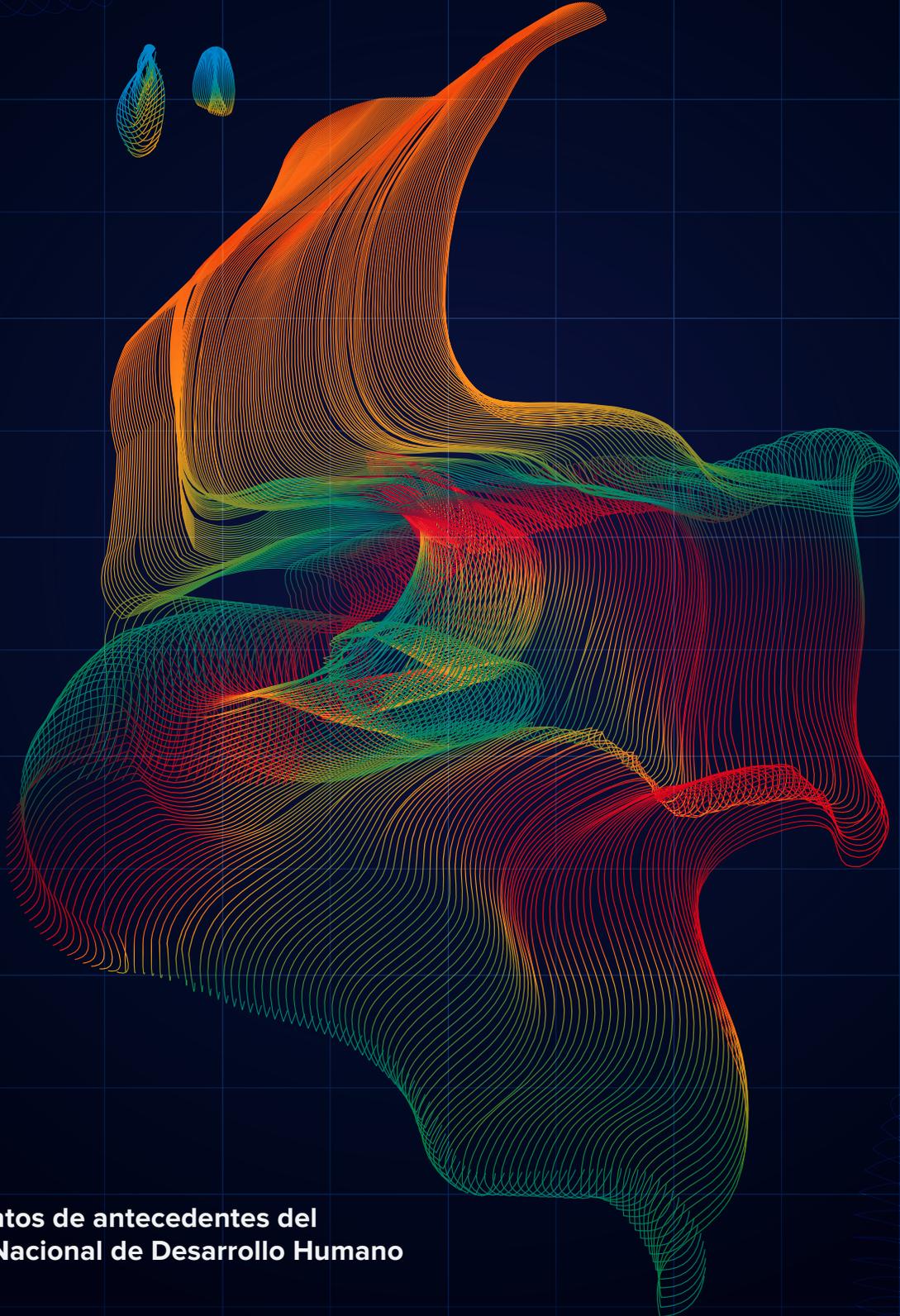


---

# Retornos a la educación superior en Colombia



**10**

Documentos de antecedentes del  
Informe Nacional de Desarrollo Humano  
2024

# **Retornos a la educación superior en Colombia**

Christian Posso  
Pablo Uribe

## Introducción

Uno de los grandes retos en educación es poder estimar sus retornos en los diferentes niveles que alcanzan las personas. Tener esta información es crucial para la formulación de políticas públicas pues permite hacer análisis de costo-beneficio y evaluar de forma más precisa los posibles resultados de las políticas educativas. Sin embargo, uno de los principales problemas es el acceso a los datos a nivel de estudiante, especialmente a aquellos que tengan información de línea de base, es decir, del momento previo a comenzar el nivel educativo al cual se le quieren calcular los retornos. Tener esta información es importante para controlar las estimaciones por posibles factores que puedan explicar los salarios y que de no incluirse, estarían confundiendo los efectos.

Tradicionalmente, los investigadores y analistas usan encuestas de hogares para tratar de capturar los retornos a la educación, pero este tipo de fuentes de datos tienen el limitante de no poder hacer análisis específicos de municipios, al igual que de no contar con información de línea de base,<sup>1</sup> específicamente de alguna medida de habilidad de las personas como sus puntajes en pruebas de Estado. Este documento pretende cerrar dicha brecha estimando los retornos a la educación superior en Colombia. Para ello se utilizan microdatos a nivel de estudiante para toda la población de individuos que presentaron la prueba Saber 11 entre 2001 y 2006, es decir, el universo de jóvenes en su último año de secundaria. Estos datos se complementan con la información de acceso a educación superior hasta 2015 y del mercado laboral formal entre 2008 y 2012.

Para calcular los retornos a la educación superior se hace una estimación con métodos de regresión lineal donde se trata de capturar el efecto más limpio al controlar por características de línea de base, incluyendo proxies de habilidad individual como el puntaje en la prueba Saber 11, y de calidad del colegio como los puntajes promedio de cada colegio el año anterior al que el individuo toma la prueba Saber 11. Es clave resaltar que ninguna de estas estimaciones se pretende analizar como causal, pues el principal objetivo del informe es mostrar los patrones de retornos asociados a la educación superior. Dada la granularidad de los datos, las estimaciones se pueden concentrar en una población muy grande (universo de jóvenes que presentan la Saber 11), y se pueden hacer estimaciones a nivel municipal.

---

<sup>1</sup> Esto se debe a la misma naturaleza de las encuestas, pues la información que tienen es del momento en que se realiza la encuesta.

En la primera sección del documento, se describe el proceso de construcción de la base de datos, incluyendo sus fuentes y la manera en que se cruzó la información y se crearon las variables más importantes. Luego, se procede a explicar la metodología que se usó para la estimación de los retornos, y finalmente se presentan los resultados.

## **Construcción de la base de datos**

Para la base de datos final se siguieron varios pasos que se detallan a continuación:

1. Identificación de los períodos de tiempo que se usarán en las estimaciones.

Tras hacer un análisis de la base consolidada de mercado laboral (PILA) disponible, se identificó que los años para los cuales se tenía la mejor calidad en los datos eran 2008 a 2012. Dado que este es el mejor período, se restringe la muestra del mercado de educación para las cohortes que presentaron la prueba Saber 11 entre 2001-2006, garantizando así poder observar el mercado laboral en los primeros años posteriores al acceso a educación superior de las personas.

2. Limpieza de la base de datos de Saber 11.

Inicialmente, se construyó un diccionario de las bases de datos para ver la calidad de las variables tanto del mercado de educación como del mercado laboral. Dado que en el período analizado las pruebas son distintas y la forma de recolectar los datos también varía, se hizo un proceso de homogeneización de variables que logró reducir el tamaño de la base y permitir definir cuáles eran las variables que presentaban mejor calidad. Esto último se hizo basado en la proporción de *missing values* que tenía cada una de ellas en cada uno de los períodos. De esta manera, se escogieron aquellas que tuvieran una cantidad suficiente de información en todos los períodos y que fueran relevantes para ser usadas como controles en las regresiones. El período 2005-1 fue excluido de la muestra dado que tenía una alta proporción de *missing* en las variables que eran de buena calidad.

Las variables que quedan de este proceso son: género del estudiante, municipio y departamento de residencia, indicador de si el estudiante paga o no paga pensión en su colegio, puntaje total de la prueba Saber 11, puntaje total promedio del colegio de cada estudiante, jornada del colegio, carácter del colegio (académico o técnico), y calendario del colegio. El puntaje de la prueba se estandariza para cada cohorte, asegurando así la comparabilidad entre períodos. Adicionalmente, se dejan únicamente las observaciones con

identificaciones válidas de tal forma que se pueda hacer seguimiento en otras bases administrativas. Por su parte, el puntaje promedio del colegio es calculado usando los puntajes individuales del año anterior a cada una de las cohortes. Por ejemplo, para las cohortes del 2003, el puntaje del colegio se calcula con los puntajes de las cohortes de 2002. Para el caso de 2001, dado que no hay un año anterior, lo que se hace es calcular el promedio sin tener en cuenta la observación de cada estudiante, para que esta no entre dos veces en la regresión (como puntaje individual y como parte del promedio del colegio).

### 3. Pega entre Saber 11 y SPADIES.

Posteriormente se hizo la pega entre Saber 11 y el universo de educación superior para identificar el acceso y graduación de nuestra población objetivo. El grupo de comparación en las estimaciones son aquellos bachilleres que no accedieron a la educación superior. Posteriormente, se hace un proceso de depuración de observaciones repetidas para garantizar que quede un solo registro por persona de su acceso a educación superior. Lo que se hace en este paso es escoger el programa del que la persona se haya graduado primero, o en caso de no haberse graduado de ninguno, el primer programa al que se matriculó. En caso de haber fechas de graduación iguales, se usa la observación del programa de mayor nivel (profesional se considera de mayor nivel que técnico y tecnológico). Las variables que se obtienen de SPADIES son: el programa al que accedió, indicador de si la persona se graduó del programa, el área del conocimiento a la que pertenece dicho programa, y su núcleo.<sup>2</sup>

### 4. Pega con el mercado laboral.

Una vez se tiene el mercado de educación consolidado para la población de interés, se procede a agregar la información del mercado laboral. Se hace la pega con cada uno de los años del mercado laboral entre 2008 y 2012. De esta manera, todas las observaciones que peguen van a estar repetidas tantas veces como años coticen en el sistema, y aquellos que no aparezcan van a seguir con una observación única. Sin embargo, para poder hacer estadísticas descriptivas y análisis econométricos, es necesario balancear el panel. Tras garantizar que cada persona aparezca tantas veces como años hay en el período de análisis, la base queda balanceada (en los años en los que la persona no cotiza, sencillamente aparece con *missing values* en las variables de mercado laboral). El producto de este proceso es lo que en adelante se considera la base final, que tiene cerca de 1,600,000 individuos y ocho millones de observaciones.

---

<sup>2</sup> El núcleo es un código que agrupa programas en categorías generales, aunque no tan generales como las áreas del conocimiento. Este núcleo sirve para identificar los programas que son STEM y no STEM, usando la guía del Higher Education Research Institute de UCLA (<https://www.herl.ucla.edu/PDFs/surveyAdmin/fac/Listing-of-STEM-Disciplines.pdf>).

Esta base tiene cuatro variables principales: formalidad, días cotizados, si la persona es dependiente (asalariado) y el salario mensual real. La primera es un indicador de si la persona aparece en PILA o no.<sup>3</sup> Los días cotizados son el número total de días trabajados por los que la persona cotiza a su seguridad social; si la persona no cotizó en algún mes, se le imputa el valor de cero. Dependiente se refiere a si la persona está contratada por una empresa con un contrato de trabajo a término fijo o indefinido y es por ende una persona asalariada. Por último, el salario mensual real se calcula con el IBC del aporte de cada persona en salud. Para los independientes se tiene en cuenta que el IBC debe ser el 40% de su salario mensual. Este IBC ajustado se divide por los días cotizados para encontrar el salario diario, y posteriormente se multiplica por 30 para encontrar el salario mensual nominal, que es dividido por el IPC y multiplicado por 100 para hallar los salarios reales con precios base de diciembre de 2018. En caso de no haber cotizado en algún mes, la observación queda en *missing*.

Ahora bien, las bases de datos originales del mercado laboral tienen la información mensual. Para disminuir la dimensionalidad de la base de datos, estas se anualizan de la siguiente manera: en el caso de formalidad, la variable va a tomar un valor de 1 si la persona cotizó al menos un mes en ese año, y si lo hizo como dependiente toma valor de 1 en asalariado. Para días cotizados y salario mensual real, se usa el último dato observado en cada año,<sup>4</sup> y se genera una nueva variable que contiene el mes en el que se observó ese último dato. Esto se hace para incluir dicha variable como control de estacionalidad en las estimaciones. Los resultados no cambian a variaciones en estas decisiones.

---

<sup>3</sup> Esta sería una proxy de formalidad, considerando que los que no aparecen en PILA pueden ser o informales o desempleados o inactivos.

<sup>4</sup> Para días cotizados sería la última observación condicional en que esta sea diferente de cero, es decir, que la persona haya cotizado por al menos un día laboral.

## Metodología

Para estimar los retornos a la educación terciaria se usan métodos de regresión lineal que capturan las correlaciones más importantes. Para mejorar la precisión de los estimadores, se controla por un conjunto de características de línea de base, en particular el puntaje de Saber 11. Esto permite que las comparaciones entre quienes acceden a la universidad y quienes no, se ajusten por las diferencias observables que existen entre ellos. Las cuatro variables de interés van a ser formalidad, días cotizados, salario mensual real y probabilidad de ser asalariado.

Así, se estiman

$$Y_i = \alpha + \beta_2 TyT_i + \beta_3 Prof_i + X'_i \delta + \phi + \gamma + \varepsilon_i \quad (1)$$

donde  $Y$  es la variable de interés.  $TyT$  y  $Prof$  son dummies que indican si la persona accedió a un programa técnico y tecnológico o a uno profesional, respectivamente. El vector  $X$  contiene controles como género, jornada, calendario y carácter del colegio, si el estudiante paga pensión donde estudia, el puntaje total del estudiante en Saber 11 y el puntaje promedio del colegio donde estudió, medido el año anterior. Adicionalmente,  $\phi$  son los efectos fijos de cohorte de Saber 11, mientras que  $\gamma$  representa los efectos fijos de año del mercado laboral. En el caso de días cotizados y salario mensual real también se incluyen efectos fijos del mes en el que se observó el dato en el año.  $\varepsilon$  es el término de error. Los errores estándar son robustos a la heterocedasticidad.

Adicional a esta estimación, también se analizan los efectos de las habilidades académicas de los estudiantes (medidas por su puntaje en Saber 11) y de la calidad de los colegios (medida por el puntaje promedio de cada colegio). Para esto, se agrega a la estimación anterior el puntaje individual y el puntaje promedio del colegio, así como las interacciones de cada tipo de programa con estos, para observar los roles que juegan las habilidades y la calidad del colegio en la explicación de los retornos de cada programa. De manera que se estima la siguiente regresión:

$$Y_i = \alpha + \beta_2 TyT_i + \beta_3 Prof_i + \beta_4 PI_i + \beta_5 TyT_i * PI_i + \beta_6 Prof_i * PI_i + \beta_7 PC_i + \beta_8 TyT_i * PC_i + \beta_9 Prof_i * PC_i + \dots + X'_i \delta + \phi + \gamma + \varepsilon_i \quad (2)$$

donde  $PI$  es el puntaje individual estandarizado y  $PC$  es el puntaje total promedio del colegio donde estudió cada estudiante. La especificación de los efectos fijos y el cálculo de errores estándar sigue la misma estructura mencionada anteriormente. El vector  $X$  contiene los mismos controles, exceptuando los puntajes individuales y de colegio.

Siguiendo las dos especificaciones mostradas anteriormente, se hacen tres cálculos adicionales donde lo único que varía son las variables explicativas de interés. Así, en el segundo ejercicio se reemplazan *TyT* y *Prof* por dos nuevas variables: *Grado* e *Incompleto*. La primera es una dummy igual a uno si la persona se graduó de educación superior y cero en caso contrario, mientras que la segunda es igual a uno si la persona accedió a la educación superior pero no se graduó y cero en caso contrario. A su vez, estas dos variables se descomponen por tipo de programa, de manera que se busca analizar el efecto de graduarse o no graduarse de un programa profesional y de un programa TyT de manera separada.

En un tercer ejercicio, se reemplazan *TyT* y *Prof* por cuatro nuevas variables: *STEM TyT*, *STEM Prof*, *No STEM TyT*, y *No STEM Prof*. Las primeras dos son iguales a uno si el núcleo al que pertenece el programa hace parte de la categoría de ciencias, tecnología, ingeniería y matemáticas (STEM, por sus siglas en inglés) y el programa es TyT o profesional, respectivamente. Las últimas dos son iguales a uno si el núcleo no hace parte de STEM y el programa es TyT o profesional, respectivamente.

En el cuarto y último ejercicio se reemplazan las dos variables por ocho variables nuevas que representan ocho áreas del conocimiento. Estas son: (i) Agronomía, veterinaria y afines; (ii) Bellas artes; (iii) Ciencias de la educación; (iv) Ciencias de la salud; (v) Ciencias sociales y humanas; (vi) Economía, administración, contaduría y afines; (vii) Ingeniería, arquitectura, urbanismo y afines; (viii) Matemáticas y ciencias naturales. En cada caso, la variable es una dummy con valor de uno si el programa se enmarca dentro de dicha área y cero en caso contrario.

Para cada uno de estos tres ejercicios adicionales, se corre tanto la regresión donde se incluyen solo las variables explicativas de interés, como aquella en donde se incluyen las interacciones de cada variable con el puntaje individual y con el puntaje promedio del colegio. Adicionalmente, se exploran las brechas de género al estimar todas las regresiones condicionando en si la persona es hombre o mujer. Esto va a mostrar las heterogeneidades que existen para las cuatro variables de interés por género. Sin embargo, dado que todas estas estimaciones son a nivel nacional, también se exploran las heterogeneidades regionales que hay en el país. Para esto se usan las 62 ciudades identificadas en O'Clery, Prieto Curiel & Lora (2019), que tienen mercados laborales integrados. En cada una de estas ciudades se estiman las regresiones para profesional y TyT, graduado e incompleto general y por tipo de programa, y las variaciones de STEM, pero usando solo la especificación sin interacciones para limitar la dimensionalidad de los resultados. Vale aclarar que las ciudades que se usan para estos cálculos son aquellas en las que vivía el estudiante alrededor de los 16 o 17 años al presentar la prueba Saber 11, previo a tomar la decisión de ir a la universidad.

## Estadísticas descriptivas y resultados

La Tabla 1 muestra las estadísticas descriptivas de la muestra. El Panel A muestra los puntajes individuales estandarizados y el puntaje del colegio calculado como se especificó anteriormente. Las observaciones difieren porque el código del colegio no se tiene para todos los estudiantes. Por su parte, el Panel B muestra las covariables usadas. Se puede notar que la mayoría de estudiantes asisten a un colegio de carácter académico y que pagan matrícula. A su vez, el 55% de la muestra son mujeres. Con respecto al acceso a educación superior, el 33% de estudiantes accedieron a programas profesionales y el 14% se graduó, mientras que solo el 11% accedieron a programas TyT y el 3% se graduó. Con respecto al mercado laboral, el 41% de la muestra tuvo un empleo formal, en el que cotizaban 10.83 días en promedio con un salario real mensual de \$1,171,000 pesos colombianos con base en el año 2018.

Tabla 1: Estadísticas descriptivas

	N	Media	SD	Min	Max
<i>Panel A: Puntajes</i>					
Puntaje total estandarizado	8,000,805	0.00	1.00	-12.91	8.03
Puntaje colegio	7,741,030	0.00	0.28	-2.09	3.75
<i>Panel B: Covariables</i>					
Jornada completa	7,985,290	0.21	0.41	0.00	1.00
Paga algo de matrícula	7,508,500	0.57	0.50	0.00	1.00
Mujer	7,997,120	0.55	0.50	0.00	1.00
Colegio académico	7,978,175	0.80	0.40	0.00	1.00
Calendario A	7,978,175	0.80	0.40	0.00	1.00
<i>Panel C: Educación superior</i>					
Acceso a programas TyT	8,001,970	0.11	0.31	0.00	1.00
Acceso profesionales	8,001,970	0.33	0.47	0.00	1.00
Profesionales STEM	8,001,970	0.12	0.33	0.00	1.00
TyT STEM	8,001,970	0.05	0.22	0.00	1.00
Profesionales no STEM	8,001,970	0.17	0.38	0.00	1.00
TyT no STEM	8,001,970	0.05	0.21	0.00	1.00
Graduación TyT	8,001,970	0.03	0.18	0.00	1.00
Incompletos TyT	8,001,970	0.07	0.26	0.00	1.00
Graduación profesionales	8,001,970	0.14	0.35	0.00	1.00
Incompletos profesionales	8,001,970	0.19	0.39	0.00	1.00
Agronomía	8,001,970	0.01	0.10	0.00	1.00
Bellas artes	8,001,970	0.01	0.11	0.00	1.00
Educación	8,001,970	0.04	0.20	0.00	1.00
Salud	8,001,970	0.03	0.17	0.00	1.00
Sociales y humanas	8,001,970	0.06	0.24	0.00	1.00
Economía	8,001,970	0.11	0.31	0.00	1.00
Ingeniería	8,001,970	0.12	0.32	0.00	1.00
Matemáticas	8,001,970	0.01	0.10	0.00	1.00
<i>Panel D: Mercado laboral</i>					
Formalidad	8,001,970	0.41	0.49	0.00	1.00
Días cotizados	8,001,970	10.83	14.03	0.00	30.00
Asalariados	8,001,970	0.35	0.48	0.00	1.00
Salario (Percentil 99.9)	3,274,317	1,171,536	965,335	0.00	11,347,854

*Nota:* La tabla muestra el número de observaciones con un valor distinto a *missing*, la media, desviación estándar, mínimo y máximo. En el caso de salario, se muestra el percentil 99.9 como valor máximo.

Los resultados que se presentan a continuación, a menos que se indique lo contrario, utilizan las especificaciones descritas en la sección de metodología donde se incluye el vector de covariables, así como un efecto fijo de cohorte de Saber 11 y de año del mercado laboral. Los resultados son robustos a la inclusión de efectos fijos de departamento o de colegio, como se puede ver en las Tablas A.1 y A.2 en el apéndice.<sup>5</sup>

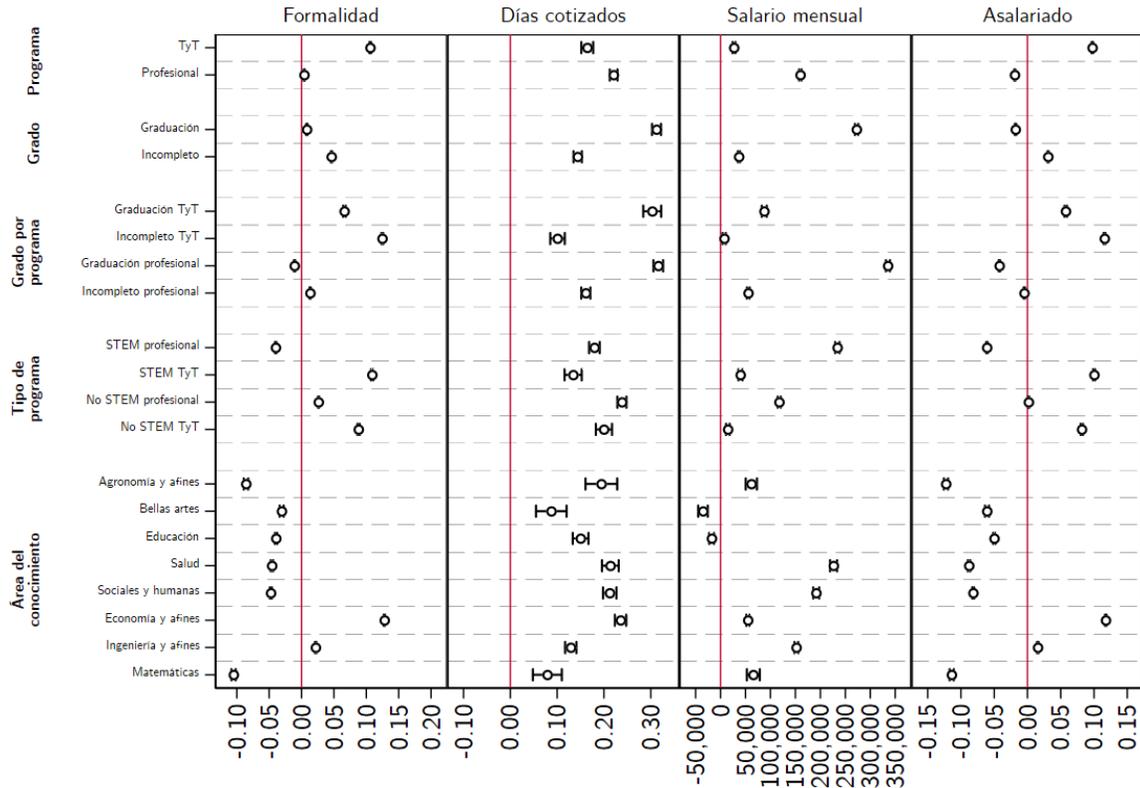
La Figura 1 muestra los resultados de todas las estimaciones para cada una de las cuatro variables de interés usando la ecuación (1). Los programas técnicos y tecnológicos parecen tener un mejor efecto en cuanto a la probabilidad de ser empleados en el sector formal, y consecuente con lo anterior, en la cantidad de días que cotizan al sistema de seguridad social y en la probabilidad de ser asalariados. Sin embargo, esto no se ve reflejado en el caso de los salarios, pues las personas con educación profesional claramente tienen retornos más altos.

Al analizar las diferencias que tienen las personas graduadas y no graduadas, se puede observar en cierta medida lo que se conoce en la literatura como el *sheepskin effect* (Layard & Psacharopoulos, 1974), o el efecto cartón. Si bien las personas que entran a la educación superior y no se gradúan tienen mejores resultados en el mercado laboral que los que no ingresan, los retornos son ampliamente superados por las personas que obtienen el título.

---

<sup>5</sup> Solo se incluyen las tablas para el caso de la regresión con acceso a TyT y acceso a profesional con el fin de ser sucintos.

**Figura 1:** Resultados de las estimaciones principales



*Nota:* La gráfica tiene cuatro paneles, uno para cada variable de interés. Cada fila hace referencia a una variable explicativa, agrupadas por las diferentes especificaciones usadas. Los círculos representan los coeficientes estimados de la respectiva variable explicativa. Dado que son estimaciones, se incluyen también los intervalos de confianza alrededor del coeficiente para mostrar la precisión del mismo.

Separando los graduados y no graduados por tipo de programa, se observa el mismo fenómeno, y nuevamente los retornos son mucho mayores para los graduados de un programa profesional (alrededor de cinco veces mayores). Adicionalmente, al descomponer aún más el tipo de programa y comparar los profesionales y TyT de STEM, contra aquellos no STEM, se ve que el primer grupo de programas está asociado con unos mayores retornos, pero no necesariamente se traduce en mayores efectos en formalidad. No obstante, los que estudian una STEM profesional tienen en promedio el doble del retorno salarial que los de otro tipo de programas. Lo mismo sucede con los que estudian una STEM TyT.

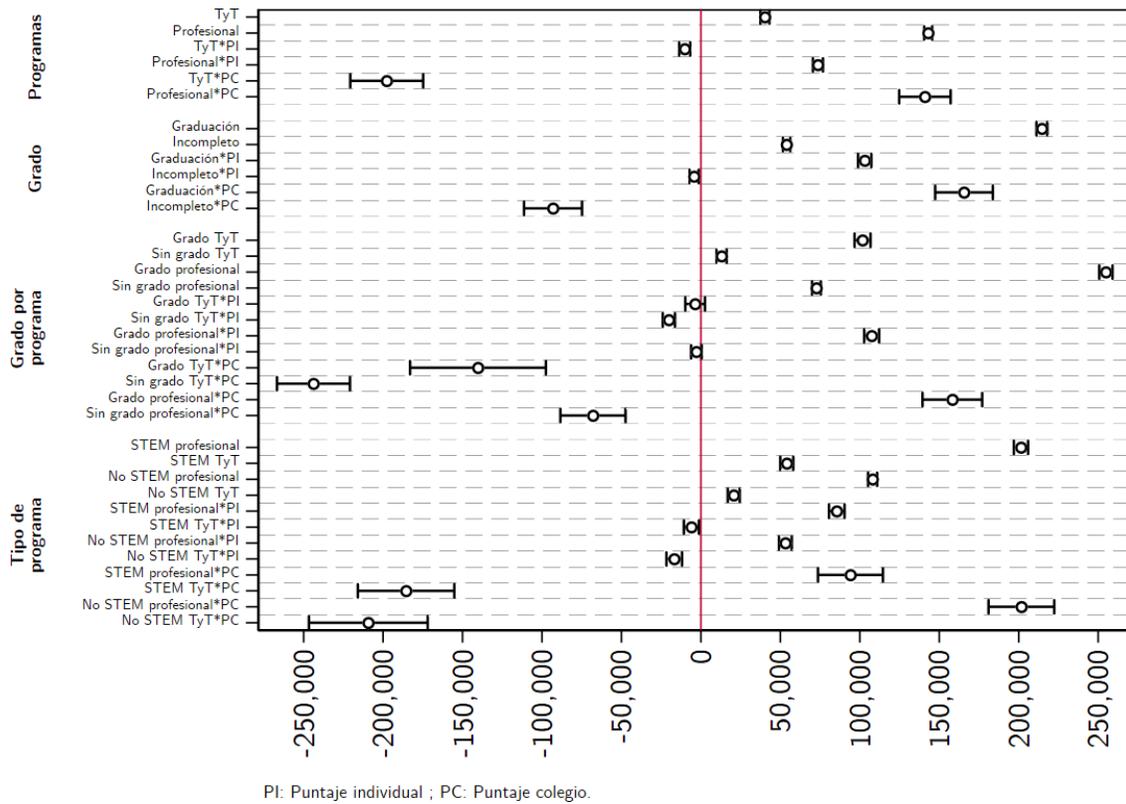
Finalmente, al separar la muestra por áreas del conocimiento se observa que hay ciertos programas que se asocian con mejores o peores retornos. Por ejemplo, los mayores retornos monetarios los tienen las personas que estudian en ciencias de la salud, ciencias sociales y humanas, e ingeniería, arquitectura, urbanismo y afines. Pero aquellos que estudian bellas artes o ciencias de la

educación, obtienen menores retornos que las personas que no acceden a la educación superior en promedio. Sin embargo, si bien todas las áreas están relacionadas con más días cotizados, el panorama en formalidad es un poco diferente, pues las únicas dos áreas del conocimiento en las que es más probable ser formal y a su vez asalariado, son economía, administración, contaduría y afines, e ingeniería, arquitectura, urbanismo y afines.

La Figura 2 muestra los resultados de las estimaciones de la ecuación (2) sobre salarios en las que se incluyen las interacciones de las variables explicativas con la medida de habilidad personal (puntaje individual) y de calidad del colegio (puntaje del colegio). En cuanto al tipo de programa, se sigue observando el mismo patrón. Los programas TyT son ampliamente superados por los programas profesionales en retornos monetarios. Sin embargo, la calidad del colegio parece ser mucho más importante que la habilidad individual como determinante de resultados en el mercado laboral para los que acceden a programas profesionales.

Por su parte, en cuanto a graduación se ve que la calidad del colegio y la habilidad individual son especialmente relevantes para los salarios de las personas con título. De hecho, los graduados tienen retornos mucho más altos si son personas más hábiles y de un colegio de mayor calidad, mientras que los no graduados en la misma posición de habilidad y calidad del colegio ven disminuidos sus retornos.

**Figura 2:** Resultados de las estimaciones con interacción

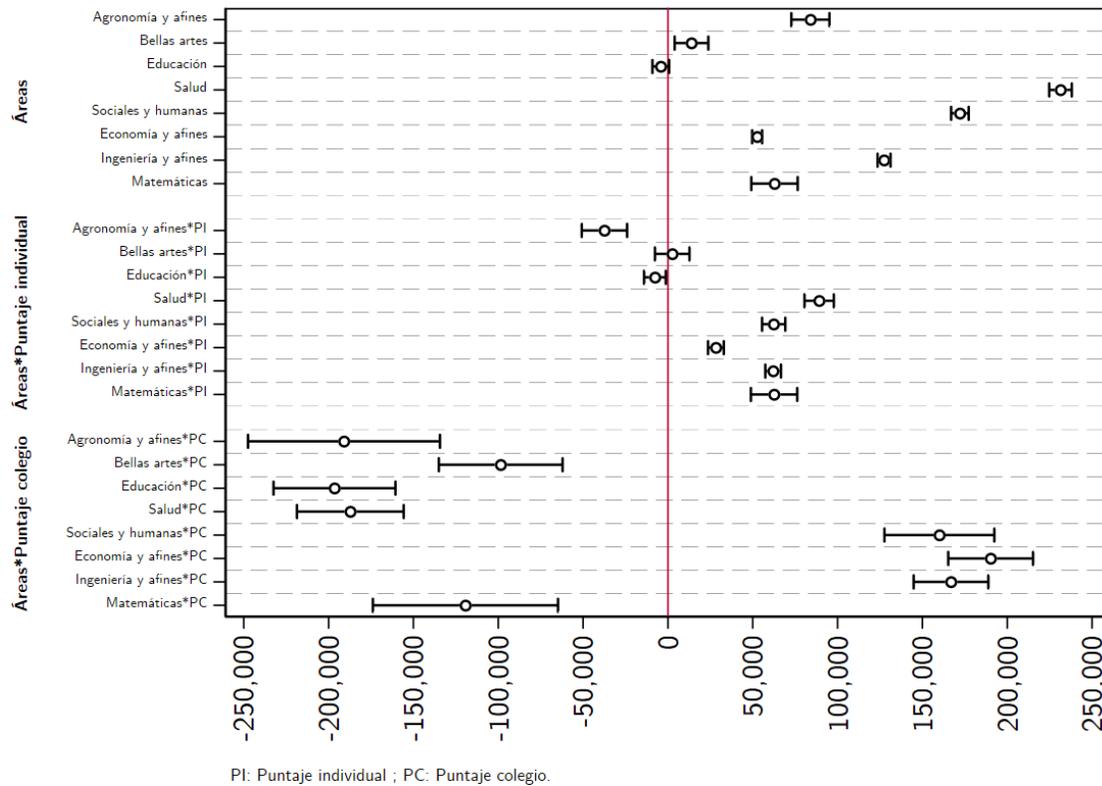


*Nota:* Cada fila hace referencia a una variable explicativa, agrupadas por las diferentes especificaciones usadas. Los círculos representan los coeficientes estimados de la respectiva variable explicativa sobre salarios. Dado que son estimaciones, se incluyen también los intervalos de confianza alrededor del coeficiente para mostrar la precisión del mismo.

De igual forma, la magnitud de los retornos está asociada al tipo de programa y a si la persona se graduó o no de dicho programa. De manera que los profesionales que tienen mayor habilidad cognitiva y que son de un mejor colegio obtienen mayores retornos monetarios, pero no sucede lo mismo con los técnicos y tecnólogos, o con las personas que abandonan sus estudios, pues estos ven sus salarios disminuidos. A su vez, en programas profesionales STEM, tener una mayor habilidad cognitiva mejora los retornos del mercado laboral. Por su parte, para el resto de programas profesionales, ser de un colegio de mayor calidad está correlacionado con mejores retornos, especialmente en los programas no STEM. En el caso de TyT, tanto para programas STEM como no STEM, los retornos se ven disminuidos por ambas medidas de calidad, pero es mucho más acentuado con la calidad del colegio.

Al ver los resultados por áreas del conocimiento, se observa que algunas áreas como ciencias sociales y humanas, economía y afines, e ingenierías y afines, tienen mejores resultados como consecuencia de la mayor calidad del colegio. En cuanto a habilidad cognitiva, a estas tres áreas se les suman ciencias de la salud y matemáticas.

**Figura 3:** Resultados de las estimaciones con interacción para áreas del conocimiento



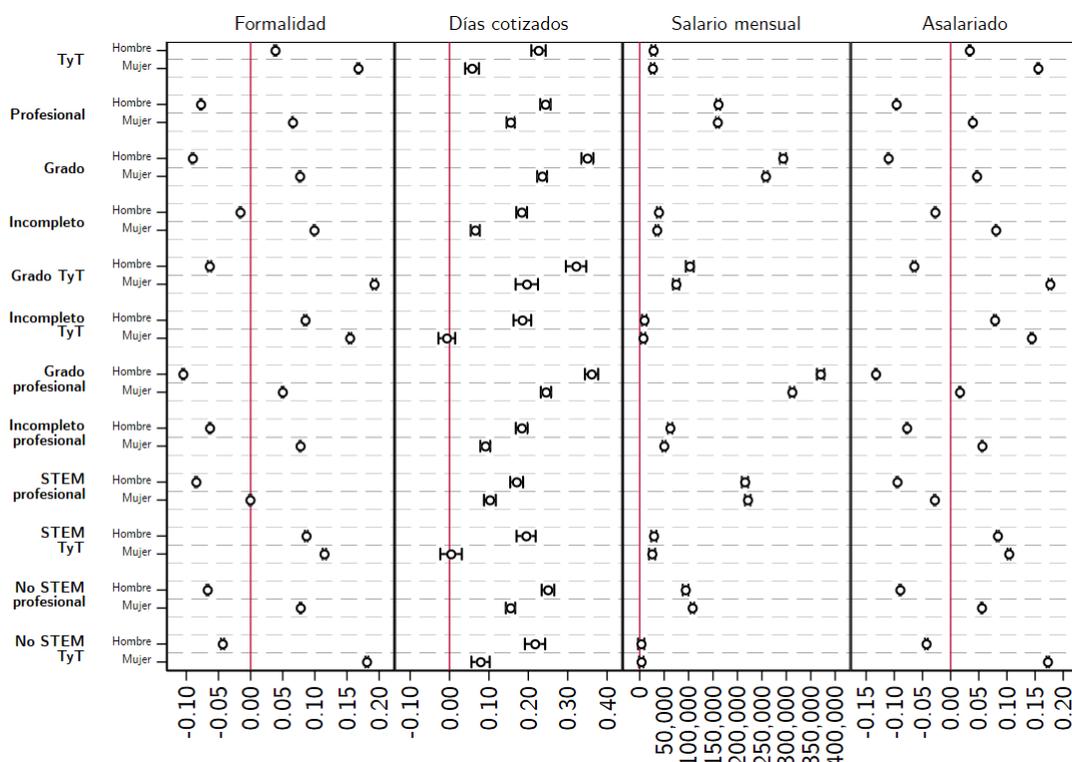
*Nota:* Cada fila hace referencia a una variable explicativa, agrupadas por las diferentes especificaciones usadas. Los círculos representan los coeficientes estimados de la respectiva variable explicativa sobre salarios. Dado que son estimaciones, se incluyen también los intervalos de confianza alrededor del coeficiente para mostrar la precisión del mismo.

## Heterogeneidad por género

La Figura 4 muestra los resultados de todas las estimaciones (con excepción de áreas del conocimiento) para las cuatro variables de interés condicionadas al género de la persona. En formalidad y probabilidad de ser asalariado se observa un patrón muy similar, donde las mujeres que acceden a la educación superior tienen una mayor probabilidad de tener trabajos formales y de ser asalariadas, independiente del tipo de programa o de si se graduaron o no, que las que no acceden. Estos efectos son mucho más pronunciados que los de los hombres, quienes en algunos casos están incluso asociados con menores tasas de formalidad y con menores probabilidades de ser asalariados.

En cuanto a días cotizados la historia cambia significativamente, pues ya son los hombres quienes a lo largo de todas las categorías tienen mayores efectos (en magnitud) sobre los días cotizados. Para el caso de las mujeres, si bien los efectos son menores que los de su contraparte, se destaca que no hay efectos negativos, lo que quiere decir que con respecto a las mujeres que no acceden a la educación superior, estas cotizan a seguridad social por una mayor cantidad de días.

**Figura 4:** Resultados de estimaciones por género



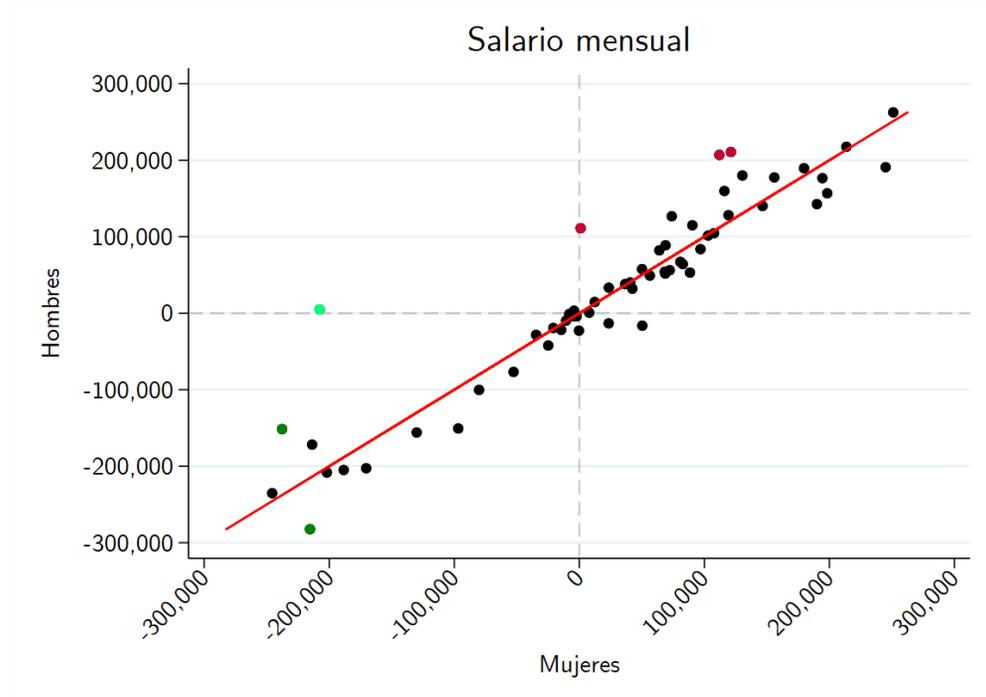
*Nota:* La gráfica tiene cuatro paneles, uno para cada variable de interés. Cada fila hace referencia al género para el cual se muestra el resultado de una variable explicativa. Los círculos representan los coeficientes

estimados de la variable explicativa para el respectivo grupo poblacional. Dado que son estimaciones, se incluyen también los intervalos de confianza alrededor del coeficiente para mostrar la precisión del mismo.

En último lugar, los retornos sobre salarios son positivos a lo largo de las diferentes divisiones de educación superior que se realizan. Sin embargo, en algunas hay una diferencia importante de género. La graduación de programas profesionales y de programas TyT muestran una brecha de género significativa en favor de los hombres. Pero esta brecha no es tan amplia para las personas que no se gradúan de sus respectivos programas. Solo parece haber un caso en el que las mujeres tienen mayores retornos que los hombres, y es el caso de las personas que acceden a programas no STEM profesionales.

Para el caso de las regresiones con interacción, dado que la dimensionalidad es muy grande para mostrar los resultados en el mismo tipo de gráfico que se usó anteriormente, se propone analizar patrones en vez de coeficientes específicos. Lo que se hace entonces es guardar todos los coeficientes de cada regresión de salarios por género (excluyendo los de puntaje individual y puntaje del colegio) y graficar la dispersión que hay entre ellos. La Figura 5 muestra esta dispersión. La línea roja es una referencia de  $45^\circ$ , y las líneas grises punteadas referencian los cuadrantes para saber si los coeficientes son positivos o negativos en cada grupo. Coeficientes por encima de la línea de  $45^\circ$  estarían reflejando mayores retornos en los hombres respecto a lo estimado para las mujeres, mientras que por debajo de la línea de  $45^\circ$  estaría sucediendo lo contrario.

**Figura 5:** Resultados de estimaciones con interacciones por género



*Nota:* La figura contiene una gráfica de dispersión en la que se relacionan todos los coeficientes estimados en las especificaciones en las que se interactúan las variables explicativas con el puntaje individual y el puntaje del colegio por género. En los ejes de cada gráfica se encuentran las magnitudes de los coeficientes estimados para cada uno de los grupos poblacionales. La recta roja es una línea de 45° que sirve como punto de referencia de dónde se ubicaría un punto de dispersión si el coeficiente estimado entre hombres y mujeres es igual. Dado que es una dispersión, no se pueden incluir los intervalos de confianza.

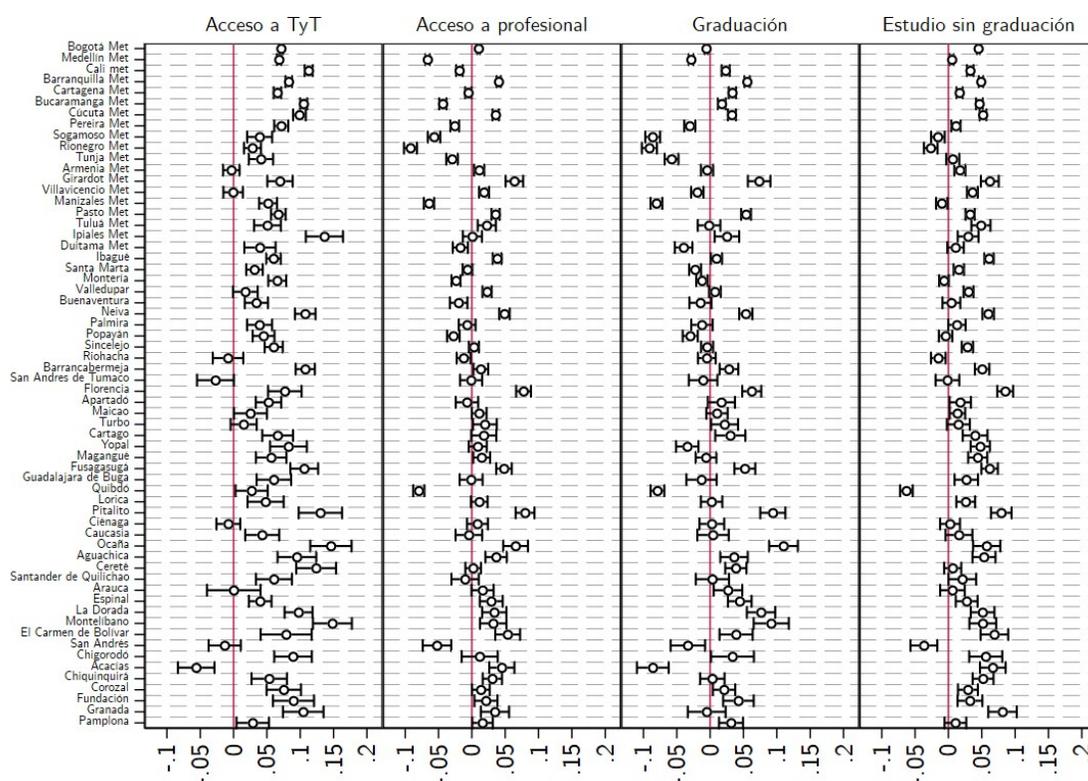
Lo que se puede observar es que en salarios no parece haber un patrón claro, pues los coeficientes están agrupados a lo largo de la línea de 45°. Sin embargo, resaltan algunos coeficientes específicos en los que hay una diferencia pronunciada entre hombres y mujeres. Por ejemplo, los puntos de color rojo oscuro muestran el caso en que los retornos de los hombres son casi el doble que los de las mujeres. Estos tres puntos hacen referencia (de izquierda a derecha) a los siguientes coeficientes: STEM profesional\*Puntaje colegio, Grado profesional\*Puntaje colegio, y Grado\*Puntaje colegio.

De igual forma, los puntos de color verde oscuro, muestran el caso en que si bien ambos coeficientes son negativos, un grupo se ve menos afectado que el otro. En este caso, el que está sobre la línea de 45°, es decir, que la brecha está en favor de los hombres, es el de Ciencias de la educación\*Puntaje colegio. El que se encuentra por debajo de la línea roja, mostrando brecha en favor de las mujeres es el de Programas no STEM TyT\*Puntaje colegio. Finalmente, el punto de color verde claro es quizás el caso más extremo, donde la brecha es muy amplia en favor de los hombres. Este punto hace referencia al coeficiente de Matemáticas\*Puntaje colegio.

## Heterogeneidad por ciudades

Las estimaciones en las 62 ciudades propuestas por O'Clery, Prieto Curiel & Lora (2019), muestran una heterogeneidad interesante al interior del país. Es importante resaltar que la ciudad de estudio es aquella en la que el individuo vivía antes de tomar la decisión de ir a la universidad, alrededor de los 16 o 17 años. Al ver los efectos sobre formalidad (Figura 6), se puede notar que los programas profesionales no necesariamente están asociados a una mayor tasa de formalidad, mientras que las técnicas y tecnológicas están correlacionadas en su mayoría con una mayor formalidad. Por su parte, al separar por graduación y no graduación, se ve que graduarse de educación superior no necesariamente está asociado a mayor formalidad, pues hay una gran heterogeneidad entre ciudades. Sin embargo, estudiar algo de educación superior parece estar correlacionado de manera más fuerte con una mayor tasa de formalidad.

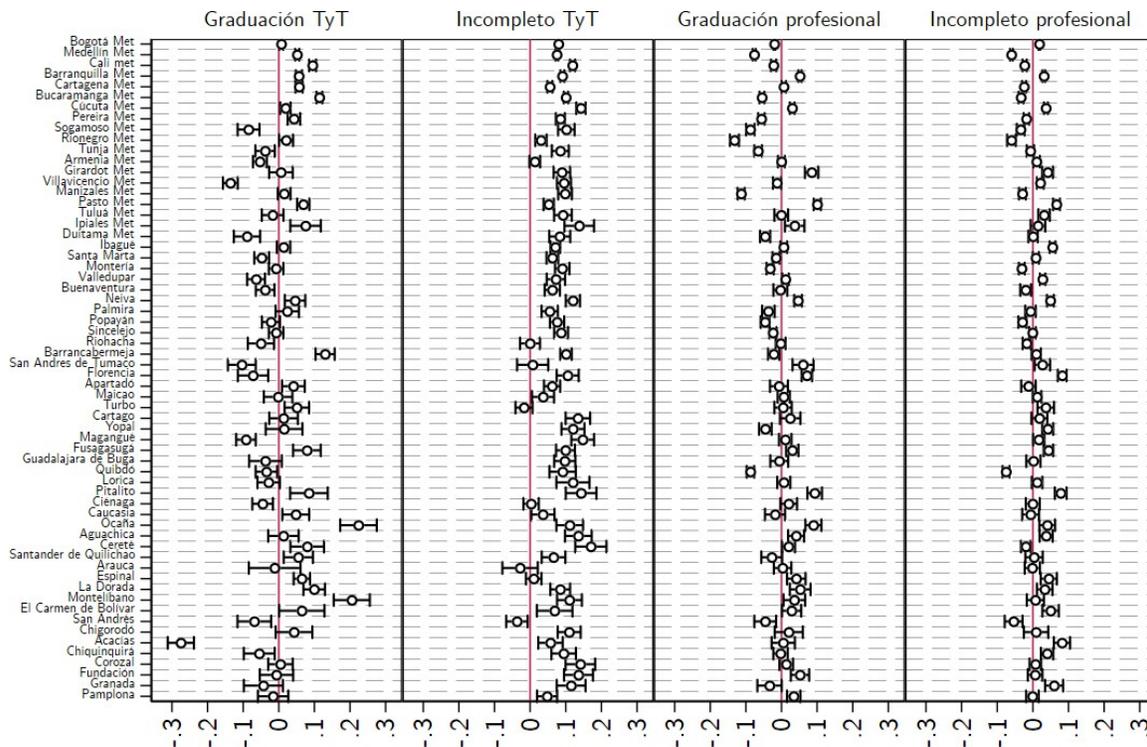
**Figura 6:** Estimaciones de tipo de programa y graduación sobre formalidad



*Nota:* La gráfica tiene cuatro paneles, uno para cada variable explicativa. Cada fila hace referencia a una ciudad específica. Los círculos representan los coeficientes estimados de la variable explicativa de cada panel en cada una de las ciudades. Dado que son estimaciones, se incluyen también los intervalos de confianza alrededor del coeficiente para mostrar la precisión del mismo.

Al detallar el tipo de programa del que se gradúan las personas (Figura 7), se llega a una conclusión similar. Graduarse de un programa, sea TyT o profesional no está relacionado con una mayor formalidad de forma generalizada, pues hay una gran heterogeneidad entre ciudades. La mayor formalidad solo se ve en aquellas personas que no completan su programa TyT.

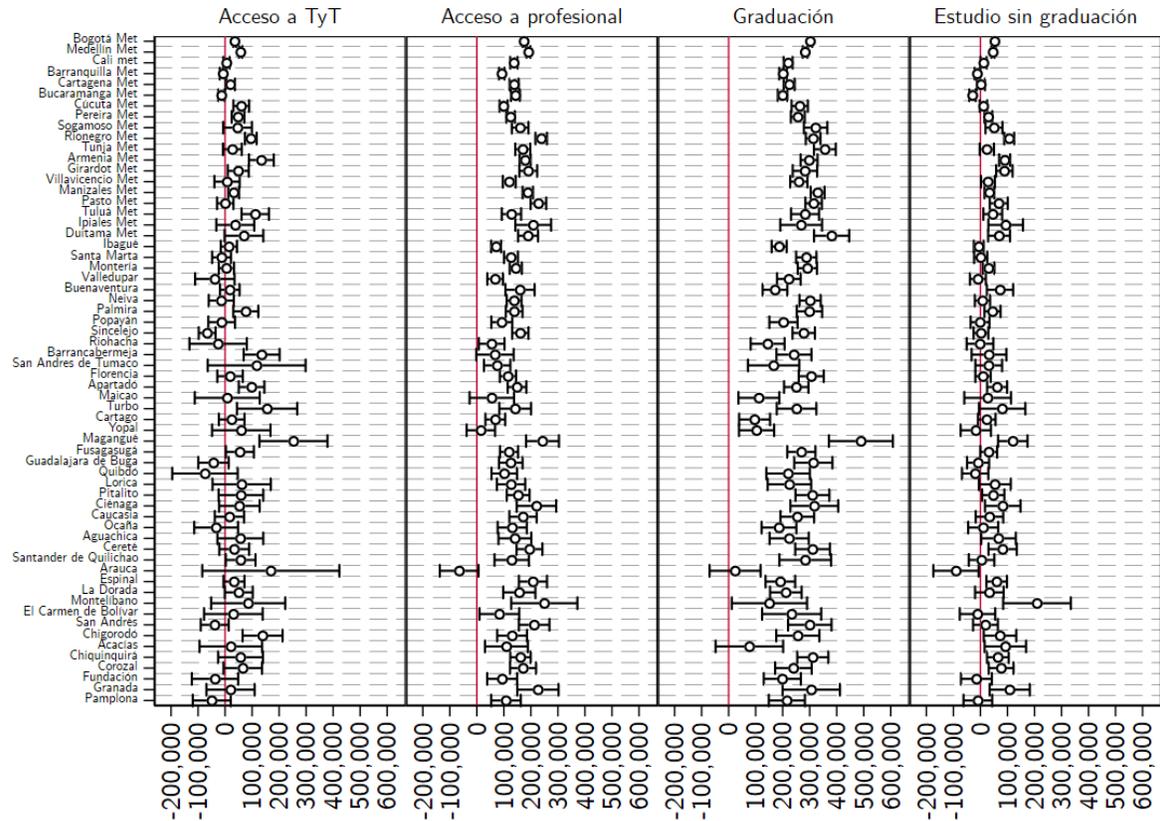
**Figura 7:** Estimaciones de graduación por programa sobre formalidad



*Nota:* La gráfica tiene cuatro paneles, uno para cada variable explicativa. Cada fila hace referencia a una ciudad específica. Los círculos representan los coeficientes estimados de la variable explicativa de cada panel en cada una de las ciudades. Dado que son estimaciones, se incluyen también los intervalos de confianza alrededor del coeficiente para mostrar la precisión del mismo.

Ahora bien, los resultados de retornos monetarios siguen un patrón completamente distinto a los de formalidad. Como se puede ver en la Figura 8, los programas profesionales están asociados con mayores salarios a lo largo de casi todo el país, con contadas excepciones. Sin embargo, los programas técnicos y tecnológicos muestran una gran heterogeneidad entre ciudades, y en muchas de ellas tienen efectos nulos. Por su parte, al observar lo que pasa con la graduación de educación superior, se nota la existencia del *sheepskin effect*, donde la graduación está asociada con salarios mucho más altos en la gran mayoría de ciudades del país. Adicionalmente, tener estudios incompletos parece también estar asociado con mayores salarios que no haber entrado a la educación superior, pero son menores en magnitud que los de graduados, y en muchos casos son nulos.

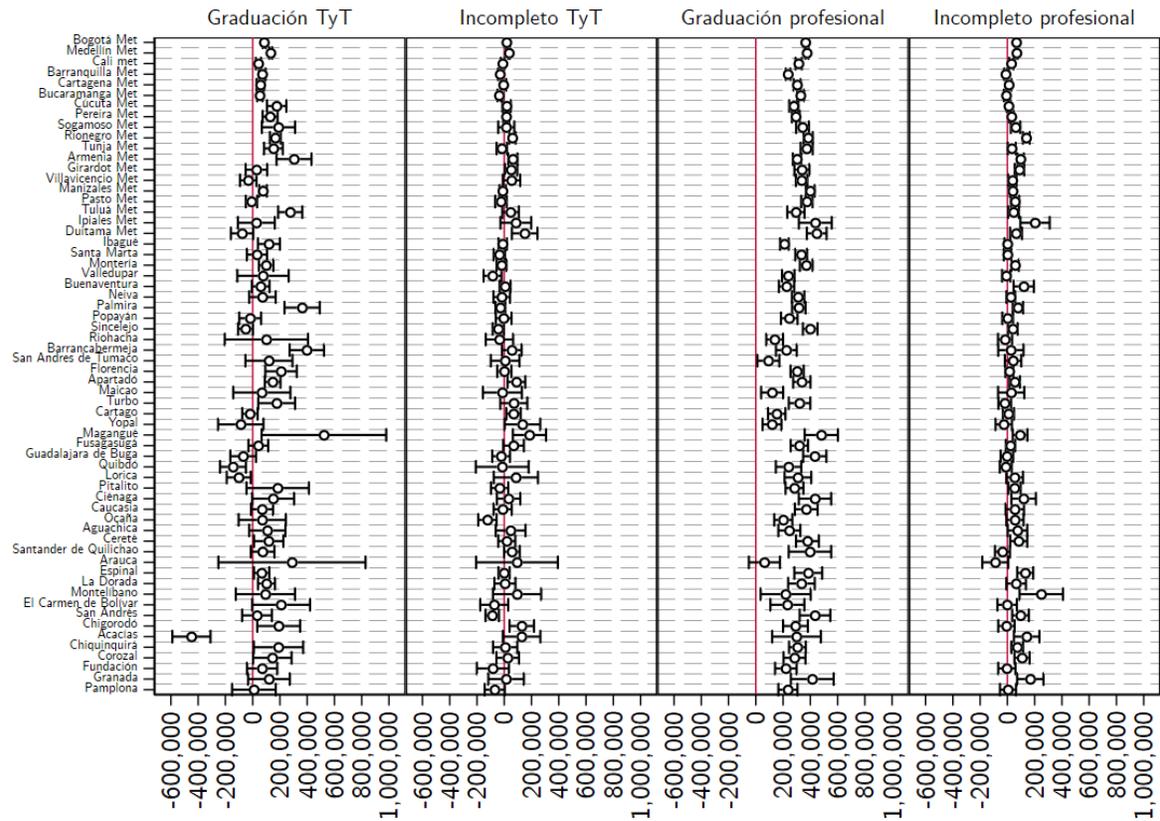
**Figura 8:** Estimaciones de tipo de programa y graduación sobre salarios



*Nota:* La gráfica tiene cuatro paneles, uno para cada variable explicativa. Cada fila hace referencia a una ciudad específica. Los círculos representan los coeficientes estimados de la variable explicativa de cada panel en cada una de las ciudades. Dado que son estimaciones, se incluyen también los intervalos de confianza alrededor del coeficiente para mostrar la precisión del mismo.

Al separar la graduación por tipo de programa (Figura 9), se observa que tener estudios incompletos de profesional o de TyT no está asociado con mayores salarios, pues en la mayoría de ciudades hay efectos nulos. Sin embargo, graduarse de un programa profesional tiene efectos positivos sobre los salarios. Esto también parece suceder con la graduación de programas TyT, aunque las estimaciones son más ruidosas.

**Figura 9:** Estimaciones de graduación por programa sobre salarios



*Nota:* La gráfica tiene cuatro paneles, uno para cada variable explicativa. Cada fila hace referencia a una ciudad específica. Los círculos representan los coeficientes estimados de la variable explicativa de cada panel en cada una de las ciudades. Dado que son estimaciones, se incluyen también los intervalos de confianza alrededor del coeficiente para mostrar la precisión del mismo.

## Referencias

- Layard, R., & Psacharopoulos, G. (1974). The screening hypothesis and the returns to education. *Journal of political economy*, 82(5), 985-998.
- O'Clery, N., Curiel, R. P., & Lora, E. (2019). Commuting times and the mobilisation of skills in emergent cities. *Applied Network Science*, 4(1), 1-27.

# Apéndice

Tabla A.1: Resultados de tipo de programa con diferentes especificaciones

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Tasa de formalidad	Tasa de formalidad	Tasa de formalidad	Días cotizados (última observación anual)	Días cotizados (última observación anual)	Días cotizados (última observación anual)	Salario con missing (última observación anual)	Salario con missing (última observación anual)	Salario con missing (última observación anual)	Asalariado	Asalariado	Asalariado
Educación TyT completa o incompleta	0.106*** (0.001)	0.069*** (0.001)	0.080*** (0.001)	0.164*** (0.006)	0.170*** (0.006)	0.168*** (0.006)	26,919*** (1,431)	36,654*** (1,406)	29,741*** (1,439)	0.098*** (0.001)	0.063*** (0.001)	0.071*** (0.001)
Educación profesional completa o incompleta	0.004*** (0.000)	0.009*** (0.000)	0.007*** (0.000)	0.221*** (0.004)	0.210*** (0.004)	0.219*** (0.004)	159,920*** (1,280)	148,065*** (1,258)	155,966*** (1,283)	-0.019*** (0.000)	-0.011*** (0.000)	-0.016*** (0.000)
Puntaje individual	0.012*** (0.000)	0.004*** (0.000)	0.003*** (0.000)	0.058*** (0.002)	0.060*** (0.002)	0.062*** (0.002)	83,871*** (725)	81,792*** (703)	86,677*** (726)	0.011*** (0.000)	0.004*** (0.000)	0.002*** (0.000)
Puntaje colegio	0.130*** (0.001)	0.115*** (0.001)	0.094*** (0.001)	0.275*** (0.007)	0.264*** (0.009)	0.296*** (0.007)	457,375*** (3,981)	278,132*** (4,868)	465,992*** (4,030)	0.080*** (0.001)	0.075*** (0.001)	0.047*** (0.001)
Jornada completa colegio	-0.015*** (0.000)		-0.033*** (0.000)	0.022*** (0.004)		0.015*** (0.005)	30,089*** (1,433)		32,870*** (1,492)	-0.018*** (0.000)		-0.037*** (0.000)
Estudiante paga pensión	-0.012*** (0.000)		-0.011*** (0.000)	0.009** (0.004)		0.009** (0.004)	21,242*** (1,033)		27,659*** (1,068)	-0.014*** (0.000)		-0.015*** (0.000)
Mujer	-0.073*** (0.000)	-0.091*** (0.000)	-0.078*** (0.000)	0.375*** (0.004)	0.376*** (0.004)	0.374*** (0.004)	-69,095*** (1,068)	-68,527*** (1,047)	-67,305*** (1,065)	-0.092*** (0.000)	-0.109*** (0.000)	-0.097*** (0.000)
Colegio académico	0.009*** (0.000)		-0.011*** (0.000)	0.021*** (0.004)		-0.010** (0.005)	15,310*** (1,234)		11,817*** (1,281)	0.005*** (0.000)		-0.015*** (0.000)
Calendario A	0.004*** (0.000)		0.016*** (0.001)	-0.002 (0.005)		-0.087*** (0.006)	-40,030*** (1,521)		-90,196*** (2,024)	0.009*** (0.000)		0.026*** (0.001)
Observaciones	7,237,215	7,719,285	7,237,215	7,237,215	7,719,285	7,237,215	2,894,311	3,167,145	2,894,311	7,237,215	7,719,285	7,237,215
Media del control	0.386	0.386	0.386	10.02	10.02	10.02	1,050,000	1,050,000	1,050,000	0.335	0.335	0.335
Controles	SI	SI	SI	SI	SI	SI	SI	SI	SI	SI	SI	SI
FE colegio	NO	SI	NO	NO	SI	NO	NO	SI	NO	NO	SI	NO
FE departamento	NO	NO	SI	NO	NO	SI	NO	NO	SI	NO	NO	SI

Errores estándar robustos a la heterocedasticidad en paréntesis.

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Tabla A.2: Resultados de tipo de programa e interacciones con diferentes especificaciones

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Tasa de formalidad	Tasa de formalidad	Tasa de formalidad	Días cotizados (última observación anual)	Días cotizados (última observación anual)	Días cotizados (última observación anual)	Salario con missing (última observación anual)	Salario con missing (última observación anual)	Salario con missing (última observación anual)	Asalariado	Asalariado	Asalariado
Educación TyT completa o incompleta	0.099*** (0.001)	0.065*** (0.001)	0.072*** (0.001)	0.159*** (0.007)	0.164*** (0.006)	0.162*** (0.007)	40,293*** (1,449)	45,118*** (1,419)	43,068*** (1,459)	0.091*** (0.001)	0.059*** (0.001)	0.065*** (0.001)
Educación profesional completa o incompleta	0.009*** (0.000)	0.012*** (0.000)	0.012*** (0.000)	0.225*** (0.004)	0.214*** (0.004)	0.224*** (0.004)	143,011*** (1,280)	133,911*** (1,236)	138,620*** (1,281)	-0.015*** (0.000)	-0.007*** (0.000)	-0.011*** (0.000)
Puntaje individual	0.028*** (0.000)	0.019*** (0.000)	0.021*** (0.000)	0.070*** (0.003)	0.071*** (0.003)	0.074*** (0.003)	53,561*** (872)	54,588*** (862)	55,397*** (873)	0.024*** (0.000)	0.016*** (0.000)	0.018*** (0.000)
Puntaje colegio	0.164*** (0.001)	0.097*** (0.001)	0.114*** (0.001)	0.280*** (0.011)	0.290*** (0.012)	0.317*** (0.012)	368,319*** (5,603)	298,226*** (6,108)	384,120*** (5,695)	0.113*** (0.001)	0.056*** (0.001)	0.067*** (0.001)
TyT*Puntaje individual	0.007*** (0.001)	0.002*** (0.001)	0.002*** (0.001)	0.019*** (0.008)	0.019*** (0.007)	0.022*** (0.008)	-10,157*** (1,763)	-6,986*** (1,711)	-9,504*** (1,758)	0.009*** (0.001)	0.003*** (0.001)	0.004*** (0.001)
Profesional*Puntaje individual	-0.036*** (0.000)	-0.035*** (0.000)	-0.040*** (0.000)	-0.032*** (0.004)	-0.030*** (0.004)	-0.031*** (0.004)	73,629*** (1,604)	65,512*** (1,536)	74,911*** (1,601)	-0.031*** (0.000)	-0.030*** (0.000)	-0.035*** (0.000)
TyT*Puntaje colegio	-0.040*** (0.003)	-0.023*** (0.003)	-0.038*** (0.003)	-0.065*** (0.035)	-0.102*** (0.030)	-0.060*** (0.035)	-197,702*** (11,700)	-184,272*** (9,598)	-203,876*** (11,649)	-0.033*** (0.003)	-0.009*** (0.003)	-0.030*** (0.003)
Profesional*Puntaje colegio	-0.049*** (0.002)	0.033*** (0.002)	-0.022*** (0.002)	0.007 (0.015)	-0.038*** (0.014)	-0.022 (0.015)	140,970*** (8,253)	-9,842 (7,421)	126,978*** (8,259)	-0.050*** (0.002)	0.033*** (0.002)	-0.024*** (0.002)
Jornada completa colegio	-0.014*** (0.000)		-0.032*** (0.000)	0.023*** (0.004)		0.016*** (0.005)	27,054*** (1,428)		30,015*** (1,487)	-0.017*** (0.000)		-0.036*** (0.000)
Estudiante paga pensión	-0.011*** (0.000)		-0.010*** (0.000)	0.010*** (0.004)		0.010*** (0.004)	18,101*** (1,033)		25,366*** (1,067)	-0.013*** (0.000)		-0.015*** (0.000)
Mujer	-0.074*** (0.000)	-0.091*** (0.000)	-0.079*** (0.000)	0.374*** (0.004)	0.376*** (0.004)	0.373*** (0.004)	-65,480*** (1,064)	-66,983*** (1,046)	-63,975*** (1,061)	-0.093*** (0.000)	-0.109*** (0.000)	-0.097*** (0.000)
Colegio académico	0.011*** (0.000)		-0.010*** (0.000)	0.022*** (0.004)		-0.009* (0.005)	10,725*** (1,230)		8,738*** (1,280)	0.007*** (0.000)		-0.014*** (0.000)
Calendario A	0.002*** (0.000)		0.013*** (0.001)	-0.003 (0.005)		-0.090*** (0.006)	-32,109*** (1,498)		-78,020*** (1,985)	0.007*** (0.000)		0.023*** (0.001)
Observaciones	7,237,215	7,719,285	7,237,215	7,237,215	7,719,285	7,237,215	2,894,311	3,167,145	2,894,311	7,237,215	7,719,285	7,237,215
Medía del control	0.386	0.386	0.386	10.02	10.02	10.02	1,050,000	1,050,000	1,050,000	0.335	0.335	0.335
Controles	SI	SI	SI	SI	SI	SI	SI	SI	SI	SI	SI	SI
FE colegio	NO	SI	NO	NO	SI	NO	NO	SI	NO	NO	SI	NO
FE departamento	NO	NO	SI	NO	NO	SI	NO	NO	SI	NO	NO	SI

Errores estándar robustos en paréntesis  
 \*\*\* p<0.01, \*\* p<0.05, \* p<0.1



Documentos de antecedentes del  
**Informe Nacional de Desarrollo Humano**  
2024

