



Efecto Par de Género y Elección de Carreras STEM: Evidencia para Chile

TESIS PARA OPTAR AL GRADO DE
Magíster en Análisis Económico

Autor:
Felipe Flores Vasquez

Profesora Guía:
Valentina Paredes Haz

Fecha de Entrega:
16 de junio de 2021

Agradecimientos

Esta tesis, junto a la segunda mitad del MAE, ha sido marcada tanto por el estallido social como por la pandemia del coronavirus. Y aunque el esfuerzo personal es importante a la hora de terminar un magíster (no conozco a nadie que no se haya desvelado estudiando o terminando trabajos, o no haya faltado a encuentros familiares debido a la carga académica del magíster), me parece importante reconocer el apoyo de todas las personas que me han ayudado a llegar a donde estoy ahora.

En primer lugar, debo agradecer la ayuda de mi familia. A riesgo de usar términos muy economicistas, la familia es una variable exógena, no podemos elegir la que nos toca y esta tiene consecuencias de largo plazo en nuestro desarrollo. En ese sentido, no puedo sino pensar que me saqué la lotería con la familia que tengo: han sido mi base moral y anímica durante toda mi vida, han sido un lugar seguro y tranquilo donde refugiarme después de cada paso en mi vida, y a pesar de que no siempre ha sido fácil, me han dado la libertad de dedicarme por completo a terminar mis estudios superiores durante 7 años que podría haber estado generando ingresos. Nunca podré agradecerles lo suficiente por todo esto.

En segundo lugar, debo agradecer a las personas que conocí gracias al colegio. Estudiar en un colegio “emblemático” me dio la base académica para estudiar en la FEN, pero fueron mis amigos y amigas que hice durante mi vida escolar quienes me ayudaron reducir en gran parte los aspectos negativos que surgen al desarrollarse en un contexto tan competitivo y sexista como lo era el Instituto Nacional. Me han ayudado más de lo que se imaginan, y nuestras juntas por discord básicamente todos los viernes de cuarentena han sido cruciales para terminar el MAE sin perder la cordura durante la pandemia.

En tercer lugar, debo resaltar la ayuda que recibí de mis amigos y amigas en la universidad. Desde la ayuda cuando no entendía alguna materia, a los carretes cuando necesitaba desconectarme de nuestras obligaciones académicas, cada uno de ustedes me ha permitido ser la persona que soy en estos momentos. Mención especial al hermoso grupo machmelou, cuyo apoyo sentí durante toda mi vida universitaria y sin los cuales habría abandonado la carrera antes de siquiera pensar en hacer un magíster.

Finalmente, pero no menos importante, debo dar las gracias a la Escuela Desarrollo de Talentos. Me ayudaron a mantener los pies en la tierra mientras terminaba una carrera que en muchos casos nos hacía quedaros con modelos teóricos. El esfuerzo y las ganas que ponían tanto alumnos, tutores, profesores y equipo de la escuela me ha inspirado hasta el día de hoy, y la temática de esta tesis surge en gran parte a la pasión por la educación que me ayudaron a desarrollar.

Resumen

La presente tesis estudia la relación entre la proporción de pares mujeres a los cuales están expuestos los estudiantes en el colegio con la probabilidad de matricularse y titularse de carreras STEM, fenómeno denominado efecto par de género.

Se estiman modelos diferenciados para hombres y mujeres, encontrándose un efecto negativo del porcentaje de mujeres tanto sobre la probabilidad de matricularse como titularse de una carrera STEM para ambos sexos, aunque al utilizar variables instrumentales para controlar por un posible ordenamiento no aleatorio de alumnos en enseñanza media se pierde la significancia en la muestra de mujeres. Al extender el análisis al desempeño en la prueba PSU, se observa una relación generalmente positiva, y más fuerte en matemáticas.

Se encuentra evidencia a favor de una relación no lineal entre las variables estudiadas al incluir el cuadrado del porcentaje de mujeres. En específico, la relación entre proporción de mujeres y ambos *outcomes* STEM es convexa, llegando a una probabilidad mínima entre el 80 % y 90 % de mujeres en el curso. Por el contrario, la relación con el rendimiento en la PSU es cóncava, llegando a un máximo del puntaje con una distribución entre 40 % y 60 % de mujeres aproximadamente.

Así, medidas que busquen cerrar brechas de género en la elección de carreras STEM mediante la creación de ambientes de un solo sexo (dentro de todo el colegio o en algunos cursos), tendrían un doble efecto negativo: por un lado, la brecha se hace más grande debido a un aumento en la probabilidad de elección de carreras STEM por parte de hombres y una disminución de la probabilidad por parte de la mujer, o falta de efecto en el mejor de los casos; por otro lado, disminuirán los puntajes PSU de los alumnos en estos ambientes monogénicos, independiente de su sexo.

Índice

| | |
|---|-----------|
| 1. Introducción | 1 |
| 2. Relevancia Económica | 2 |
| 2.1. Relevancia Microeconómica: Participación Laboral y Brecha Salarial | 2 |
| 2.2. Relevancia Macroeconómica: Productividad y Crecimiento Económico | 3 |
| 3. Revisión de Literatura | 4 |
| 3.1. Efecto Par de Género y Elección de Carrera | 4 |
| 3.2. Efecto Par de Género y Rendimiento Académico | 4 |
| 3.3. Mecanismos detrás del Efecto Par de Género | 6 |
| 4. Estrategia Empírica | 7 |
| 4.1. Modelo Base | 7 |
| 4.2. Modelo Ampliado | 9 |
| 4.3. ¿Cómo identificar a los pares? | 10 |
| 5. Datos | 11 |
| 5.1. Fuentes de Información | 11 |
| 5.1.1. PSU y <i>outcomes</i> STEM | 11 |
| 5.1.2. SIMCE | 12 |
| 5.2. Estadística Descriptiva | 13 |
| 6. Resultados Empíricos | 16 |
| 6.1. Elección de Carrera: Matrícula y Titulación STEM | 16 |
| 6.2. Desempeño Académico | 19 |
| 7. Análisis de Robustez | 24 |
| 7.1. Test de Falsificación | 24 |
| 7.2. Análisis de Heterogeneidad: Tipo de Establecimiento | 26 |
| 7.3. Análisis de heterogeneidad: Escolaridad del Padre y la Madre | 28 |
| 7.4. Efectos No Lineales | 29 |
| 7.5. ¿En que curso se debe medir el porcentaje de mujeres? | 32 |
| 8. Conclusiones | 35 |

| | |
|---|-----------|
| 9. Anexos | 40 |
| Anexo 1: Descomposición de la varianza: % de mujeres en el curso | 40 |
| Anexo 2: Detalle de controles utilizados: PSU y STEM | 41 |
| Anexo 3: Detalle de controles utilizados: SIMCE | 43 |
| Anexo 4: Distribución de titulación STEM, según sexo del alumno y porcentaje de mujeres en IV medio | 45 |
| Anexo 5: Heterogeneidad según dependencia del colegio - PSU | 46 |
| Anexo 6: Heterogeneidad según escolaridad de la madre - Titulación STEM | 47 |
| Anexo 7: Heterogeneidad según escolaridad de la madre - PSU | 48 |
| Anexo 8: Heterogeneidad según escolaridad del padre - PSU | 49 |
| Anexo 9: Efecto par no lineal - puntaje PSU | 50 |
| Anexo 10: Puntaje PSU predicho - efecto par de género cuadrático | 51 |
| Anexo 11: Efecto par de género - robustez ante cambios en el curso donde se mide el porcentaje de mujeres | 52 |

1. Introducción

Las brechas de género en el mercado laboral son un problema que afecta a gran parte de los países, independiente de su nivel de desarrollo. Datos para países OCDE en el 2018 muestran que mientras la tasa de participación laboral de los hombres era en promedio de 69 %, la de mujeres llegaba solo a 52,5 %. Las diferencias también se observan en los salarios, llegando al 13 % en la mediana de la distribución de salarios en contra de las trabajadoras. Lo anterior ocurre a pesar de los esfuerzos realizados para promover un acceso más equitativo a la educación superior, observándose en la actualidad que las mujeres representan más del 50 % de los matriculados y graduados de la educación superior (OECD, 2020).

¿Por qué persisten entonces estas disparidades en el mercado laboral? Gran parte de la brecha explicada se debería a una fuerte segregación ocupacional (Blau & Kahn, 2017), la que a su vez se produce por fuertes desigualdades en la elección de carrera en la educación superior (Joy, 2006). Es precisamente esto lo que se observa al desagregar las cifras en el contexto chileno, ya que a pesar de que las mujeres son el sexo dominante al observar la matrícula total, representan solo un 22,6 % en carreras ligadas a ciencias, mientras que en ingeniería el porcentaje baja a un 17,4 % (CONICYT, 2018).

El objetivo de la presente investigación es evaluar el impacto de una variable que podría explicar la elección de carreras por parte de los y las jóvenes: la proporción de género a las que están expuestos en su etapa escolar. En específico, se estimará el impacto del efecto par de género, medido mediante el porcentaje de mujeres a los que está expuesto el estudiante, sobre la matrícula y titulación de carreras STEM. Además, se estudiará uno de los posibles mecanismos que explican la relación entre ambas variables: el desempeño en la PSU. Para identificar el efecto se utilizará un modelo de efectos fijos, aprovechando la variación idiosincrática que se genera al comparar cohortes adyacentes de un mismo nivel y colegio. Así, será posible encontrar un coeficiente limpio de no observables fijos en el tiempo a nivel de establecimiento y de sesgo de selección por parte de los colegios.

Una ventaja de utilizar la proporción de género como un instrumento para reducir brechas en la elección de carreras, en caso de encontrar evidencia a favor, es que su costo de implementación es virtualmente cero. Y es que, a pesar de que realizar un ordenamiento de alumnos entre distintos colegios podría resultar costoso en términos económicos, hacerlo entre cursos dentro de un mismo establecimiento no implicaría un gasto considerable. Así, por ejemplo, la creación de cursos de un solo sexo dentro de liceos coeducacionales, aunque sea solo para asignaturas específicas, podría ayudar a reducir las diferencias en la elección de carrera, reduciendo en el largo plazo la segregación ocupacional y las brechas salariales.

El aporte a la literatura existente surge de dos dimensiones principales. En primer lugar, se entregará evidencia para un país en desarrollo como Chile, donde no se ha estudiado directamente el efecto par de género sobre la elección de carreras. La presencia de mayores brechas de género, tanto en el desempeño académico como en la elección de carreras, puede llevar a encontrar efectos de distinta magnitud a los vistos en países desarrollados. Por otro

lado, se propone una mejora metodológica a la hora de estudiar este fenómeno al controlar por el efecto par tradicional, y diferenciarlo del efecto par de género.

El informe se estructura de la siguiente forma: en la sección 2 se detallará la relevancia económica del tema. En la sección 3 se examina la literatura existente sobre el efecto par de género, tanto internacional como nacional, detallando sus principales conclusiones. En la sección 4 se expone la estrategia empírica de la investigación, identificando los posibles problemas de endogeneidad que podrían sesgar las estimaciones y las soluciones propuestas. En la sección 5 se detallan las fuentes de datos con las que se trabaja, presentando estadística descriptiva que permite caracterizar la población estudiada. La sección 6 muestra los resultados del estudio, diferenciando según el género de los estudiantes, mientras que en la sección 7 se analizan posibles heterogeneidades en el efecto. Finalmente, la sección 8 expone las principales conclusiones que se deducen de los resultados, posibilidades de mejora, y posibles políticas públicas que se pueden seguir a futuro.

2. Relevancia Económica

La relevancia económica de estudiar el efecto par de género viene dada, principalmente, por su potencial como instrumento para cerrar las brechas de género en la elección de carrera. La existencia de diferencias entre hombres y mujeres en esta decisión, especialmente aquellas del área STEM, tendrá consecuencias tanto microeconómicas como macroeconómicas, las cuales serán discutidas a continuación:

2.1. Relevancia Microeconómica: Participación Laboral y Brecha Salarial

Históricamente se ha pensado en los años de educación como un mecanismo para reducir las brechas salariales existentes en el mercado laboral. Sin embargo, estudios recientes muestran que es la segregación ocupacional, y no la escolaridad, lo que explica en gran parte la diferencia salarial entre hombres y mujeres (Blau & Kahn, 2017). Brechas en la elección de carreras explican en gran parte la segregación ocupacional (Joy, 2006), por lo que no es de extrañar que la sobrerrepresentación de hombres en carreras STEM se traspase al mercado laboral.

Sumado a lo anterior, se observa que las carreras STEM, y en general, las carreras masculinizadas¹, reciben un mayor retorno monetario a lo largo de todo el ciclo de vida (Arcidiacono, 2004). Así, lo que inicialmente se observa como una brecha en la elección de carreras tendría efectos de largo plazo, siendo una de las principales variables explicativas de las diferencias

¹Se tiende a clasificar una carrera como masculinizada cuando existe una brecha de género importante en la elección de esta carrera, la cual se sostiene en el tiempo. De manera similar se definen carreras femeninas

salariales entre hombres y mujeres con educación superior. Esta brecha salarial no solo afectará el bienestar individual de las mujeres mediante la pérdida de ingreso, sino que también afectará su poder de negociación en modelos de *intrahousehold allocation* y podría cambiar sus decisiones sobre fertilidad (Brenøe & Zölitz, 2020).

En conclusión, la importancia de investigar el efecto par de género sobre la elección de carreras STEM proviene de su potencial utilización como instrumento para cerrar brechas de género en el mercado laboral. En un escenario ideal, el efecto par de género sería positivo y significativo únicamente para mujeres, por lo que se podría aumentar tanto el desempeño académico como la probabilidad de elegir carreras STEM de las mujeres sin afectar negativamente a los hombres. Aún en el caso de que sea significativo para ambos sexos, se podría encontrar una distribución de género óptima al interior de las salas de clases.

2.2. Relevancia Macroeconómica: Productividad y Crecimiento Económico

La equidad de género, además de un fin en sí mismo, tiene efectos importantes a la hora de generar un crecimiento inclusivo y sustentable. Por un lado, el aumento de la participación de mujeres en el área STEM puede impactar directamente el crecimiento de los países, debido al efecto positivo de esta área sobre la productividad, sobre todo la de trabajadores con educación superior (Peri, Shih, & Sparber, 2015).

Por otro lado, situándonos dentro del contexto del modelo de Solow, el crecimiento de largo plazo depende del trabajo efectivo, el capital físico y la productividad total de los factores. La reducción de brechas de participación laboral, mediante el aumento de la proporción de mujeres en carreras más demandadas como lo son las STEM, permite aumentar tanto el trabajo efectivo como la PTF (Devadas & Kim, 2020). El primero de estos efectos es directo, ya que el aumento de la participación laboral aumenta el número de trabajadores sin afectar el tamaño de la población, aumentando por tanto la producción. Este efecto podría ser especialmente importante en países en transición demográfica, evitando problemas asociados al envejecimiento de la población.

El efecto sobre la productividad se explica por la mayor eficiencia en el *matching* entre puesto de trabajo e individuo. Si se asume que no hay diferencias de habilidad entre hombres y mujeres, entonces la existencia de brechas en la participación de mujeres en áreas STEM implica una distribución ineficiente de los trabajadores, existiendo mujeres de alta habilidad que están quedando fuera del sector. Cuberes & Teigner (2016) estiman que las inequidades de género generan en promedio una pérdida de ingresos de 14% de países OCDE, donde cerca de un 44% de esta pérdida se debe exclusivamente a la segregación ocupacional.

Otra razón por la que el aumento en la participación laboral de las mujeres afectaría la productividad, y por tanto el crecimiento económico, viene dada por la complementariedad del trabajo de hombres y mujeres. Modelos clásicos de crecimiento, al utilizar un individuo representativo, asumen una perfecta sustitución entre trabajadores de ambos sexos. Sin em-

bargo, estudios empíricos suelen encontrar cierto grado de complementariedad entre mano de obra masculina y femenina (Ostry, Alvarez, Espinoza, & Papageorgiou, 2018). Así, los autores encuentran que cerrar las brechas de género en la participación laboral podría aumentar el PIB de los países entre 10 % y 80 %, dependiendo de la calibración utilizada.

3. Revisión de Literatura

En esta sección se presenta una breve revisión de la literatura relevante. El análisis se dividirá en tres partes: por un lado, se analizará la evidencia que relaciona el efecto par de género y la elección de carrera; en segundo lugar, se examinará la literatura que se centra en el impacto de la distribución de género sobre el desempeño académico. Finalmente, se discutirán los posibles mecanismos que estarían detrás del efecto par de género.

3.1. Efecto Par de Género y Elección de Carrera

No existe un consenso claro sobre la dirección y magnitud del efecto par de género sobre la matrícula en carreras STEM, y más generalmente, en la matrícula en carreras masculinizadas. Por un lado, se ha encontrado que un mayor porcentaje de mujeres en la cohorte reforzaría los estereotipos de género de la sociedad, provocando que mujeres elijan carreras femeninas y hombres carreras masculinizadas con mayor probabilidad (Brenøe & Zölitz, 2020), ampliando por tanto las brechas de género existentes. Lo anterior contrasta con otras investigaciones que efecto positivo para las mujeres (Schneeweis & Zweimüller, 2012), o incluso la ausencia de un efecto significativo en promedio (Anelli & Peri, 2019).

La evidencia de efectos sobre la graduación de carreras “masculinizadas” también es mixta. Se ha encontrado un efecto significativo para hombres y mujeres (Feld & Zölitz, 2018), significativo solo para hombres (Hill, 2017), o ningún efecto (Anelli & Peri, 2019).

En Chile esta dimensión del efecto par de género solo ha sido estudiada de forma indirecta. Bordón, Canals & Mizala (2020) muestran que tanto hombres como mujeres tienen una mayor probabilidad de elegir carreras de áreas femeninas (como salud y pedagogía) al provenir de colegios con más de 60 % de mujeres. Por el contrario, estos alumnos tienen una menor probabilidad de postular a carreras masculinizadas como ingeniería. Un resultado de especial interés expuesto por las autoras es que cerrar estas brechas de género en la postulación a carreras no genera una reducción en el *pool* de habilidad, al menos al medirlo por el puntaje PSU de los alumnos aceptados.

3.2. Efecto Par de Género y Rendimiento Académico

La investigación del efecto par de género sobre el desempeño académico es algo más amplia, y a diferencia de la dimensión anterior, sí ha logrado cierto grado de consenso. Los trabajos de

Hoxby (2000) y Lavy & Schlosser (2011) corresponden a los primeros esfuerzos econométricos de medirlo, encontrando ambos un efecto positivo y significativo del porcentaje de mujeres sobre el desempeño académico para alumnos de ambos géneros. Mientras el primero encuentra que el efecto aumentaría cuando el alumno tiene más edad, el segundo encuentra evidencia de que el porcentaje de mujeres también ayuda a reducir la interrupción y violencia en la sala, mejora las relaciones con los profesores, y disminuye su fatiga, lo que funcionaría como potencial mecanismo para explicar el efecto encontrado.

Investigaciones más recientes han resaltado la importancia de aspectos omitidos anteriormente. Por un lado, los colegios pueden actuar tomando en cuenta el efecto par de género, generando acciones que mejoren el ambiente educativo en cursos con menor cantidad de mujeres, lo que sesgaría hacia abajo el coeficiente estimado (Iversen & Bonesrønning, 2015). Por otro lado, este efecto podría ser especialmente fuerte en contextos que parten con una baja proporción de mujeres (Huntington-Klein & Rose, 2018). Además, el efecto par de género podría tener un impacto tanto en las habilidades cognitivas como no cognitivas (Gong, Lu, & Song, 2019), lo que reforzaría aún más el efecto de largo plazo al observar a los individuos en el mercado laboral (Heckman & Rubinstein, 2001).

En el contexto de nuestro país, Cabezas (2010) utiliza la misma metodología para analizar el impacto del porcentaje de mujeres en la cohorte sobre el puntaje en pruebas SIMCE de 4to básico. La autora encuentra que un aumento de 10 puntos porcentuales en la proporción de mujeres de la cohorte aumenta entre 0,2 y 0,8 el puntaje obtenido. Este efecto sería mayor en establecimientos de bajo índice socioeconómico, lo que la autora asocia a que en estos contextos el ambiente es de por sí más disruptivo, por lo que la “productividad” de las mujeres en mejorar el ambiente académico sería mayor. Paredes (2018) también investiga el efecto par de género en nuestro país, pero centrándose en cursos de un solo sexo dentro de colegios coeducacionales. Se encuentra que los cursos de un solo sexo reducen la brecha de género en matemáticas en casi un 50 %, lo que se explica por un aumento en el rendimiento de las mujeres, y un nulo efecto sobre los hombres.

La literatura que estudia colegios de un solo sexo tiene una relación cercana con lo expuesto anteriormente. Estudios iniciales solían encontrar un efecto positivo de este tipo de establecimientos sobre el desempeño académico, pero en general no se hacían cargo del sesgo de selección presente en este tipo de investigaciones. Trabajos más recientes que buscan controlar por este sesgo, incluso aprovechando la asignación aleatoria de los alumnos a los distintos colegios, encuentran evidencia mixta sobre el signo y magnitud del coeficiente estimado (Dustmann, Ku, & Kwak, 2018; Hayes, Pahlke, & Bigler, 2011; Park, Behrman, & Choi, 2018). Sin embargo, este tipo de estudios siguen teniendo problemas para separar el efecto par de género de otras variables específicas de los establecimientos no mixtos.

3.3. Mecanismos detrás del Efecto Par de Género

Desde una perspectiva teórica, existe un amplio número de razones por las cuales la composición de género podría afectar tanto la elección de carrera como el rendimiento académico de los estudiantes. Por un lado, los hombres podrían ser más disruptivos que las mujeres. Por lo tanto, una mayor proporción de pares femeninas mejoraría el comportamiento de los cursos, creando una mejor atmósfera para el aprendizaje (Lavy & Schlosser, 2011). Así, el efecto sobre el desempeño académico sería directo, e incluso se reforzaría por un aumento en el esfuerzo realizado por los alumnos (Gong, Lu, & Song, 2019). Por otro lado, el impacto sobre la elección de carreras STEM podría explicarse por el hecho de que esta decisión se basa en la afinidad del alumno con ciertas materias. Dado que las materias asociadas a las carreras STEM suelen considerarse más difíciles (matemáticas y ciencias), es necesario un mejor ambiente en las salas para motivar el interés de los alumnos por estas asignaturas.

La presencia de hombres también puede intimidar a las mujeres, sobre todo en ramos masculinizados, reduciendo su participación. Existe basta evidencia que relaciona la presencia de hombres con una reducción tanto en el desempeño como en la probabilidad de elegir escenarios competitivos por parte de las mujeres (Booth & Nolen, 2012a, 2012b; Gneezy, Niederle, & Rustichini, 2003; Niederle & Vesterlund, 2007). Así, mientras los hombres aumentan su desempeño en ambientes competitivos, mujeres lo harían únicamente al competir con otras mujeres. Lo anterior tendría un impacto directo sobre la elección de carreras STEM, que por lo general son consideradas como competitivas (Buser, Niederle, & Oosterbeek, 2014).

Ligado con lo anterior, la presencia de varones podría afectar la percepción de las mujeres sobre sus habilidades (Sax, Shapiro, & Eagan, 2011), y generar visiones más tradicionales sobre los roles de género (Lee & Marks, 1990; Maccoby, 1990, 1999). Esto impactaría en la predisposición de las alumnas hacia ramos y carreras masculinizadas, y por tanto, afectará el rendimiento académico y la elección de áreas STEM.

Otro posible mecanismo tiene que ver con la presencia de modelos a seguir. Cuando hay una mayor proporción de mujeres en la cohorte, es más probable que los alumnos de mejor rendimiento sean mujeres, lo que genera modelos a seguir positivos para otras alumnas (Thompson, 2003). Brenøe & Zölitz (2020) entregan evidencia a favor de este mecanismo, al mostrar que el efecto par de género no es significativo cuando las alumnas tienen una madre que estudio una carrera STEM. Los autores asocian este resultado a que la madre ya es un modelo lo suficientemente fuerte, lo que disminuye el impacto de las pares en el curso. Cabe destacar que la dirección de este efecto, al menos teóricamente, no es clara. Una mayor proporción de mujeres podría llevar a que haya más mujeres que siguen un comportamiento estereotipado, o que sean malos modelos, lo que afectará negativamente la probabilidad de que otras mujeres entren a carreras STEM.

La proporción de mujeres también afectará los estereotipos de género que tienen los alumnos. La elección de una carrera se basa, entre otras cosas, en nociones de congruencia de esta con la identidad de género propia, existiendo un beneficio no pecuniario cuando la elección se

condice con las normas sociales (Akerlof & Kranton, 2002). Debido a esto, mujeres no elegirán carreras tradicionalmente masculinas debido a la desutilidad de contravenir estas normas. La mayor presencia de mujeres afectará la percepción de los roles de género, reduciendo la influencia de los estereotipos, y por tanto, aumentando la probabilidad de elegir carreras STEM (Favara, 2012).

Si existen distintas formas de aprendizaje asociadas al género del alumno, la composición de género también afectará la forma en que los profesores desarrollan las clases y las expectativas que estos tienen sobre sus alumnos (Jones & Wheatley, 1990). La presencia de estereotipos de género por parte docente también sería un potencial mecanismo (Lee, Marks, & Byrd, 1994). Sin embargo, no es claro si el signo del efecto par de género en este caso será positivo, debido a que la presencia de mujeres ayuda a romper con esta visión sesgada, o negativa, debido a que el profesor prepara clases menos desafiantes ante la mayor proporción de alumnas.

Finalmente, un mecanismo que no ha recibido demasiada atención en la literatura es el posible efecto *spillover* debido precisamente a las brechas de género observadas en las variables estudiadas. Generalmente se observa que las mujeres tienen una brecha de género positiva en el desempeño en lenguaje, y una negativa en matemáticas. Dado esto, una cohorte con mayor porcentaje de mujeres será también uno con pares de mayor “calidad” en lenguaje, y peor “calidad” en matemáticas. La no consideración de este efecto *spillover* sesgará a los coeficientes observados, habiendo una sobreestimación del efecto par de género cuando la brecha de género es en favor de las mujeres (lenguaje), y una subestimación en caso contrario.

4. Estrategia Empírica

4.1. Modelo Base

En general el estudio del efecto par sufre de un problema de selección, ya que la distribución de alumnos entre colegios no es aleatoria. En el caso específico del efecto par de género, el problema surgiría si esta distribución dependiera de la composición de género de estos últimos. Llevado al contexto chileno, donde el sistema de voucher otorga una libertad de elección relativamente alta a los padres, el problema podría ser incluso más grave. Sin embargo, incluso en sistemas como el de nuestro país, las familias no pueden predecir a la perfección la composición de género de la cohorte donde estudiará su hijo (a menos que elijan matricularlos en un colegio de un solo sexo). Así, la estrategia planteada utiliza la variación idiosincrática del porcentaje de mujeres en cohortes adyacentes de un mismo colegio, siendo posible encontrar un efecto libre de este sesgo de selección con la utilización de efectos fijos por colegio y periodo. La ecuación base a estimar es la siguiente:

$$Y_{iet} = M_{et}\beta + X_{iet}\alpha_1 + E_{et}\alpha_2 + \lambda_e + \lambda_t + \varepsilon_{iet} \quad (1)$$

Donde i representa al individuo, e a la escuela y t al periodo. Las variables dependientes

principales que se estudiarán son: una variable *dummy* que toma valor 1 cuando el estudiante se matricula en una carrera STEM y 0 cuando se matricula en otro tipo de carreras; y una variable *dummy* que toma valor 1 cuando el estudiante se titula de una carrera STEM y 0 cuando se titula en otra área. Esta última variable se incluye para comprobar que el efecto observado no desaparezca debido a una mayor atrición. Además, se estudiará el impacto de la proporción de mujeres sobre el desempeño académico, medido con el puntaje en matemáticas y lenguaje en las pruebas SIMCE y PSU. Estas últimas variables se analizan como un posible mecanismo para explicar el efecto par de género en la elección de carrera.

M_{et} es la variable de interés, definida como el porcentaje de mujeres a las que está expuesto el individuo i en el colegio e y periodo t . Cuando se estudia el desempeño académico se utiliza la proporción en el grado que rinde la prueba, mientras que al estudiar la matrícula y titulación de carreras STEM se utiliza el porcentaje del último año de educación secundaria (IV medio). Una medida alternativa para este último caso es el porcentaje de mujeres en toda enseñanza media, sin embargo, en este caso existirá una fuente de variación incluso dentro de la misma cohorte dentro del colegio debido a alumnos que se cambian de colegio. La anterior no es una fuente exógena, lo que podría sesgar los coeficientes encontrados. La utilización del porcentaje en IV medio, junto al efecto fijo por establecimiento, permite utilizar únicamente la variación idiosincrática que surge al comparar cohortes adyacentes, la cual se asume exógena. La alta correlación entre ambas formas de medir el porcentaje de mujeres le da aún mayor sustento a la variable elegida. De todos modos, se comprobará la robustez de los resultados ante cambios en el curso donde se mide la distribución de género.

X_{iet} es un vector de controles socioeconómicos del individuo i y su familia, y E_{et} es un vector de controles a nivel de establecimiento que varían en el tiempo, el cual incluye variables del colegio, profesores, y pares del alumno i . λ_e y λ_t son efectos fijos a nivel de colegio y año, los cuales buscan controlar por no observables comunes dentro del colegio y fijos en el tiempo. ε_{iget} es el término de error de la regresión.

La ecuación anterior se estima por MCO tanto para las variables dicotómicas como para el desempeño académico. La utilización de un modelo de probabilidad lineal al utilizar la matrícula y titulación de carreras STEM hace sentido dado que el foco de la investigación es el efecto marginal de la proporción de mujeres, no realizar predicciones sobre la probabilidad de elegir este tipo de carreras.

A pesar de utilizar una fuente exógena de variación, las estimaciones anteriores podrían verse afectadas por las siguientes fuentes de endogeneidad: sesgo de selección; omisión de variables relevantes; y error de medición. El sesgo de selección puede originarse debido a las familias o los colegios. En el primer caso, las familias pueden elegir el establecimiento según las características del colegio o sus alumnos, incluyendo la distribución de género de este, mientras que en el segundo caso son los colegios los que seleccionan a sus alumnos según sus características. La omisión de variables relevantes ocurre cuando no se puede controlar por variables que afectan el rendimiento académico y correlacionan con el porcentaje de mujeres. Finalmente, el error de medición podría surgir en caso de que la medida de la proporción de

género de las cohortes sea de mala calidad.

Para evitar sesgo de selección de parte de las familias se excluyen de la muestra colegios que en algún momento del periodo estudiado fueron de un solo sexo, ya que en ese caso los padres sabrán a ciencia cierta la distribución de género en la cohorte. Para este fin, se considerarán como colegios no mixtos aquellos que tienen más de un 95 % de su matrícula del mismo sexo. También se eliminan de la muestra a los alumnos de colegios que no aparecen en todos los años del periodo analizado, dado que la estrategia se basa en la variación del porcentaje de mujeres entre cohortes adyacentes. La inclusión de efectos fijos a nivel de colegio permitirá controlar por la selección de estos, además de no observables fijos en el establecimiento. El error de medición podría surgir si se calcula el porcentaje de mujeres en la cohorte a partir de los alumnos que rinden las pruebas, ya que es posible que haya ausentes. Sin embargo, se cuenta con datos sobre la matrícula total del establecimiento, desagregada por curso, por lo que es posible tener una medida mucho más limpia de la variable.

4.2. Modelo Ampliado

Además de lo detallado anteriormente, el modelo se amplía en las siguientes dimensiones para encontrar un coeficiente lo más limpio posible:

- El modelo estará libre de variables omitidas relevantes solo si estas son invariables en el tiempo. Para controlar por posibles no observables que varíen en el tiempo, se incluirá una tendencia lineal para cada establecimiento ($\gamma_e \cdot t$).
- Dada la naturaleza de la variable de interés, cualquier control omitido que explique Y_{iet} y tenga una brecha de género sesgará el coeficiente estimado. El vector E_{et} incluye las variables socioeconómicas de los pares como controles, lo que permite controlar en parte este problema. Sin embargo, dado que existen brechas de género en el rendimiento de pruebas estandarizadas, la proporción de mujeres en la cohorte correlacionará con la “calidad” de los pares, sesgando los coeficientes según la dirección de la brecha. Para solucionar este problema, omitido en la literatura hasta ahora, se incluye la variable \bar{Y}_{et-1} que incluye el promedio rezagado de la variable dependiente para los pares del alumno i . Es posible incluir esta variable rezagada dado que en el periodo estudiado los individuos rindieron una prueba estandarizada cada dos años. Lo anterior, combinado con el modelo de efectos fijos, permite evitar problemas de doble causalidad y omisión de variables relevantes (Hanushek, Kain, Markman, & Rivkin, 2001).

Por lo tanto, la ecuación a estimar para el modelo ampliado será:

$$Y_{iet} = M_{et}\beta + X_{iet}\alpha_1 + E_{et}\alpha_2 + \lambda_e + \lambda_t + \gamma_e \cdot t + \bar{Y}_{et-1} + \varepsilon_{iet} \quad (2)$$

Cabe destacar que al analizar la elección y titulación de carreras STEM, el modelo solo se ampliará en la primera dimensión señalada anteriormente. Aunque se podría pensar que el

porcentaje de pares que elige una carrera STEM afecta la decisión individual², no es posible tener esta variable rezagada. La inclusión de la variable contemporánea provocaría lo que Manski (1993) identificó como *reflection problem*, sesgando los coeficientes obtenidos en las regresiones.

4.3. ¿Cómo identificar a los pares?

Un último detalle que se debe considerar en la estrategia empírica es el nivel al cual se identifican los pares del individuo i para calcular el porcentaje de mujeres. La medida ideal es la distribución de género dentro del curso, dado que son estas personas con las que más compartirá. Sin embargo, el coeficiente estimado con esta variable podría estar sesgado en caso de que el ordenamiento de alumnos entre distintos cursos de un mismo colegio no sea aleatorio. La literatura evita este problema utilizando el porcentaje de mujeres de toda la cohorte dentro del colegio, el cuál correlaciona fuertemente con el porcentaje a nivel de curso, pero no estaría sesgado debido a la utilización de efectos fijos por colegio. Sin embargo, la utilización de esta variable presenta dos problemas: uno empírico y otro teórico.

Por el lado empírico, para estimar el efecto par de género se necesita una alta variación del porcentaje de mujeres dentro del colegio (variación *within*), ya que la variación entre colegios (*between*) es capturada por el efecto fijo a nivel de establecimiento. Por lo tanto, para poder capturar el efecto es necesario que la variación *within* sea lo suficientemente alta.

Cuadro 1: Análisis de descomposición de la varianza - % de mujeres en la cohorte

| | 4to | | | 6to | | |
|---------|----------------|---------|--------|----------------|---------|--------|
| | Sum of Squares | % Total | DF | Sum of Squares | % Total | DF |
| Between | 2268.2569 | 55 % | 4,849 | 2654.6162 | 59 % | 4,821 |
| Within | 1861.3365 | 45 % | 532152 | 1874.5219 | 41 % | 533605 |
| Total | 4129.5934 | | 537001 | 4529.1381 | | 538431 |
| | II | | | PSU | | |
| | Sum of Squares | % Total | DF | Sum of Squares | % Total | DF |
| Between | 6452.3372 | 85 % | 2,615 | 8896.0451 | 78 % | 2,288 |
| Within | 1131.5717 | 15 % | 488150 | 2451.5176 | 22 % | 714493 |
| Total | 7583.9089 | | 490765 | 11347.563 | | 716781 |

El Cuadro 1 muestra la descomposición de la varianza para el porcentaje de mujeres en la cohorte, para todos los niveles analizados. Como se puede observar, en los cursos iniciales una gran parte de la variación se origina dentro de los colegios. Sin embargo, el porcentaje disminuye estrepitosamente a partir de II medio. Lo anterior se podría explicar si los colegios

²Funcionaría como *proxy* de la afinidad de los pares hacia esta área

generan un proceso de selección u ordenamiento mucho más fuerte en enseñanza media. Si los establecimientos buscan llegar a una distribución de género objetivo, entonces se explicaría el hecho de que la variación entre cohortes adyacentes sea menor. Se debe dejar claro que lo anterior es solo una suposición, dado que no existen investigaciones empíricas que estudien este fenómeno, el cual escapa a los objetivos del presente trabajo. El Anexo 1 muestra el mismo ANOVA para el porcentaje de mujeres en el curso, observándose nuevamente una caída en la proporción de la varianza *within* a partir de II medio, pero manteniendo una proporción de la variación *within* mucho mayor a lo observado en el Cuadro 1, lo que permitiría estimar un coeficiente de forma más precisa.

Por otro lado, desde el punto de vista teórico, gran parte de los mecanismos que permiten explicar la relación entre porcentaje de mujeres y los *outcomes* estudiados surgen de la interacción constante entre individuos, lo que ocurre sobre todo dentro de las salas. Así, y aunque ambas medidas tienen una alta correlación, es imposible asegurar que posibles diferencias al utilizar ambas medidas se deban al sesgo por el ordenamiento de los alumnos, y no a que la medida para toda la cohorte no permite capturar los mecanismos importantes. Debido a lo anterior, se procederá a mostrar todos los resultados dos veces: uno para cada versión del porcentaje de mujeres. La inclusión un amplio vector de características de los pares, tanto a nivel de curso como cohorte³, permite controlar por gran parte de las posibles variables que generarían un ordenamiento no aleatorio, haciendo relativamente confiable el coeficiente estimado cuando se considera a los pares dentro del curso.

5. Datos

5.1. Fuentes de Información

5.1.1. PSU y *outcomes* STEM

La prueba de selección universitaria (PSU) es una evaluación que deben rendir los alumnos que deseen postular a una institución de educación superior mediante el sistema único de admisión, del cual participan todas las instituciones del consejo de rectores y gran parte de las universidades privadas de Chile. Además, y aunque no seleccionan por el puntaje obtenido, también hay instituciones de educación superior técnico profesional (carreras de 4 años o menos) que exigen la rendición de esta. La PSU se realiza una vez al año, y evalúa la materia vista en los últimos cuatro años de educación media (secundaria) en los subsectores de lenguaje, matemáticas, ciencias e historia, siendo estas últimas dos optativas. Desde las bases de datos PSU se obtiene el puntaje en las pruebas, además de controles socioeconómicos de la familia y características del colegio de procedencia.

³La forma en que se calcularán las variables de los pares dependerá directamente de cómo se define el porcentaje de mujeres. Si se calcula en toda la cohorte dentro del colegio, entonces las variables de los pares también se medirán así.

Para identificar si la carrera estudiada corresponde al área STEM se utilizan bases del portal de datos abiertos del Ministerio de Educación (Mineduc). Aquí es posible identificar la carrera en la cual se matricula el estudiante, las cuales son agrupadas según la clasificación ISCED 1997 en ocho campos: educación; humanidades y artes; ciencias sociales, negocios y leyes; ciencia; ingeniería, manufactura y construcción; agricultura; salud y bienestar; y servicios. Siguiendo la metodología utilizada por la UNESCO, se identifican como carreras STEM aquellas provenientes del área de ciencia e ingeniería, industria y construcción (UNESCO, 2014, 2017). Además, desde este portal se extraen datos sobre la titulación en la educación superior (con el mismo identificador por área del conocimiento), y controles a nivel de la institución. El detalle con todas las variables de control se encuentra en el anexo 2 del presente trabajo.

Al analizar los *outcomes* STEM se utilizan datos sobre quienes se matriculan (titulan) entre los años 2012 y 2019. Al estudiar el puntaje PSU, sin embargo, se utilizan datos desde el 2014 debido a la inclusión del puntaje rezagado como control. Este puntaje rezagado corresponde a la prueba SIMCE de II medio. Finalmente, y tal como se detalló anteriormente, se eliminan de la muestra todos los alumnos provenientes de colegios que no aparecen o que son de un solo sexo en alguno de los años estudiados. Para evitar problemas con errores de medida, los colegios son clasificados como no mixtos cuando un solo sexo concentra al menos el 95 % de la matrícula en todo el establecimiento. Los resultados son robustos al cambiar este porcentaje.

5.1.2. SIMCE

El SIMCE es una serie de pruebas estandarizadas que mide el conocimiento con respecto al currículum vigente en distintas asignaturas, las cuales son rendidas por todos los alumnos del grado en que se aplica⁴. La prueba se realiza todos los años, aunque los grados y asignaturas evaluadas cambian⁵: hasta el 2011 solo los alumnos de 4to básico la rendían todos los años, mientras que estudiantes de 8vo básico y II medio se intercalaban. A partir del 2012 aumentan los grados a los que se les aplica la prueba, lo que permite ampliar la cantidad de cursos estudiados. En específico, se analizarán los siguientes cursos y periodos: alumnos de 4to básico que rindieron prueba SIMCE entre los años 2015 y 2017; alumnos de 6to básico que rindieron pruebas SIMCE entre los años 2014 y 2016; y alumnos de II medio que pruebas SIMCE entre los años 2015 y 2017.

La elección de los periodos se debe a la necesidad de observar al mismo individuo en dos periodos para calcular el puntaje promedio rezagado de los pares.⁶ Entre los años 2013 y 2015

⁴En la práctica los alumnos pueden faltar a la prueba, por lo que la muestra del SIMCE no es censal.

⁵Generalmente se evalúa el conocimiento en lenguaje, matemáticas y una tercera prueba que varía entre ciencias naturales e historia. Además, algunos cursos rinden un SIMCE de inglés y otro de educación física.

⁶En el sentido estricto de la palabra, es incorrecto decir que es el puntaje rezagado. Las pruebas SIMCE solo son comparables dentro del grado en que se rinden, por lo que instrumentos de cursos diferentes se pueden considerar pruebas independientes. Lo anterior no es un problema para la estimación, incluso simplificando la metodología al no necesitar utilizar un modelo de panel dinámico.

la prueba SIMCE se rindió en todos los grados, lo que permite encontrar este rezago para cohortes adyacentes. La prueba también es rendida en 2do y 8vo básico, sin embargo, en el primero de estos cursos no hay rezagos del puntaje, mientras que en el segundo solo hay dos cohortes adyacentes con rezago. Se estudiará el rendimiento en lenguaje y matemáticas, dado que son las materias donde se cuenta con cohortes adyacentes. En 4to básico solo es posible utilizar el puntaje de lenguaje, dado que no se cuenta con un rezago para matemáticas.

Los datos del SIMCE son complementados con datos administrativos del Mineduc, obtenidos a través de su plataforma de datos abiertos. El detalle de los controles utilizados en las estimaciones del puntaje SIMCE se encuentra en el anexo 3. Tal como se hizo anteriormente, se eliminan de la muestra todos los alumnos provenientes de colegios clasificados como de un solo sexo en al menos uno de los años analizados, o que no aparecen en todos los periodos.

5.2. Estadística Descriptiva

La Figura 1 muestra la distribución de la matrícula de educación superior en áreas STEM y no STEM, según el sexo del alumno y la proporción de mujeres en su cohorte de cuarto medio. Para esta última categoría se definen tres escenarios: cursos mixtos, donde las mujeres son entre un 40 % y 60 % de los alumnos; cursos mayoritariamente de hombres (menos de 40 % de mujeres); y cursos mayoritariamente de mujeres (más de un 60 % de mujeres).

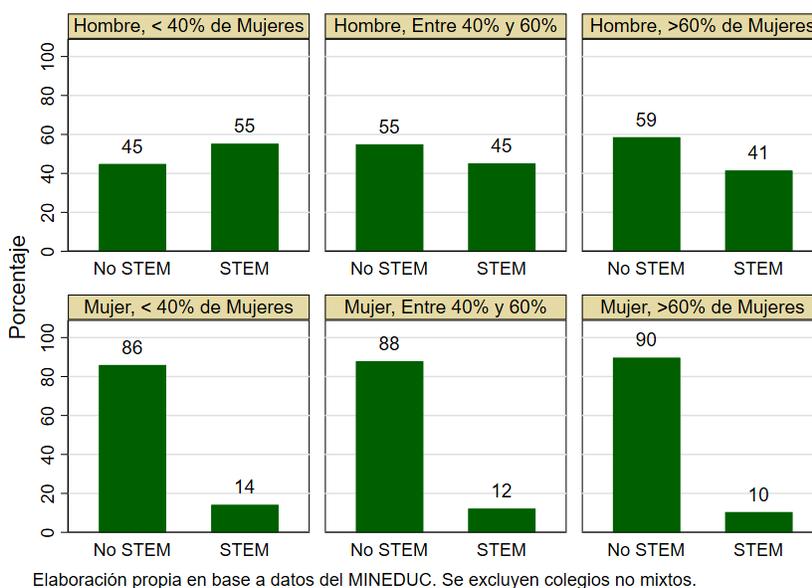


Figura 1: Distribución de la matrícula STEM, según sexo del alumno y porcentaje de mujeres en su cohorte de IV medio

Las mujeres eligen carreras STEM con una probabilidad extremadamente baja en comparación a los hombres. Aún así, para ambos sexos se observa el mismo patrón: la probabilidad de entrar a una carrera STEM es más alta en cohortes con menos de 40% de mujeres. Los varones provenientes de cohortes mayoritariamente de hombres, y que se matriculan en la educación superior, lo harán en el área STEM con un 55% de probabilidades, mientras que las mujeres de estas cohortes lo harán con un 14% de probabilidades. El Anexo 4 muestra un gráfico similar para la titulación STEM, observándose el mismo patrón

La figura 2 es similar a la figura 1, solo que muestra la evolución del desempeño en la PSU. Cabe destacar que el puntaje se estandariza para cada año con media cero y varianza 1. El panorama expuesto por esta figura es distinto a lo observado para los *outcomes* STEM: las cohortes mayoritariamente de hombres, donde hay una mayor probabilidad de estudiar carreras del área STEM, son las que tienen un peor desempeño promedio en la PSU. Por el contrario, son los ambientes mixtos (entre 40% y 60% de mujeres), los que muestran un mayor puntaje en las pruebas de acceso.

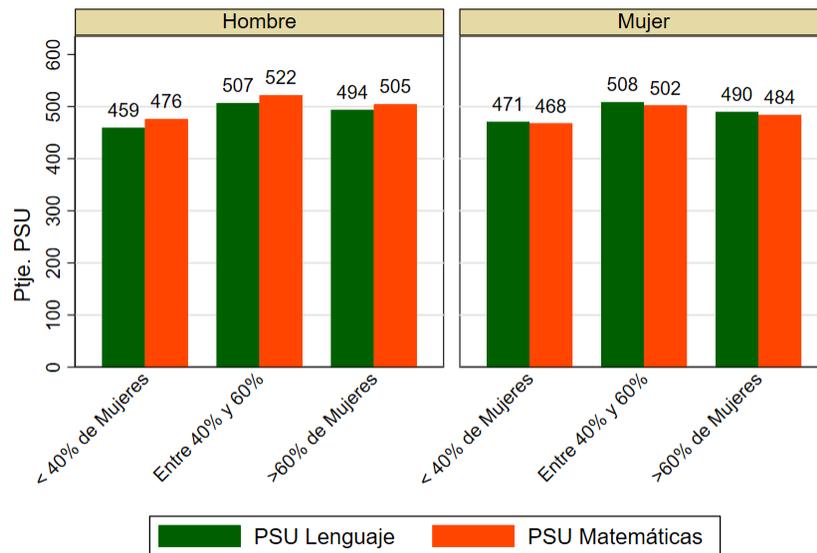


Figura 2: Puntaje estandarizado PSU, según sexo del alumno y porcentaje de mujeres en su cohorte de IV medio

Así, y aunque la información entregada por estos gráficos no es causal, sí es posible entregar algunas conclusiones preliminares. Por un lado, hay una correlación clara negativa entre la proporción de mujeres en la cohorte y la probabilidad de que los alumnos se matriculen en carreras del área STEM. Por otro lado, la relación entre la proporción de género y el

rendimiento en la PSU no permite explicar el efecto sobre la elección de carreras⁷.

El Cuadro 2 muestra estadística descriptiva para las principales variables utilizadas en las regresiones, distinguiendo según el curso analizado y el sexo del estudiante. Lo primero que es posible notar es la existencia de una clara brecha contra las mujeres tanto en la matrícula como titulación de carreras STEM. Condicional a rendir la PSU y matricularse en alguna carrera, un 46 % de los hombres lo hace en el área STEM, mientras que para las mujeres este porcentaje es de solo 11 %. Al observar la titulación los porcentajes son 43 % y 5,3 % respectivamente para cada sexo. Ambas brechas son estadísticamente significativas.

Cuadro 2: Estadística Descriptiva de Variables Principales

| | Mujeres | | Hombres | | Diferencia | |
|---|----------|-----------|----------|-----------|-------------|-----------|
| | Media | Desv. Est | Media | Desv. Est | Media | Desv. Est |
| a. Universidad | | | | | | |
| % Matrícula STEM | 0.11 | 0.31 | 0.46 | 0.50 | -0.35*** | (0.00069) |
| % Titulo STEM | 0.053 | 0.22 | 0.43 | 0.49 | -0.37*** | (0.0014) |
| b. IV Medio | | | | | | |
| Ptje. Est. PSU Lenguaje ¹ | 0.020 | 0.98 | -0.023 | 1.02 | 0.042*** | (0.0020) |
| Ptje. Est. PSU Matemáticas ¹ | -0.065 | 0.96 | 0.075 | 1.03 | -0.14*** | (0.0020) |
| % Padres con Ed. Superior | 0.18 | 0.38 | 0.20 | 0.40 | -0.023*** | (0.00073) |
| % Madres con Ed. Superior | 0.19 | 0.39 | 0.21 | 0.41 | -0.019*** | (0.00074) |
| Ingresos ² | 572273.4 | 540711.9 | 602362.1 | 556184.1 | -30088.7*** | (1026.0) |
| c. II Medio | | | | | | |
| Ptje. Est. SIMCE Lenguaje ¹ | 0.16 | 0.97 | -0.16 | 1.00 | 0.32*** | (0.0026) |
| Ptje. Est. SIMCE Matemáticas ¹ | -0.023 | 0.99 | 0.023 | 1.01 | -0.046*** | (0.0026) |
| % Padres con Ed. Superior | 0.21 | 0.41 | 0.21 | 0.41 | -0.0013 | (0.0010) |
| % Madres con Ed. Superior | 0.21 | 0.41 | 0.21 | 0.41 | -0.0011 | (0.0010) |
| Ingresos ² | 598087.6 | 563885.9 | 611813.5 | 564737.7 | -13725.9*** | (1438.8) |
| % con Expec. de Ed. Superior | 0.94 | 0.23 | 0.92 | 0.27 | 0.025*** | (0.00064) |
| d. 6to Básico | | | | | | |
| Ptje. Est. SIMCE Lenguaje ¹ | 0.12 | 0.96 | -0.12 | 1.02 | 0.24*** | (0.0025) |
| Ptje. Est. SIMCE Matemáticas ¹ | -0.014 | 0.99 | 0.013 | 1.01 | -0.027*** | (0.0025) |
| % Padres con Ed. Superior | 0.21 | 0.41 | 0.20 | 0.40 | 0.0032** | (0.00099) |
| % Madres con Ed. Superior | 0.22 | 0.42 | 0.22 | 0.41 | 0.0033** | (0.0010) |
| Ingresos ² | 568015.0 | 569435.0 | 568890.9 | 566353.1 | -876.0 | (1399.3) |
| % con Expec. de Ed. Superior | 0.92 | 0.27 | 0.89 | 0.31 | 0.026*** | (0.00071) |
| e. 4to Básico | | | | | | |
| Ptje. Est. SIMCE Lenguaje ¹ | 0.099 | 0.96 | -0.098 | 1.03 | 0.20*** | (0.0025) |
| Ptje. Est. SIMCE Matemáticas ¹ | -0.015 | 0.99 | 0.015 | 1.01 | -0.030*** | (0.0025) |
| % Padres con Ed. Superior | 0.23 | 0.42 | 0.23 | 0.42 | 0.0030** | (0.0010) |
| % Madres Ed. Superior | 0.25 | 0.43 | 0.25 | 0.43 | 0.0039*** | (0.0011) |
| Ingreso ² | 617873.3 | 587577.9 | 619751.5 | 586584.9 | -1878.2 | (1452.8) |
| % con Expec. de Ed. Superior | 0.94 | 0.24 | 0.92 | 0.26 | 0.015*** | (0.00062) |

¹ Los puntajes de todas las pruebas son estandarizados con promedio cero y desviación estándar 1 en cada año.

² Los ingresos son variables categóricas en las bases de datos. El valor expuesto en la tabla se obtiene al utilizar la marca de clase de cada intervalo como el ingreso de las familias.

⁷Las carreras del área STEM suelen ser altamente demandadas, por lo que sus puntajes de corte son más altos que en la carrera promedio.

Al observar el desempeño académico, también es posible notar brechas de género. Independiente del curso analizado, siempre existe una brecha de género a favor de las mujeres en lenguaje, mientras que en matemáticas es a favor de los hombres. La brecha tiende a aumentar al avanzar de grado, excepto en lenguaje PSU, donde disminuye al compararla con el SIMCE de II medio.⁸

En IV medio también se encuentran diferencias en las variables de control: los alumnos hombres provienen, en promedio, de familias con padres y madres de mayor nivel educativo y de mayores ingresos. Cabe destacar que en enseñanza básica las brechas en los controles suelen ser a favor de las mujeres, aunque la magnitud de estas es mucho menor.

En conclusión, existen brechas de género importantes tanto en la elección de carreras STEM como en el desempeño académico. También existirían brechas de género en los controles individuales. Lo anterior hace especialmente importante controlar por las variables de los pares a la hora de estudiar el efecto par de género. Una cohorte con mayor proporción de mujeres en IV medio será, en promedio, una con pares que tienen padres y madres con un menor nivel educativo completado y de menos ingresos. Además, será una cohorte con pares de menor calidad en matemáticas y mayor calidad en lenguaje. Si no se controla por estas variables, entonces su efecto se confundirá con el de un mayor porcentaje de mujeres, sesgando los coeficientes encontrados.

6. Resultados Empíricos

A continuación, se presentan los resultados de las regresiones, tanto para la matrícula y titulación de carreras STEM, como para el desempeño en las pruebas PSU y SIMCE.

6.1. Elección de Carrera: Matrícula y Titulación STEM

El Cuadro 3 muestra los resultados al estudiar el impacto de la distribución de género sobre la probabilidad de matricularse o titularse de una carrera STEM. Las primeras dos columnas consideran como pares a todos los alumnos de la misma cohorte dentro del colegio, mientras que las últimas dos columnas restringen a los pares dentro del mismo curso. Cabe destacar que, debido a los datos utilizados, el efecto sobre la matrícula STEM es condicional en dar la PSU y matricularse en una carrera, mientras que el efecto en la titulación es condicional a titularse de alguna carrera. El error estándar de la regresión está corregido por conglomerados a nivel de colegio y la variable del porcentaje de mujeres está multiplicada por 100, por lo que sus coeficiente se interpretan como el cambio en la probabilidad que se produce al aumentar en un punto porcentual (pp) la proporción de mujeres.

Lo primero que es posible notar es que el coeficiente asociado al porcentaje de mujeres generalmente es negativo, aunque solo es significativo al considerar como pares a los alumnos

⁸los puntajes de todas las pruebas se estandarizan con media cero y varianza 1 en cada año.

del mismo curso. Lo anterior se podría deber, como se mostró anteriormente, a que la variación de esta variable al considerar como pares a todos los alumnos de la cohorte proviene en gran parte de la comparación entre colegios, mientras que la estrategia empírica se basa en la variación dentro del mismo establecimiento para estimar los parámetros. Aunque los coeficientes de las columnas 3 y 4 son susceptibles al sesgo de selección por un ordenamiento no aleatorio de estudiantes entre cursos, la inclusión de efectos fijos por establecimiento y un amplio vector de controles tanto de los individuos como de sus pares en el curso permite reducirlo considerablemente. Dado que no hay cambios importantes en los otros coeficientes estimados, el análisis posterior se centrará en estos últimos modelos.

Cuadro 3: Efecto par de género sobre la matrícula y titulación STEM

| | Pares: Cohorte | | Pares: Curso | |
|---------------------------|------------------------|--------------------------|-------------------------|-------------------------|
| | (1) | (2) | (3) | (4) |
| a. Matrícula STEM | | | | |
| % Mujeres | 0.000057 (0.000086) | -0.0000096 (0.000090) | -0.0025*** (0.00017) | -0.0025*** (0.00018) |
| Mujer | -0.30*** (0.0029) | -0.30*** (0.0029) | -0.27*** (0.0022) | -0.27*** (0.0022) |
| Ptje. PSU Lenguaje | -0.050*** (0.0012) | -0.050*** (0.0012) | -0.048*** (0.0012) | -0.048*** (0.0012) |
| Ptje. PSU Matemáticas | 0.12*** (0.0023) | 0.12*** (0.0023) | 0.11*** (0.0022) | 0.11*** (0.0022) |
| Observaciones | 888282 | 888282 | 888282 | 888282 |
| b. Titulación STEM | | | | |
| % Mujeres | -0.000042 (0.00015) | -0.000081 (0.00016) | -0.0028*** (0.00019) | -0.0028*** (0.00020) |
| Mujer | -0.31*** (0.0042) | -0.31*** (0.0043) | -0.28*** (0.0032) | -0.28*** (0.0032) |
| Ptje. PSU Lenguaje | -0.042*** (0.0016) | -0.042*** (0.0017) | -0.039*** (0.0016) | -0.039*** (0.0016) |
| Ptje. PSU Matemáticas | 0.071*** (0.0020) | 0.071*** (0.0020) | 0.070*** (0.0019) | 0.070*** (0.0019) |
| Observaciones | 244997 | 244997 | 244997 | 244997 |
| Tendencia-Colegio | No | Sí | No | Sí |

Coefficiente del porcentaje de mujeres se interpreta como el cambio en la probabilidad ante un aumento de 1pp en la proporción de mujeres en cuarto medio. Las primeras dos columnas consideran como pares a todos los estudiantes de la cohorte dentro del colegio, mientras que las últimas dos columnas solo consideran a los alumnos del mismo curso. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

La inclusión de tendencias específicas por colegio no genera cambios importantes en los coeficientes estimados, por lo que de existir no observables relevantes que varían en el tiempo, estos no lo hacen siguiendo un patrón definido. Antes de hablar de la magnitud de los coeficientes estimados, es necesario hacer una pequeña transformación, dado que un aumento de 1pp en la proporción de mujeres implica agregar, en promedio, menos de una alumna al curso. Una mejor alternativa es multiplicar los coeficientes por 10, lo que implica estudiar el impacto de un aumento de 10pp. Así, este aumento en la proporción de alumnas provocará una disminución de 2,5pp en la probabilidad de que los alumnos se matriculen en una carrera STEM. El efecto se mantiene en el tiempo, provocando una disminución de 2,8pp en la probabilidad de titularse de una carrera STEM.

Las variables de control presentan los signos esperados. Mujeres tienen una brecha claramente negativa tanto en la matrícula como en la titulación de carreras STEM. El signo negativo encontrado en el efecto del puntaje PSU de lenguaje, junto con el positivo del puntaje en matemáticas, da a entender que los alumnos observarían su habilidad relativa a la hora de elegir una carrera. Así, y a pesar de que un aumento en el puntaje de lenguaje tiene un efecto positivo automático en la probabilidad de entrar a una carrera (al aumentar el puntaje ponderado), tendrá un efecto negativo a la hora de elegir carreras de áreas relativamente alejadas de las habilidades y conocimientos medidos en esta prueba, como es el caso de las STEM.

El Cuadro 4 muestra como el efecto par de género difiere entre hombres y mujeres. Para esto, se separa la muestra original en dos, dependiendo del sexo del estudiante. Para realizar la estimación se utiliza el modelo más completo posible (columnas 2 y 4 del Cuadro 3). Como se puede observar, el efecto negativo es de mayor magnitud para los hombres: un aumento de 10pp en la proporción de mujeres dentro del curso disminuirá en 4,1pp (5,4pp) la probabilidad de que se matriculen (titulen) en una carrera, comparado con una disminución de 1,4pp (1pp) en el caso de las mujeres.

Aunque la magnitud absoluta es claramente superior para los hombres, es importante tomar en consideración la base desde la que parten ambos sexos. Una disminución de 4,1pp en la probabilidad de matricularse en una carrera STEM equivale a un 8,9% de la situación inicial de los hombres (46% se matriculan en una carrera STEM). Por otro lado, la disminución de 1,4pp de las mujeres equivale a un 12,7% de la situación inicial de las mujeres (un 11% se matricula en carreras STEM). Del mismo modo, el efecto sobre la probabilidad de titulación para los hombres es un 11,7% de la situación base, mientras que para las mujeres es de un 9,1%. Así, aunque la magnitud absoluta del efecto es mayor para los hombres, al analizarla en relación a la proporción de cada sexo que está eligiendo estas carreras las diferencias disminuyen o derechamente se invierten.

La influencia del puntaje PSU de lenguaje y matemáticas es más fuerte para los hombres. De todos modos, se mantienen los signos encontrados anteriormente: el rendimiento en matemáticas tiene un efecto positivo en la probabilidad de matricularse y titularse de una carrera STEM, mientras que el puntaje en lenguaje tiene la dirección opuesta.

Cuadro 4: Efecto par de género sobre la matrícula y titulación STEM - Análisis por sexo del alumno

| | Pares: Cohorte | | Pares: Curso | |
|---------------------------|------------------------|-------------------------|-------------------------|--------------------------|
| | Hombres | Mujeres | Hombres | Mujeres |
| a. Matrícula STEM | | | | |
| % Mujeres | -0.000085 (0.00015) | -0.000061 (0.000092) | -0.0041*** (0.00015) | -0.0014*** (0.000090) |
| Ptje. PSU Lenguaje | -0.072*** (0.0017) | -0.034*** (0.0013) | -0.068*** (0.0017) | -0.033*** (0.0012) |
| Ptje. PSU Matemáticas | 0.16*** (0.0031) | 0.074*** (0.0019) | 0.16*** (0.0030) | 0.070*** (0.0018) |
| Observaciones | 414961 | 473320 | 414961 | 473320 |
| b. Titulación STEM | | | | |
| % Mujeres | -0.000013 (0.00036) | -0.00032** (0.00013) | -0.0054*** (0.00022) | -0.0010*** (0.000095) |
| Ptje. PSU Lenguaje | -0.084*** (0.0031) | -0.016*** (0.0014) | -0.078*** (0.0030) | -0.015*** (0.0014) |
| Ptje. PSU Matemáticas | 0.12*** (0.0038) | 0.037*** (0.0015) | 0.12*** (0.0038) | 0.035*** (0.0015) |
| Observaciones | 94735 | 150231 | 94735 | 150231 |
| Prop. matriculado STEM | 0.46 | 0.11 | 0.46 | 0.11 |
| Prop. titulado STEM | 0.43 | 0.053 | 0.43 | 0.053 |

Coeficiente del porcentaje de mujeres se interpreta como el cambio en la probabilidad ante un aumento de 1pp en la proporción de mujeres en cuarto medio. Las primeras dos columnas consideran como pares a todos los estudiantes de la cohorte dentro del colegio, mientras que las últimas dos columnas solo consideran a los alumnos del mismo curso. Se utiliza modelo más completo en cada estimación. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

6.2. Desempeño Académico

Una variable que tendrá una influencia directa sobre la decisión de los estudiantes de estudiar en la educación superior es su rendimiento en la PSU. Es por esto que en esta sección se estudiará si la relación entre proporción de mujeres y los *outcomes* STEM puede ser explicada, al menos en parte, debido a la existencia de un efecto par de género sobre el rendimiento en estas pruebas.

En primera instancia, el Cuadro 5 muestra los resultados al estimar utilizando toda la muestra. A diferencia del Cuadro 3, en esta oportunidad es posible ampliar el modelo incluyendo

el puntaje rezagado de los pares (columnas 3 y 6), para así diferenciar el efecto par de género del efecto par académico. La inclusión de tendencias para cada establecimiento nuevamente no genera grandes cambios en los coeficientes estimados.

La inclusión del puntaje rezagado confirma la hipótesis planteada anteriormente: en pruebas con brechas de género a favor de las mujeres, como lenguaje, el efecto par de género se confunde con el efecto de tener pares de mayor “calidad”, lo genera una sobreestimación del efecto si no se controla. Lo contrario ocurre con pruebas donde la brecha de género es contra las mujeres, como matemáticas. Así, la inclusión del puntaje de los pares hace más negativo el efecto par de género en lenguaje, mientras que en matemáticas se mueve hacia los positivos.

Cuadro 5: Impacto de la composición de género sobre el desempeño académico (PSU)

| | Pares: Cohorte | | | Pares: Curso | | |
|---------------------------|------------------------|-----------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| a. PSU Lenguaje | | | | | | |
| % Mujeres | -0.000028 (0.00023) | -0.00015 (0.00024) | -0.00056** (0.00022) | 0.00094*** (0.00014) | 0.00098*** (0.00015) | -0.00027** (0.00011) |
| Mujer | 0.031*** (0.0027) | 0.031*** (0.0027) | 0.030*** (0.0027) | 0.022*** (0.0026) | 0.022*** (0.0026) | 0.021*** (0.0026) |
| Ptje. Rezagado (Pares) | | | 0.22*** (0.0081) | | | 0.45*** (0.0090) |
| Observaciones | 721571 | 721571 | 721209 | 721326 | 721326 | 718266 |
| R^2 Ajustado | 0.394 | 0.397 | 0.398 | 0.398 | 0.401 | 0.418 |
| b. PSU Matemáticas | | | | | | |
| % Mujeres | 0.00014 (0.00025) | 0.00015 (0.00026) | 0.00057** (0.00023) | 0.000058 (0.00015) | 0.000083 (0.00015) | 0.0011*** (0.00011) |
| Mujer | -0.14*** (0.0028) | -0.15*** (0.0028) | -0.15*** (0.0028) | -0.14*** (0.0028) | -0.14*** (0.0028) | -0.14*** (0.0028) |
| Ptje. Rezagado (Pares) | | | 0.30*** (0.014) | | | 0.60*** (0.013) |
| Observaciones | 717155 | 717155 | 716806 | 716915 | 716915 | 714145 |
| R^2 Ajustado | 0.426 | 0.430 | 0.431 | 0.430 | 0.433 | 0.457 |
| Tendencia por Colegio | No | Sí | Sí | No | Sí | Sí |

Coefficiente del porcentaje de mujeres se interpreta como el cambio, en desviaciones estándar, del puntaje ante un aumento de 1pp en la proporción de mujeres. Las primeras tres columnas consideran a todos los estudiantes del cohorte dentro del colegio como pares del individuo i , mientras que las últimas tres columnas solo consideran a los alumnos del mismo curso. Puntajes PSU se encuentran estandarizados. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

La significancia y el signo del coeficiente asociado al porcentaje de mujeres no cambia con la definición de pares utilizada, aunque la dirección del efecto si depende de la prueba analizada.

Así, un aumento de 10pp en la proporción de mujeres disminuirá entre 0,0056 y 0,0027 desviaciones estándar el puntaje PSU de lenguaje, y lo aumentará entre 0,0057 y 0,011 en matemáticas. Recordando los efectos encontrados en el Cuadro 3, es posible concluir que los efectos pares de género sobre los puntajes PSU en ambas asignaturas se complementan para aumentar de forma indirecta la probabilidad de que los estudiantes se matriculen y titulen de carreras STEM, atenuando por tanto el efecto par de género negativo encontrado. Esta atenuación, sin embargo, es bastante tenue: aproximadamente un 5% del efecto par de género directo sobre la matrícula STEM y 3% del efecto directo sobre la titulación⁹.

El análisis anterior se amplía en el Cuadro 6 en dos dimensiones: por un lado, se divide la muestra según el sexo del alumno, y por otro, se amplía la cantidad de pruebas analizadas al incluir las evaluaciones SIMCE de II medio, 6to y 4to básico. Esta última dimensión es de especial importancia para saber si es posible generalizar los resultados obtenidos para la PSU, desprendiendo de ellos un efecto par de género sobre el desempeño académico.

Observando los coeficientes sobre el desempeño en lenguaje, es posible notar que la significancia se concentra en los modelos que consideran como pares a los compañeros de curso, sin embargo, el efecto no es consistente. En la PSU se observa un signo positivo para los hombres y negativo para las mujeres, mientras que en 6to básico los signos se invierten y en II medio ambos son positivos. En 4to básico, en tanto, no se obtienen efectos significativos. La magnitud de los coeficientes tiende a aumentar al avanzar por los cursos.

Al analizar los coeficientes sobre el desempeño en matemáticas es posible notar mayor consistencia tanto en los signos (son siempre positivos) como en la significancia (en general son estadísticamente significativos) independiente de la definición de pares utilizada. El efecto sobre el puntaje PSU es claramente mayor para los hombres, sin embargo, esta diferencia es menos clara al analizar el puntaje Simce. Finalmente, la magnitud de los coeficientes tiende a aumentar al avanzar por los cursos, excepto en la PSU, donde el efecto es más pequeño al compararlo con el puntaje SIMCE de II medio.

En conclusión, el efecto par de género encontrado al utilizar los puntajes PSU difieren a los estimados usando el puntaje SIMCE. El aumento en la magnitud a medida que se avanza en los cursos ya ha sido documentado en otras investigaciones. La explicación que se suele dar a esta evolución es que los alumnos, al acercarse a la adolescencia, comienzan a generar redes más fuertes y toman más en consideración la opinión de otras personas, lo que los hace más susceptibles a la influencia de sus pares.

Por otro lado, la diferencia tanto en signo como en magnitud al analizar el desempeño en la PSU se podría explicar por la naturaleza misma de esta prueba. La PSU es rendida únicamente por los alumnos que desean entrar a la educación superior, generalmente los de mayor desempeño académico, y tiene consecuencias directas sobre las carreras e instituciones

⁹Utilizando coeficientes encontrados en el Cuadro 3. Un aumento de 10pp en la proporción de pares mujeres provocará un efecto indirecto positivo 0,13396pp sobre la probabilidad de matrícula STEM y de 0,08753 sobre la probabilidad de titulación STEM, debido a la influencia de este mismo cambio sobre el puntaje PSU.

a las que podrán postular. En cambio, la prueba Simce es rendida por todos los alumnos de un determinado nivel, y no tiene consecuencias directas sobre los alumnos. Así, el contexto en el que se prepara la PSU es mucho más competitivo al de la prueba SIMCE, lo que podría afectar al efecto par de género tal como se discutió al analizar los mecanismos detrás de este efecto.

Cuadro 6: Efecto par de género sobre el desempeño académico - Análisis por sexo y nivel

| | Pares: Cohorte | | Pares: Curso | |
|-----------------------|--------------------------------------|------------------------------------|-------------------------------------|-------------------------------------|
| | Hombres | Mujeres | Hombres | Mujeres |
| a. Lenguaje | | | | |
| PSU | -0.00099*** (0.00035) [325458] | -0.00039 (0.00030) [395748] | 0.0018*** (0.00018) [323998] | -0.0016*** (0.00014) [394265] |
| Simce II medio | 0.00048 (0.00083) [248102] | 0.00070 (0.00079) [242632] | 0.00086*** (0.00030) [247476] | 0.00090*** (0.00030) [242121] |
| Simce 6to básico | -0.00077 (0.00058) [277559] | -0.00042 (0.00053) [260870] | -0.00061* (0.00035) [277187] | 0.00055* (0.00033) [260522] |
| Simce 4to básico | -0.000089 (0.00055) [274876] | -0.00024 (0.00054) [262099] | 0.00020 (0.00034) [274720] | -0.000089 (0.00033) [261950] |
| b. Matemáticas | | | | |
| PSU | 0.00028 (0.00034) [323592] | 0.00058** (0.00028) [393211] | 0.0016*** (0.00018) [322251] | 0.00022* (0.00013) [391891] |
| Simce II medio | 0.0012** (0.00059) [252307] | 0.0019*** (0.00055) [247168] | 0.0021*** (0.00025) [251659] | 0.0023*** (0.00024) [246651] |
| Simce 6to básico | 0.0012** (0.00048) [272493] | 0.00098** (0.00049) [257860] | 0.0012*** (0.00028) [272124] | 0.0014*** (0.00030) [257515] |

Coeficientes se interpretan como el cambio, en desviaciones estándar, del puntaje ante un aumento de 1pp en la proporción de mujeres. Cada celda es una regresión distinta, dependiendo del sexo del alumno, definición de pares, y prueba utilizada. Cada estimación utiliza el modelo más completo. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. Número de observaciones se encuentra entre corchetes. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

Además, dada la importancia de la prueba muchos alumnos deciden prepararla en preuniversitarios, los cuales son instituciones privadas que se dedican a preparar a alumnos para la PSU. Si el porcentaje de mujeres afecta la probabilidad de asistir a un preuniversitario, o en general de preparar la PSU, entonces habrá un nuevo sesgo que afecta a estos resultados. En definitiva, no es correcto generalizar el efecto par de género encontrado en el puntaje PSU a un efecto sobre el desempeño académico. Aun así, esto no le resta importancia los resultados encontrados, ya que es la prueba PSU la que tendrá consecuencias directas sobre el ingreso a la educación superior, y por tanto, sobre la matrícula y titulación de carreras STEM.

La influencia de la proporción de pares mujeres sobre el puntaje PSU sigue atenuando el efecto par de género sobre la matrícula y titulación STEM al separar por sexo. Sin embargo, para los hombres este efecto indirecto es incluso menor al encontrado anteriormente: un 3,3 % (4,9 %) del efecto par de género sobre la matrícula STEM de los hombres (mujeres) se ve atenuada por la influencia indirecta sobre el puntaje PSU, mientras que el efecto sobre la titulación STEM se ve atenuado en un 1 % (3,2 %) ¹⁰. La disminución en el caso de los hombres se explica por el cambio de signo de la relación entre porcentaje de mujeres y puntaje PSU de lenguaje, lo que provoca que los efectos de ambas pruebas ya no se complementen.

Finalmente, y para comprobar si existe sesgo de selección en los coeficientes al estudiar el puntaje PSU, el Cuadro 7 muestra el efecto que el % de mujeres tiene sobre la probabilidad de rendir la PSU. Dado que las variables socioeconómicas y de los profesores solo se conocen para los alumnos que rinden la prueba, el modelo solo incluye como controles el género del alumno y los efectos fijos por año, colegio, y tendencias específicas por establecimiento. Se diferencian las regresiones al utilizar toda la muestra (columnas 1 y 2), solo hombres (columnas 3 y 4), o solo mujeres (columnas 5 y 6).

Cuadro 7: Efecto par de género sobre la probabilidad de rendir la PSU

| | Muestra Total | | Hombres | | Mujeres | |
|---------------|-------------------------|--------------------------|------------------------|------------------------|-------------------------|-------------------------|
| | % Cohorte | % Curso | % Cohorte | % Curso | % Cohorte | % Curso |
| % de Mujeres | -0.000014 (0.000058) | 0.00089*** (0.000086) | -0.00012 (0.000075) | 0.0016*** (0.00012) | 0.00013** (0.000065) | 0.00017** (0.000078) |
| Observaciones | 1416697 | 1416697 | 682413 | 682413 | 734283 | 734283 |

Coefficientes se interpretan como el cambio en la probabilidad de rendir la PSU ante un aumento de 1pp en la proporción de mujeres. Primeras dos columnas utilizan la muestra completa, mientras que en las siguientes se separa según el sexo del alumno. Para cada muestra, se diferencian los modelos que utilizan el porcentaje de mujeres en la cohorte dentro del colegio o en el curso. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

¹⁰Los hombres (mujeres) aumentan en 0,1336pp (0,0685pp) su probabilidad de matricularse en carreras STEM ante un aumento de 10pp en la proporción de mujeres, debido al efecto indirecto sobre el puntaje PSU. En el caso de la titulación, el mismo cambio aumenta en 0,0516pp (0,0317pp) la probabilidad.

Como se puede observar, solo en dos de los 6 modelos estimados el efecto no es significativo (muestra total y de hombres utilizando el % de mujeres en la cohorte). Un aumento de 10pp en la proporción de mujeres en el curso aumenta en 1,6pp la probabilidad de que un alumno hombre rinda la PSU, mientras que en una alumna mujer el efecto es de 0,17pp.

En conclusión, y aunque las magnitudes no son demasiado grandes, existiría un efecto par de género sobre la probabilidad de rendir la prueba PSU, lo que sesgará los resultados obtenidos en esta prueba que se observan en los Cuadros 5 y 6.

7. Análisis de Robustez

En esta sección se analizará la robustez de los resultados expuestos anteriormente. En primera instancia, se ejecutará un test de falsificación para asegurar que los efectos encontrados se deban realmente al porcentaje de mujeres y no simplemente una tendencia a nivel de colegio no capturada por los controles. A continuación, y dado que con los datos disponibles no es posible investigar empíricamente mecanismos alternativos detrás de la relación encontrada, se analizará la heterogeneidad del efecto en cuanto a la dependencia del colegio de origen del estudiante, y el máximo nivel educativo del padre y de la madre. Cada una de estas dimensiones permitirá evaluar posibles mecanismos que explican la existencia de una relación entre la distribución de género, el desempeño académico y la elección de carreras STEM. Finalmente, se estudiará la existencia de un efecto par de género no lineal.

7.1. Test de Falsificación

Una posible amenaza a la credibilidad de los coeficientes obtenidos es que estos no estén capturando la relación entre el porcentaje de mujeres y los *outcomes* analizados, sino más bien el impacto de otras variables no observables a nivel del establecimiento. Incluso, sería posible que las relaciones encontradas sean espurias. La utilización de efectos fijos y tendencias específicas a nivel de colegio, además de una serie de controles tanto de los profesores como de los pares, debería reducir en gran parte la amenaza de no observables. Sin embargo, es imposible rechazar por completo su existencia.

Debido a esto, se propone la utilización de un test de falsificación, donde el tratamiento “placebo” es el porcentaje de mujeres del mismo nivel analizado, pero rezagado un año. Así, por ejemplo, a la cohorte de IV medio del 2017 se le asigna como tratamiento placebo la proporción de mujeres del mismo grado y colegio, pero en el 2016. Dado que el número de cursos puede variar entre años, no es posible generar tratamientos placebos para la distribución de género dentro de estos, utilizando por tanto únicamente el porcentaje en la cohorte.

El Cuadro 8 muestra la prueba de falsificación al utilizar la matrícula y titulación de carreras STEM, y el puntaje PSU como variables dependientes. Las primeras dos columnas utilizan toda la muestra, mientras que las siguientes columnas separan a los individuos según su sexo.

Cabe recordar que, al utilizar el porcentaje de mujeres dentro de la cohorte, muchos de los coeficientes estimados no eran estadísticamente significativos. Por lo tanto, lo que se espera del test es que está no significancia se mantenga.

Tal como se puede observar, en general los coeficientes de las medidas placebo no son significativos, disminuyendo incluso su magnitud. La excepción es el efecto par de género sobre el puntaje PSU de matemáticas al utilizar toda la muestra. Sin embargo, de todos modos disminuye su significancia y magnitud.

Cuadro 8: Test de Falsificación - Outcomes STEM

| | Total | | Hombres | | Mujeres | |
|---------------------------|-------------------------------------|-------------------------------------|--------------------------------------|-----------------------------------|-------------------------------------|------------------------------------|
| | % Real | % Placebo | % Real | % Placebo | % Real | % Placebo |
| a. Matrícula STEM | | | | | | |
| % Mujeres | -0.000015 (0.000089) [893374] | -0.000011 (0.000088) [859827] | -0.000085 (0.00015) [417596] | 0.000023 (0.00015) [401866] | -0.000061 (0.000092) [475777] | -0.00010 (0.000092) [457958] |
| b. Titulación STEM | | | | | | |
| % Mujeres | -0.000081 (0.00016) [244997] | 0.000062 (0.00017) [225163] | -0.000013 (0.00036) [94735] | 0.00045 (0.00038) [86794] | -0.00032** (0.00013) [150231] | -0.000010 (0.00014) [138297] |
| c. PSU lenguaje | | | | | | |
| % Mujeres | -0.00056** (0.00022) [721209] | 0.000086 (0.00022) [710528] | -0.00099*** (0.00035) [325458] | 0.00036 (0.00034) [320454] | -0.00039 (0.00030) [395748] | -0.000086 (0.00029) [390072] |
| e. PSU matemáticas | | | | | | |
| % Mujeres | 0.00057** (0.00023) [716806] | -0.00037* (0.00022) [706175] | 0.00028 (0.00034) [323592] | -0.00044 (0.00032) [318610] | 0.00058** (0.00028) [393211] | -0.00028 (0.00028) [387562] |

Coefficientes del efecto de un aumento de 1pp en la proporción de mujeres sobre la variable indicada. Primeras dos columnas utilizan la muestra completa, mientras que en las siguientes se separa según el género del alumno. % Real corresponde al porcentaje de mujeres del cohorte dentro del establecimiento. % Placebo corresponde al porcentaje de mujeres del cohorte anterior en el mismo nivel y colegio. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. Número de observaciones se encuentra entre corchetes. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

Por lo tanto, la evidencia expuesta con estas pruebas refuerza la credibilidad de los coeficientes encontrados en la sección anterior. Las medidas placebo suelen no ser significativas, e inequívocamente disminuyen en su magnitud al comparar los coeficientes con el porcentaje de mujeres real.

7.2. Análisis de Heterogeneidad: Tipo de Establecimiento

Los datos con los que se cuenta no permiten estudiar mecanismos que permitan explicar la relación entre porcentaje de mujeres y elección de carreras STEM más allá de su influencia sobre el puntaje PSU. Dado lo anterior, en esta y la siguiente sección se presenta un análisis de robustez que permitirá dar luces sobre posibles mecanismos.

En primera instancia, se presenta un análisis de robustez según la dependencia del establecimiento de origen del alumno: municipal, particular subvencionado y particular privado. El sistema educativo chileno es extremadamente segregado, la gran parte de los alumnos de contextos vulnerables asisten a colegios municipales, y colegios particulares poseen muchos más recursos para entregar educación de calidad. Dado lo anterior, es de esperar que la ocurrencia de hechos disruptivos en el contexto educacional sea mucho más probable en colegios municipales. Por lo tanto, un mayor efecto sobre alumnos de colegios municipales daría señales de que un mecanismo importante detrás del efecto par de género sería la mejora en el ambiente dentro de las salas/colegios producto de una mayor proporción de mujeres.

Para testear si el efecto par de género es heterogéneo en esta dimensión, se separa la muestra según la dependencia del establecimiento de origen, estimando un coeficiente diferente para cada submuestra. Lo anterior se hace de forma separada para hombres y mujeres. El Cuadro 9 muestra los resultados de este análisis al utilizar como variables dependientes la matrícula y titulación de carreras STEM. Se utiliza el modelo más completo para todas las estimaciones. Como se puede observar, el coeficiente asociado a los colegios particulares privados es el de menor magnitud. Además, la significancia estadística se encuentra sobre todo al considerar como pares a los compañeros de curso. La diferencia en magnitud entre la muestra de colegios municipales y subvencionados no es tan amplia, aunque se suelen encontrar un coeficiente levemente mayor al enfocarnos en los primeros. En esta muestra, un aumento de 10pp en la proporción de mujeres dentro del curso aumenta entre 4,5pp la probabilidad de que los hombres se matriculen en carreras STEM, efecto mayor al observado en el Cuadro 4. Para las mujeres, el efecto provocado por este cambio en la distribución de género es de 1,5pp, levemente mayor al encontrado anteriormente.

Los resultados obtenidos al utilizar la titulación STEM como variable dependiente son similares: el efecto en la muestra de colegios municipales es levemente mayor al compararlo con el Cuadro 4, y la magnitud cae abruptamente en la muestra de alumnos provenientes de colegios particulares privados.

En el Anexo 5 se muestran los resultados de este mismo análisis al utilizar como variables dependientes el puntaje en la PSU. Las conclusiones que se puede obtener de estas tablas son similares a lo discutido anteriormente: el coeficiente es de una magnitud mucho menor en la muestra de colegios particulares privados.

Estos resultados refuerzan la hipótesis de que un mecanismo importante para explicar el efecto par de género es el mejor ambiente dentro de las salas de clase cuando hay una mayor proporción de mujeres (Lavy & Schlosser, 2011). Es de esperar que la ocurrencia de hechos

disruptivos sea más probable en colegios municipales, dado que asisten alumnos de contextos más vulnerables y no son capaces de atraer a los mejores profesores, los cuales podrían estar más capacitados para mantener el orden dentro del aula. Así, la “productividad” de tener más mujeres sería mayor en estos contextos, provocando por tanto un mayor coeficiente.

Cuadro 9: Heterogeneidad STEM - Dependencia del Establecimiento

| | Pares: Cohorte | | Pares: Curso | |
|---------------------------|------------------------------------|------------------------------------|-------------------------------------|-------------------------------------|
| | Hombres | Mujeres | Hombres | Mujeres |
| a. Matrícula STEM | | | | |
| Municipal | -0.00032 (0.00033) [120983] | -0.00035* (0.00019) [147701] | -0.0045*** (0.00019) [120983] | -0.0015*** (0.00014) [147701] |
| Subvencionado | -0.000020 (0.00019) [244487] | 0.0000054 (0.00011) [277801] | -0.0040*** (0.00022) [244487] | -0.0013*** (0.00013) [277801] |
| Privado | -0.00027 (0.00034) [52123] | -0.000087 (0.00027) [50274] | -0.0012*** (0.00034) [52123] | -0.00085*** (0.00025) [50274] |
| b. Titulación STEM | | | | |
| Municipal | -0.00028 (0.00078) [29325] | -0.00053** (0.00024) [49936] | -0.0052*** (0.00030) [29325] | -0.0011*** (0.00014) [49936] |
| Subvencionado | 0.00018 (0.00044) [55854] | -0.00033* (0.00017) [86682] | -0.0056*** (0.00031) [55854] | -0.0010*** (0.00013) [86682] |
| Privado | -0.00019 (0.00098) [9554] | -0.00011 (0.00049) [13610] | -0.0017** (0.00071) [9554] | -0.00057 (0.00040) [13610] |

Coeficientes del efecto de un aumento de 1pp en la proporción de mujeres sobre la prob. de matricularse o titularse de carreras STEM, diferenciando por el sexo del alumno. Se divide la muestra según la dependencia del colegio de origen. Las primeras dos columnas miden el porcentaje de mujeres en la cohorte dentro del colegio, mientras que las últimas dos lo miden en el curso. Se utiliza el modelo más completo para cada estimación. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. Número de observaciones se encuentra entre corchetes. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

La explicación anterior, aunque originada para explicar la relación entre porcentaje de mujeres y desempeño académico, también permite explicar el patrón encontrado al utilizar la matrícula y titulación de carreras STEM como variables dependientes. Al elegir una carrera,

la afinidad con ramos asociados a esta juega un rol fundamental. Esto sería especialmente importante para carreras STEM, ya que ramos como matemáticas y ciencias son vistos como más complicados, y por tanto es más difícil que los alumnos tengan afinidad con ellos. Un mejor ambiente en las clases podría ayudar a que profesores realicen mejores clases, pudiendo por tanto incentivar el interés de los alumnos por estos ramos, aumentando por tanto la probabilidad de que posteriormente elijan carreras STEM.

7.3. Análisis de heterogeneidad: Escolaridad del Padre y la Madre

Una última dimensión en la cual se analizará la heterogeneidad del efecto par de género es según la escolaridad del padre y la madre. Uno de los posibles mecanismos que permitirían explicar la relación entre la proporción de pares mujeres y la elección de carreras STEM por parte de los alumnos es la existencia de modelos a seguir de su mismo sexo. Si el *role model* de las madres es sustituto del de las compañeras, entonces hijas de madres con estudios STEM deberían verse menos afectadas por el efecto par de género a la hora de elegir carreras en esta área. Lo mismo debería observarse con alumnos de padres con estudios superiores STEM.

Las bases de datos con las que se cuentan no permiten identificar el área de trabajo o estudio del padre y la madre, por lo que no es posible analizar lo anterior de forma directa. Sin embargo, se aproximará esta posible heterogeneidad mediante la escolaridad de estos. En específico, separaremos la muestra en dos subgrupos: aquellos con madres(padres) que asistieron a la educación superior, y aquellos con madres(padres) que no asistieron.

El Cuadro 10 muestra los resultados de este análisis al enfocarnos sobre la matrícula y titulación STEM, utilizando la escolaridad de la madre para separar la muestra. Como se puede observar, las diferencias son claras: el efecto al utilizar la muestra de alumnos con madres sin educación superior es aproximadamente el doble al que se encuentra al utilizar a los hijos de madres con algún tipo de educación superior.

En el Anexo 6 muestra el mismo análisis, pero ahora diferenciando por la escolaridad del padre. Los coeficientes estimados en este caso son prácticamente idénticos a los del Cuadro 10. La similitud entre los coeficientes de ambos cuadros, tanto para alumnas como alumnos, tiene una explicación más práctica que teórica. Al analizar la educación del padre y de la madre en conjunto, se observa que solo un 17,5% de los individuos tienen padres que se encuentran en grupos distintos, es decir, una madre que asistió a educación superior y un padre que no o viceversa. Debido a esto, no es posible distinguir claramente la influencia individual del nivel educativo de cada progenitor, dificultando generar conclusiones claras de estos resultados.

En los Anexos 7 y 8 se muestra el mismo análisis, pero utilizando el desempeño en la PSU como variable dependiente. Nuevamente se encuentra que el efecto es mayor en la muestra cuyos padres no asistieron a la educación superior. Sin embargo, para el desempeño académico se suele observar que el efecto par de género no es significativo en la muestra de estudiantes cuyos padres sí asistieron, algo no observado en la matrícula y titulación de carreras STEM.

Esta falta de significancia podría explicarse por el hecho que los alumnos de este grupo tienen una mayor probabilidad de asistir a establecimientos privados. En efecto, mientras solo un 10 % de la muestra total de alumnos asiste a colegios particulares pagados, el porcentaje aumenta a 30,3 % cuando analizamos únicamente alumnos cuyos padres asistieron a la educación superior. Así, estos resultados podrían estar reproduciendo lo observado en el análisis por dependencia del establecimiento.

Cuadro 10: Heterogeneidad STEM - Escolaridad de la Madre

| | Pares: Cohorte | | Pares: Curso | |
|------------------------------|-----------------------------------|------------------------------------|-------------------------------------|--------------------------------------|
| | Hombres | Mujeres | Hombres | Mujeres |
| a. Matrícula STEM | | | | |
| Madre sin educación superior | -0.00027 (0.00019) [273393] | -0.000061 (0.00011) [329516] | -0.0046*** (0.00015) [273393] | -0.0015*** (0.000099) [329516] |
| Madre con educación superior | 0.000045 (0.00023) [144145] | -0.00011 (0.00017) [146188] | -0.0021*** (0.00018) [144145] | -0.00089*** (0.00013) [146188] |
| b. Titulación STEM | | | | |
| Madre sin educación superior | -0.000087 (0.00045) [67517] | -0.00022 (0.00015) [111305] | -0.0058*** (0.00023) [67517] | -0.0011*** (0.00011) [111305] |
| Madre con educación superior | -0.00017 (0.00064) [26947] | -0.00045 (0.00030) [38722] | -0.0033*** (0.00040) [26947] | -0.00054*** (0.00019) [38722] |

Coeficientes del efecto de un aumento de 1pp en la proporción de mujeres sobre la prob. de matricularse o titularse de carreras STEM, diferenciando por género del alumno. Se separa entre alumnos cuyas madres asistieron a educación superior y aquellos que no. Las primeras dos columnas miden el porcentaje de mujeres en la cohorte dentro del colegio, mientras que las últimas dos lo miden en el curso. Se utiliza el modelo más completo para cada estimación. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. Número de observaciones se encuentra entre corchetes. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

7.4. Efectos No Lineales

Una última dimensión que será estudiada en esta tesis es la existencia de un efecto par de género no lineal. En específico, se agregará el porcentaje de mujeres al cuadrado como una nueva variable explicativa. El Cuadro 11 muestra los resultados de estas estimaciones, ocupando en cada caso el modelo más completo. Nuevamente, al considerar a todos los alumnos de la cohorte como pares no se obtienen efectos estadísticamente significativos. Sin embargo, para el caso en que se consideran como pares los alumnos del mismo curso,

se obtiene una clara no linealidad: el porcentaje de mujeres tiene un efecto negativo, sin embargo, el porcentaje al cuadrado es positivo. Es decir, el efecto par de género tiene un comportamiento convexo, tanto sobre la matrícula como titulación de carreras STEM.

Cuadro 11: Efecto par de género sobre matrícula y titulación STEM - no linealidad

| | Pares: Cohorte | | Pares: Curso | |
|---------------------------|--------------------------|---------------------------|----------------------------|----------------------------|
| | Hombres | Mujeres | Hombres | Mujeres |
| a. Matrícula STEM | | | | |
| % Mujeres | -0.00043 (0.00075) | 0.00061 (0.00055) | -0.0085*** (0.00046) | -0.0042*** (0.00036) |
| % Mujeres ² | 0.0000034 (0.0000075) | -0.0000061 (0.0000049) | 0.000052*** (0.0000048) | 0.000024*** (0.0000028) |
| <i>N</i> | 417596 | 475777 | 417596 | 475777 |
| b. Titulación STEM | | | | |
| % Mujeres | -0.0028 (0.0018) | -0.00073 (0.00087) | -0.010*** (0.00069) | -0.0037*** (0.00040) |
| % Mujeres ² | 0.000028 (0.000017) | 0.0000038 (0.0000077) | 0.000060*** (0.0000075) | 0.000022*** (0.0000030) |
| <i>N</i> | 94735 | 150231 | 94735 | 150231 |

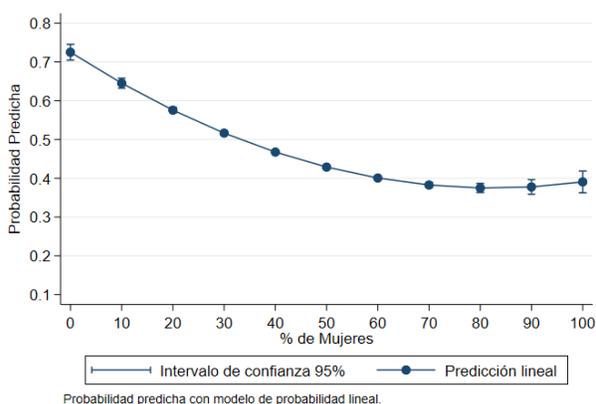
Coefficientes se interpretan como el cambio en la probabilidad ante un aumento de 1pp en la variable indicada, diferenciando por el sexo del alumno. Las primeras dos columnas miden el porcentaje de mujeres en la cohorte dentro del colegio, mientras que las últimas dos lo miden en el curso. Se utiliza el modelo más completo para cada estimación. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

La Figura 3 permite ver esta convexidad de forma gráfica: en el eje x se muestra el porcentaje de pares mujeres¹¹, mientras que en el eje y se incluye la probabilidad predicha de matrícula (figuras 3a y 3b) y titulación (figuras 3c y 3d), diferenciando por el sexo del alumno. Como se puede observar, ambas probabilidades son decrecientes con el porcentaje de pares mujeres en el curso, sin embargo, a tasas decrecientes. Entre el 80 % y 90 % de mujeres la probabilidad llega a un mínimo, luego de lo cual el efecto marginal se vuelve positivo.

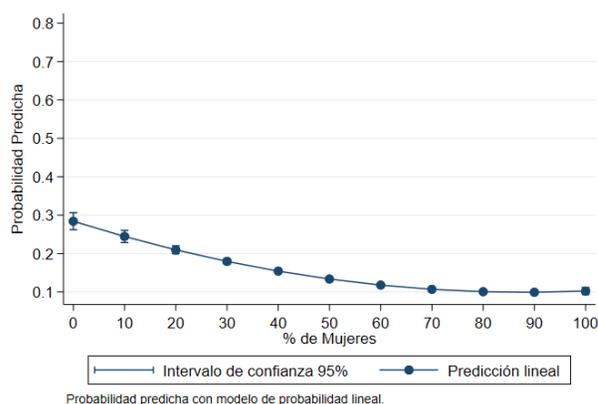
De lo anterior se puede concluir que en gran parte de la distribución el efecto par de género será negativo para los *outcomes* STEM. Solo en contextos con una proporción de mujeres extremadamente alta (más del 90 %), se tendrá un efecto marginal positivo. Cabe destacar que, debido a la estrategia empírica utilizada, estos resultados no se pueden aplicar a los escenarios extremos (colegios de un solo sexo). Se ha documentado que la mera presencia

¹¹Se utiliza el modelo que considera como pares a los alumnos del mismo curso, dado que es el único modelo en el que se obtienen efectos significativos.

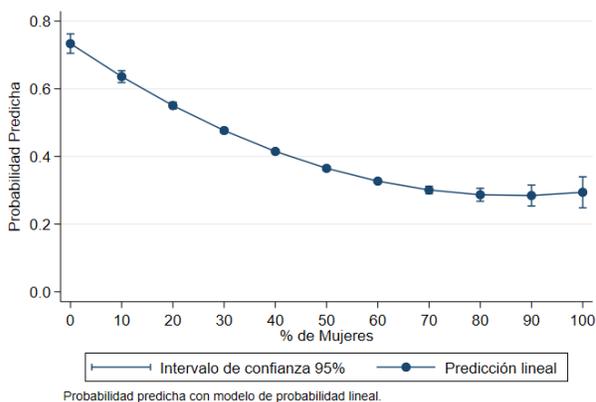
de hombres en escenarios competitivos podría ser perjudicial para las mujeres, por lo que el efecto par de género al momento de elegir una carrera podría comportarse de forma distinta a lo predicho por el modelo.



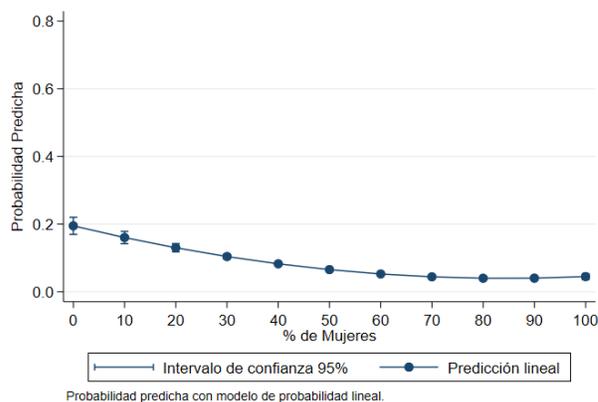
(a) Matrícula STEM: hombres



(b) Matrícula STEM: mujeres



(c) Titulación STEM - hombres



(d) Titulación STEM - mujeres

Figura 3: Probabilidad de matrícula y titulación STEM - efecto par de género no lineal

Los Anexos 9 y 10 muestra el mismo análisis de no linealidad, pero estudiando el puntaje PSU. Nuevamente se encuentra evidencia a favor de un efecto no lineal al considerar como pares a los alumnos del mismo curso. Sin embargo, ahora la naturaleza de la relación entre porcentaje de mujeres y *outcome* es cóncava¹²: el puntaje, tanto para lenguaje como para matemáticas, es creciente con el porcentaje de mujeres en el curso, pero a tasas decrecientes.

¹²El único caso donde esto no ocurre es el efecto sobre la PSU de matemáticas en la muestra de mujeres. Lo anterior hace sentido si se toma en cuenta que este efecto no es significativo en el cuadro 6.

Entre el 40 % y 60 % de mujeres el puntaje llega a un máximo, a partir del cual la relación entre ambas variables se vuelve negativa.

Es necesario destacar el efecto sobre el puntaje PSU de lenguaje para las mujeres. Cuando se asume un efecto par lineal se encuentra un coeficiente negativo, mientras que para los hombres es positivo. Sin embargo, al agregar el porcentaje de mujeres al cuadrado, se obtiene un efecto cóncavo similar para ambos sexos. Lo que estaría ocurriendo es que el punto de inflexión de la concavidad es cercano al 50 % de pares mujeres, y las alumnas estudiarán, por construcción, en mayor porcentaje en cohortes con más mujeres. Por lo tanto, el efecto negativo al asumir un efecto lineal es simplemente una consecuencia de que las estudiantes se encuentran en una mayor proporción en la parte de decreciente de la curva, y no a que los efectos para ambos sexos sean de distinta naturaleza.

7.5. ¿En que curso se debe medir el porcentaje de mujeres?

Hay una particularidad dentro del ciclo de enseñanza media que podría afectar los coeficientes estimados: la elección de electivos. En el caso de la enseñanza media técnico-profesional, al pasar de segundo a tercero medio los alumnos comienzan a tener ramos propios de su especialidad, lo que provoca un reordenamiento de los alumnos entre distintos cursos. Lo mismo ocurre en muchos colegios científico-humanistas, solo que en este caso la elección viene dada por los electivos (más enfocados en ciencias naturales, matemáticas, etc.). Este reordenamiento de los alumnos en tercero medio podría constituir una fuente de sesgo para las estimaciones del efecto par de género sobre la PSU y los *outcomes* STEM, ya que el porcentaje de mujeres en estos casos se mide en cuarto medio, posterior a la elección.

Por lo anterior, en el cuadro 12 se muestra el cambio en el efecto par de género estimado ante cambios en el curso donde se mide la distribución de género. Cabe destacar que solo se muestran los resultados al definir como pares a los alumnos del mismo curso, ya que es la definición que ha entregado efectos significativos. Además, el efecto fijo por establecimiento está determinado por el nivel en que se calcula el porcentaje de mujeres, para así solo ocupar la variación *within* de la variable.

Como se puede observar, el coeficiente estimado es de mayor magnitud al medir el porcentaje de mujeres en tercero o cuarto medio, precisamente los niveles posteriores a la elección de electivos. Esto ocurre tanto para el efecto sobre la probabilidad de matricularse como titularse desde una carrera STEM, y es especialmente importante para las mujeres: el efecto al utilizar el porcentaje de tercero o cuarto medio tiene una magnitud casi diez veces mayor al estimado utilizando las otras medidas.

El Anexo 11 muestra una tabla similar, pero al estudiar el puntaje PSU. En este caso, tanto el efecto para los hombres como mujeres tiende a hacerse más positivo. Esto es especialmente importante para la estimación con la submuestra de mujeres: el efecto par de género negativo sobre las mujeres en la PSU de lenguaje se observa únicamente con las medidas de tercero y cuarto medio, siendo positivo y significativo con las otras medidas, mientras que el efecto

sobre la prueba de matemáticas se hace estadísticamente significativo.

Cuadro 12: Efecto par de género - robustez ante cambios en el curso donde se mide el porcentaje de mujeres

| | Hombres | | | | Mujeres | | | |
|---------------------------|-------------------------|-------------------------|-------------------------|-------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| a. Matrícula STEM | | | | | | | | |
| % Mujeres IV Medio | -0.0041*** (0.00015) | | | | -0.0014*** (0.000090) | | | |
| % Mujeres III Medio | | -0.0041*** (0.00015) | | | | -0.0014*** (0.000089) | | |
| % Mujeres II Medio | | | -0.0011*** (0.00019) | | | | -0.00012 (0.000079) | |
| % Mujeres I Medio | | | | -0.0010*** (0.00019) | | | | -0.00016* (0.000080) |
| <i>N</i> | 417596 | 411483 | 396621 | 381595 | 475777 | 470091 | 457505 | 443701 |
| b. Titulación STEM | | | | | | | | |
| % Mujeres IV Medio | -0.0054*** (0.00022) | | | | -0.0010*** (0.000095) | | | |
| % Mujeres III Medio | | -0.0055*** (0.00022) | | | | -0.0010*** (0.000092) | | |
| % Mujeres II Medio | | | -0.0019*** (0.00033) | | | | -0.00016** (0.000083) | |
| % Mujeres I Medio | | | | -0.0019*** (0.00033) | | | | -0.00017** (0.000081) |
| <i>N</i> | 94735 | 93423 | 90232 | 86892 | 150231 | 148445 | 144415 | 139737 |

Coefficientes se interpretan como el cambio en la probabilidad ante un aumento de 1pp en la proporción de mujeres en el curso indicado, diferenciando por el sexo del alumno. Todos los modelos estimados miden el porcentaje de mujeres dentro del curso. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

A fin de encontrar un efecto lo más limpio posible, el cuadro 13 muestra los resultados al utilizar el porcentaje de mujeres en I y II medio como instrumentos para el porcentaje en IV medio. Cada sección de la tabla muestra el efecto par de género para una variable dependiente distinta: matrícula STEM; titulación STEM; PSU de lenguaje; y PSU de matemáticas. Nuevamente, se muestran los resultados únicamente al utilizar el porcentaje de mujeres dentro del curso. Se incluyen también el estadístico F de Kleibergen-Paap para instrumentos débiles y el test LM de Kleibergen-Paap para subidentificación, los cuales permiten comprobar que los instrumentos son correctos en todos los modelos.

Al centrarse los primeros dos paneles de la tabla (carreras STEM), es posible notar que el coeficiente asociado a un aumento en la proporción de mujeres dentro del curso reduce su magnitud al ser instrumentalizado. La reducción es mucho más importante en el efecto sobre la matrícula STEM: el coeficiente disminuye en más de un 50% para los hombres, mientras que el de las mujeres pierde significancia estadística. La reducción en la magnitud del coeficiente al analizar la titulación STEM aún existe, aunque es menor y no se pierde significancia. Así, un aumento de 10 puntos porcentuales en la proporción de mujeres dentro del curso disminuirá entre 1,6 y 2 puntos porcentuales la probabilidad de matrícula en carreras STEM de los hombres, mientras que el efecto sobre las mujeres del mismo cambio no

es estadísticamente significativo. La reducción de la probabilidad de titulación STEM ante este mismo cambio será de entre 1 y 3,5 puntos porcentuales para los hombres, y de entre 0,53 y 0,66 puntos porcentuales para las mujeres.

Cuadro 13: Efecto par de género - variables instrumentales

| | Hombres | | | | Mujeres | | | |
|---------------------------|-------------------------|-------------------------|-------------------------|-------------------------|--------------------------|-------------------------|--------------------------|--------------------------|
| | MCO | VI | VI | VI | MCO | VI | VI | VI |
| a. Matrícula STEM | | | | | | | | |
| % Mujeres del Curso | -0.0041*** (0.00015) | -0.0020*** (0.00042) | -0.0016*** (0.00036) | -0.0016*** (0.00036) | -0.0014*** (0.000090) | -0.000065 (0.00024) | -0.00016 (0.00018) | -0.00018 (0.00018) |
| N | 417596 | 381670 | 396724 | 380468 | 475777 | 443781 | 457588 | 442259 |
| Kleibergen-Paap (LM) | | 177.1 | 232.1 | 304.0 | | 281.1 | 369.5 | 391.3 |
| Kleibergen-Paap (F) | | 379.9 | 686.4 | 403.9 | | 555.3 | 946.4 | 498.5 |
| b. Titulación STEM | | | | | | | | |
| % Mujeres del Curso | -0.0054*** (0.00022) | -0.0043*** (0.00073) | -0.0035*** (0.00060) | -0.0035*** (0.00061) | -0.0010*** (0.000095) | -0.00066** (0.00029) | -0.00053*** (0.00020) | -0.00059*** (0.00020) |
| N | 94735 | 86982 | 90326 | 86722 | 150231 | 139837 | 144532 | 139392 |
| Kleibergen-Paap (LM) | | 104.9 | 136.2 | 151.8 | | 181.3 | 239.0 | 243.5 |
| Kleibergen-Paap (F) | | 210.4 | 376.9 | 197.9 | | 333.6 | 582.2 | 295.3 |
| c. PSU Lenguaje | | | | | | | | |
| % Mujeres del Curso | 0.0018*** (0.00018) | 0.0024*** (0.00065) | 0.00085* (0.00048) | 0.0011** (0.00048) | -0.0016*** (0.00014) | 0.0086*** (0.00095) | 0.0044*** (0.00062) | 0.0045*** (0.00062) |
| N | 323998 | 297335 | 309739 | 296276 | 394265 | 368888 | 381217 | 367569 |
| Kleibergen-Paap (LM) | | 143.2 | 184.9 | 229.7 | | 277.7 | 354.5 | 363.7 |
| Kleibergen-Paap (F) | | 303.1 | 539.9 | 308.7 | | 476.0 | 770.4 | 396.7 |
| d. PSU Matemáticas | | | | | | | | |
| % Mujeres del Curso | 0.0016*** (0.00018) | 0.0031*** (0.00055) | 0.0022*** (0.00042) | 0.0023*** (0.00041) | 0.00022* (0.00013) | 0.0075*** (0.00076) | 0.0052*** (0.00052) | 0.0052*** (0.00051) |
| N | 322251 | 295905 | 308152 | 294857 | 391891 | 366841 | 379022 | 365538 |
| Kleibergen-Paap (LM) | | 143.8 | 182.2 | 221.2 | | 298.0 | 375.9 | 382.5 |
| Kleibergen-Paap (F) | | 317.3 | 558.3 | 316.1 | | 527.2 | 845.6 | 432.2 |
| Instrumento: % I medio | No | Sí | No | Sí | No | Sí | No | Sí |
| Instrumento: % II medio | No | No | Sí | Sí | No | No | Sí | Sí |

Coefficientes se interpretan como el cambio en la variable dependiente indicada ante un aumento de 1pp en la proporción de mujeres en el curso indicado, diferenciando por el sexo del alumno. Todos los modelos estimados miden el porcentaje de mujeres dentro del curso. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

Los cambios al utilizar VI al analizar el efecto sobre la PSU de lenguaje son distintos para hombres y mujeres. Para los primeros la magnitud disminuye al instrumentalizar el porcentaje de mujeres, mientras que para las segundas el efecto no solo aumenta, sino que se hace positivo. En matemáticas se observa para ambos sexos un aumento en la magnitud del coeficiente estimado al instrumentalizar la proporción de mujeres, aunque nuevamente el cambio es más importante para estas últimas.

En resumen, los resultados al instrumentalizar la proporción de mujeres van en la línea de lo observado en el cuadro 12 y en el anexo 11, los cuales se pueden entender como modelos reducidos al incluir los instrumentos utilizados en el cuadro 13 directamente como controles en la regresión. El efecto par de género sobre la elección de carreras STEM de las mujeres pierde magnitud y significancia al ser instrumentalizado, mientras que el efecto par de género sobre el rendimiento en la prueba PSU de las mujeres aumenta su magnitud, e incluso cambia de signo para el caso de lenguaje.

Cabe destacar que, como consecuencia de los cambios provocados al instrumentalizar el porcentaje de mujeres, el efecto par de género sobre el rendimiento académico ahora es más fuerte sobre las mujeres. Además, el signo del efecto anterior permite concluir que esta canal sigue atenuando de forma indirecta el impacto negativo de un aumento de la proporción de mujeres sobre la probabilidad de elección de carreras STEM.

8. Conclusiones

La presente tesis busca entregar nueva evidencia sobre como el efecto par de género, medido mediante el porcentaje de compañeras mujeres, afecta la elección de carreras STEM por parte de los estudiantes. Para encontrar el efecto se utiliza un modelo de efectos fijos, aprovechando la variación idiosincrática que se genera al comparar cohortes adyacentes, controlando por características de los establecimientos y socioeconómicas de los individuos y sus pares.

Los resultados muestran una relación negativa entre porcentaje de mujeres y elección de carreras STEM, y generalmente positiva con el desempeño académico (pruebas SIMCE y PSU), tanto para alumnos varones como mujeres. Así, el mecanismo del desempeño académico no permitiría explicar el efecto par de género sobre la elección de carreras STEM, incluso atenuándolo, aunque en una proporción muy baja.

Además, se encuentra evidencia robusta sobre la no linealidad del efecto: el puntaje PSU tanto de matemáticas como lenguaje se maximiza en distribuciones de género equitativas (cercasas al 50%), mientras que la elección de carreras STEM se minimiza en escenarios con una proporción excesiva de mujeres en el curso (entre 80% y 90%).

Finalmente, el hecho de que los alumnos en tercero medio se reordenen debido a la elección de electivos y especialidad técnico-profesionales provoca sesgo en las estimaciones al analizar el desempeño en la PSU, la matrícula y la titulación STEM. La utilización de variables instrumentales provoca una disminución en la magnitud del efecto sobre los *outcomes* STEM. En el desempeño PSU se observa un aumento de la magnitud del efecto para las mujeres, y una disminución para los hombres. Es de especial importancia el cambio en el primer grupo para el desempeño de lenguaje, ya que el efecto se hace positivo, en línea con lo que se encuentra generalmente en la literatura.

Los resultados de la presente tesis entregan evidencia en contra de la utilización de ambientes segregados por sexo para aumentar la proporción de mujeres estudiando en el área STEM. Dada la no linealidad, este tipo de políticas provocarían un aumento en las brechas de género dentro de estas carreras: los hombres tendrían probabilidad máxima de matricularse y titularse en esta área, mientras que las mujeres se encontrarían cerca del mínimo. Incluso en el mejor escenario (efectos encontrados con VI), un aumento en la proporción de mujeres dentro del curso no tendría un efecto significativo sobre la probabilidad de que alumnas se matriculen o titulen de carreras STEM.

Al incluir el efecto sobre el desempeño en la PSU, a juicio del autor la evidencia permite

defender ambientes mixtos dentro de las salas de clase, idealmente paritarios. Por un lado, habría un efecto positivo sobre las mujeres en comparación a cursos de un solo sexo. Por otro lado, se maximizaría el rendimiento en la prueba de selección universitaria tanto para hombres como para mujeres.

Aunque en primera instancia el efecto negativo sobre la probabilidad de entrar a carreras STEM para los hombres podría verse como una consecuencia negativa, para el autor esto no es así. La brecha de género en el área STEM puede verse tanto como una subrepresentación de mujeres como una sobrerrepresentación de hombres. Así, si los alumnos que dejan de entrar a estas carreras terminan estudiando en áreas tradicionalmente feminizadas¹³, como pedagogía, el efecto agregado en la economía podría ser incluso mayor. Como se explicó anteriormente, son las brechas de género en general, no solo aquellas contra las mujeres, las que afectarían la productividad.

Es necesario destacar, sin embargo, que los resultados expuestos anteriormente presentan una serie de limitaciones. En primer lugar, el análisis ANOVA mostró que la variación del porcentaje de mujeres entre cohortes adyacentes de un mismo colegio baja bastante a partir de enseñanza media. Lo anterior nos obligó a utilizar también el porcentaje de mujeres en el curso, lo que podría provocar sesgo en los estimadores.

Por otro lado, el trabajo se enfocó en la elección de carreras STEM, lo cual se podría ampliar a un modelo más complejo de decisión que incluyan más carreras como lo hecho por Bordón, Canals, & Mizala, (2020). Saber a ciencia cierta a qué tipo de carreras postularán los hombres que dejarían de postular al área STEM ante una hipotética transición desde colegios de un solo sexo a mixtos ayudaría a generar conclusiones más completas. Esta dimensión queda para ser estudiada en futuros trabajos.

Finalmente, los datos con los que se cuentan no permiten explorar mecanismos alternativos que expliquen la relación entre porcentaje de pares mujeres y elección de carreras STEM. El hecho de encontrar un efecto generalmente negativo para ambos sexos permitiría descartar un gran número de candidatos. Un canal que permitiría explicar el signo del efecto es que un mayor porcentaje de mujeres reduzca la competitividad en los cursos, y el gusto por entrar a futuro a ambientes competitivos, como el de las carreras STEM. Sin embargo, esto es solo una hipótesis, y queda como labor para futuros trabajos ahondar más en ella.

Un resultado necesario de destacar es el hecho de que, al estudiar el efecto par de género sobre el desempeño académico, es vital controlar por el efecto par académico, sobre todo en contextos donde las brechas en los resultados de hombres y mujeres son amplias. De lo contrario, se estaría sobreestimando el efecto de una mayor proporción de alumnas al estudiar asignaturas con brechas de género a favor de ellas, y subestimando cuando la brecha es en contra de ellas. Esta dimensión del análisis ha sido, hasta ahora, omitida en la literatura.

¹³La reducción en la probabilidad de los hombres no debe entenderse que estos dejan de entrar a la educación superior, al fin y al cabo, las estimaciones son condicionales a matricularse en una carrera

Bibliografía

- Akerlof, G., Kranton, R. (2002).** Identity and Schooling: Some Lessons for the Economics of Education. *Journal of Economic Literature*, 1167-1201.
- Anelli, M., Peri, G. (2019).** The Effects of High School Peers' Gender on College Major, College Performance and Income. *The Economic Journal*, 129(618), 553-602.
- Arcidiacono, P. (2004).** Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2), 343-375.
- Blau, F., Kahn, L. (2017).** The Gender Wage Gap: Extent, Trends, and Explanations. *Journal of Economic Literature*, 55(3), 789-865.
- Booth, A., Nolen, P. (2012a).** Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization*, 81(2), 542-555.
- Booth, A., Nolen, P. (2012b).** Gender differences in risk behaviour: does nurture matter? *The Economic Journal*, 122, F56-78.
- Bordón, P., Canals, C., Mizala, A. (2020).** The gender gap in college major choice in Chile. *Economics of Education Review*, 77, 1-27.
- Brenøe, A. A., Zölitz, U. (2020).** Exposure to More Female Peers Widens the Gender Gap in STEM Participation. *Journal of Labor Economics*, 38(4), 1009-1054.
- Buser, T., Niederle, M., Oosterbeek, H. (2014).** Gender, Competitiveness and Career Choices. *The Quarterly Journal of Economics*, 129(3), 1409-1448.
- Cabezas, V. (2010).** Gender Peer Effects in School: Does the Gender of School Peers Affect Student's Achievement? Tesis Doctoral, Universidad de Columbia.
- CONICYT. (2018).** Reporte de Participación Femenina 2018. Departamento de Estudios y Gestión Estratégica, Santiago.
- Cuberes, D., Teignier, M. (2016).** Aggregate Effects of Gender Gaps in the Labor Market: A Quantitative Estimate. *Journal of Human Capital*, 10(1), 1-32.
- Devadas, S., Kim, Y. E. (2020).** Exploring the Potential of Gender Parity to Promote Economic Growth. World Bank, Research Policy Briefs No. 39, Washington, DC.
- Dustmann, C., Ku, H., Kwak, D. W. (2018).** Why Are Single-Sex Schools Successful? *Labour Economics*, 54, 79-99.
- Favara, M. (2012).** The Cost of Acting "Girly": Gender Stereotypes and Educational Choices. IZA Discussion Papers No. 7037.
- Feld, J., Zölitz, U. (2018).** The Effect of Peer Gender on Major Choice. Universidad de Zurich, Departamento de Economía. Working Paper No. 270.

- Gneezy, U., Niederle, M., Rustichini, A. (2003).** Performance in competitive environments: Gender differences. *The Quarterly Journal of Economics*, 118(3), 1049-1074.
- Gong, J., Lu, Y., Song, H. (2019).** Gender Peer Effects on Students' Academic and Noncognitive Outcomes: Evidence and Mechanisms. *Journal of Human Resources*, 1-49.
- Hanushek, E., Kain, J., Markman, J., Rivkin, S. (2001).** Does Peer Ability Affect Student Achievement? National Bureau of Economic Research. Cambridge: NBER Working Paper 8502.
- Hayes, A., Pahlke, E., Bigler, R. (2011).** The Efficacy of Single-Sex Education: Testing for Selection and Peer Quality Effects. *Sex Roles*, 65, 693-703.
- Heckman, J., Rubinstein, Y. (2001).** The Importance of Noncognitive Skills: Lessons from the GED Testing Program. *The American Economic Review*, 91(2), 145-149.
- Hill, A. (2017).** The positive influence of female college students on their male peers. *Labour Economics*, 44, 151-160.
- Hoxby, C. (2000).** Peer Effects in the Classroom: Learning From Gender and Race Variation. National Bureau of Economic Research. Cambridge: NBER Working Paper 7867.
- Huntington-Klein, N., Rose, E. (2018).** Gender Peer Effects in a Predominantly Male Environment: Evidence from West Point. *AEA Papers and Proceedings*, 108, 392-395.
- Iversen, J., Bonesrønning, H. (2015).** Conditional gender peer effects? *Journal of Behavioral and Experimental Economics*, 55, 19-28.
- Jones, M., Wheatley, J. (1990).** Gender differences in teacher-student interactions in science classrooms. *Journal of Research in Science Teaching*, 27(9), 861-874.
- Joy, L. (2006).** Occupational differences between recent male and female college graduates. *Economics of Education Review*, 25(2), 221-231.
- Lavy, V., Schlosser, A. (2011).** Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics*, 3(2), 1-33.
- Lee, V., Marks, H. (1990).** Sustained effects of the single-sex secondary school experience on attitudes, behaviors, and values in college. *Journal of Educational Psychology*, 82(3), 578-592.
- Lee, V., Marks, H., Byrd, T. (1994).** Sexism in Single-Sex and Coeducational Independent Secondary School Classrooms. *Sociology of Education*, 67(2), 92-120.
- Maccoby, E. (1990).** Gender and relationships: A developmental account. *American Psychologist*, 45(4), 513-520.
- Maccoby, E. (1999).** The two sexes: Growing up apart, coming together. Family and public policy Vol. 4. Harvard University Press.

- Manski, C. (1993).** Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60(3), 531-542.
- Niederle, M., Vesterlund, L. (2007).** Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics*, 122(3), 1067-1101.
- OECD. (2020).** Education at a Glance 2020: OECD Indicators. Paris: OECD Publishing.
- Ostry, J., Alvarez, J., Espinoza, R., Papageorgiou, C. (2018).** Economic Gains from Gender Inclusion: New Mechanisms, New Evidence. Fondo Monetario Internacional. Washington, DC: IMF Staff Discussion Note 18/06.
- Paredes, V. (2018).** Mixed but not Scrambled Gender Gaps in Single-Sex Classrooms. Universidad de Chile, Departamento de Economía. Serie Documentos de Trabajo.
- Park, H., Behrman, J., Choi, J. (2018).** Do single-sex schools enhance students' STEM (science, technology, engineering, and mathematics) outcomes? *Economics of Education Review*, 62, 35-47.
- Peri, G., Shih, K., Sparber, C. (2015).** STEM Workers, H-1B Visas, and Productivity in US Cities. *Journal of Labor Economics*, 33(1), 225-255.
- Sax, L., Shapiro, C., Eagan, K. (2011).** Promoting Mathematical and Computer Self-Concept among Female College Students: is There a Role of Single-Sex Secondary Education? *Journal of Women and Minorities in Science and Engineering*, 17(4), 325-355.
- Schneeweis, N., Zweimüller, M. (2012).** Girls, girls, girls: Gender composition and female school choice. *Economics of Education Review*, 31, 482-500.
- Thompson, J. (2003).** The Effect of Single-sex Secondary Schooling on Women's Choice of College Major. *Sociological Perspectives*, 46(2), 257-278.
- UNESCO. (2014).** ISCED Fields of Education and Training 2013 (ISCED-F 2013). UNESCO Institute for Statistics, Paris.
- UNESCO. (2017).** Measuring Gender Equality in Science and Engineering: the SAGA Toolkit. Paris: SAGA Working Paper 2.

9. Anexos

Anexo 1: Descomposición de la varianza: % de mujeres en el curso

| | 4to | | | 6to | | |
|---------|----------------|---------|--------|----------------|---------|--------|
| | Sum of Squares | % Total | DF | Sum of Squares | % Total | DF |
| Between | 2286.9708 | 33 % | 4,849 | 2694.5178 | 37 % | 4,826 |
| Within | 4829.0812 | 67 % | 532152 | 4628.6043 | 63 % | 533605 |
| Total | 7116.052 | | 537001 | 7323.1221 | | 538431 |

| | II | | | PSU | | |
|---------|----------------|---------|--------|----------------|---------|--------|
| | Sum of Squares | % Total | DF | Sum of Squares | % Total | DF |
| Between | 6475.2357 | 51 % | 2,615 | 8721.7808 | 31 % | 2,288 |
| Within | 6131.2566 | 49 % | 488150 | 19728.654 | 69 % | 714493 |
| Total | 12606.492 | | 490765 | 28450.435 | | 716781 |

Anexo 2: Detalle de controles utilizados: PSU y STEM

| Variable | Definición | Fuente |
|-------------------------------|---|---------|
| <i>Individuales</i> | | |
| Género | Dicotómica que toma valor 1 para hombres y 2 para mujeres | PSU |
| Asistencia | Porcentaje de asistencia del alumno | MINEDUC |
| Hrs. Trabajadas (STEM) | Número de horas que trabaja por semana. Solo se utiliza en regresiones de outcomes STEM | PSU |
| <i>Socioeconómicas</i> | | |
| Escolaridad Padre | Variable categórica que indica el nivel de escolaridad máximo alcanzado por el padre: básica incompleta o menos; básica completa; media incompleta; media completa; superior técnica incompleta; universitaria incompleta; superior técnica completa; universitaria completa | PSU |
| Escolaridad Madre | Variable categórica que indica el nivel de escolaridad máximo alcanzado por la madre. Mismas categorías que variable del padre. | PSU |
| Ingresos | Variable categórica que indica nivel de ingresos de la familia: menos de \$150.000; entre \$150.000 y \$300.000; entre \$300.000 y \$450.000; entre \$450.000 y \$600.000; entre \$600.000 y 750.000; entre \$750.000 y \$1.000.000; entre \$1.000.000 y \$1.550.000; entre \$1.550.000 y \$2.100.000; más de \$2.100.000 | PSU |
| Familiares que trabajan | Número de integrantes del grupo familiar que tienen trabajo remunerado | PSU |
| Hermanos que estudian | Número de hermanos que estudian en la educación superior | PSU |
| <i>Colegio</i> | | |
| Nº de Profesores | Número de profesores contratados por el colegio | MINEDUC |
| Horas Profesores | Cantidad de horas de profesores contratadas por el colegio | MINEDUC |
| Nº de Asistentes | Número de asistentes de la educación contratados por el colegio | MINEDUC |
| Horas Asistentes | Cantidad de horas de asistentes de la educación contratadas por el colegio | MINEDUC |
| % de Aprobados | Porcentaje de alumnos que aprobaron en el establecimiento y periodo del alumno | MINEDUC |

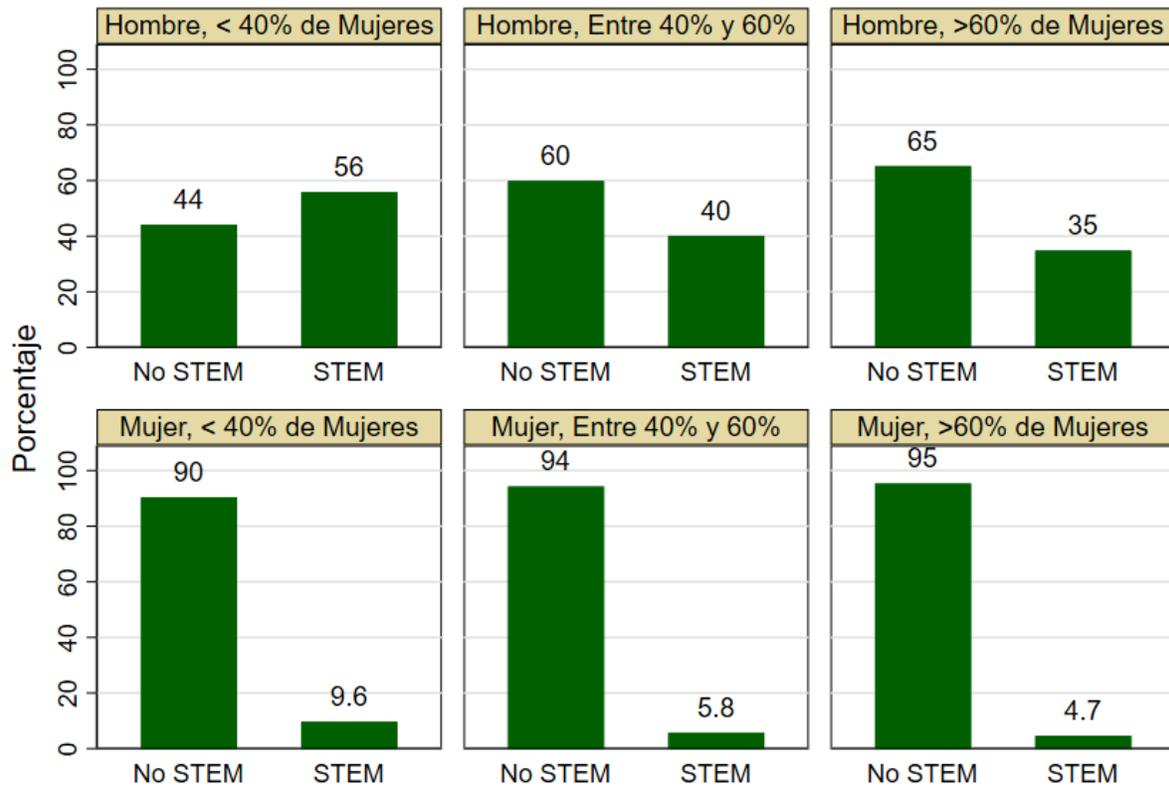
| | | |
|-----------------------------------|---|---------|
| Matrícula Colegio | Número de alumnos matriculados en el colegio del alumno. | MINEDUC |
| Tamaño del curso | Número de alumnos en el curso del alumno | MINEDUC |
| <i>Carrera (solo STEM)</i> | | |
| Tipo de Institución | Categoría que indica si el estudiante estudia en un centro de formación técnica, instituto profesional, universidad estatal, universidad privada del CRUCH, o universidad privada fuera del CRUCH. Solo se utiliza al estudiar la titulación STEM | MINEDUC |
| <i>PSU (solo STEM)</i> | | |
| Puntaje Lenguaje | Puntaje estandarizado en la prueba PSU de lenguaje | PSU |
| Puntaje Matemáticas | Puntaje estandarizado en la prueba PSU de Matemáticas | PSU |
| Puntaje Optativa | Puntaje estandarizado en la prueba PSU de optativa rendida. En caso de rendir las dos, se utiliza el puntaje mayor | PSU |
| Ciencias | Dummy que indica si el individuo rindió la prueba optativa de ciencias | PSU |
| Puntaje NEM | Puntaje estandarizado asignado a las notas de enseñanza media | PSU |
| Puntaje Ranking | Puntaje estandarizado asignado al ranking de egreso | PSU |
| <i>Pares</i> | | |
| Carac. Socioeconómicas | Promedio de las variables socioeconómicas de los pares del alumno. La definición de pares es la misma utilizada en el cálculo de la proporción de mujeres. | PSU |
| Calidad de los pares (PSU) | Promedio del puntaje de los pares en el simce de II medio. Solo se utiliza en el modelo ampliado al estudiar el desempeño PSU | SIMCE |

Anexo 3: Detalle de controles utilizados: SIMCE

| Variable | Definición | Fuente |
|------------------------|---|---------|
| Individuales | | |
| Género | Dicotómica que toma valor 1 para hombres y 2 para mujeres | SIMCE |
| Asistencia | Porcentaje de asistencia del alumno | MINEDUC |
| Socioeconómicas | | |
| Escolaridad Padre | Variable categórica que indica el nivel de escolaridad máximo alcanzado por el padre: no estudió; básica incompleta; básica completa; media incompleta; media TP completa; media humanista completa; superior técnica incompleta; universitaria incompleta; superior técnica completa; universitaria completa; postgrado | SIMCE |
| Escolaridad Madre | Variable categórica que indica el nivel de escolaridad máximo alcanzado por la madre. Mismas categorías que variable del padre. | SIMCE |
| Ingresos | Variable categórica que indica nivel de ingresos de la familia: menos de \$100.000; entre \$100.000 y \$200.000; entre \$200.000 y \$300.000; entre \$300.000 y 400.000; entre \$400.000 y \$500.000; entre \$500.000 y 600.000; entre \$600.000 y \$800.000; entre \$800.000 y \$1.000.000; entre \$1.000.000 y \$1.200.000; entre \$1.200.000 y \$1.400.000; entre \$1.400.000 y \$1.600.000; entre \$1.600.000 y \$1.800.000; entre \$1.800.000 y \$2.000.000; entre \$2.000.000 y \$2.200.000; más de \$2.200.000 | SIMCE |
| Expectativas Padres | Mayor nivel educativo del estudiante esperado por los padres: no completará IV medio; IV medio técnico profesional; IV medio humanista; superior técnico; universidad; postgrado. | SIMCE |
| Profesores | | |
| Profesor con Título | Dummy que indica si el profesor tiene un título en pedagogía. | SIMCE |
| Profesor con Postgrado | Dummy que indica si el profesores tiene algún estudio de postgrado. | SIMCE |
| Colegio | | |
| Nº de Profesores | Número de profesores contratados por el colegio | MINEDUC |
| Horas Profesores | Cantidad de horas de profesores contratadas por el colegio | MINEDUC |

| | | |
|------------------------|--|---------|
| Nº de Asistentes | Número de asistentes de la educación contratados por el colegio | MINEDUC |
| Horas Asistentes | Cantidad de horas de asistentes de la educación contratadas por el colegio | MINEDUC |
| % de Aprobados | Porcentaje de alumnos que aprobaron en el establecimiento y periodo del alumno | MINEDUC |
| Matrícula Colegio | Número de alumnos matriculados en el colegio del alumno. | MINEDUC |
| Tamaño del curso | Número de alumnos en el curso del alumno | MINEDUC |
| <i>Pares</i> | | |
| Carac. Socioeconómicas | Promedio de las variables socioeconómicas de los pares del alumno. La definición de pares es la misma utilizada en el cálculo de la proporción de mujeres. | SIMCE |
| Calidad de los pares | Promedio del puntaje de los pares en el simce rezagado. Para 4to, el rezago es en 2do, para 6to el rezago es en 4to, y para II medio el rezago es en 8vo. La definición de pares es la misma utilizada en el cálculo de la proporción de mujeres. Solo se utiliza en el modelo ampliado. | SIMCE |

Anexo 4: Distribución de titulación STEM, según sexo del alumno y porcentaje de mujeres en IV medio



Elaboración propia en base a datos del MINEDUC. Se excluyen colegios no mixtos.

Anexo 5: Heterogeneidad según dependencia del colegio - PSU

| | Pares: Cohorte | | Pares: Curso | |
|-----------------------|-------------------------------------|------------------------------------|------------------------------------|-------------------------------------|
| | Hombres | Mujeres | Hombres | Mujeres |
| a. Lenguaje | | | | |
| Municipal | -0.000046 (0.00069) [103131] | 0.000075 (0.00061) [128559] | 0.0019*** (0.00023) [101971] | -0.0016*** (0.00020) [127324] |
| Subvencionado | -0.0013*** (0.00044) [184405] | -0.00047 (0.00038) [227267] | 0.0017*** (0.00026) [184262] | -0.0017*** (0.00020) [227208] |
| Privado | -0.0010 (0.00089) [37922] | -0.0011 (0.00077) [39922] | -0.00077 (0.00068) [37765] | -0.00068 (0.00059) [39733] |
| b. Matemáticas | | | | |
| Municipal | 0.00062 (0.00070) [102207] | -0.000054 (0.00060) [127281] | 0.0016*** (0.00026) [101127] | 0.00022 (0.00021) [126170] |
| Subvencionado | 0.00017 (0.00043) [183500] | 0.00077** (0.00035) [226053] | 0.0019*** (0.00024) [183340] | 0.00042** (0.00018) [225964] |
| Privado | 0.00014 (0.00087) [37885] | 0.00069 (0.00078) [39877] | -0.0025*** (0.00065) [37784] | -0.0012** (0.00059) [39757] |

Coefficientes del efecto sobre el puntaje PSU, en desviaciones estándar, debido a un aumento de 1pp en la proporción de mujeres, diferenciando por sexo del alumno y la asignatura evaluada. Se divide la muestra según la dependencia del colegio de origen. Las primeras dos columnas miden el porcentaje de mujeres en la cohorte dentro del colegio, mientras que las últimas dos lo miden en el curso. Se utiliza el modelo más completo para cada estimación. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. Número de observaciones se encuentra entre corchetes. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

Anexo 6: Heterogeneidad según escolaridad de la madre - Titulación STEM

| | Pares: Cohorte | | Pares: Curso | |
|----------------------------------|------------------------------------|------------------------------------|-------------------------------------|--------------------------------------|
| | Hombres | Mujeres | Hombres | Mujeres |
| <i>a. Matrícula STEM</i> | | | | |
| Padre sin educación superior | -0.00021 (0.00019) [267036] | -0.00012 (0.00011) [323660] | -0.0046*** (0.00015) [267036] | -0.0015*** (0.000098) [323660] |
| Padre con educación superior | 0.0000062 (0.00023) [150486] | -0.000023 (0.00017) [152035] | -0.0021*** (0.00019) [150486] | -0.00088*** (0.00013) [152035] |
| <i>b. Titulación STEM</i> | | | | |
| Padre sin educación superior | 0.000092 (0.00045) [66306] | -0.00024 (0.00015) [109808] | -0.0058*** (0.00023) [66306] | -0.0011*** (0.00010) [109808] |
| Padre con educación superior | -0.00043 (0.00063) [28142] | -0.00060** (0.00029) [40188] | -0.0032*** (0.00039) [28142] | -0.00064*** (0.00019) [40188] |

Coefficientes del efecto de un aumento de 1pp en la proporción de mujeres sobre la prob. de matricularse o titularse de carreras STEM, diferenciando por sexo del alumno. Se divide entre alumnos cuyos padres asistieron a educación superior y aquellos que no. Las primeras dos columnas miden el porcentaje de mujeres en la cohorte dentro del colegio, mientras que las últimas dos lo miden en el curso. Se utiliza el modelo más completo para cada estimación. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. Número de observaciones se encuentra entre corchetes. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

Anexo 7: Heterogeneidad según escolaridad de la madre - PSU

| | Pares: Cohorte | | Pares: Curso | |
|------------------------------|-------------------------------------|-----------------------------------|------------------------------------|-------------------------------------|
| | Hombres | Mujeres | Hombres | Mujeres |
| a. Lenguaje | | | | |
| Madre sin educación superior | -0.00089** (0.00041) [224153] | -0.00030 (0.00036) [281650] | 0.0020*** (0.00018) [222956] | -0.0017*** (0.00015) [280436] |
| Madre con educación superior | -0.0013** (0.00059) [101233] | -0.00036 (0.00052) [114010] | 0.00071** (0.00033) [100969] | -0.0011*** (0.00029) [113743] |
| b. Matemáticas | | | | |
| Madre sin educación superior | 0.00028 (0.00039) [222527] | 0.00046 (0.00034) [279374] | 0.0019*** (0.00018) [221390] | 0.00034** (0.00013) [278255] |
| Madre con educación superior | 0.00015 (0.00059) [100996] | 0.00095* (0.00049) [113746] | -0.000029 (0.00034) [100791] | -0.00035 (0.00030) [113547] |

Coefficientes del efecto sobre el puntaje PSU, en desviaciones estándar, debido a un aumento de 1pp en la proporción de mujeres, diferenciando por sexo del alumno y la asignatura evaluada. Se divide la muestra entre alumnos cuyas madres asistieron a educación superior y aquellos que no. Las primeras dos columnas miden el porcentaje de mujeres en la cohorte dentro del colegio, mientras que las últimas dos lo miden en el curso. Se utiliza el modelo más completo para cada estimación. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. Número de observaciones se encuentra entre corchetes. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

Anexo 8: Heterogeneidad según escolaridad del padre - PSU

| | Pares: Cohorte | | Pares: Curso | |
|------------------------------|-------------------------------------|-----------------------------------|------------------------------------|--------------------------------------|
| | Hombres | Mujeres | Hombres | Mujeres |
| Lenguaje | | | | |
| <i>a. Lenguaje</i> | | | | |
| Padre sin educación superior | -0.00063 (0.00043) [220227] | -0.00032 (0.00036) [276679] | 0.0020*** (0.00018) [219026] | -0.0017*** (0.00015) [275458] |
| Padre con educación superior | -0.0015*** (0.00059) [105121] | -0.00042 (0.00050) [118955] | 0.00049 (0.00034) [104862] | -0.00077*** (0.00030) [118692] |
| <i>b. Matemáticas</i> | | | | |
| Padre sin educación superior | 0.00059 (0.00041) [218592] | 0.00044 (0.00034) [274424] | 0.0019*** (0.00017) [217454] | 0.00034** (0.00013) [273297] |
| Padre con educación superior | -0.00024 (0.00059) [104891] | 0.0011** (0.00048) [118669] | -0.00017 (0.00035) [104688] | -0.00026 (0.00030) [118475] |

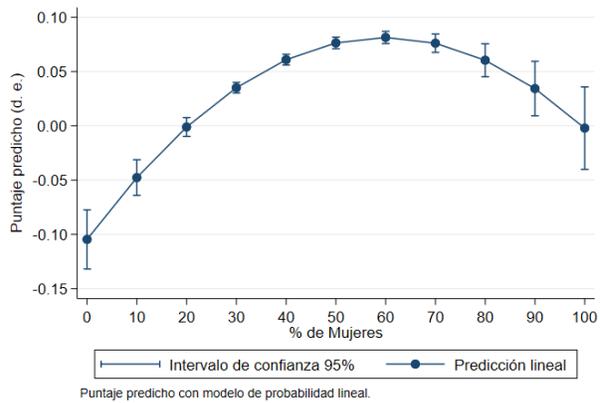
Coeficientes del efecto sobre el puntaje PSU, en desviaciones estándar, debido a un aumento de 1pp en la proporción de mujeres, diferenciando por sexo del alumno y la asignatura evaluada. Se divide la muestra entre alumnos cuyos padres asistieron a educación superior y aquellos que no. Las primeras dos columnas miden el porcentaje de mujeres en la cohorte dentro del colegio, mientras que las últimas dos lo miden en el curso. Se utiliza el modelo más completo para cada estimación. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. Número de observaciones se encuentra entre corchetes. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

Anexo 9: Efecto par no lineal - puntaje PSU

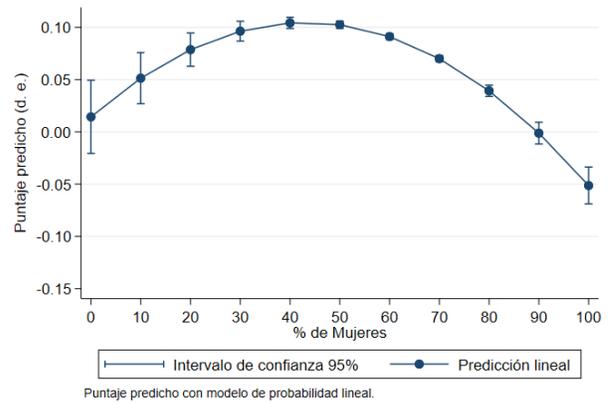
| | Pares: Cohorte | | Pares: Curso | |
|---------------------------|----------------|------------|--------------|--------------|
| | Hombres | Mujeres | Hombres | Mujeres |
| c. PSU Lenguaje | | | | |
| % Mujeres | -0.0030* | -0.0017 | 0.0062*** | 0.0042*** |
| | (0.0018) | (0.0016) | (0.00065) | (0.00061) |
| % Mujeres ² | 0.000020 | 0.000012 | -0.000052*** | -0.000048*** |
| | (0.000018) | (0.000015) | (0.0000068) | (0.0000051) |
| <i>N</i> | 325458 | 395748 | 323998 | 394265 |
| d. PSU Matemáticas | | | | |
| % Mujeres | -0.0024 | -0.00021 | 0.0052*** | 0.00021 |
| | (0.0016) | (0.0015) | (0.00055) | (0.00062) |
| % Mujeres ² | 0.000028* | 0.0000071 | -0.000043*** | 0.00000072 |
| | (0.000016) | (0.000014) | (0.0000060) | (0.0000051) |
| <i>N</i> | 323592 | 393211 | 322251 | 391891 |

Coeficientes se interpretan como el cambio en la probabilidad ante un aumento de 1pp en la variable indicada, diferenciando por el sexo del alumno. Las primeras dos columnas miden el porcentaje de mujeres en la cohorte dentro del colegio, mientras que las últimas dos lo miden en el curso. Se utiliza el modelo más completo para cada estimación. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$

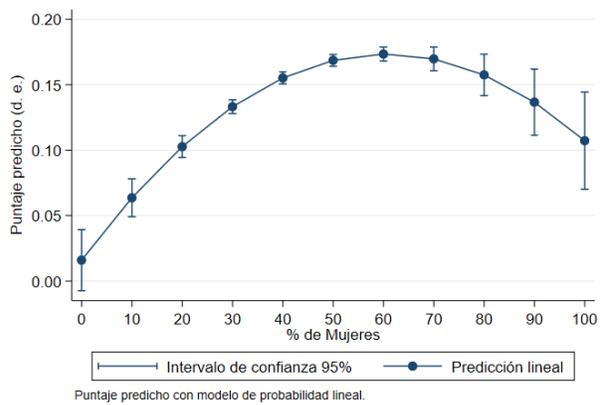
Anexo 10: Puntaje PSU predicho - efecto par de género cuadrático



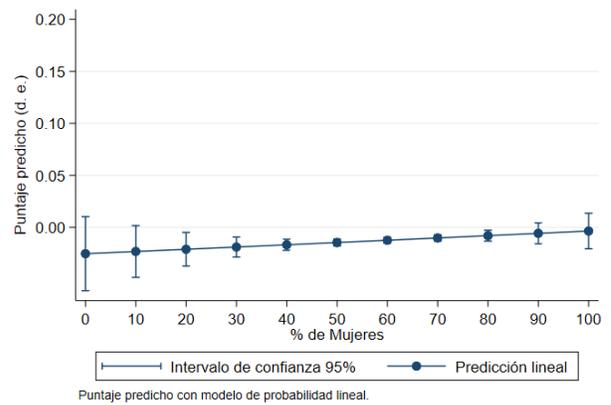
(a) Lenguaje: hombres



(b) Lenguaje: mujeres



(c) Matemáticas: hombres



(d) Matemáticas: mujeres

Anexo 11: Efecto par de género - robustez ante cambios en el curso donde se mide el porcentaje de mujeres

| | Hombres | | | | Mujeres | | | |
|---------------------------|------------------------|------------------------|------------------------|------------------------|-------------------------|-------------------------|-------------------------|------------------------|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| a. PSU Lenguaje | | | | | | | | |
| % Mujeres IV Medio | 0.0018*** (0.00018) | | | | -0.0016*** (0.00014) | | | |
| % Mujeres III Medio | | 0.0019*** (0.00018) | | | | -0.0014*** (0.00014) | | |
| % Mujeres II Medio | | | 0.0012*** (0.00022) | | | | 0.00082*** (0.00022) | |
| % Mujeres I Medio | | | | 0.0017*** (0.00024) | | | | 0.0015*** (0.00024) |
| <i>N</i> | 323998 | 319793 | 309651 | 297285 | 394265 | 390049 | 381160 | 368836 |
| b. PSU Matemáticas | | | | | | | | |
| % Mujeres IV Medio | 0.0016*** (0.00018) | | | | 0.00022* (0.00013) | | | |
| % Mujeres III Medio | | 0.0015*** (0.00019) | | | | 0.00012 (0.00013) | | |
| % Mujeres II Medio | | | 0.0019*** (0.00020) | | | | 0.0014*** (0.00017) | |
| % Mujeres I Medio | | | | 0.0020*** (0.00021) | | | | 0.0015*** (0.00019) |
| <i>N</i> | 322251 | 318100 | 308066 | 295855 | 391891 | 387712 | 378965 | 366790 |

Coefficientes se interpretan como el cambio en desviaciones estándar del puntaje PSU ante un aumento de 1pp en la proporción de mujeres en el curso indicado, diferenciando por el sexo del alumno. Todos los modelos estimados miden el porcentaje de mujeres dentro del curso. Error estándar, corregido por conglomerados a nivel de colegio, se encuentra entre paréntesis. * $p < 0,1$, ** $p < 0,05$, *** $p < 0,01$