudiaron en colegios públicos o subvencionados.

¿Cuál es, entonces, la contribución de la educación superior a la movilidad social en Chile? En este estudio abordamos esta pregunta utilizando datos del Registro de Información Social (RIS) del Ministerio de Desarrollo Social y Familia (MDS). Para entender la contribución de la educación superior a la movilidad social es necesario estudiar el efecto causal de las distintas combinaciones de carreras e instituciones (en adelante, *programas*) para individuos de distinto origen social.

La estimación de efectos causales en este contexto supone desafíos importantes. Por una parte, es necesaria una definición precisa de qué entendemos por retornos a la educación superior en un contexto en que existen múltiples programas de estudio. En segundo lugar, los estudiantes no son asignados en forma aleatoria a los distintos programas, sino que se auto-seleccionan vía preferencias y son seleccionados en base a criterios académicos. Así, es probable que una comparación directa de ingresos u otros resultados entre estudiantes de distintos programas refleje una combinación de efectos causales y diferencias en características de base.

Para abordar estos desafíos, seguimos una estrategia de identificación que continúa una línea metodológica habitual en estudios sobre educación superior. Específicamente, la estrategia saca ventaja de las reglas discontinuas de admisión originadas en el Sistema Único de Admisión que centraliza el proceso de admisión a programas de educación superior en

25 universidades chilenas. Bajo supuestos convencionales de continuidad, nuestra estrategia nos permite estimar el retorno salarial para 680 programas universitarios.

Nuestro análisis revela que, a nivel agregado, la educación superior logra revertir parcialmente la brecha inicial en ingresos laborales entre estudiantes de primera y segunda generación universitaria. En ausencia de educación secundaria, individuos de primera generación tendrían ingresos laborales 9,6% inferiores a los de individuos de segunda generación. Como consecuencia de la educación superior, esta brecha se reduce a 5,2%. Lo anterior se debe a que los retornos a la educación superior son mayores para estudiantes de primera generación (24,9%) que para estudiantes de segunda generación (19%).

Las brechas por origen social en los retornos a la educación son a su vez el resultado de dos efectos. Por una parte, existen diferencias en los retornos que un mismo programa tiene para individuos de primera y segunda generación. Si bien existe un nivel considerable de heterogeneidad, con algunos programas más rentables para la primera generación y otros más rentables para la segunda, en promedio los programas tienen retornos mayores para la primera generación. Por otra parte, estudiantes de primera y segunda generación tienden a estudiar en programas distintos y, potencialmente, con distintos retornos (a este factor le llamamos *sorting*). En efecto, encontramos que estudiantes de primera generación suelen ingresar a programas con retornos algo mayores. Para determinar la importancia relativa de ambos factores, realizamos una descomposición del tipo Kitagawa-Oaxaca-Blinder. A partir de este análisis concluimos que 70-80% de la brecha en retornos es atribuible a las brechas intra-carrera, mientras que el *sorting* sería responsable del resto. Finalmente, mostramos que el efecto *sorting* es mayor cuando se condiciona en puntajes de admisión. Esto es, estudiantes de primera generación ingresan a programas considerablemente más rentables que estudiantes de segunda generación con puntajes de admisión comparables, efecto que es en parte compensado por el hecho de que los puntajes son menores en promedio para la primera generación.

2. Datos y selección de muestra

2.1. Fuentes de datos

Nuestro análisis utiliza información administrativa a nivel individual a través de distintos años, combinando datos de postulaciones y matrícula universitaria con registros de ingresos, fertilidad y matrimonio. Particularmente, nos centramos en las cohortes que ingresan a la educación superior entre los años 2004 y 2008, para quienes podemos observar resultados en un plazo de 10 años (2014-2018), cuando éstas cohortes tienen en torno a 28 años de edad.

Toda la información administrativa es obtenida del Registro de información Social (RIS) del Ministerio de Desarrollo Social y Familia (MDSF), el que recopila datos provenientes de distintas fuentes. En el caso de los datos sobre educación, la información proviene del Departamento de Evaluación, Medición y Registro Educacional (DEMRE) para los años entre 2004 y 2008. Aquí se obtuvo información sobre los puntajes en pruebas estandarizadas de selección universitaria (PSU), notas en enseñanza media, postulaciones a programas a través del Sistema Único de Admisión (SUA) a la educación superior y programa estudiados. Siguiendo las categorías recomendadas por la OECD, agrupamos los programas en nueve áreas: negocios; salud y servicios sociales; servicios; ingeniería, industria y construcción; ciencia; educación; ciencias sociales y derecho; agricultura; y humanidades y arte. El DEMRE proporciona además características socioeconómicas de los estudiantes y sus hogares reportadas al momento del registro para tomar la prueba de admisión. Estos datos incluyen el nivel educacional de ambos padres, lo que nos habilita para determinar si un individuo pertenece a la primera generación de su familia en asistir a la educación superior. La información anterior es complementada con datos del Ministerio de Educación (MINEDUC) sobre matrícula y titulación para cada año e institución, lo que nos permite observar resultados de graduación después de 10 años desde el ingreso a la educación superior.

Los datos de ingreso combinan información proveniente de las Administradora de Fon-

dos de Pensiones (AFP) y de la Administradora de Fondos de Cesantía (AFC). La información de las AFP ofrece una ventaja en términos de cobertura, por cuanto incluye a empleados del sector público y, para años recientes, a algunos trabajadores independientes. Si bien ambos datos de ingreso están censurados por arriba debido a la existencia de topes imponibles, el tope para la AFC es aproximadamente un 50% más alto. Por este motivo, en aquellos casos en que ambos datos están disponibles, nos quedamos con el de la AFC. La información sobre ingresos es recolectada a nivel individual para todos los meses de los años 2014 a 2018.

Nuestra variable de ingreso laboral anual es construida a partir de la suma de los ingresos anuales, dividida por el número de meses trabajados y llevada a dólares del año 2018. A los individuos que no presentan registro de ingresos durante el año les imputamos un ingreso laboral igual a cero. Nuestra variable de empleo, por su parte, toma el valor 1 si el individuo registró ingresos durante al menos un mes en el año.

Finalmente, obtenemos datos de fertilidad y estado civil para el período 2014-2018 del Registro Civil. Para cada individuo de la muestra podemos obtener los datos sobre su historial de estado civil y el nacimiento de cada uno de sus descendientes. Nuestra definición de matrimonio incluye a personas casadas y convivientes legales, pero se amplía además a parejas que tienen un hijo en común, esto con el fin de incorporar a convivientes legalmente solteros.

2.2. Sistema de admisión y selección de la muestra

La selección de la muestra para el análisis está directamente relacionada a nuestra estrategia de identificación de efectos causales. Dicha estrategia, desarrollada en detalle en la sección 3, explota las reglas discontinuas de admisión del Sistema Único de Admisión (SUA), un sistema centralizado de admisión en base a exámenes de selección. Por esto, antes de detallar el proceso de selección de la muestra es necesario describir brevemente el proceso de admisión universitaria a través del SUA.

De las 60 instituciones que imparten carreras universitarias, 25 seleccionan a sus estudiantes a través del SUA. Las instituciones que no participan de este sistema son principalmente instituciones privadas y tienden a recibir estudiantes de menor rendimiento académico. Las 25 universidades que participan del SUA son todas sin fines de lucro, pero pueden ser públicas o privadas.

Los egresados de educación secundaria que aspiran a estudiar en una universidad que selecciona a través del SUA deben tomar un examen estandarizado llamado Prueba de Selección Universitaria (PSU)¹ que se rinde una vez al año. El examen tiene secciones obligatorias de matemáticas y lenguaje y secciones opcionales de ciencia e historia. Los puntajes de cada sección son re-escalados a una distribución con rango 150-850 y con una mediana de 500.

Una vez que un estudiante recibe sus puntajes en las distintas secciones de la PSU, puede postular a la universidad a través de una plataforma en línea. A diferencia de sistemas de educación como el de Estados Unidos, los estudiantes universitarios en Chile se matriculan directamente en una carrera específica al interior de una institución. En este trabajo llamaremos *programa* a una carrera específica ofrecida en una institución de educación superior. En su postulación, cada estudiante incluye una lista de hasta 10 programas ordenados de más a menos preferido.

Al inicio de cada proceso de admisión, las instituciones participantes deben definir el número de cupos disponibles para cada uno de los programas que ofrece, así como la ponderación que cada programa dará a las distintas secciones de la PSU y a las notas de enseñanza secundaria. Estas ponderaciones son públicamente conocidas al momento de la postulación.

A partir de la información sobre preferencias, cupos, puntajes y ponderaciones, el sistema implementa un algoritmo de aceptación diferida [Gale and Shapley \(1962\)](#) para determinar qué estudiantes son admitidos en cada programa. El algoritmo genera puntajes de corte para cada carrera. Si bien los estudiantes pueden tener una idea acerca de la probabilidad de

¹Recientemente reemplazada por la PTU para el año 2021 y la PAES a partir del año 2022. Todas las cohortes consideradas en este trabajo rindieron la PSU.

admisión en los distintos programas a los que postula, los puntajes de corte varían de forma impredecible de un año a otro debido a shocks de demanda y capacidad. Esta impredecibilidad es crucial para la identificación de efectos causales a través de estrategias de regresión discontinua [Lee and Lemieux \(2010\)](#).

La muestra para nuestro análisis considera a postulantes cercanos al un margen de admisión. Esto es, postulantes cuyo puntaje ponderado para alguno de los programas incluidos en su postulación está cerca del puntaje de corte. En nuestra especificación base nos quedamos con observaciones a menos de 50 puntos ponderados de distancia al puntaje de corte, aunque los resultados son robustos a definiciones más y menos restrictivas. Si, de acuerdo al criterio anterior, un individuo está cerca del margen de admisión para más de un programa, nos quedamos exclusivamente con el que tenga el margen más estrecho. A este programa le llamamos *programa objetivo*.

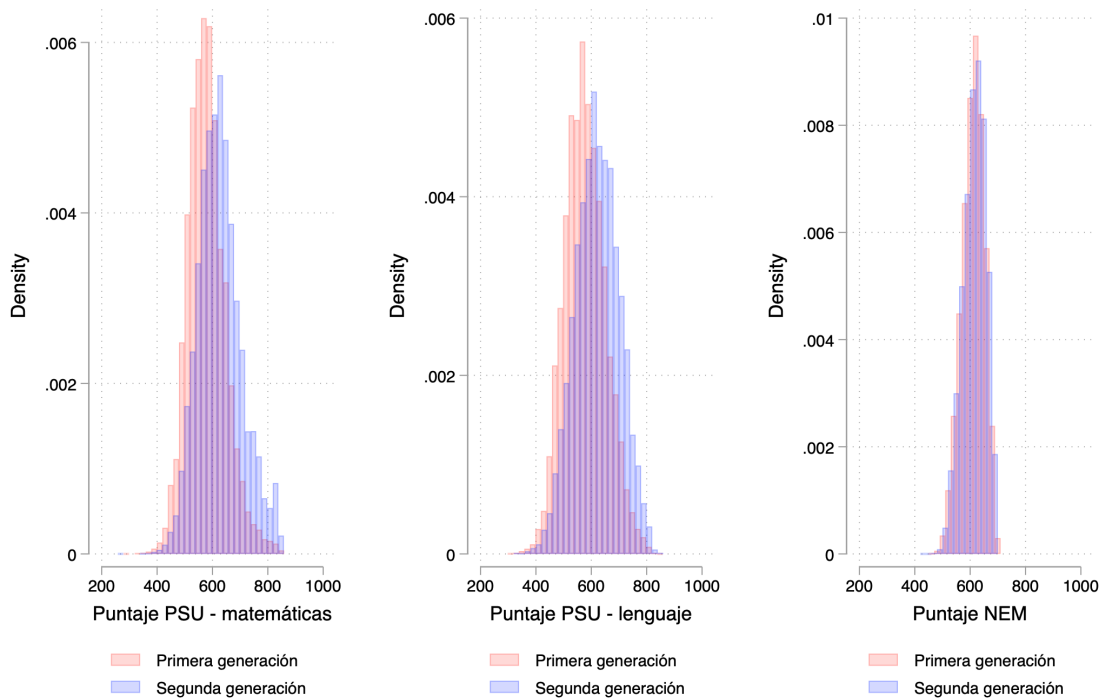
Puesto que observamos la postulación completa del individuo, podemos identificar también el programa al que éste sería admitido en caso de obtener un puntaje ponderado por debajo del puntaje de corte para el programa objetivo. A este programa le llamamos *programa de reserva*.

3. Evidencia descriptiva sobre brechas por origen social

En esta sección presentamos brevemente una comparación descriptiva de características y variables de resultado entre individuos de distinto origen social. Nuestra definición operativa de origen social es un indicador que toma el valor 1 para individuos cuyos padres no asistieron a la educación superior (*primera generación*) y 0 para individuos con al menos uno de sus padres con estudios superiores (*segunda generación*). De acuerdo a esta definición, un 45 % de nuestra muestra son estudiantes de primera generación.

En primer lugar, vale la pena comparar a estudiantes de primera y segunda generación en términos de su desempeño en el examen de admisión. La Figura 1 muestra histogramas

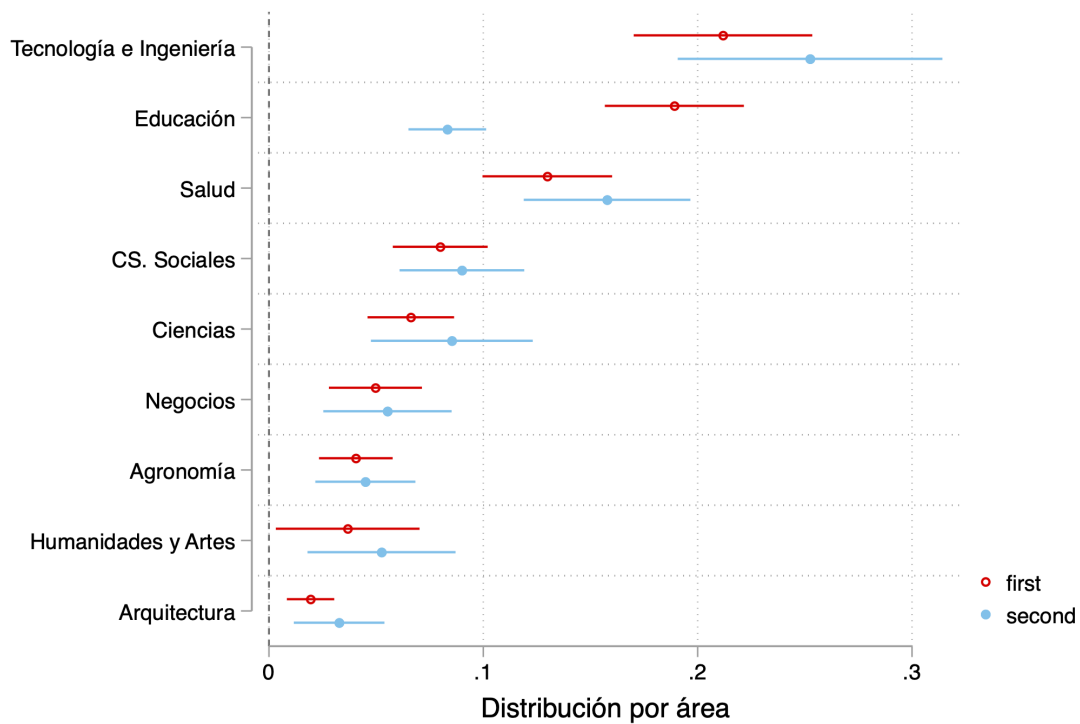
Figura 1: Distribución de puntajes PSU por origen social



para los puntajes en matemáticas y lenguaje, así como para los puntajes asociados a las notas de educación secundaria. En todos los casos, aunque especialmente en matemáticas y lenguaje, se observa una distribución de puntajes mayores para estudiantes de segunda generación que para estudiantes de primera generación. Esto es particularmente relevante si consideramos que el sistema de admisión otorga mayor prioridad a estudiantes con mejores puntajes. Esto quiere decir que para un mismo orden de preferencias sobre programas, un estudiante de segunda generación tiene una probabilidad mayor de ser admitido a programas más preferidos que un estudiante de primera generación.

Una segunda comparación relevante concierne los programas de estudio. La Figura 2 muestra la distribución de estudiantes de primera y segunda generación entre áreas de estudio. La principal diferencia se observa en educación, área en la que estudia un 19% de los estudiantes de primera generación y solo un 8,3% de los estudiantes de segunda generación. Condicional en estudiar un programa en un área distinta de educación, la distribución de

Figura 2: Áreas de estudio por origen social



áreas es similar para estudiantes de primera y segunda generación.

Concluimos mostrando comparaciones de medias entre estudiantes de primera y segunda generación en variables que potencialmente podrían ser afectadas por la educación superior: ingresos laborales, empleo, graduación, fertilidad y matrimonio. Todas estas variables son observadas luego de transcurridos 10 años desde el ingreso a la universidad, cuando los individuos tienen aproximadamente 28 años. Los resultados se muestran en el Cuadro 1. Las columnas (1) y (2) muestran las medias para estudiantes de primera y segunda generación, mientras que la columna (3) muestra la diferencia entre ambos. El ingreso laboral es de 1.206 USD para estudiantes de primera generación y de 1.272 USD para estudiantes de segunda generación. La diferencia entre ambos, de 66,6 USD en favor de la segunda generación representa un 5.2% del ingreso laboral de estudiantes de segunda generación.

En términos de empleo, vemos que un 82,8% de los estudiantes de primera generación

Cuadro 1: Brechas en variables de resultado por origen social

| | Mean 1st gen. (1) | Mean 2nd gen. (2) | Gap (1)-(2) (3) |
|------------|-------------------------|-------------------------|--------------------|
| Earnings | 1,206*** (36,87) | 1,272*** (61,54) | -66,58 (70,92) |
| Employment | 0,828*** (0,02) | 0,776*** (0,02) | 0,052 (0,03) |
| Graduation | 0,593*** (0,02) | 0,669*** (0,02) | -0,077** (0,03) |
| Fertility | 0,271*** (0,01) | 0,175*** (0,00) | 0,096*** (0,01) |
| Marriage | 0,176*** (0,01) | 0,171*** (0,01) | 0,005 (0,01) |

aparecen en nuestros registros como trabajando al menos un mes en el año, un porcentaje algo mayor que el 77,6% para estudiantes de segunda generación. Puesto que los ingresos laborales incluyen ceros para individuos que no trabajan, podemos obtener los salarios mensuales de los individuos que trabajan dividiendo el ingreso laboral promedio en la tasa de empleo. Dicho salario es de 1.456 USD para estudiantes de primera generación y de 1.639 USD para estudiantes de segunda generación, lo que corresponde a una brecha del 11,2%. La brecha es mayor que la brecha en ingresos laborales debido a que la tasa de empleo es mayor para individuos de primera generación.

Las tasas de graduación de la educación superior 10 años después del ingreso son de un 59,3% para individuos de primera generación y de un 66,9% para individuos de segunda generación. Así, es probable que parte de la brecha en ingresos laborales responda a que individuos de primera generación tienen una menor probabilidad de graduarse de la educación superior. En términos de fertilidad, por otra parte, vemos tasas considerablemente mayores entre individuos de primera generación (27,1% vs. 17,5%). Finalmente, las tasas de matrimonio son similares para individuos de distinto origen social.

4. Estimación de retornos a la educación superior

Las brechas en ingresos y otros resultados entre estudiantes de primera y segunda generación pueden deberse a diferencias en habilidad (como sugieren las diferencias en puntajes de PSU) o a diferencias en retornos a la educación superior. Para determinar qué parte de la brecha corresponde a cada factor, es necesario estimar los efectos causales de la educación superior para estudiantes de distinto origen social. En esta sección desarrollamos el instrumental para obtener dichas estimaciones.

4.1. Modelo empírico

A continuación describimos brevemente el marco conceptual que servirá de base para la estimación de retornos a la educación superior.

Resultados potenciales

Nos interesa modelar el ingreso salarial de los individuos como función de los programas que estudian. Para esto consideramos un conjunto J de programas y definimos a y_{ij} como el resultado potencial del individuo i en caso de estudiar el programa $j \in J$. Supondremos además que el resultado potencial es aditivamente separable en un componente individual y un componente específico al programa:

$$y_{ij} = \mu_i + \theta_j, \quad (1)$$

El modelo admite heterogeneidad en la habilidad individual capturada por el término μ_i , pero supone que el retorno salarial de estudiar el programa j es el mismo para todos los individuos. Si bien nuestro análisis permitiremos que θ_j difiera para individuos de distinto origen social, por el momento será conveniente abstraernos del grupo social.

Programas de comparación

Nuestro análisis supondrá que es posible identificar un grupo de programas $J_0 \subseteq J$ que no agregan valor, es decir:

$$y_{ij} = \mu_i, \quad \forall j \in J_0$$

De este modo, podemos interpretar θ_j como el efecto causal de estudiar en el programa j en lugar en una programa perteneciente a J_0 :

$$\theta_j = y_{ij} - y_{ik}, \quad \forall k \in J_0$$

En la práctica, incluimos en este grupo a los programas de menor salario promedio (medido a los 28 años de edad y considerando solo a quienes trabajan) y que reciben a un 5% de la matrícula total. La Figura 3 muestra un histograma del salario promedio de los programas y destaca (en rojo) a la parte de la distribución que incluimos en nuestro grupo de comparación. Para tener una mejor idea del tipo de de carreras incluídas en el grupo de comparación, la Figura 4 compara la distribución por áreas de estudio en éstas carreras con la distribución general por áreas. Se aprecia, por ejemplo, que las artes y humanidades están fuertemente sobre-representadas en el grupo de comparación. Por su parte, los programas en áreas como ingeniería, salud o negocios tienen una baja representación en este grupo. Finalmente, la Figura 5 muestra que los programas que forman parte del benchmark son programas con bajos puntajes promedios en matemáticas.

En caso que el supuesto de retornos nulos para programas en este grupo resulte poco plausible, es posible re-interpretar los parámetros θ_j como el efecto causal de estudiar en el programa j en lugar de estudiar en una programa cuyos egresados reciben salarios bajos.

Figura 3: Programas de comparación

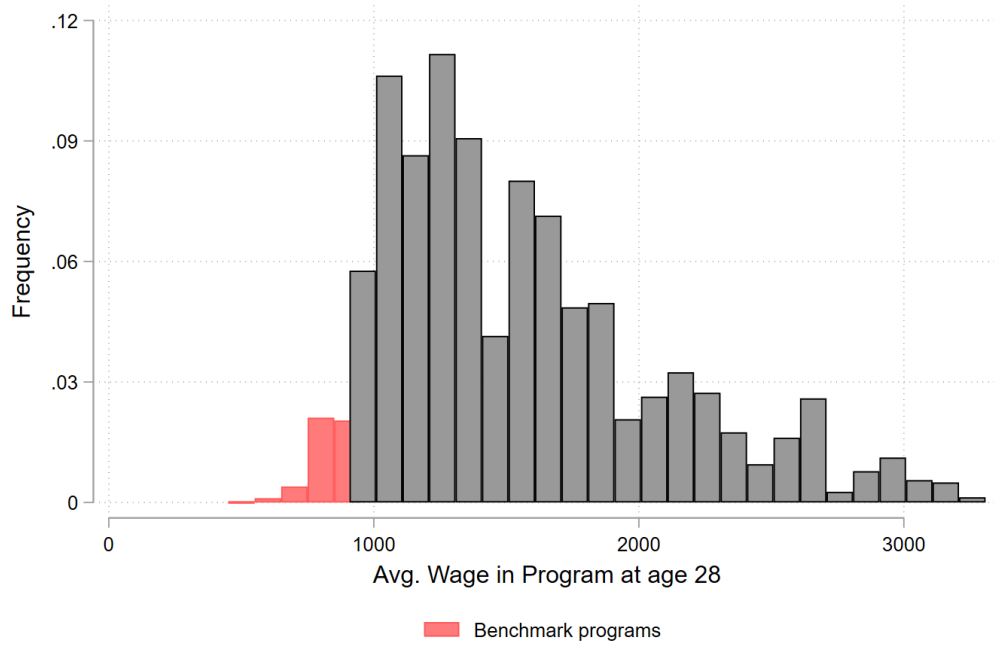


Figura 4: Áreas de estudio en programas de comparación

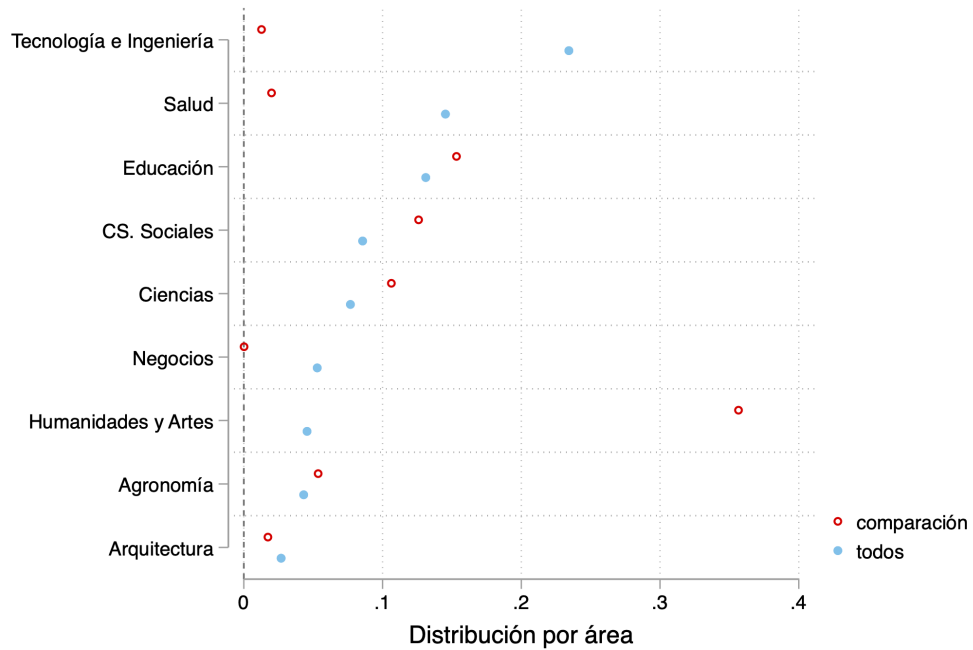
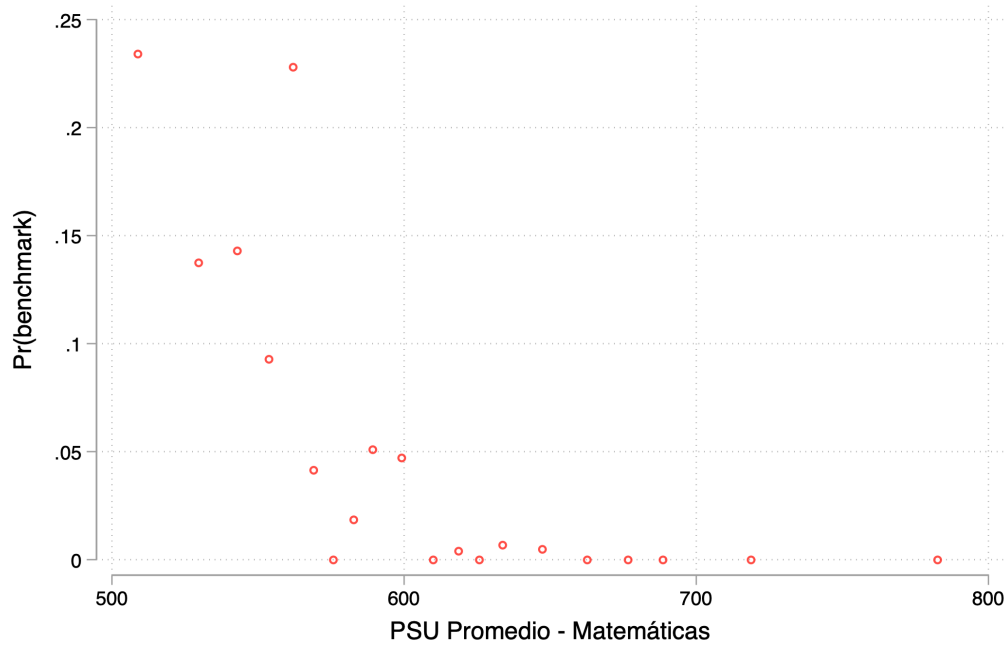


Figura 5: Programas de comparación



Ingreso salarial observado

Llamemos $d_i \in J$ al programa que efectivamente estudia el individuo i y definamos el indicador $d_{ij} = \mathbb{1}(d_i = j)$ que toma el valor 1 si i estudia el programa j y 0 en caso contrario. A partir de las definiciones anteriores, podemos expresar el ingreso salarial efectivo del individuo i como:

$$y_i = \mu_i + \sum_{j \in J_0} \theta_j \cdot d_{ij} = \mu_i + \theta_i, \quad (2)$$

donde el término $\theta_i = \sum_{j \in J_0} \theta_j \cdot d_{ij}$ es el efecto causal de la educación superior para el individuo i , dada el programa que estudia.

4.2. Estimación por mínimos cuadrados ordinarios

Si tomamos el valor esperado de (2) condicionando en el programa estudiado, obtenemos:

$$E[y_i|d_i] = \mu + \sum_{j \neq J_0} \{\theta_j + \psi_j\} \cdot d_{ij},$$

donde $\mu \equiv E[\mu_i]$ es la habilidad promedio y $\psi_j = E[\mu_i|d_i = j] - \mu$ captura la selección en habilidad hacia el programa j . A partir de lo anterior es fácil verificar que la estimación por mínimos cuadrados ordinarios de un modelo con dummies por programa identificaría al verdadero efecto causal, más un sesgo de selección:²

$$plim(\hat{\theta}_j^{ols}) = \theta_j + \psi_j$$

4.3. Admisión discontinua e identificación de retornos

Debido al sesgo de selección, la identificación de retornos salariales requiere de una estrategia de identificación causal. Nuestra estrategia aprovecha discontinuidades en la admisión que se originan en el Sistema Único de Admisión (SUA) a la educación superior en Chile, según lo descrito en la sección 2.2.

Margen de admisión

Un postulante marginal es un postulante con un puntaje ponderado cercano al puntaje de corte para alguna de los programas incluidas en su postulación. A dicho programa le llamamos *programa objetivo*, lo identificamos como $t_i \in J$ y definimos el indicador $t_{ij} \in \{0, 1\}$ que toma el valor 1 si j es el programa objetivo de i . Puesto que podemos observar la postulación completa de un estudiante, para cada postulante marginal es posible identificar el programa en el que sería admitido en caso de que su puntaje ponderado se ubicara por debajo del puntaje de corte para el programa objetivo. A este programa le llamamos *programa de reserva*, lo identificamos como $f_i \in J$ y definimos el indicador $f_{ij} \in \{0, 1\}$ que toma el valor

²En un modelo con efectos heterogéneos, el sesgo de OLS incluirá además un término ξ_j para la selección en retornos extraordinarios al programa j . En este caso, θ_j se puede concebir como un ATE y $\theta_j + \xi_j$ como un ATT.

1 si j es el programa de reserva del postulante i .

Adicionalmente, definimos la siguiente *running variable* para el postulante marginal i :

$$r_i = \sum_j (s_{ij} - c_j) \cdot t_{ij}$$

Esto es, la *running variable* corresponde a la diferencia entre el puntaje ponderado y el puntaje de corte para el programa objetivo del postulante.

Admisión

Sea $a_i \in J$ el programa en la que es admitido el postulante i y $a_{ij} \in \{0, 1\}$ un indicador de admisión al programa j . La admisión de un postulante estará determinada por:

$$a_{ij} = t_{ij} \cdot \mathbb{1}(r_i \geq 0) + f_{ij} \cdot \mathbb{1}(r_i < 0), \quad \forall j \in J$$

Esta expresión nos dice que, para ser admitido al programa j , el postulante i debe, o bien tener a j como su programa objetivo y obtener un puntaje ponderado mayor al puntaje de corte, o bien tener a j como su programa de reserva y obtener un puntaje menor que el puntaje de corte para el programa objetivo.

Efectos de la admisión en la matrícula efectiva

Nuestra estrategia de identificación se basa en la comparación de postulantes marginales con el mismo programa objetivo y el mismo programa de reserva, pero cuyos puntajes ponderados los ubican a lados opuestos del puntaje de corte para el programa objetivo, siendo uno asignado al programa objetivo y el otro al programa de reserva.

Para este efecto, estimaremos un modelo de mínimos cuadrados en dos etapas (2SLS) usando la asignación como instrumento para la matrícula efectiva. Para la primera etapa

supondremos un modelo de probabilidad lineal como el siguiente para cada programa j :

$$Pr(d_i = j|t_i, f_i, r_i) = \sum_{k \notin J_0} \pi_{jk}^+ \cdot t_{ik} \cdot \mathbb{1}(r_i \geq 0) + \sum_{k \notin J_0} \pi_{jk}^- \cdot f_{ik} \cdot \mathbb{1}(r_i < 0) + \gamma_j(t_i, f_i, r_i) \quad (3)$$

donde $\gamma_j(t, f, r)$ es una función continua de r . Dada esta especificación, la diferencia $\pi_{jt}^+ - \pi_{jf}^-$ identifica la discontinuidad en torno al puntaje de corte en la probabilidad de ingreso al programa j para individuos con programa objetivo t y programa de reserva f .

Vale la pena notar que en el caso particular en que $\pi_{jk}^+ = \pi_{jk}^- = \pi_{jk}$, la ecuación (3) se reduce a:

$$Pr(d_i = j|t_i, f_i, r_i) = \sum_{k \notin J_0} \pi_{jk} \cdot a_{ik} + \gamma_j(t_i, f_i, r_i)$$

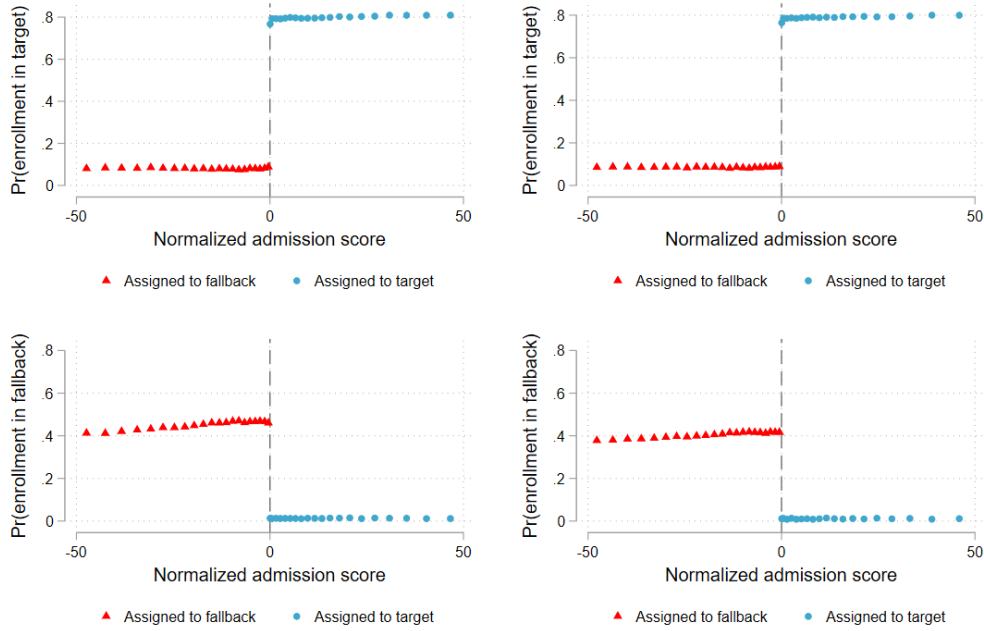
Puesto que el programa objetivo está por arriba del programa de reserva en el orden de preferencias del postulante, parece razonable suponer que el cumplimiento con la asignación difiera dependiendo de si se es asignado a una u otra, razón por la cual preferimos la especificación más flexible representada por la ecuación (3).

El modelo (3) es estimado por mínimos cuadrado ordinarios para cada programa j , asumiendo una forma funcional específica para $\gamma_j(t, f, r)$ y usando exclusivamente a postulantes con r_i cercano a cero.³ En base a los parámetros estimados, es posible obtener una predicción \hat{d}_{ij} para la probabilidad de un postulante i de matricularse en el programa j . Además, podemos estimar la probabilidad de matricularse en el programa objetivo, $\hat{p}_i^+ = \sum_{\forall j} \hat{p}_{ij} \cdot t_{ij}$ y la probabilidad de matricularse en el programa de reserva, $\hat{p}_i^- = \sum_{\forall j} \hat{p}_{ij} \cdot f_{ij}$.

La Figura 6 muestra estas probabilidades predichas como una función de la *running variable*. Se aprecia un salto discontinuo y positivo en la probabilidad de matricularse en el programa objetivo \hat{p}_i^+ cuando el puntaje ponderado supera el puntaje de corte. De igual manera, se aprecia que cruzar el umbral de admisión para el programa objetivo produce una caída discontinua, aunque de menor magnitud, en la probabilidad de matricularse en

³Nuestra especificación base supone que $\gamma_j(t, f, r) = \rho_{jt} + \lambda_{jf} + \beta_j r$ y utiliza un ancho de banda de 50 puntos. Los principales resultados son robustos a especificaciones y anchos de banda alternativos.

Figura 6: Efectos de la admisión en la matrícula efectiva



el programa de reserva, \hat{p}_i^- . Estas discontinuidades proveen la fuente de identificación para estimar el efecto causal de estudiar en distintos programas.

Efectos de la matrícula en los ingresos laborales

A partir de la ecuación (2), es posible expresar el ingreso laboral y_i como una función del margen de admisión y de la *running variable* del siguiente modo:

$$y_i = \mu(t_i, f_i, r_i) + \sum_{j \notin J_0} \theta_j \cdot d_{ij} + \xi_i, \quad (4)$$

donde $\mu(t, f, r) \equiv E[\mu_i | t_i = t, f_i = f, r_i = r]$ y $\xi_i = \mu_i - \mu(t_i, f_i, r_i)$.

Esta ecuación anterior corresponde a la segunda etapa de nuestra estimación por 2SLS. El modelo puede ser estimado por mínimos cuadrados ordinarios, utilizando las predicciones \hat{d}_{ij} obtenidas de la primera etapa en lugar de d_{ij} y suponiendo una forma funcional continua

para $\mu(t, f, r)$.⁴

Supuesto de identificación

La validez de nuestra estrategia de identificación descansa en el siguiente supuesto de continuidad:

$$\lim_{r \rightarrow 0^+} \mu(t, f, r) = \lim_{r \rightarrow 0^-} \mu(t, f, r)$$

Si $\mu(t, f, r)$ es una función continua de r en torno a $r = 0$, entonces cualquier discontinuidad en $E[y_i | t_i, f_i, r_i]$ debe ser consecuencia de cambios discontinuos en las probabilidades de matricularse en los distintos programas, discontinuidades que identificamos en la primera etapa.

El supuesto anterior se puede re-interpretar como un supuesto de exogeneidad local, con $a_i \perp \xi_i$ en torno a $r_i = 0$, o equivalentemente:

$$a_i | t_i, f_i, r_i \simeq 0 \perp \mu_i, \quad \forall j \notin J_0$$

Es decir, la estrategia supone que, para postulantes en un mismo margen de admisión (t_i, f_i) y con un puntaje ponderado cercano al puntaje de corte ($r_i \simeq 0$), la oferta de admisión a_i es independiente de la habilidad individual μ_i . En cualquiera de sus versiones, el supuesto garantiza una estimación consistente de los retornos a los distintos programas:

$$plim(\hat{\theta}_j^{2sls}) = \theta_j$$

4.4. Bayes Empírico

El supuesto de identificación recién discutido garantiza estimadores consistentes para los parámetros θ_j . Sin embargo, es probable que estos estimadores sean ruidosos, puesto

⁴En nuestra especificación base, suponemos $\mu(t, f, r) = \tau_t + \phi_f + \delta r$

que en cada margen contamos con un número limitado de postulantes. Con el propósito de reducir el ruido en las estimaciones, y siguiendo una práctica habitual en la literatura sobre estimación de valor agregado (ver por ejemplo, Angrist et al., 2022), computaremos contrapartes empíricas para las medias de la distribución posterior de θ_j , en base al siguiente modelo jerárquico:

$$\theta_j \sim N(\theta, \sigma^2)$$

$$\hat{\theta}_j \sim N(\theta_j, s_j^2)$$

Los hiperparámetros θ y σ^2 pueden ser estimados como:

$$\hat{\theta} = \frac{1}{J} \sum_j \hat{\theta}_j$$

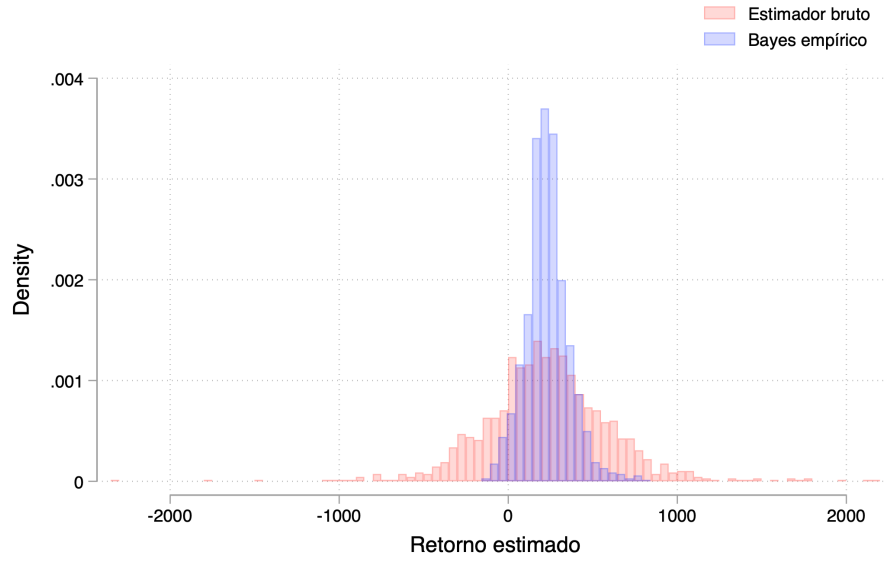
$$\hat{\sigma}^2 = \frac{1}{J} \sum_j \{(\hat{\theta}_j - \hat{\theta})^2 - s_j^2\}$$

Dado el modelo jerárquico anterior, la media de la distribución posterior de θ_j corresponde a:

$$\theta_j^* = \left(\frac{s_j^2}{s_j^2 + \sigma^2} \right) \theta + \left(\frac{\sigma^2}{s_j^2 + \sigma^2} \right) \hat{\theta}_j$$

Para obtener la contraparte empírica de θ_j^* , basta con reemplazar los hiperparámetros por sus valores estimados. La Figura 7 muestra histogramas de los retornos brutos estimados por 2SLS y de los estimadores de bayes empírico. La figura permite apreciar que el procedimiento anterior reduce considerablemente la varianza de las estimaciones. Los análisis de la sección de resultados se basan en nuestras estimaciones de θ_j^* para distintos programas y grupos sociales.

Figura 7: Efectos estimados: 2SLS vs. Bayes Empírico



5. Resultados

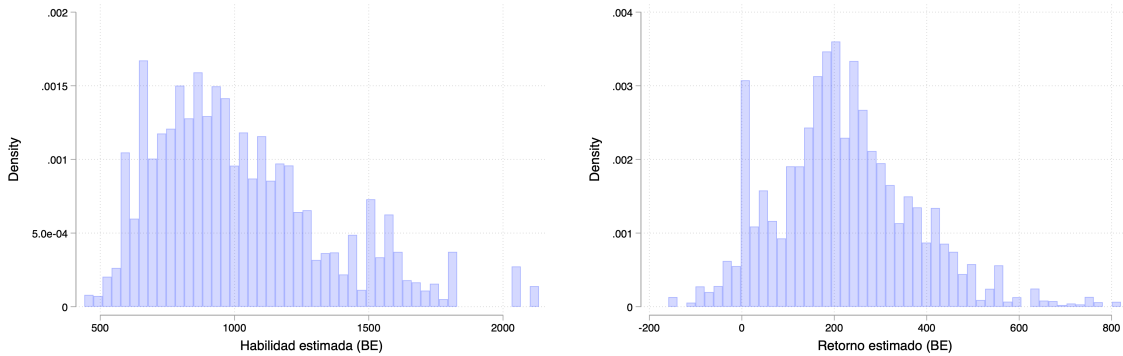
En esta sección presentamos los principales hallazgos de nuestro análisis. Los métodos descritos en la sección precedente son aplicados por separado a estudiantes de distinto origen social, de modo que para cada programa j obtenemos una estimación de los retornos para estudiantes de primera y segunda generación.

5.1. Caracterización de retornos

Comenzamos nuestro análisis caracterizando la distribución de retornos a la educación superior, ignorando por ahora las diferencias según el origen social. Para este efecto, definimos el retorno de la educación superior para la persona i como:

$$\theta_i = \sum_{j \notin J_0} \{\theta_{j0} \cdot (1 - g_i) + \theta_{j1} \cdot g_i\} \cdot d_{ij} = \sum_{j \notin J_0} \{\theta_{j0} + \Delta\theta_j \cdot g_i\} \cdot d_{ij}, \quad (5)$$

Figura 8: Distribución de habilidad y retornos estimados



donde $g_i \in \{0, 1\}$ indica si es un estudiante de primera generación en la educación superior, θ_{jg} es el retorno salarial de estudiar el programa j para un individuo del grupo social g_i y $\Delta\theta_j = \theta_{j1} - \theta_{j0}$ es la brecha social en los retornos al programa j .

La Figura 8 muestra histogramas de la habilidad promedio y los retornos estimados por programa. Si promediamos para todos los programas, obtenemos una habilidad promedio de 1021 USD. Este corresponde al ingreso promedio que observaríamos si la educación superior no agregara valor. El retorno promedio, por su parte, es de 220,4 USD, lo que representa un aumento de un 21,6% con respecto al ingreso laboral promedio en el escenario contrafactual.

La Figura 9 muestra la relación entre los retornos estimados y el grado de selectividad de las carreras, medido por los puntajes promedios en la PSU de matemáticas y lenguaje, así como por el puntaje promedio asociado a las notas de educación secundaria (NEM). La figura revela que el retorno salarial está positivamente asociado a la selectividad en matemáticas y notas de educación secundaria, pero negativamente asociado a la selectividad en lenguaje.

La metodología empleada para estimar el retorno salarial de las distintas carreras también puede ser utilizada para estimar el efecto causal de las carreras sobre otros resultados como empleo, graduación, fertilidad y matrimonio. La figura 10 muestra la relación existente entre cada uno de estos efectos y el retorno salarial estimado. Este análisis muestra que existe una relación positiva y estrecha entre los retornos salariales y los efectos en empleo y en graduación. En otras palabras, carreras con mayores efectos en empleabilidad y con mayores

Figura 9: Retornos estimados vs. selectividad

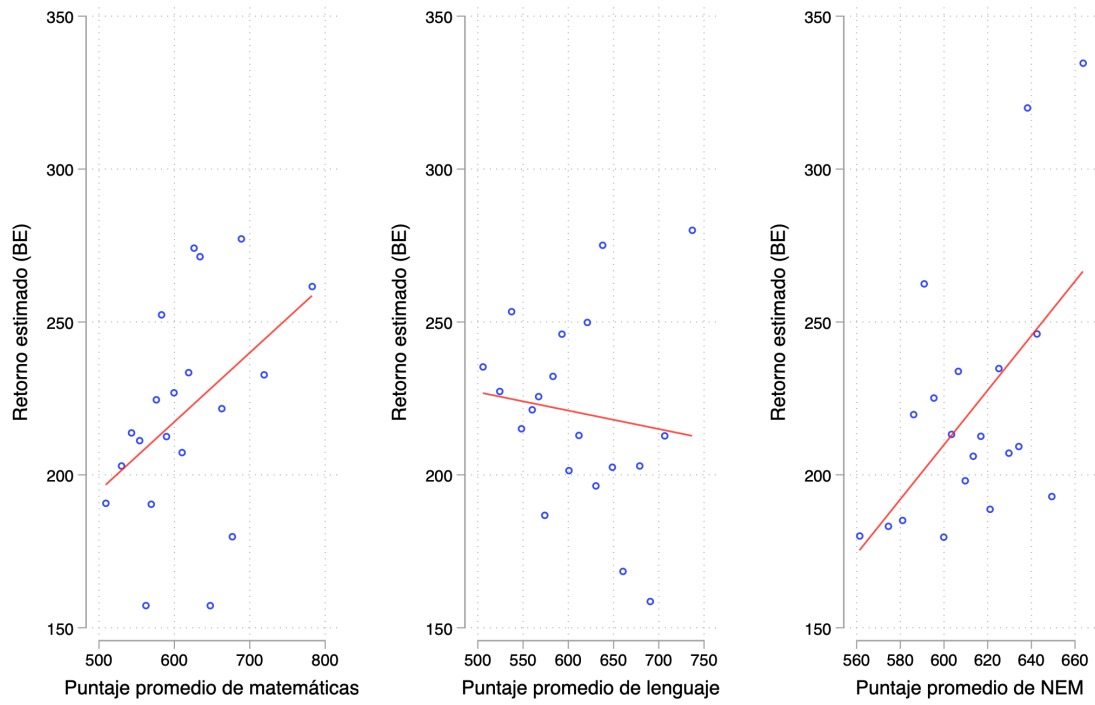
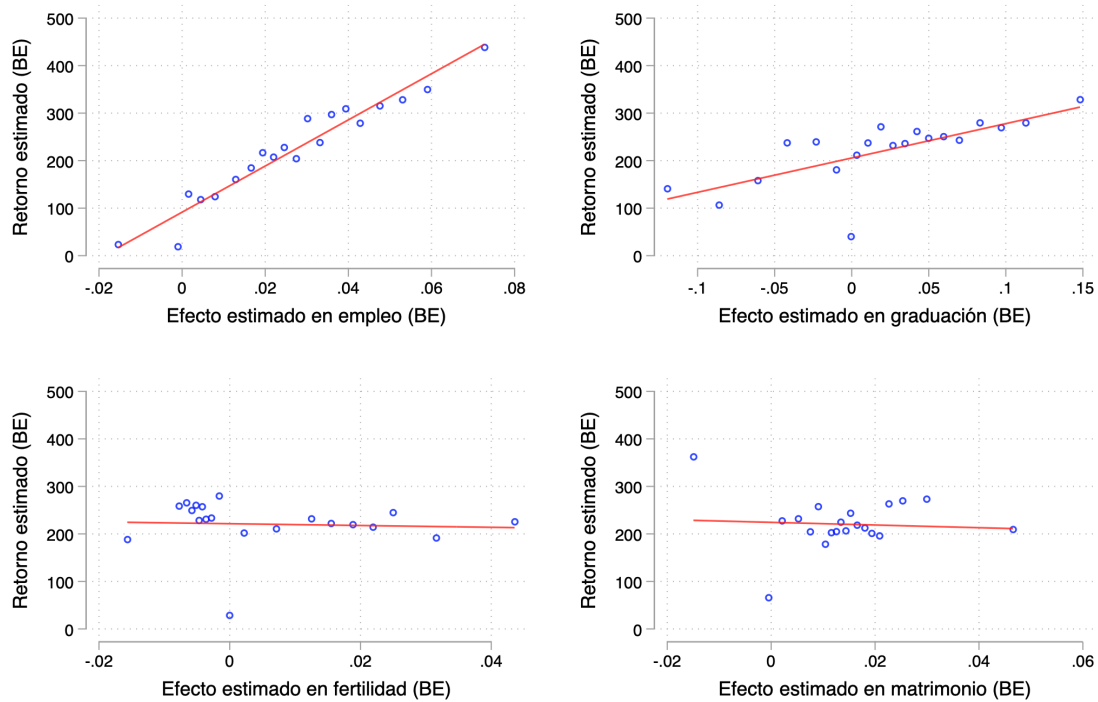


Figura 10: Retornos estimados vs. efectos estimados en otras variables



tasas de graduación son a su vez carreras con mayores efectos sobre los ingresos laborales. En contraste, no se observa una relación clara entre los retornos salariales y los efectos en fertilidad y matrimonio.

6. Descomposición de brechas por origen social

6.1. Efectos de la educación superior sobre las brechas

El análisis descriptivo presentado en la sección 3 mostraba que el ingreso laboral mensual promedio de individuos de primera generación es 66 USD menor que el de individuos de segunda generación, una brecha de un 5,2% medida como porcentaje del ingreso laboral de estudiantes de segunda generación. Teniendo en cuenta que $y_i = \mu_i + \theta_i$, la brecha anterior

se puede descomponer en brechas de habilidad y brechas en retornos:

$$E[y_i|g_i = 1] - E[y_i|g_i = 0] = \underbrace{\{E[\mu_i|g_i = 1] - E[\mu_i|g_i = 0]\}}_{\text{brecha de habilidad}} + \underbrace{\{E[\theta_i|g_i = 1] - E[\theta_i|g_i = 0]\}}_{\text{brecha en retornos}} \quad (6)$$

Ahora que disponemos de estimaciones para θ_i , es posible identificar cuánto aporta cada componente a la mencionada brecha. La primera fila del Cuadro 2 muestra esta descomposición. Las columnas (2) a (3) muestran que estudiantes de segunda generación tienen una habilidad promedio mayor que estudiantes de primera generación. En concreto, esto quiere decir que, en el escenario contrafactual de no ir a la educación superior (o ir a una programa del grupo de comparación), los estudiantes de segunda generación obtendrían ingresos laborales mensuales en promedio 103 USD mayores que los de estudiantes de primera generación. Esto representa una brecha de un 9,6% medida como porcentaje de la habilidad de estudiantes de segunda generación. Puesto que la brecha efectiva de ingresos es de un 5,2%, es posible atribuir a la educación superior una reducción de 4,4 puntos porcentuales en las brechas de ingreso por origen social.

La reducción en las brechas se explica por el hecho de que los estudiantes de primera generación tienen retornos mayores que los de segunda generación, como se observa en las columnas (5) a (7) de la tabla. Los retornos para estudiantes de primera generación son de 240,4 USD; un 24,9% de retorno con respecto al escenario contrafactual. Para estudiantes de segunda generación, por su parte, el retorno es de 204 USD, o un 19% con respecto al escenario contrafactual. La brecha en retornos es de 36.4 USD lo que permite cerrar un 35,3% de la brecha inicial en habilidad de 103 USD.

La segunda fila del Cuadro 2 muestra resultados análogos, utilizando el empleo como variable de resultado. Estos resultados muestran que los estudiantes de primera generación tienen una probabilidad algo mayor de trabajar en el escenario contrafactual (79,1% vs. 75,9%). Si dividimos el ingreso contrafactual por el empleo contrafactual, podemos obtener una medida del salario mensual promedio de aquellos individuos que trabajarían en dicho escenario. Para estudiantes de primera generación, el salario promedio es de 1.220 USD,

mientras que para estudiantes de segunda generación es de 1.407 USD. Esta brecha corresponde a un 13,3 %.

La educación superior tiene un efecto positivo sobre el empleo de ambos grupos de estudiantes, aunque éste es mayor para estudiantes de primera generación (3,7 p.p. vs. 1,8 p.p.). Esto muestra que parte del mayor retorno estimado para estudiantes de primera generación responde a un efecto mayor sobre el margen extensivo. En efecto, la brecha en salarios mensuales promedio efectivamente observada es de un 11,2 %, algo menor pero no muy distinta que la brecha contrafactual de 13,3 %. Todo esto sugiere que la educación superior contribuye a generar movilidad social, principalmente por la vía de mejorar la empleabilidad de estudiantes de primera generación.

Las últimas dos filas del Cuadro 2 muestran la descomposición de las brechas en fertilidad y matrimonio. En primer lugar, se observa una diferencia considerable entre las tasas de fertilidad a los 28 años de estudiantes de primera y segunda generación en el escenario contrafactual (27,5 % vs. 16 %). Esta brecha se reduce en 1,9 puntos porcentuales como consecuencia de la educación superior, debido a la combinación de un efecto negativo en la fertilidad de estudiantes de primera generación y un efecto positivo en la fertilidad de estudiantes de segunda generación. En segundo lugar, se observa un efecto positivo en ambos grupos de estudiantes sobre la probabilidad de haber contraído matrimonio a los 28 años. Esto es consistente la hipótesis de que la educación superior aumenta el atractivo de los individuos en el mercado del matrimonio.

6.2. Descomposición de brechas en retornos

Como hemos visto, los retornos a la educación superior son mayores para estudiantes de primera generación, lo que contribuye a cerrar un 35,3 % de las brechas de ingresos laborales. Esta brecha en retornos puede a su vez descomponerse en dos factores. Por una parte, estudiantes de distinto origen social podrían estudiar en programas con distintos retornos. A este factor le llamamos *sorting entre programas*. Por otra parte, es posible que un mismo

Cuadro 2: Descomposición de brechas

| | Overall Gap (1) | Ability 1st gen. (2) | Ability 2nd gen. (3) | Ability Gap (4) | Returns 1st gen. (5) | Returns 2nd gen. (6) | Returns Gap (7) |
|------------|-----------------------|----------------------------|----------------------------|-----------------------|----------------------------|----------------------------|-----------------------|
| Earnings | -66,58 (70,92) | 965,3*** (23,40) | 1,068*** (52,04) | -103,0* (62,03) | 240,4*** (15,51) | 204,0*** (13,25) | 36,44*** (12,40) |
| Employment | 0,052 (0,03) | 0,791*** (0,02) | 0,759*** (0,02) | 0,032 (0,03) | 0,037*** (0,00) | 0,018*** (0,00) | 0,020*** (0,00) |
| Graduation | -0,077** (0,03) | 0,556*** (0,01) | 0,662*** (0,02) | -0,106*** (0,03) | 0,036*** (0,00) | 0,007 (0,01) | 0,029*** (0,00) |
| Fertility | 0,096*** (0,01) | 0,275*** (0,01) | 0,160*** (0,00) | 0,115*** (0,01) | -0,004*** (0,00) | 0,014*** (0,00) | -0,019*** (0,00) |
| Marriage | 0,005 (0,01) | 0,163*** (0,01) | 0,156*** (0,01) | 0,007 (0,01) | 0,013*** (0,00) | 0,015*** (0,00) | -0,002** (0,00) |

programa le signifique retornos distintos a estudiantes de distinto origen social. A este segundo factor le llamamos *brechas intra-programas*. Antes de presentar una descomposición cuantitativa de la contribución de ambos factores, evaluaremos cada uno por si solo.

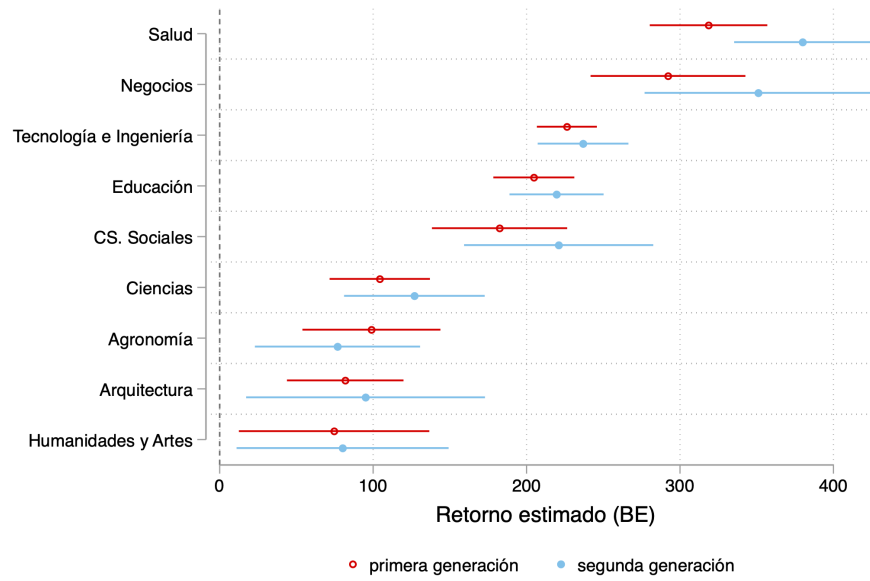
Sorting entre programas

Una primera manera de evaluar la contribución del sorting entre programas a las brechas en retornos es comparar los retornos de los programas estudiados entre individuos de primera y segunda generación. Esta comparación se puede hacer usando los retornos para la primera o para la segunda generación. Dicho análisis sugiere que individuos de primera generación asisten a programas con retornos para la primera generación 11,5 USD mayores que aquellos a los que asisten individuos de segunda generación. Asimismo, los retornos para la segunda generación son 4,5 USD mayores en los programas que estudian individuos de primera generación. En ambos casos, la diferencia es estadísticamente no significativa.

Brechas en retornos intra-programa

Hemos dicho que a nivel agregado existe una brecha de habilidad de 103 USD a favor de estudiantes de segunda generación y una brecha en retornos de 36,4 USD a favor de estudiantes de primera generación. En consecuencia, la brecha en retornos logra cerrar en un

Figura 11: Habilidad y retornos estimados por área de estudios



35,3% la brecha pre-existente. El siguiente análisis pretende estudiar la variación de estas brechas al interior de cada programa.

Podemos definir la brecha de ingresos laborales entre estudiantes de primera y segunda generación al interior del programa j como:

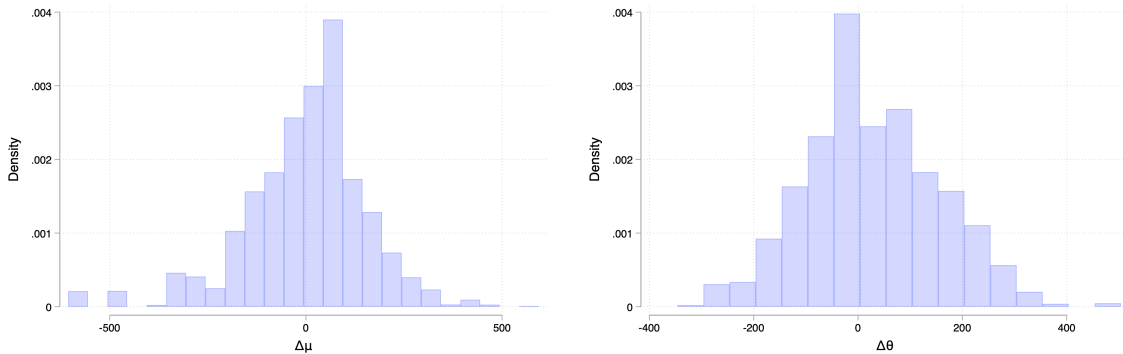
$$\Delta y_j \equiv E[y_i | d_i = j, g_i = 1] - E[y_i | d_i = j, g_i = 0]$$

Combinando esta definición con la ecuación (2), obtenemos la siguiente descomposición:

$$\Delta y_j = \underbrace{\Delta \mu_j}_{\text{brecha en habilidad}} + \underbrace{\Delta \theta_j}_{\text{brecha en retornos}},$$

donde $\Delta \mu_j = E[\mu_i | d_i = j, g_i = 1] - E[\mu_i | d_i = j, g_i = 0]$ y $\Delta \theta_j = \theta_{j1} - \theta_{j2}$. El primer término identifica a aquella parte de la brecha salarial atribuible a diferencias de habilidad entre estudiantes de primera y segunda generación que ingresan al programa j . El segundo término, por su parte, identifica la brecha en los retornos al programa j entre estudiantes de

Figura 12: Distribución de brechas en habilidad y retorno por programa



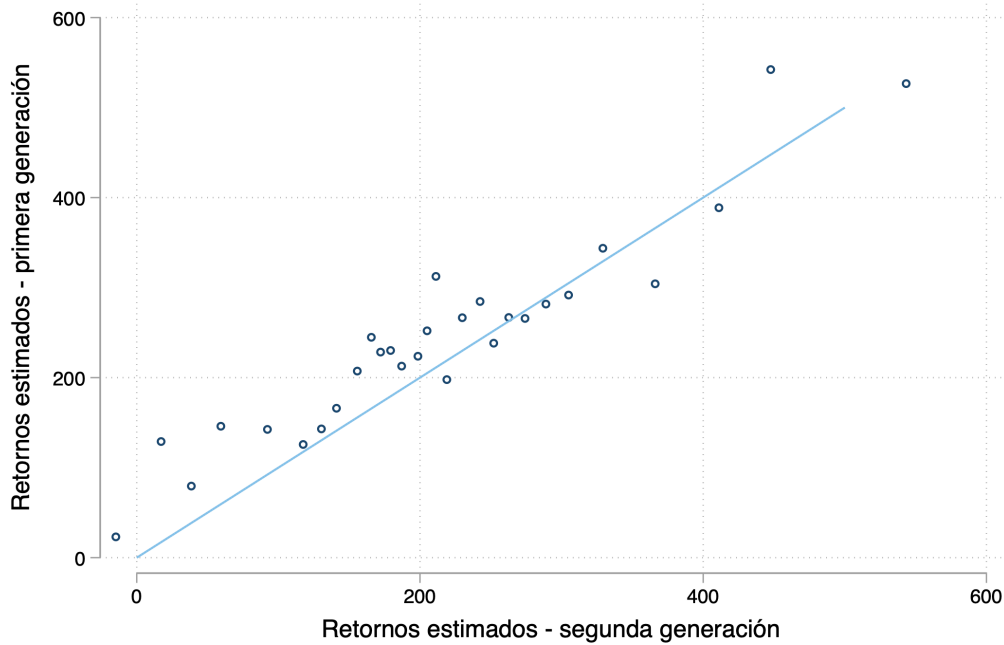
primera y segunda generación.

Los histogramas de la Figura 12 permiten apreciar la dispersión en $\Delta\mu_j$ y $\Delta\theta_j$ entre programas. La brecha de habilidad intra-programa promedio de todos los programas es cercana a cero y estadísticamente no significativa. Es decir, si bien en promedio los estudiantes de primera generación tienen menores ingresos en el escenario contrafactual, al interior de cada programa estas diferencias desaparecen. Por otra parte, la brecha intra-programas en retornos, promediada para todos los programas, es de 28,7 USD. Es decir, pese a que en promedio los programas reciben a estudiantes de primera y segunda generación con el mismo nivel de habilidad, los retornos son mayores para los primeros.

La Figura 13 muestra la relación entre los retornos para la primera y la segunda generación para un mismo programa. La línea continua es la recta de 45 grados, por lo que programas por sobre esta línea son más rentables para la primera generación que para la segunda. Si bien no se observan diferencias demasiado pronunciadas, las carreras de menores retornos parecen ser más rentables para la primera generación que para la segunda. Esta diferencia parece cerrarse para carreras más rentables.

La Figura 14 muestra la relación entre la brecha de habilidad y la brecha efectiva en ingresos laborales al interior de los programas. La línea continua es una línea de 45 grados. Los programas que están por sobre esta línea tienen $\Delta\theta_j > 0$ y las que están por debajo de ella tienen $\Delta\theta_j < 0$. La línea discontinua representa la línea de regresión y la diferencia entre ésta

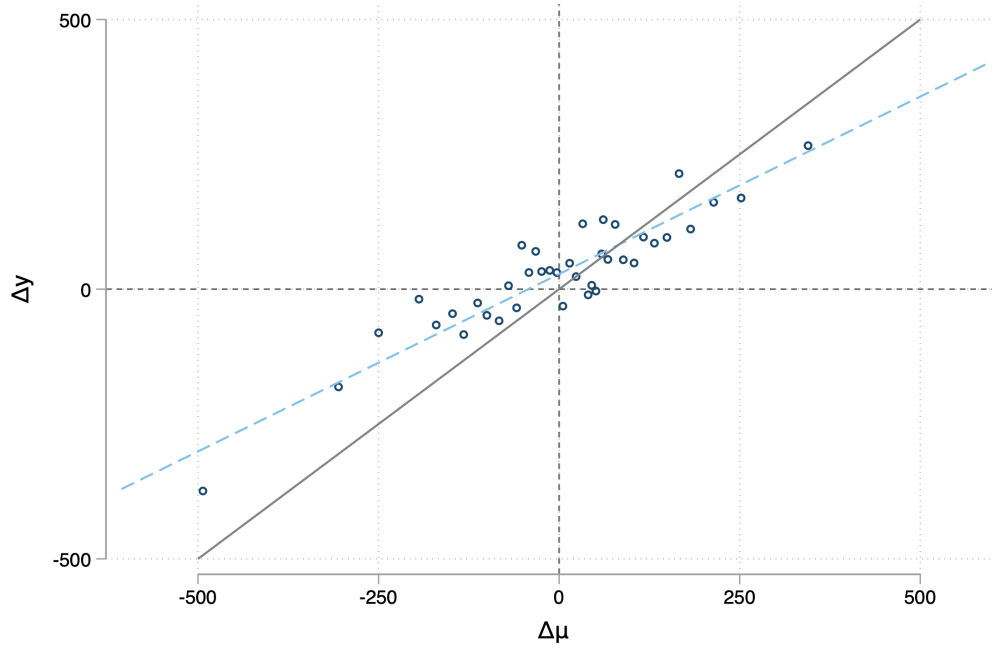
Figura 13: Retornos para primera vs. segunda generación por programa



y la línea de 45 grados representa el retorno condicional en $\Delta\mu_j$. El intercepto de la línea de regresión es de 28,1 USD, que corresponde a la brecha promedio en retornos para programas que reciben estudiantes de primera y segunda generación con el mismo nivel de habilidad. La pendiente de 0.66 indica que ambas curvas se interceptan cuando $\Delta\mu_j = 85,2$ USD. Esto es, programas con brechas de habilidad a favor de la segunda generación o levemente a favor de la primera tienden a ser más rentables para la primera generación, mientras que aquellas con brechas de habilidad lo suficientemente grandes (mayores que 85 USD) en favor de la primera generación son en promedio más rentables para estudiantes de segunda generación. En otras palabras, los programas no solo tienden a cerrar las brechas pre-existentes cuando éstas favorecen a la primera generación, sino también cuando favorecen (lo suficiente) a la segunda.

La Figura 15 es análoga a la Figura 14, pero muestra promedios por área de estudios. A excepción de agronomía, todas las áreas reportan retornos mayores para estudiantes de primera generación. El área de negocios llama la atención por recibir estudiantes de primera

Figura 14: Brecha de habilidad vs. brecha en ingresos laborales por programa

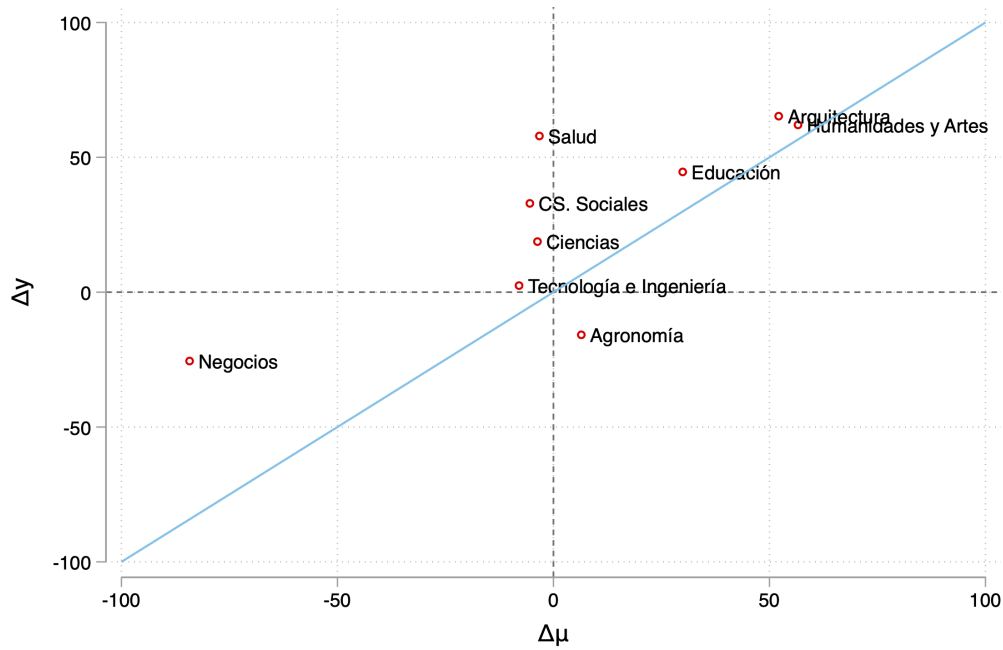


generación con un nivel de habilidad considerablemente menor que los de segunda generación. Sin embargo, la brecha en retornos es también considerable en favor de los estudiantes de primera generación. Otra área que destaca por su brecha de retornos en favor de la primera generación es el área de la salud. Si bien el punto de partida es una brecha en habilidad cercana a cero, los retornos son considerablemente más altos para estudiantes de primera generación.

Descomposición entre sorting y brechas intra-programas

Finalizamos nuestro análisis con una descomposición cuantitativa de las brechas sociales en ingresos laborales y otras variables de resultado entre factores de sorting y brechas intra-carreras en retornos. Formalmente, a partir de la ecuación (5), podemos expresar la brecha en retornos como:

Figura 15: Brecha de habilidad vs. brecha en ingresos laborales por área de estudio



$$\Delta\theta \equiv E[\theta_i|g_i = 1] - E[\theta_i|g_i = 0] = \underbrace{\sum_{j \notin J_0} \theta_{j0} \cdot \Delta\phi_j}_{\text{sorting entre programas}} + \underbrace{\sum_{j \notin J_0} \Delta\theta_j \cdot \phi_{j1}}_{\text{brechas intra-programas}}, \quad (7)$$

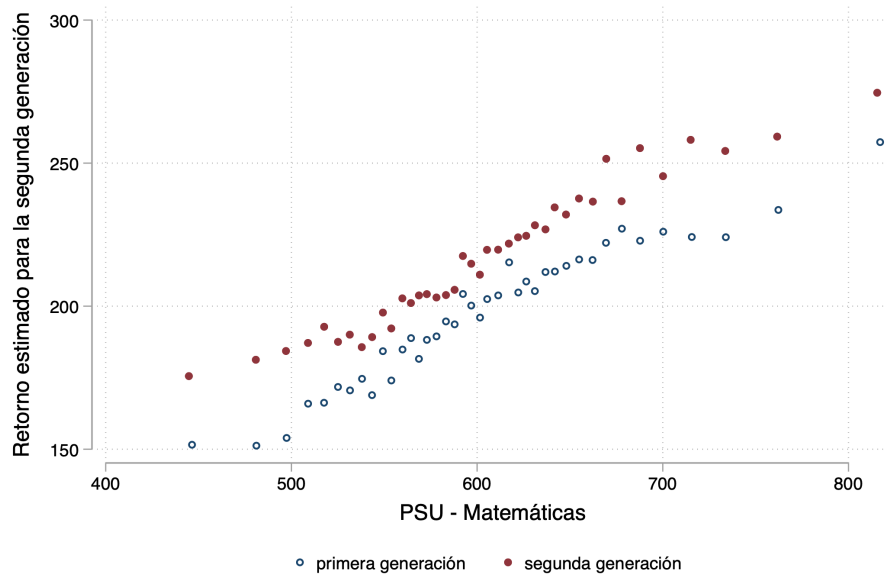
donde $\phi_{jg} \equiv Pr(d_i = j|g_i = g)$ y $\Delta\phi_j \equiv \phi_{j1} - \phi_{j0}$.

Esta descomposición, análoga a las descomposiciones de Oaxaca-Blinder, expresa la brecha en retornos como la suma de ambos componentes.⁵

Es posible demostrar que el primer componente equivale a $E[\theta_{i0} \cdot \sigma_i]$ y el segundo a $E[\Delta\theta_i \cdot \sigma_i \cdot g_i]$, donde θ_{ig} es el retorno para el grupo social $g \in \{0, 1\}$ del programa estudiado por el individuo i , $\Delta\theta_i = \theta_{i1} - \theta_{i0}$ y $\sigma_i = \frac{g_i - Pr(g_i=1)}{Pr(g_i=1)(1 - Pr(g_i=1))}$. Para estimar componente, entonces, basta con computar las contrapartes muestrales.

⁵Como cualquier descomposición de Oaxaca-Blinder, existe una formulación alternativa; en este caso: $\Delta\theta \equiv E[\theta_i|g_i = 1] - E[\theta_i|g_i = 0] = \sum_{j \notin J_0} \theta_{j1} \cdot \Delta\phi_j + \sum_{j \notin J_0} \Delta\theta_j \cdot \phi_{j0}$.

Figura 16: Sorting condicional en PSU de matemáticas



Sorting condicional

Como vimos en la sección descriptiva, los estudiantes de primera generación tienen en promedio menores puntajes en las pruebas de admisión. Puesto que los puntajes determinan en buena medida las posibilidades de acceso a los distintos programas, sería conveniente distinguir entre el sorting que se explica por diferencias en puntajes y el *sorting condicional* en puntajes. Este último nos dice la medida en que estudiantes de primera generación estudian en programas más o menos rentables que estudiantes de segunda generación con sus mismos puntajes. Este tipo de sorting condicional en rendimiento académico puede ser el resultado de diferencias en preferencias o de restricciones presupuestarias o de liquidez.

Sea $\phi_{jg}(x) = Pr(d_i = j | g_i = g, x_i = x)$, y $\Delta\phi_j(x) = \phi_{j1}(x) - \phi_{j0}(x)$, donde x_i es un vector con los puntajes obtenidos en las distintas secciones de la PSU. Podemos definir el sorting condicional en $x_i = x$ como:

$$E[\theta_{i0} | g_i = 1; x_i = x] - E[\theta_{i0} | g_i = 0; x_i = x] = \sum_{j \neq J_0} \theta_{j0} \cdot \Delta\phi_j(x)$$

La Figura 16 representa gráficamente una estimación de $E[\theta_{i0}|g_i = 1; x_i = x]$ y $E[\theta_{i0}|g_i = 0; x_i = x]$, condicionando solo en el puntaje de matemáticas. El sorting condicional corresponde a la diferencia entre ambas curvas, integrado sobre la distribución marginal de x_i , $F(x)$:

$$\int_{x_i \in \mathcal{X}} \{E[\theta_{i0}|g_i = 1; x_i = x] - E[\theta_{i0}|g_i = 0; x_i = x]\} dF(x) = \sum_{j \neq J_0} \theta_{j0} \cdot \left(\int_{x \in \mathcal{X}} \Delta \phi_j(x) dF(x) \right)$$

Se puede demostrar que, este último término equivale a $E[\theta_{i0} \cdot \tilde{\sigma}_i]$, donde $\tilde{\sigma}_i = \frac{g_i - Pr(g_i=1|x_i)}{Pr(g_i=1|x_i)(1 - Pr(g_i=1|x_i))}$. Para computar $\tilde{\sigma}_i$, estimamos un modelo probit para la probabilidad de ser un estudiante de primera generación condicional en el vector de puntajes x_i y reemplazamos la probabilidad predicha en $Pr(g_i = 1|x_i)$. El sorting condicional se obtiene de promediar $\theta_{i0} \cdot \tilde{\sigma}_i$ en la población.

La parte del sorting que se explica por los puntajes es la diferencia entre el sorting total y el sorting condicional. Éste componente se puede estimar promediando $\theta_{i0} \cdot (\sigma_i - \tilde{\sigma}_i)$ en la población.

Resultados de la descomposición de retornos

Los resultados de la descomposición se presentan en el Cuadro 3. La primera fila muestra la descomposición de la brecha en retornos salariales. De acuerdo con esta descomposición, la brecha en retornos de 36.44 USD mensuales en favor de los estudiantes de primera generación se explica en buena medida por las brechas de retornos intra-programas (26 a 32 USD, dependiendo de la versión de la descomposición que se considere, es decir, 70-80 % de la brecha). El sorting, por su parte, explica entre 4,5 y 10,5 USD de la brecha (20-30 %). Este componente se divide a su vez en un sorting condicional de 17,9 a 20,8 USD y un sorting explicado por puntajes de -13,5 a -10,4 USD. Es decir, si bien a nivel agregado el sorting no explica una parte importante de la brecha en retornos, esto es el resultado de dos efectos que se contraponen: los estudiantes de primera generación tienden a estudiar en programas más

Cuadro 3: Descomposición de retornos a la educación superior

| | Returns Gap (1) | Sorting on 1st gen. returns | | Sorting on 2nd gen. returns | | Within-program returns gap | |
|------------|---------------------|-----------------------------------|---------------------------------|-----------------------------------|---------------------------------|----------------------------|----------------------------|
| | | Conditional on Test scores (2) | Explained by Test scores (3) | Conditional on Test scores (4) | Explained by Test scores (5) | 1st gen. enrollment (6) | 2nd gen. enrollment (7) |
| Earnings | 36,44*** (12,40) | 20,82*** (5,96) | -10,36 (10,52) | 17,93*** (4,77) | -13,47 (8,63) | 31,98*** (5,85) | 25,98*** (8,15) |
| Employment | 0,020*** (0,00) | 0,003*** (0,00) | 0,002 (0,00) | 0,002*** (0,00) | 0,000 (0,00) | 0,017*** (0,00) | 0,015*** (0,00) |
| Graduation | 0,029*** (0,00) | 0,005*** (0,00) | 0,012*** (0,00) | 0,005*** (0,00) | 0,018*** (0,00) | 0,006** (0,00) | 0,012*** (0,00) |
| Fertility | -0,019*** (0,00) | -0,000 (0,00) | 0,000 (0,00) | 0,000 (0,00) | -0,001* (0,00) | -0,018*** (0,00) | -0,019*** (0,00) |
| Marriage | -0,002** (0,00) | -0,000 (0,00) | -0,002 (0,00) | -0,000 (0,00) | -0,001** (0,00) | -0,001 (0,00) | 0,000 (0,00) |

rentables que estudiantes de segunda generación con el mismo puntaje que ellos, pero sus menores puntajes los llevan a estudiar en programas menos rentables. Dicho de otro modo, si eliminásemos las brechas en puntajes, la brecha en retornos sería de 54-57 USD en lugar de 36,44 USD.

7. Conclusión

En este trabajo estudiamos la contribución de la educación superior a la movilidad social en Chile. Combinando datos administrativos de educación, ingresos, fertilidad y matrimonio con una estrategia de identificación que explota reglas discontinuas de admisión del SUA, estimamos el retorno a las distintas carreras de educación superior para estudiantes de primera y segunda generación.

Nuestro análisis muestra que la educación superior contribuye parcialmente a borrar las huellas del origen social en el desempeño laboral por la vía de mayores retornos para estudiantes de primera generación. Este mayor retorno responde principalmente al hecho de que los retornos al interior de las carreras suelen ser mayores para la primera generación, aunque también contribuye el hecho de que estudiantes de primera generación suelen ingresar a programas más rentables, dado un mismo puntaje en la PSU.

Referencias

- Angrist, J. D., Hull, P., and Walters, C. R. (2022). Methods for measuring school effectiveness.
- Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.
- Hastings, J. S., Neilson, C. A., and Zimmerman, S. D. (2013). Are some degrees worth more than others? evidence from college admission cutoffs in chile. Working Paper 19241, National Bureau of Economic Research.
- Hoxby, C. and Turner, S. (2013). Expanding college opportunities for high-achieving, low income students. *Stanford Institute for Economic Policy Research Discussion Paper*, (12-014):1–57.
- Hoxby, C. and Turner, S. (2015). What High-Achieving Low-Income Students Know About College. *NBER Working Paper Series*, (w20861).
- Hoxby, C. M. and Avery, C. (2012). The Missing "One-Offs": The Hidden Supply of High-Achieving, Low Income Students. Technical Report w18586.
- Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016). Field of study, earnings, and self-selection *. *The Quarterly Journal of Economics*, 131(3):1057.
- Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of economic literature*, 48(2):281–355.
- Rodríguez, J., Urzúa, S., and Reyes, L. (2016). Heterogeneous economic returns to post-secondary degrees: Evidence from Chile. *Journal of Human Resources*, 51(2):416–460.
- Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32(4):711–754.

Zimmerman, S. D. (2019). Elite colleges and upward mobility to top jobs and top incomes.
American Economic Review, 109(1):1–47.