



FACULTAD DE
CIENCIAS ECONÓMICAS
Y DE ADMINISTRACIÓN

DEPARTAMENTO DE
ECONOMÍA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Preferencias heterogéneas y segregación escolar: el caso de la educación media en Uruguay

Ignacio Cabrera Bonino

Programa de Maestría en Economía de la Facultad de Ciencias Económicas y de Administración
Universidad de la República

Montevideo - Uruguay

Febrero de 2024

Preferencias heterogéneas y segregación escolar: el caso de la educación media en Uruguay

Ignacio Cabrera Bonino

Tesis de Maestría presentada al Programa de Maestría en Economía de la Facultad de Ciencias Económicas y de Administración, Universidad de la República, como parte de los requisitos para la obtención del título de Magíster en Economía.

Director de tesis:

Juan Pereyra Barreiro

Director académico:

Javier Alejo

Montevideo - Uruguay

Febrero de 2024

Agradecimientos

Agradezco a mi tutor Juan Pereyra por su apoyo y guía constante, en el proceso de elaboración de esta tesis y más allá. A Verónica Amarante y Carolina Román por sus valiosos aportes en el seminario de tesis. A Matías Faggetti por brindarme la geolocalización de los estudiantes y por los intercambios sobre el uso del paquete de estimación en R. A Rodrigo González por el procesamiento de los datos de denuncias de delitos. Al Ministerio del Interior y en particular a Edgardo García por brindarme los datos de denuncias de delitos. A Andrés Peri de la Dirección Sectorial de Planificación Educativa, ANEP, por ayudarme a acceder a los datos de secundaria y UTU. A Maximiliano Machado y Sebastián Fleitas por sus valiosos comentarios y aportes en la definición del problema de investigación. A Guillermo Sánchez y María Eugenia Echeberría por su compañía y aportes. A la Universidad de Montevideo por brindarme el espacio para trabajar junto a mi tutor. Al IECON y a la Universidad de la República por ofrecerme el espacio y equipo para trabajar en esta tesis. Agradezco especialmente a Paula Carrasco, Andrés Dean y el equipo de economía laboral del IECON, por su enorme paciencia. A Paulo Somaini por guiarme en la comprensión de los métodos de estimación de preferencias en el contexto de elección escolar y por poner a disposición el acceso a un servidor para ejecutar las estimaciones. A Thilo Klein y Robert Aue, por desarrollar el paquete utilizado como base para la estimación y por su incansable disposición a intercambiar sobre el mismo. A Milton Silveira del Área de Gestión de Información de ANEP y Luana Méndez por poner a disposición los datos de la preinscripción y el algoritmo de asignación, así como por sus aportes para comprender el mecanismo. A Carolina García del Departamento de Estadística de DGES y Damián Vales del Programa de Gestión de Datos y Tecnología de DGETP por poner a disposición los datos de liceos y UTU, así como su disposición para ayudarme a interpretar y procesar estos datos. A mis amigos, mi familia y a Maru por la paciencia, el apoyo y el amor.

Este trabajo fue realizado en el marco del Proyecto de investigación Aplicada Fondo María Viñas FMV-1-2021-1-166576.

Preferencias heterogéneas y segregación escolar: el caso de la educación media en Uruguay

Ignacio Cabrera Bonino*

Resumen

Uno de los objetivos de los sistemas centralizados de elección escolar es reducir la desigualdad en el acceso a los centros educativos de calidad. En Uruguay, la asignación de estudiantes a centros de educación media se realiza mediante un sistema centralizado que considera las preferencias de las familias y prioriza a los estudiantes más vulnerables en todos los centros educativos, más allá de su lugar de residencia. Sin embargo, el acceso a centros de calidad es más desigual que si la asignación dependiera únicamente de la proximidad. El objetivo de este trabajo es conocer en qué medida la heterogeneidad en las preferencias de las familias por características de los centros educativos explica la segregación y el acceso desigual a los centros de mayor calidad. Para esto, se estiman las preferencias por calidad utilizando los rankings enviados por las familias en un procedimiento bayesiano de aumento de datos. Las estimaciones muestran que los estudiantes más vulnerables valoran en menor medida los indicadores de calidad. El análisis de escenarios contrafactuales muestra que si las preferencias por calidad fueran homogéneas, tomarlas en cuenta para asignar a los estudiantes no aumentaría la segregación en centros de distinta calidad con respecto a una asignación por proximidad, aunque tampoco la reduciría. Además, aumentaría el nivel agregado de estudiantes asignados a centros de mayor calidad.

Palabras clave: educación; elección escolar; desigualdad; preferencias heterogéneas; modelos de elección discreta.

JEL: C78, I24, I28.

Abstract

One of the main goals of centralized school choice systems is to reduce unequal access to high-quality schools. In Uruguay, students are assigned to public secondary schools using a centralized system that takes into account families' preferences and prioritizes the most vulnerable students in every school, regardless of their place of residence. However, access to quality schools is more unequal than if the assignment depended solely on proximity. This study asks to what extent the heterogeneity in families' preferences explains segregation and unequal access to higher-quality schools. To this end, preferences for quality are estimated using data on the rankings submitted by families in a bayesian data augmentation procedure. Estimates show that the most vulnerable students value quality indicators less than non-vulnerable students. Counterfactual analysis shows that if preferences for quality were homogeneous, taking them into account in a centralized mechanism would not increase proximity assignment's segregation in terms of school quality, although it would not reduce it either. Also, more students would be assigned to high-quality schools overall.

Keywords: education; school choice; inequality; heterogeneous preferences; discrete choice models.

JEL: C78, I24, I28.

*IECON, Universidad de la República, Uruguay. Email: ignacio.cabrera@fcea.edu.uy

Índice

1. Introducción	1
2. Antecedentes	5
3. Breve descripción del sistema educativo uruguayo	9
3.1. Asignación a centros de educación media	10
4. Marco teórico	22
4.1. Modelos de elección discreta	22
4.2. Un modelo de comportamiento para el caso uruguayo	24
4.3. Preferencias heterogéneas y el rol de la información	27
5. Estrategia empírica	29
5.1. Procedimiento de estimación	29
5.2. Especificación del modelo	31
5.3. Identificación e interpretación de los parámetros.	35
5.4. Escenarios contrafactuales	37
6. Resultados	40
6.1. Heterogeneidad en las preferencias	40
6.2. Segregación	42
6.3. Dinámica de los cambios	44
7. Comentarios finales	47
Referencias	49
A. Anexo	53
A.1. Preparación de los datos	53
A.2. Construcción del índice de vulnerabilidad	56
A.3. Características de los centros	58
A.4. Análisis de convergencia	61
A.5. Robustez	65

1. Introducción

Los sistemas centralizados de elección y asignación a centros educativos han sido ampliamente adoptados por los países en las últimas décadas (Neilson, 2019). Estos sistemas generan asignaciones a partir de tres insumos que caracterizan la demanda y oferta en el mercado. Desde el punto de vista de la demanda, se recogen las preferencias de los padres pidiéndoles que presenten un ranking de centros educativos preferidos. Desde el punto de vista de la oferta, la asignación tiene en cuenta las capacidades de los centros educativos y las prioridades (si las hay) sobre los estudiantes, que son decididas por las autoridades. Esto contrasta con los mecanismos tradicionales que asignaban a los estudiantes al centro más cercano a su hogar. La asignación tradicional puede ser criticada por motivos de igualdad de oportunidades de acceso y de segregación, ya que las familias en mejor situación socioeconómica tienen mayor capacidad de elegir dónde vivir de forma de asegurar que sus hijos sean asignados a los centros públicos de mejor calidad, mientras que las familias de menor nivel socioeconómico no. En los mecanismos de asignación centralizada, los gobiernos pueden incluir sus objetivos y preferencias de política mediante la definición de prioridades. El objetivo principal de estas políticas es disminuir la segregación educativa producto de la segregación residencial. La eficacia de los sistemas centralizados para alcanzar este objetivo depende de las preferencias heterogéneas de los padres y los estudiantes por los centros (Laverde, 2022). Por ejemplo, la desigualdad en el acceso a los centros de alta calidad difícilmente se superaría sólo con el mecanismo de asignación si las familias desfavorecidas valoran más la distancia que la calidad y los mejores centros están concentrados geográficamente, o tienen restricciones para acceder a la información sobre la calidad de los centros, aún cuando se les dé prioridad en la elección.

En Uruguay, aumentar la finalización de la educación media es uno de los principales objetivos de las autoridades educativas, siendo que la tasa es baja en la comparación internacional y a la vez altamente desigual (INEEd, 2023). La decisión de abandono del sistema educativo está, entre otras cosas, relacionada con la calidad del centro educativo al que asiste un estudiante (Hanushek *et al.*, 2008), por lo que garantizar el acceso igualitario a centros de calidad es un aspecto relevante para los objetivos de política. La *Administración Nacional de Educación Pública* (ANEP) introdujo en 2018 un mecanismo centralizado de elección para asignar a los estudiantes desde las escuelas primarias a los centros de educación media. A grandes rasgos, el mecanismo de la ANEP requiere que cada estudiante que desee inscribirse en el sistema público de educación media (ya sea en un liceo o en un centro técnico-profesional) presente un ranking de sus tres opciones preferidas a modo de preinscripción. A continuación, un algoritmo calcula la asignación de estudiantes teniendo en cuenta las preferencias, las capacidades de los

centros y las puntuaciones de prioridad de los estudiantes. De manera singular, estas prioridades se definen exclusivamente con un índice de “vulnerabilidad” que es creciente en ciertos indicadores de pobreza y decreciente en el rendimiento en la escuela primaria. Las prioridades son las mismas en todos los centros, no teniéndose en cuenta la ubicación de los estudiantes. De esta forma, el diseño del mecanismo favorece en todos los centros a los estudiantes más vulnerables. Aunque el objetivo de esta prioridad no fue explicitado por las autoridades, puede interpretarse como una forma de evitar que el lugar de residencia condicione el acceso a centros educativos de calidad a los alumnos más propensos a abandonar el sistema educativo.

Sin embargo, este trabajo muestra que la introducción del mecanismo aumenta la segregación si se toma como referencia un sistema que asigne únicamente por cercanía. En particular, la segregación en centros clasificados según sus resultados educativos se duplica. Esto sucede porque cuando los estudiantes menos vulnerables pueden expresar sus preferencias por calidad pasan a centros de mayor calidad, mientras la asignación de estudiantes vulnerables es similar. Comprender las razones de este resultado es relevante para el diseño de política educativa. Podría suceder que las familias más vulnerables no cuenten con información sobre la calidad de los centros educativos, su relevancia para los resultados educativos o los retornos de la educación, y por lo tanto elijan centros de menor calidad. Esto señalaría que difundir información sobre la calidad de los centros y sobre la relevancia de los mismos sobre los resultados educativos e ingresos futuros es una opción de política para reducir las brechas en el acceso a la educación. Además, resulta interesante evaluar qué potencial tendría la priorización de los estudiantes vulnerables para reducir la desigualdad en el acceso a centros de calidad si las preferencias fueran homogéneas entre familias.

El objetivo de este trabajo es conocer en qué medida la heterogeneidad en las preferencias de las familias por características de los centros educativos explica la segregación y el acceso desigual a los centros de mayor calidad. Para esto, es necesario estimar las preferencias por calidad considerando la posibilidad de que sean heterogéneas entre grupos y considerar el escenario contrafactual en el que las preferencias fueran homogéneas.

La introducción del mecanismo centralizado de elección permite acceder a una medida revelada de estas preferencias, contenida en los rankings que envían en forma de preinscripción. En este trabajo se utilizan las preferencias declaradas por las familias para estimar un modelo estructural de demanda por características de los centros de educación media como la distancia, la calidad educativa, los recursos disponibles, la composición de los compañeros y la seguridad en el entorno, distinguiendo las preferencias de las familias según sean beneficiarias de la *Tarjeta Uruguay Social* (TUS), un programa de transfe-

rencias no contributivas que busca atender a los hogares en situación de vulnerabilidad socioeconómica extrema.

El modelo desarrollado explicita en qué medida las familias tienen incentivos para enviar rankings que revelen sus verdaderas preferencias, a partir de las propiedades teóricas del mecanismo de asignación utilizado por ANEP y de la información disponible para las familias. Los parámetros del modelo se estimaron utilizando un procedimiento de aumento de datos bayesiano basado en [McCulloch y Rossi \(1994\)](#). Se asume que la utilidad de un estudiante por un centro es una variable latente de la decisión de rankearlo en determinada posición. De esta forma, los rankings de centros enviados por las familias se utilizan a la luz de los supuestos de comportamiento para definir truncamientos en la distribución posterior de la utilidad de cada centro. Asumiendo una forma funcional de la distribuciones de la utilidad y los parámetros, se utiliza la técnica de *Gibbs Sampling* para tomar simulaciones de forma iterativa de la distribución posterior de la utilidad dados los parámetros y simulaciones de la distribución posterior de los parámetros dada la utilidad. Las simulaciones de los parámetros tomadas en cada iteración forman Cadenas de Markov, y el promedio de cada cadena es asintóticamente equivalente al estimador de máxima verosimilitud ([Van der Vaart, 2000](#)).

Para la estimación se construyó una base de datos combinando diversas fuentes. Los datos de la preinscripción incluyen los rankings que envían las familias y el resultado de la asignación, así como la dirección del estudiante, sus resultados en la escuela primaria y los planes sociales que recibe. Adicionalmente, se obtuvo información sobre la plantilla docente y los resultados educativos de los centros de educación media. Estos datos provienen de forma separada de la dirección de educación secundaria y educación técnico-profesional, por lo que se realizó un proceso de homogeneización de ambas bases. Por último, se obtuvieron datos de denuncias de delitos georreferenciadas por parte del Ministerio del Interior, con lo que se identifica la criminalidad en el entorno de los centros educativos. Más allá de su uso en este trabajo, el conjunto de datos que se obtuvo y procesó ofrece un enorme potencial de uso para futuras investigaciones.

Las preferencias estimadas muestran que los estudiantes más vulnerables valoran en menor medida las características de los centros relacionadas con los resultados educativos y los recursos de los centros, una vez que se controla la distancia, la composición de pares y la criminalidad en el entorno. Este resultado es robusto a la definición del grupo de estudiantes vulnerables y a la especificación del modelo.

En una segunda etapa, se utilizan las preferencias estimadas para construir un escenario contrafactual en el que las preferencias por calidad sean homogéneas entre grupos. Los resultados de este ejercicio contrafactual muestran que todo el aumento de la desigualdad en el acceso a centros de calidad intro-

ducida por el mecanismo desaparecería si las preferencias de los estudiantes fueran homogéneas. Aún más, se encuentra que al homogeneizar las preferencias por calidad se alcanzaría un aumento agregado en el acceso a centros de mayor calidad con respecto a la asignación actual, aunque implicaría un costo para algunos estudiantes no vulnerables. Esto se explica porque los centros identificados como de mayor calidad tienen espacio disponible en la asignación actual. Sin embargo, aún cuando las preferencias fueran homogéneas y se priorizara a los estudiantes más vulnerables, el uso del mecanismo centralizado no lograría reducir la segregación que existe por la distribución geográfica de los centros de calidad. Este escenario podría reflejar el resultado potencial de políticas que intervengan la información de las familias sobre la calidad de los centros así como su relevancia para la trayectoria educativa y los ingresos futuros.

Los resultados de este trabajo deben interpretarse como una estimación de los efectos parciales de eliminar la heterogeneidad en las preferencias, y no como resultados de equilibrio general. El modelo desarrollado no permite evaluar cómo cambiaría la calidad de los centros educativos ante la presión competitiva causada por el aumento de la demanda por calidad de los estudiantes, ni ante los cambios en la composición estudiantil. Tampoco permite evaluar en qué medida cambiaría la elección de las familias entre centros públicos y privados. Estas cuestiones podrían abordarse en trabajos posteriores que incorporen un modelo de la oferta por parte de los centros educativos y amplíen el modelo de demanda para considerar también la decisión entre el sistema público y privado.

Explorar los efectos de la segregación residencial sobre el acceso igualitario a los centros de calidad es una opción interesante para investigaciones futuras. Adicionalmente, la singular definición de prioridades de la ANEP, junto con la falta de información sobre el mecanismo, presentan un escenario ideal para un experimento de suministro de información destinado a estimar los efectos de las creencias de aceptación en los ranking enviados por los padres, así como los efectos de la información sobre los centros en las preferencias. En particular, la intervención controlada de la información de las familias permitiría distinguir el efecto de la información sobre las características de los centros del efecto de la información sobre los retornos educativos y la relevancia de la calidad de los centros, algo que no puede distinguirse en este trabajo. El modelo estructural de demanda desarrollado puede utilizarse para estos fines.

2. Antecedentes

Este trabajo está relacionado con varios grupos de literatura que giran en torno a los mecanismos de elección de escuela (*school choice*). Un primer grupo proviene de la teoría de matching y diseño de mercados y se centra en las propiedades teóricas de los mecanismos de asignación. Estos trabajos aportan fundamentos teóricos sobre los algoritmos de asignación, así como pruebas empíricas de las propiedades cuando la teoría no otorga respuestas concluyentes. [Gale y Shapley \(1962\)](#) y [Abdulkadiroğlu y Sönmez \(2003\)](#) fueron algunos de los primeros en aplicar los desarrollos del diseño de mecanismos a la asignación de estudiantes. El primero se enfoca en las admisiones de universidades y propone el mecanismo de aceptación diferida (*Deferred Acceptance*), demostrando teóricamente sus propiedades de estabilidad y optimalidad de Pareto dentro del conjunto de algoritmos estables. El segundo profundiza esta línea relevando los principales algoritmos de asignación escolar utilizados en situaciones reales y desarrollando sus propiedades teóricas en términos de estabilidad, eficiencia y manipulabilidad.

[Pathak y Sönmez \(2008\)](#) es otro ejemplo de aportes teóricos sobre el funcionamiento de los algoritmos de asignación, centrado en las diferencias entre mecanismos manipulables y no manipulables. En particular, demuestra que la diferente capacidad que tienen los padres de manipular los mecanismos que no son *strategy proof* favorece a los padres más *sofisticados*. [Abdulkadiroglu et al. \(2006\)](#) provee evidencia empírica del comportamiento estratégico heterogéneo utilizando datos de la asignación escolar en Boston. [Abdulkadiroglu y Sönmez \(2013\)](#) presenta una revisión de los aportes teóricos en torno a los mecanismos de asignación, resumiendo las principales propiedades de los mismos. En esta tesis, conocer las propiedades del algoritmo de asignación permite tomar supuestos sobre la estrategización de los padres al realizar los rankings y de esa forma inferir correctamente las preferencias reales.

Una discusión adicional se centra en los resultados empíricos y teóricos de las políticas de acción afirmativa integradas en los mecanismos de elección escolar, en lo que se denomina *controlled school choice*. [Kojima \(2012\)](#) muestra que la aplicación de este tipo de políticas en Nueva York y Boston llevó a pérdidas de bienestar para los estudiantes minoritarios. Además, demuestra teóricamente que este tipo de políticas (ya sea con cuotas u otorgando prioridades) no son compatibles con la propiedad de estabilidad de los mecanismos. Por otro lado, [Echenique y Yenmez \(2015\)](#) discute estos resultados desarrollando teóricamente mecanismos que incluyen preferencias por diversidad y generan asignaciones estables. Este trabajo proporciona un análisis empírico novedoso de un mecanismo único caracterizado por el criterio de priorización, con objetivos de equidad similares a los de las reservas y las cuotas, mostrando que la heterogeneidad en las preferencias de las familias juega un rol importante en los objetivos de diversidad

de las autoridades.

Un conjunto diferente de literatura se centra en la estimación de la demanda de los padres por las características de las escuelas. Los primeros trabajos en este grupo surgieron de la economía de la educación e intentaron identificar la valoración de los padres por la calidad de los centros educativos a través del efecto de las características de estos en el precio de los inmuebles cercanos. [Black \(1999\)](#) explota la variación en los precios de los inmuebles en los límites geográficos que determinan la asignación a una escuela u otra, encontrando un efecto positivo de la calidad de la escuela sobre el precio de los inmuebles. [Black y Machin \(2011\)](#) revisa los hallazgos y alternativas metodológicas utilizadas en esta literatura, mostrando cierto nivel de consenso en el efecto de la calidad escolar en el precio de los inmuebles.

Con la introducción de los mecanismos de elección escolar, comenzaron a aplicarse desarrollos de la literatura de organización industrial que aprovechan las preferencias declaradas por los padres para estimar la demanda por características de centros educativos. [Hastings et al. \(2005\)](#) aparece como una referencia central en esta área. Estima las preferencias utilizando un modelo *mixed multinomial logit* (MMNL) y los rankings enviados por los padres al inscribir a sus hijos en el distrito de Charlotte-Mecklenburg. Es el primer trabajo en mostrar empíricamente la heterogeneidad en las preferencias de los padres por calidad y distancia. Más recientemente, [Allende \(2019\)](#) desarrolla un modelo de equilibrio entre oferta y demanda educativa para la educación primaria en Perú. Para la estimación de la demanda propone un modelo que incorpora las preferencias por pares, encontrando que las familias más ricas valoran la composición de pares en mayor medida y que valoran negativamente la presencia de estudiantes pobres. [Abdulkadiroğlu et al. \(2020\)](#) también estima la demanda de los padres por la composición de los pares y la calidad académica dentro de un mecanismo de elección escolar, encontrando que la última no es relevante para los padres una vez que se considera la composición de pares.

Una parte de esta literatura se ha concentrado en integrar los desarrollos teóricos de la teoría de *matching* al análisis de demanda, desarrollando modelos estructurales que incorporan la estrategización de los padres al rankear escuelas. [Agarwal y Somaini \(2018\)](#), [Agarwal et al. \(2021\)](#), [Fack et al. \(2019\)](#) y [Ortega et al. \(2020\)](#) proponen modelos estructurales para estimar las preferencias usando datos provenientes de mecanismos de asignación manipulables, demostrando la identificación de los parámetros en base a supuestos sobre las propiedades del algoritmo. [Agarwal y Somaini \(2020\)](#) presentan una guía unificada para el análisis de preferencia revelada bajo diferentes propiedades teóricas de los algoritmos y supuestos sobre el comportamiento de los padres. Estos trabajos son incorporados como referencia para la estimación de preferencias en el contexto del algoritmo de asignación de ANEP.

Existen algunos trabajos que estiman las preferencias por centros educativos para comprender su

relación con la segregación. [Laverde \(2022\)](#) estima las preferencias de los padres y muestra que la concentración geográfica de las escuelas prekindergarten de mejores resultados explica una mayor parte de la segregación étnica de los estudiantes que la heterogeneidad en las preferencias. Para esto genera escenarios contrafactuales imponiendo a las familias de grupos minoritarios las preferencias del resto y modificando la ubicación de las familias. [Oosterbeek et al. \(2021\)](#) analiza el caso de la elección de escuela secundaria en Ámsterdam, donde muestran una alta segregación étnica y por ingresos. Encuentran que en ese caso la heterogeneidad en las preferencias es efectivamente el principal factor que explica la segregación.

[Hastings y Weinstein \(2008\)](#) y [Allende et al. \(2019\)](#) se centran en los efectos de la información sobre las preferencias de los padres. [Hastings y Weinstein \(2008\)](#) realiza un experimento otorgando información sobre la calidad de las escuelas a familias de bajos ingresos y encuentra que las familias tratadas eligen centros de mejor calidad. [Allende et al. \(2019\)](#) analizan los efectos sobre la elección escolar de una política que intervino la información sobre la calidad de los centros e incluyó una sensibilización sobre la importancia de la elección escolar y de los esfuerzos de búsqueda de centros educativos para los ingresos futuros. Adicionalmente, incorpora las estimaciones del efecto de la información que surgen del experimento en un modelo estructural que permite predecir los efectos de reescalar el experimento a todo el sistema. Si bien en este trabajo no se realizan intervenciones sobre la información de las familias, estos trabajos otorgan una hipótesis plausible para la interpretación de los resultados que está asociada a una recomendación de política. Adicionalmente, la particularidad de que el mecanismo de ANEP priorice explícitamente a los estudiantes con peores resultados educativos previos supone un escenario interesante para analizar los efectos de intervenir la información de sus familias sobre su elección y sus resultados educativos futuros. Esto podrá realizarse en un trabajo posterior que aproveche el modelo estructural desarrollado en este trabajo.

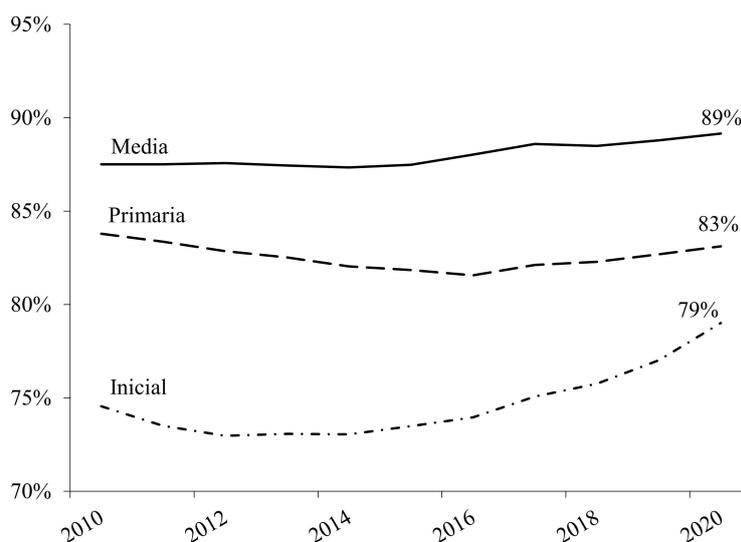
En Uruguay, los determinantes de las trayectorias educativas han sido estudiados ampliamente. [de Melo y Machado \(2018\)](#) estudia los determinantes del abandono en educación media usando datos de panel, mostrando por ejemplo la correlación del abandono con el contexto socioeconómico desfavorable y los malos resultados educativos en primaria. Además, encuentran que la experiencia y dedicación horaria del director de la escuela reduce la probabilidad de abandono posterior. [Failache et al. \(2018\)](#) muestra los efectos de la nutrición y situación socioeconómica en edades tempranas sobre la probabilidad de repetición y abandono en educación media, mostrando que son importantes predictores de ambos resultados. [Méndez-Errico y Ramos \(2022\)](#) modela la trayectoria educativa como un conjunto de elecciones secuenciales para identificar la influencia de distintos factores sobre el nivel educativo alcanzado,

encontrando que los factores de largo aliento como la educación de los padres o el grupo étnico son los mayores determinantes. Si bien estos trabajos muestran cierto consenso en torno a los determinantes que se mencionan, este trabajo busca aportar a la discusión sobre las trayectorias educativas y el abandono en Uruguay en dos sentidos. Por un lado, la heterogeneidad en las preferencias por calidad de los centros educativos puede dar luz a uno de los mecanismos a través de los cuales el contexto socioeconómico determina la probabilidad de abandono del sistema educativo. Por otro lado, es una aproximación a medir el impacto de las características de los centros y las preferencias de las familias en las trayectorias educativas, algo que no ha sido estudiado aun para el caso uruguayo. Si se logra incorporar información sobre las trayectorias posteriores de los estudiantes que participan del sistema de preinscripción, puede explotarse la variación en la asignación a distintos centros para medir su efecto sobre los resultados posteriores.

3. Breve descripción del sistema educativo uruguayo

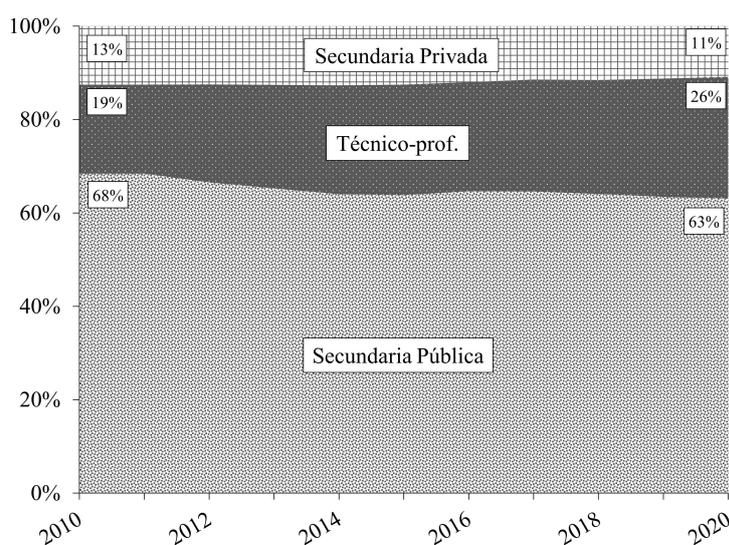
La educación en Uruguay es obligatoria en los niveles inicial, primario y medio. El nivel inicial comprende dos años a partir de los cuatro años, la educación primaria incluye seis años y la educación media se divide en dos ciclos de tres años. Dentro del sistema de educación media coexisten una vía académica y otra técnico-profesional, por lo que los alumnos pueden elegir entre escuelas secundarias (liceos) y escuelas técnico-profesionales (“UTU”) respectivamente. Todos los subsistemas, incluida la enseñanza superior, tienen oferta pública y ésta representa una porción mayoritaria de la matrícula. En el nivel primario, la provisión pública alcanza alrededor del 83% de la matrícula total, con una tendencia estable en la última década. En el nivel medio, el porcentaje asciende a un 89%. Esto incluye la porción pública de la matrícula de secundaria pública y el total de la vía técnico-profesional (UTUs), que es exclusivamente pública.

Figura 1: Provisión pública en el total de inscriptos (ANEP)



Elaboración propia con base en datos de ANEP.

Figura 2: Distribución de la matrícula de educación media



Elaboración propia con base en datos de ANEP.

Al interior de la educación media, puede observarse que la educación técnico-profesional ha ganado relevancia en la matrícula durante la última década. La secundaria privada ha reducido mínimamente su porción de la matrícula, con lo que el aumento en la matrícula de escuelas técnico-profesionales está acompañado principalmente de una menor porción de la matrícula total que elige la educación secundaria pública. Sin embargo, la matrícula de la educación secundaria pública no ha descendido sino que se ha mantenido estable en términos absolutos. El aumento en la relevancia de las escuelas técnico-profesionales está asociado a un aumento de la matrícula total de educación media que parece haberse canalizado hacia esta opción educativa.

La finalización de la escuela primaria es casi universal. En el nivel secundario, el aumento de los niveles de finalización es una de las principales metas del sistema educativo, al menos en los tres últimos períodos de gobierno¹. En 2022, sólo el 50,9% de los jóvenes entre 21 y 23 años completó la educación media. Los resultados también son fuertemente desiguales. Los jóvenes del primer quintil de ingresos la completaron en el 22,5% de los casos, mientras que los del quintil más alto alcanzaron el 82% (INEEd, 2023).

3.1. Asignación a centros de educación media

Algoritmo de asignación centralizada. Hasta 2018, la asignación de los estudiantes procedentes de la escuela primaria a los centros de enseñanza media estaba repartida entre la administración de los

¹Como se sostiene explícitamente en los “*Informes sobre el estado de la educación en Uruguay*” de 2021-2022, 2019-2020, 2015-2016 y 2014, publicados por el *Instituto Nacional de Evaluación Educativa* (INEED).

centros de enseñanza media de la vía académica (*Consejo de Educación Secundaria*) y la administración de los centros de enseñanza técnico-profesional (*Consejo de Educación Técnico Profesional*). No había criterios definidos ni transparentes para la asignación. Desde 2016 los estudiantes empezaron a matricularse presentando un ranking de 3 centros a modo de preinscripción, pero las asignaciones siguieron haciéndose por separado entre las vías académica y técnico-profesional, todavía sin criterios transparentes. Desde 2019 la asignación se realiza mediante un algoritmo centralizado que tiene en cuenta las preferencias declaradas por las familias y las prioridades definidas por las autoridades. Las preinscripciones se realizan en setiembre y las asignaciones se informan en diciembre del último año escolar. Este es un aspecto clave, siendo que en general las inscripciones a instituciones privadas se realizan antes de la publicación de los resultados de la asignación. De esta forma, es esperable que las familias que participan de la preinscripción ya tomaron la decisión de asistir a un centro público por sobre uno privado y puede asumirse que no quedar asignados implica no asistir a educación media. Como existe un único sistema para la vía académica y la técnico-profesional, los rankings pueden mezclar centros de ambos subsistemas. En el caso de los liceos, la opción elegida refiere al centro pero no al turno dentro del centro (matutino o vespertino). En el caso de las UTU, la opción refiere a un programa en específico, existiendo centros con más de un programa. En el contexto de este trabajo, se considera únicamente el Ciclo Básico Tecnológico (CBT) dentro de los programas de UTU, por ser el único que no tiene requerimientos de edad especiales. Un aspecto importante es que el egreso tanto del Ciclo Básico del liceo como del CBT habilitan a ingresar a la educación media superior sin distinciones.

El algoritmo asigna a los estudiantes en distintas etapas, asignando primero a los grupos considerados de mayor prioridad. Además, dentro de cada grupo se utilizan el índice de vulnerabilidad y la asignación de liceos predeterminados para definir prioridades sobre los estudiantes. El índice de vulnerabilidad se calcula teniendo en cuenta los indicadores de pobreza de los alumnos y su rendimiento en la escuela primaria. Se considera que los estudiantes son más vulnerables si tienen extraedad, más ausencias y peores calificaciones en los últimos tres años y reciben más prestaciones sociales (*Asignaciones Familiares y Tarjeta Uruguay Social*). Por otro lado, los liceos predeterminados se definen a nivel de escuela primaria (todos los alumnos de la misma escuela tienen el mismo centro predeterminado), pero el criterio no es transparente. Si un estudiante no pone su centro predeterminado entre las 3 opciones, su predeterminado se imputa como cuarta opción. No todos los estudiantes tienen un liceo predeterminado. Por último, en caso de necesitar desempatar entre estudiantes con el mismo nivel de prioridad, se prioriza según quién haya realizado antes la preinscripción.

Un aspecto relevante para el análisis de las preferencias es que los padres desconocen los detalles del

mecanismo. No se les informa del centro predeterminado ni de los criterios de priorización. Además, la implementación del algoritmo no fue difundida mediáticamente en la sociedad uruguaya². Las maestras de las escuelas primarias desempeñan un papel importante asesorando a los padres en el proceso. Los padres deben presentar su ranking a través de una plataforma en línea llamada “GURI” y en algunos casos lo hacen en la escuela con ayuda de la maestra o directora. De todas formas, las inscripciones ejecutadas por algún funcionario escolar se registran y no alcanzan el 1 % del total de preinscripciones en 2022.

A continuación se detallan los pasos del algoritmo.

Paso 1. Se consideran los estudiantes que tienen un Centro Educativo Asociado de UTU (CEA) asociado a la escuela primaria a la que concurrieron. Si el estudiante tiene como primera opción el CEA asociado a su escuela, se lo asigna allí sin considerar la capacidad del CEA.

Paso 2. Se consideran los estudiantes especiales ordenados por su índice de vulnerabilidad utilizando el algoritmo conocido en la literatura como *dictador serial*: se considera el primer estudiante en el orden de prioridad y se consideran sus opciones (desde la primera a la última) hasta asignarlo a una opción que tenga capacidad disponible. Si no hay capacidad disponible en alguna de sus opciones, se lo deja sin asignar en este paso. Luego, se considera el segundo estudiante, y así sucesivamente.

Paso 3. Se asignan los estudiantes de Programas Priorizados utilizando el *dictador serial* ordenando los estudiantes aleatoriamente ya que no cuentan con índice de vulnerabilidad.

Paso 4. Se asignan los estudiantes de escuelas rurales utilizando el *dictador serial* ordenando a los estudiantes por el índice de vulnerabilidad.

Paso 5. Se consideran los estudiantes no asignados y sus cuatro opciones (la cuarta opción refiere a liceo predeterminado en caso de tenerlo). Se los ordena según el índice de vulnerabilidad y se los asigna usando el *dictador serial*.

Paso 6 Se consideran los estudiantes no asignados que tengan un liceo predeterminado y se los asigna a su predeterminado sin considerar el sobrecupo.

Paso 7 Se consideran los estudiantes no asignados ordenados según valor del índice de vulnerabilidad. Se asigna cada estudiante a la opción rankeada con menor sobrecupo. Si el mínimo de sobrecupo se da en más de una opción, se lo asigna a la opción más preferida.

²Búsquedas de noticias en Google de “inscripciones liceo ANEP” e “inscripciones UTU” entre 2015 y 2022 solo muestran un resultado relacionado con la puesta en marcha del nuevo mecanismo, pero no incluye ningún detalle sobre su funcionamiento (<https://ladiaria.com.uy/articulo/2016/12/fenapes-denuncia-caos-en-las-inscripciones-a-secundaria-por-el-sistema-guri/>).

Cuadro 1: Estudiantes asignados por paso en 2022

Paso	No. estudiantes	% del total
1	707	1,7 %
2	2.185	5,2 %
3	397	0,9 %
4	2.043	4,8 %
5	35.426	84,1 %
6	1.312	3,1 %
7	62	0,1 %
Total	42.132	100 %

Elaboración propia con base en datos de ANEP.

Cuadro 2: Estudiantes asignados por número de opción

	2022		2021	
	No. estudiantes	% del total	No. estudiantes	% del total
Primera	38.464	91 %	36.467	88 %
Segunda	2.566	6 %	2.796	7 %
Tercera	1.060	3 %	2.029	5 %
Cuarta	42	<1 %	210	<1 %
Total	42.132	100 %	41.502	

	2020		2019	
	No. estudiantes	% del total	No. estudiantes	% del total
Primera	38.392	92 %	37.880	89 %
Segunda	2.268	5 %	3.053	7 %
Tercera	883	2 %	1.324	3 %
Cuarta	234	<1 %	300	<1 %
Total	41.777		42.557	

Elaboración propia con base en datos de ANEP.

Como se observa en el Cuadro 1, la mayor parte de los estudiantes son asignados en el paso 5. A la vez, solo un pequeño porcentaje de los estudiantes son asignados en los pasos 6 y 7. Además, el Cuadro 2 muestra que en todos los años donde se asignó usando el mecanismo centralizado, cerca del 90 % de los estudiantes fueron asignados a su primera opción y menos de un 1 % a su cuarta opción. Tomando en cuenta el desconocimiento que tienen los padres sobre el mecanismo de asignación, estos resultados sugieren que las familias que participan de la preinscripción podrían asumir una alta probabilidad de entrar a los centros que rankean, y en particular a su primera opción, a partir de las experiencias “exitosas” de otras familias.

Oferta: características de los centros. En 2022 se ranearon un total de 937 opciones: 276 liceos, 238 UTU, 194 instituciones privadas, 179 centros de educación no formal y 50 cursos de séptimo grado de primaria en escuelas rurales³. 6 opciones de liceos son en programas extraedad que admiten mayores de 15 años, y 102 programas de UTU son del programa de Formación Profesional Básica, que admite estudiantes mayores de 14 años. Además, 1 liceo tiene un plan especial para estudiantes hipoacústicos. A efectos del análisis en este trabajo se considerarán únicamente los 266 liceos y 137 UTU con planes que no cuentan con ningún requerimiento particular.

Cuadro 3: Oferta de liceos y UTU sin requerimientos especiales en preinscripción 2022

	Montevideo			Resto del país			Total país		
	Total	Liceos	UTU	Total	Liceos	UTU	Total	Liceos	UTU
Centros	84	58	26	319	208	111	403	266	137
Centros en cupo	50 %	52 %	46 %	31 %	29 %	35 %	35 %	34 %	37 %
Cupo									
Promedio	123	143	81	88	106	53	95	114	59
Acumulado	10.372	8.269	2.103	27.991	22.068	5.923	38.363	30.337	8.026
Sobredemanda									
Promedio (*)	24	29	14	6	6	4	9	11	6
Máximo	218	218	91	108	108	50	218	218	91
Acumulada	2.037	1.675	362	1.748	1.316	432	3.785	2.991	794

(*) Entre centros en cupo. Elaboración propia con base en datos de ANEP.

El Cuadro 3 presenta algunas estadísticas sobre la oferta de liceos y UTU considerada en este trabajo. Se observa que tanto el cupo acumulado como la cantidad de opciones es mayor en liceos que en UTU. Los centros en cupo son aquellos en los que todos los cupos disponibles son ocupados. La sobredemanda es la diferencia entre el cupo disponible en un centro y la cantidad de estudiantes que lo ranearon como primera opción, siendo un reflejo de la popularidad de los centros. Si quisiera asignarse a todos los estudiantes a su primera opción, la sobredemanda es la cantidad que debería aumentar el cupo de los centros que están llenos. En línea con el alto porcentaje de estudiantes asignados a su primera opción, la oferta es elástica y la sobredemanda no es un problema. De todas formas, sí existen algunos centros que son muy demandados. Hay 23 centros que tendrían que aumentar más de 50 plazas para acomodar a todos los que los ranearon como primera opción, y 4 centros tendrían que hacerlo en más de 100. Por su parte, 307 centros (un 76 % del total) tienen entre 0 y 10 estudiantes de sobredemanda.

³Si bien el mecanismo asigna únicamente a liceos y UTU, los estudiantes pueden rellenar opciones por fuera del sistema

Cuadro 4: Características de los centros

	Total	Secundaria	UTU	Montevideo	Resto del país
Cantidad de docentes	63 (36)	74 (38)	44 (23)	73 (31)	61 (37)
Titulados (%)	60 (14)	67 (10)	48 (11)	58 (16)	61 (13)
Efectivos (%)	38 (15)	42 (15)	29 (12)	35 (17)	38 (15)
Mayores de 40 años (%)	34 (13)	28 (8)	46 (11)	38 (10)	33 (13)
Dedicación alta (%)	3 (3)	3 (3)	2 (2)	3 (2)	3 (3)
Estudiantes por docente	5 (2)	5 (2)	5 (2)	6 (2)	4 (2)
Matrícula estudiantes	315 (222)	367 (234)	214 (153)	434 (225)	283 (211)
Aprobaciones (%)	67 (13)	67 (12)	67 (15)	62 (13)	68 (13)
Desvinculación (%)	7 (4)	6 (4)	9 (5)	8 (4)	6 (4)
Hurtos	341 (503)	380 (567)	266 (336)	929 (780)	187 (214)
Homicidios	1,0 (1,8)	1,0 (1,9)	0,9 (1,7)	2,9 (2,9)	0,4 (0,9)
Rapiñas	74 (142)	79 (147)	63 (131)	301 (170)	14 (28)

Promedio (desvío estándar). Elaboración propia con base en datos de ANEP y Ministerio del Interior.

El Cuadro 4 muestra estadísticas descriptivas de las características disponibles para todos los centros considerados. Las características de docentes (cantidad, porcentaje de titulados, porcentaje de mayores de 40 años, porcentaje de efectivos y estudiantes por docente), la matrícula de estudiantes y los resultados educativos (aprobaciones y desvinculación) están medidos a nivel de todo el ciclo básico en los liceos y a nivel de todo el programa en las UTU, tomándose el promedio entre 2021 y 2022. Se consideran docentes titulados los egresados de IPA, CERP, IFD, INET e ISEF⁴. Los cargos efectivos se obtienen por concurso y están asociados a un centro en particular. Los docentes pueden aplicar a un traslado del cargo a otro centro una vez al año, con el requisito de haber permanecido al menos 2 años en el último centro. Los docentes interinos, por su parte, aplican cada año pudiendo ser asignados a centros distintos cada vez. De esta forma, el porcentaje de docentes efectivos en el centro refleja la estabilidad del plantel docente. Vale la pena notar que la cantidad de estudiantes por docente representa la matrícula total del centro entre la cantidad de docentes y no la cantidad de estudiantes que debe atender cada docente en un grupo, lo cual no está disponible. Esta medida busca capturar el mismo efecto, pero es un poco más difusa porque no toma en cuenta la dedicación horaria del docente, y por lo tanto cuántos grupos tiene a su cargo. Más allá de esta consideración, se observa una baja cantidad de docentes con una dedicación alta en un mismo centro, sugiriendo que la medida utilizada podría estar muy correlacionada con la cantidad de estudiantes por docente en el aula. El Cuadro muestra que en promedio solo un 3% de los docentes de un centro tienen más de 30 horas asignadas en el mismo centro. Los datos de delitos refieren a denuncias policiales registradas a menos de 1 kilómetro del centro. Un aspecto relevante para la estimación es que

⁴Instituto de Profesores Artigas, Centros Regionales de Profesores, Instituto de Formación Docente, Instituto Normal de Educación Técnica e Instituto Superior de Educación Física respectivamente.

se observa variación en todas las variables. En el Anexo A.3 se presentan más estadísticas descriptivas y en el Anexo A.1 se describe en detalle la construcción de la base de datos.

Demanda: características de los estudiantes. En 2022 participaron de la preinscripción un total de 42.132 estudiantes en todo el país. Con el objetivo de analizar un conjunto más uniforme de estudiantes y que los supuestos del modelo puedan sostenerse de forma plausible, se eliminaron de la consideración los estudiantes con 15 años o más, estudiantes de escuelas especiales, estudiantes que no rankearon ninguno de los centros considerados, y estudiantes que fueron asignados a centros privados. Es esperable que este conjunto particular de estudiantes pueda tener otra forma de tomar las decisiones de elección escolar. Este filtrado resultó en un total de 37.259 estudiantes (88 % del total) para los cuales se realiza el análisis ⁵. El Cuadro 5 muestra algunas estadísticas descriptivas de los estudiantes considerados.

Cuadro 5: Características de los estudiantes considerados

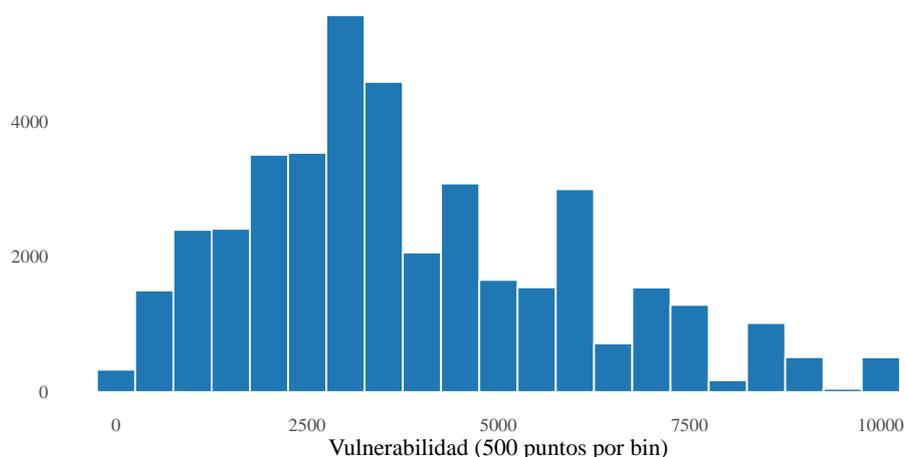
	Total	Montevideo	Resto del país	Con TUS	Sin TUS
N	37.259	11.356	25.894	10.847	26.412
Edad (*)	12,7 (0,6)	12,8 (0,7)	12,7 (0,6)	12,9 (0,7)	12,7 (0,6)
Notas (†)	6,5 (1,3)	6,3 (1,3)	6,6 (1,4)	5,8 (1,2)	6,7 (1,3)
Faltas (†)	39 (15)	42 (15)	38 (15)	45 (16)	37 (14)
Mujeres	50%	50%	49%	50%	49%
AFAM-PE	59%	59%	59%	99%	43%
TUS	29%	34%	27%	100%	0%
Doble TUS	19%	25%	16%	66%	0%
Vulnerabilidad	3.667 (2.011)	4.078 (2.163)	3.486 (1.912)	5.784 (1.697)	2.797 (1.386)
Tiene pred.	86%	98%	81%	86%	86%
Pred. rankeado 1	57%	47%	62%	48%	60%
Pred. rankeado 2	16%	14%	18%	19%	16%
Pred. rankeado 3	20%	29%	15%	23%	19%
Pred. no rankeado	7%	10%	5%	10%	6%

Promedio (desvío estándar). (*) Edad al 30 de abril de 2023 (define admisión en planes extraedad). (†) Promedio últimos 4 años de escuela. *Pred.* refiere al liceo predeterminado. Elaboración propia con base en datos de ANEP.

Prioridades: índice de vulnerabilidad. Las prioridades de asignación en el mecanismo están dadas por el índice de vulnerabilidad construido por ANEP. Los detalles del cálculo del índice de vulnerabilidad se presentan en el Anexo A.2. El índice pretende priorizar a los estudiantes con peores resultados educativos y a aquellos que tienen un contexto socioeconómico desfavorable. En particular, un estudiante es considerado más vulnerable cuanto mayor sea la cantidad de inasistencias que registre en los últimos 3 años escolares, cuanto peores notas haya tenido en esos años, cuanto más extraedad tenga y cuanto mayor sea la carga de planes sociales (AFAM-PE, TUS y doble TUS) que recibe.

⁵ Adicionalmente, en el Anexo A.5 se presentan estimaciones realizadas eliminando a los estudiantes que no tuvieran centros cercanos tomando diferentes criterios de cercanía.

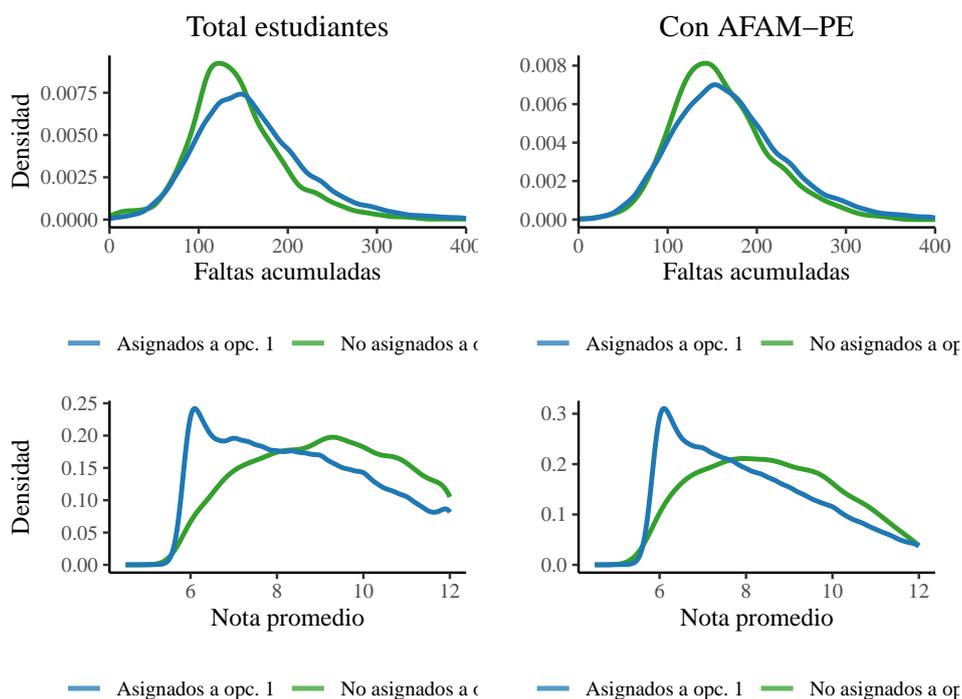
Figura 3: Histograma índice de vulnerabilidad



Elaboración propia con base en datos de ANEP.

La priorización de estudiantes con mayor cantidad de inasistencias y peores notas en la escuela es un aspecto distintivo del mecanismo de asignación de ANEP. En general, los mecanismos que consideran los resultados educativos previos de los estudiantes lo hacen para priorizar a los estudiantes con mejor desempeño. Esto genera que en algunos casos se perjudique en la asignación a estudiantes que reciben planes sociales por tener buenas notas o un nivel bajo de inasistencias. En la Figura 4 se observa que tanto para el total de estudiantes como para el subconjunto de estudiantes que reciben AFAM-PE, los estudiantes con mejores notas y menor cantidad de inasistencias en la escuela tienen una mayor probabilidad de no ser asignados a su primera opción. En 2022, 1.142 estudiantes que reciben AFAM-PE no fueron asignados a su primera opción. Esta condición motiva la reflexión sobre opciones alternativas de priorización que no penalicen a los estudiantes de contexto socioeconómico negativo por tener resultados educativos positivos en la escuela.

Figura 4: Densidad de nota promedio y faltas acumuladas según asignado a primera opción



Elaboración propia con base en datos de ANEP.

Distancias. La distancia al centro educativo es un elemento central de la elección escolar, como es de consenso en la literatura sobre el tema. El Cuadro 6 muestra la distancia de los estudiantes hacia un conjunto de centros relevantes. En primer lugar, se observa que en general los estudiantes deberán desplazarse más al ingresar a la educación media, siendo que el centro asignado y los centros rankeados están más lejos que la escuela a la que fueron. Esto parece ser resultado de cierta disposición de los estudiantes a desplazarse para obtener alguna característica deseada, porque el centro más cercano de la mayoría de los estudiantes está más cerca que la escuela. El centro predeterminado suele estar más cerca que las opciones rankeadas en segundo y tercer lugar. Por último, los estudiantes que reciben TUS rankean opciones más cercanas en relación a los estudiantes que no la reciben.

Cuadro 6: Distancia al estudiante (km)

	Total	Montevideo	Resto del país	Sin TUS	Con TUS
Escuela	1,3 [0,6-5,7]	1,1 [0,6-4,0]	1,3 [0,6-8,7]	1,2 [0,6-5,4]	1,3 [0,6-5,8]
Opción 1	1,7 [0,8-5,4]	2,2 [1,0-5,2]	1,5 [0,8-5,7]	1,9 [0,9-5,1]	1,7 [0,8-5,6]
Opción 2	2,5 [1,3-8,6]	2,7 [1,5-5,8]	2,4 [1,2-13,5]	2,6 [1,3-7,4]	2,5 [1,2-9,1]
Opción 3	3,1 [1,5-11,8]	3,0 [1,7-6,1]	3,2 [1,5-19,9]	3,1 [1,6-9,2]	3,1 [1,5-13,1]
Asignado	1,7 [0,8-5,5]	2,3 [1,1-5,4]	1,5 [0,8-5,6]	1,9 [0,9-5,1]	1,7 [0,8-5,8]
Predeterminado	1,6 [0,8-4,3]	2,1 [1,1-4,6]	1,4 [0,7-3,9]	1,8 [0,9-4,2]	1,6 [0,8-4,3]
Más cercano	0,8 [0,4-1,2]	0,8 [0,5-1,1]	0,7 [0,4-1,3]	0,8 [0,5-1,2]	0,7 [0,4-1,2]
UTU más cercana	1,5 [0,8-2,4]	1,5 [1,0-2,2]	1,4 [0,8-2,6]	1,5 [0,9-2,4]	1,4 [0,8-2,5]
Lic. más cercano	0,9 [0,5-1,4]	0,9 [0,6-1,3]	0,9 [0,5-1,6]	0,9 [0,6-1,5]	0,9 [0,5-1,4]

Mediana [25 %-75 %]. Elaboración propia con base en datos de ANEP.

Asignación y acceso a centros de calidad. La asignación resultante del mecanismo marca una segregación de los estudiantes más vulnerables hacia los centros de menor calidad. Una forma de ver esto es agrupar a los estudiantes según su vulnerabilidad y medir la proporción cada grupo en centros con distintos niveles de calidad. Se consideran cuatro grupos de vulnerabilidad según reciben doble TUS, únicamente TUS, únicamente Asignaciones Familiares (AFAM-PE) o no reciben ningún plan social⁶. En la Figura 5 se presenta la proporción de estudiantes de cada grupo en los centros que se encuentran en cada cuartil de la distribución de características que aproximan la calidad educativa, en comparación con la proporción de estudiantes por grupo en la población de estudiantes. Se puede observar que los estudiantes que reciben planes sociales están sobre representados en los centros que se encuentran en el primer cuartil de calidad y sub representados en los centros del mayor cuartil de calidad, para cualquiera de las medidas de calidad seleccionadas. Lo contrario sucede con los estudiantes que no reciben ningún plan social. A la vez, el Cuadro 7 muestra que estas características cambian sustancialmente entre los centros que se encuentran en diferentes cuartiles.

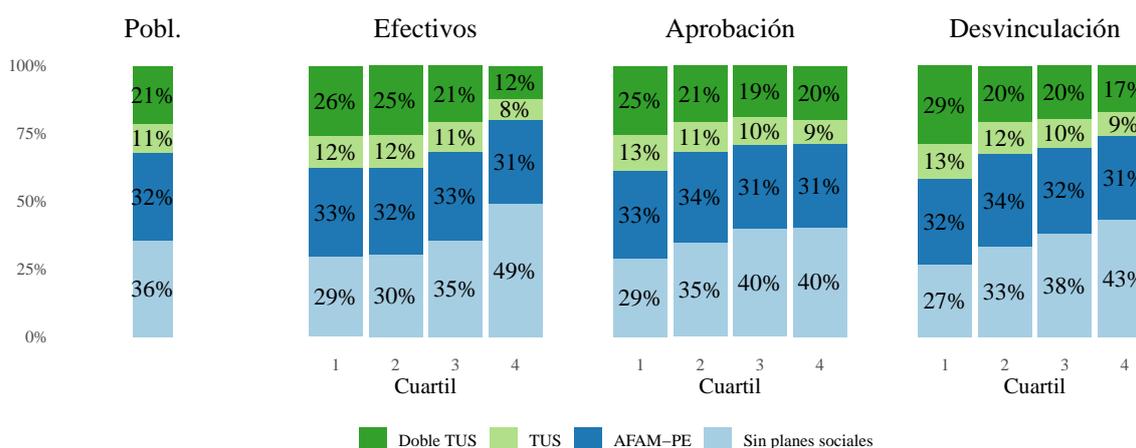
Cuadro 7: Características promedio por cuartil

	Cuartil 1	Cuartil 2	Cuartil 3	Cuartil 4
Efectivos	20 %	34 %	45 %	56 %
Aprobación	51 %	65 %	72 %	82 %
Desvinculación	13 %	7 %	5 %	3 %

Promedio dentro del cuartil de cada característica. Elaboración propia con base en datos de ANEP.

⁶La TUS doble está dirigida al 10% de la población de menores ingresos y la TUS al 20%. Su otorgamiento se define en base a un índice de carencias críticas. Las Asignaciones Familiares abarcan una población más grande. Están dirigidas a menores de 18 años o personas en situación de discapacidad de hogares en situación de vulnerabilidad socioeconómica. Las Asignaciones Familiares tienen como requisito la inscripción y asistencia a alguna institución educativa habilitada.

Figura 5: Proporción de estudiantes según grupo de ingresos



Pobl. representa la composición por grupo en la población de estudiantes. *Efectivos* es el porcentaje de docentes efectivos. *Aprobación* es la tasa de aprobación y *Desvinculación* la tasa de desvinculación. Todas las variables están medidas a nivel de Ciclo Básico tomando el promedio 2021-2022 (salvo para la tasa de aprobación). Los cuartiles se calcularon dentro de cada departamento. Elaboración propia con base en datos de ANEP.

Un aspecto relevante de la asignación es que la mayor parte de los centros de mayor calidad cuentan con cupo disponible para aceptar más estudiantes, como se observa en el Cuadro 8. Esto señala que podría mejorarse el acceso a los mismos para los estudiantes más vulnerables sin necesariamente incurrir en costos altos para los estudiantes de menor vulnerabilidad.

Cuadro 8: Porcentaje de centros que agotan su cupo en cada cuartil de calidad

	Q1	Q2	Q3	Q4
Efectivos	25 %	37 %	39 %	41 %
Aprobación	31 %	33 %	37 %	41 %
Desvinculación	27 %	29 %	41 %	42 %

Elaboración propia con base en datos de ANEP.

Por otro lado, la calidad de los centros no está distribuida de forma equitativa en el espacio. El Cuadro 9 muestra que los estudiantes más vulnerables deben desplazarse más que los menos vulnerables para acceder a centros del cuartil más alto de calidad. Por ejemplo, en Montevideo los estudiantes con doble TUS deben desplazarse en promedio 1,1 km más que los estudiantes que no reciben planes sociales para encontrar un centro del cuartil más alto de proporción de docentes efectivos. Además, los estudiantes más vulnerables deben desplazarse en promedio más para encontrar un centro de mayor calidad que para encontrar uno de menor calidad. Para los estudiantes que no reciben planes sociales esto es al revés, deben desplazarse más para encontrar un centro del primer cuartil de calidad.

Cuadro 9: Distancia promedio al centro más cercano en cada cuartil de calidad (km)

	Montevideo				Resto del país			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Efectivos								
Doble TUS	2,0	1,7	2,0	3,2	9,0	6,7	6,3	9,5
TUS	2,0	1,7	1,9	3,2	9,9	7,3	6,6	9,3
AFAM-PE	2,1	1,8	1,9	2,7	11,2	8,0	7,5	10,0
Sin PS	2,5	1,9	1,9	2,1	12,4	7,1	6,2	7,7
Aprobados								
Doble TUS	1,9	1,9	2,7	1,7	6,3	6,3	7,1	7,4
TUS	1,8	1,9	2,7	1,7	6,8	6,9	7,6	7,6
AFAM-PE	1,9	1,8	2,3	1,6	7,4	7,8	8,4	8,4
Sin PS	2,0	1,7	2,0	1,6	6,6	9,3	7,0	6,6
Desvinculación								
Doble TUS	1,7	1,7	1,6	2,0	8,8	6,9	5,8	6,4
TUS	1,7	1,7	1,6	2,0	9,2	7,2	6,2	6,9
AFAM-PE	1,7	1,6	1,6	1,9	9,3	7,9	7,2	7,6
Sin PS	1,8	1,5	1,6	1,8	9,4	7,5	6,0	6,8

Elaboración propia con base en datos de ANEP.

4. Marco teórico

4.1. Modelos de elección discreta

Los modelos de elección discreta para la estimación estructural de las elasticidades de la demanda son una de las principales herramientas en la literatura de organización industrial. Este enfoque se basa en asumir una forma funcional para las utilidades del consumidor y en la idea, en términos generales, de que los consumidores eligen un bien cuando éste reporta una mayor utilidad que el resto de los bienes disponibles. Este conjunto de opciones entre las que puede elegir se define como *choice set*. La identificación de los parámetros de sustitución entre características de los bienes resulta de la variación exógena en los *choice set*. Intuitivamente, esto significa que observamos cómo cambian las decisiones de los individuos cuando cambian exógenamente las opciones que tienen disponibles. [Berry y Haile \(2021\)](#) y [Train \(2009\)](#), por ejemplo, proporcionan una amplia discusión sobre los desarrollos de estimación de la demanda dentro de la literatura de organización industrial.

Como los datos a nivel de consumidor no son típicos en la literatura de organización industrial, la mayoría de los modelos se esfuerzan por estimar la demanda partiendo de la utilidad a nivel individual pero utilizando datos agregados. Este es el caso de modelos populares como el propuesto por [Berry et al. \(1995\)](#). Sin embargo, el escenario de elección de escuelas y los datos disponibles a nivel de estudiante dan la oportunidad de especificar modelos de demanda a nivel de consumidor al estilo de [McFadden \(1973\)](#). Adicionalmente, la disponibilidad de rankings con tres opciones preferidas enriquece las posibilidades de identificación de la demanda. Utilizar datos de más de una opción por consumidor otorga una fuente de identificación adicional, reduciendo la relevancia de la variación exógena en los *choice sets* ([Berry et al., 2004](#); [Train, 2009](#)).

Un modelo simple usando rankings. Una forma de explotar los rankings es utilizando un modelo como el desarrollado por [Train \(2009, cap. 7\)](#). Se observa un mercado escolar donde I estudiantes eligen entre J centros educativos. Cada estudiante i envía una lista $L_i = (j_i^1, j_i^2, \dots, j_i^{K_i})$. El ranking que el estudiante asigna a cada centro es $rk_i(j)$, siendo $1 \leq rk_i(j) < K_i$ si $j \in L_i$ y $rk_i(j) = \emptyset$ si j no está en la lista. Se asume que la utilidad de cada estudiante i por cada centro j tiene la forma $U_i(j) = \beta X_{ij} + \varepsilon_{ij}$, donde X_{ij} es un vector de características de los centros (que pueden estar interactuadas con características de cada estudiante), y ε_{ij} es un término no observado. Si por ejemplo un estudiante envía el ranking $L_1 = (C, B, D, A)$, entonces puede inferirse que

$$U_i(C) > U_i(B) > U_i(D) > U_i(A) \quad (1)$$

La literatura de elección discreta propone diferentes modelos que toman supuestos sobre la estructura del término no observado ε_{ij} , a partir de lo cual pueden estimarse los parámetros del modelo.

Comportamiento estratégico en el contexto de elección escolar. En el contexto de la elección escolar, la consistencia de la estimación de las preferencias reveladas depende de los supuestos de comportamiento sobre las decisiones de los padres (Agarwal y Somaini, 2020). La principal preocupación es si los rankings presentados por las familias representan sus verdaderas preferencias o incluyen algún nivel de estrategización, involucrando de ese modo sus creencias sobre las probabilidades de asignación. Por ejemplo, podrían omitirse opciones muy demandadas por considerar que las chances de admisión son bajas. En ese caso, no se cumpliría necesariamente la condición de la ecuación (1): rankear una opción no implicaría necesariamente que ésta sea preferida. Como el interés suele ser recuperar las preferencias reales sobre las opciones, la literatura ha utilizado las propiedades de los mecanismos de asignación para tomar supuestos sobre el comportamiento estratégico e incorporarlo en el modelo estructural. Conocer las propiedades del mecanismo es entonces un aspecto central de la estimación.

Agarwal y Somaini (2020) destacan dos propiedades de los mecanismos de asignación que establecen condiciones favorables para utilizar determinadas técnicas de estimación: la **estabilidad** y la *strategy-proofness*. Se dice que un mecanismo genera asignaciones estables cuando no hay envidia justificada. En el contexto de la elección de escuela, esto significa que no se asigna a ningún estudiante a una escuela A si éste prefiere otra escuela B , y un estudiante menos prioritario es asignado a esa escuela B . Un mecanismo es *strategy-proof* o a prueba de estrategia cuando decir la verdad es una estrategia dominante para todos los jugadores (padres). Este no sería el caso si los padres pudieran aumentar las posibilidades de que su hijo sea asignado a una escuela preferida presentando rankings que no representan sus verdaderas preferencias.

En este sentido, Fack *et al.* (2019) y Ortega *et al.* (2020) proponen diferentes posibilidades de expansión del modelo simple presentado según se tomen ciertos supuestos sobre el comportamiento estratégico:

1. *Weak Truth Telling (WTT)*. Implica asumir que los estudiantes rankean sus K_i opciones preferidas, y las opciones no rankeadas son menos valoradas que las rankeadas: $U_i(j) \geq U_i(j')$ si $j, j' \in L_i$ y $rk_i(j) < rk_i(j')$, o si $j \in L_i$ y $j' \notin L_i$.
2. *Estrategias no dominadas*. Implica solamente que los estudiantes ordenan según sus preferencias reales dentro de las opciones que rankean: $U_i(j) \geq U_i(j')$ si $j, j' \in L_i$ y $rk_i(j) < rk_i(j')$. Este supuesto admite que los estudiantes omitan opciones preferidas por encima de la rankeada en

primer lugar y en medio de las opciones rankeadas, ya que no asume ninguna relación entre la utilidad de las opciones rankeadas y las no rankeadas. Es un supuesto más general, pero también tiene condiciones más débiles de identificación.

3. *Estabilidad*. Si el mecanismo de asignación es estable, puede calcularse el puntaje de prioridad mínimo (*cutoff*) para ser aceptado en cada centro educativo. De esta forma, el supuesto implica que los estudiantes rankean sus K_i opciones preferidas omitiendo las opciones “inalcanzables”. Así, $U_i(j) \geq U_i(j')$ si $rk_i(j) < rk_i(j')$ o $j' \notin L_i$ y $cutoff_j < prioridad_i$, donde $prioridad_i$ es el puntaje de prioridad del estudiante i .

En general, el supuesto de WTT es difícil de sostener, en particular cuando los estudiantes pueden rankear una cantidad limitada de opciones, por lo que en ambos trabajos es utilizado como un *benchmark*. [Fack et al. \(2019\)](#) demuestra que las condiciones que impone el supuesto de estabilidad son suficientes para la identificación puntual, mientras que [Ortega et al. \(2020\)](#) presenta evidencia de simulaciones de Monte Carlo a favor de utilizar los supuestos de estrategias no dominadas y estabilidad en conjunto.

4.2. Un modelo de comportamiento para el caso uruguayo

Como se mencionó, las familias no cuentan con detalles sobre el funcionamiento del mecanismo de asignación. Sin embargo, en base a los resultados de la asignación en años anteriores podría esperarse que estimen una alta probabilidad de entrar a alguno de los centros que rankean, siendo que menos del 1% de los estudiantes es asignado a una opción no rankeada cada año. Aunque no conozcan los resultados agregados de la asignación, es muy poco probable que en su escuela o entre las familias que conocen encuentren casos de estudiantes asignados a un centro no rankeado. Este punto representa evidencia sugerente de que tienen incentivos a rankear sus opciones realmente preferidas en lugar de incurrir en comportamientos estratégicos. Adicionalmente, aquí se argumentará que aún cuando las familias sí conocieran gran parte del funcionamiento del algoritmo, el mismo cuenta con propiedades particulares que también sostienen el cumplimiento del supuesto de WTT, al menos para la mayor parte de los estudiantes, y descartan el cumplimiento del supuesto de estabilidad.

Un primer aspecto clave es el uso del algoritmo de **Dictador Serial** dentro de cada paso de la asignación. El Dictador Serial considera a los estudiantes uno a uno, ordenados desde el de mayor prioridad, y los va asignando a su opción más preferida con cupo disponible. Se diferencia de otros mecanismos porque no existe una penalización por rankear en las primeras posiciones centros altamente demandados

⁷. El mecanismo es estable y *strategy-proof* si los estudiantes pueden rankear centros ilimitadamente, pero deja de serlo si tienen un tope de centros que pueden rankear. Intuitivamente, si el largo del ranking está limitado a 3, los estudiantes de baja prioridad podrían quedar sin ser asignados si rankean solamente opciones muy demandadas. Esto supone incentivos a rankear una opción segura aunque realmente no esté entre sus 3 preferidas.

Sin embargo, una particularidad del mecanismo utilizado por ANEP es que en el paso 7 se asignan a todos los estudiantes no asignados a alguna de sus opciones rankeadas, sin importar que ya no tengan cupo disponible. Este paso garantiza que los estudiantes queden asignados a una de sus opciones rankeadas, con excepción de los estudiantes que tienen un liceo predeterminado y no lo rankean en sus tres opciones, en cuyo caso son asignados a su cuarta opción. Al mismo tiempo, el paso 7 genera envidia justificada, haciendo que la asignación no sea estable. Tomando esto en cuenta, pueden distinguirse tres tipos de estudiantes con diferentes incentivos a manipular estratégicamente su ranking:

1. Estudiantes que no tienen un liceo predeterminado
2. Estudiantes que tienen un liceo predeterminado y está entre sus tres centros preferidos
3. Estudiantes que tienen un liceo predeterminado pero no está entre sus tres centros preferidos

Para los estudiantes de tipo 1 el mecanismo garantiza la asignación a alguno de los centros rankeados. De esta forma, el estudiante no tiene incentivos para excluir su primera opción preferida. Sin embargo, podría tomar como estrategia rankear en primer lugar su opción más preferida y completar el segundo y tercer lugar con opciones altamente demandadas. El objetivo de esta estrategia sería, en caso de no quedar asignado a su primera opción en el paso 5, tampoco ser asignado a sus otras dos opciones y ser asignado a su primera opción en el paso 7, por ser la de menor sobrecupo. Esta estrategia supondría un conocimiento detallado del algoritmo de asignación, y a la vez implica el riesgo para el estudiante de quedar asignado a una opción no preferida si su creencia sobre los *cutoffs* no es correcta, por lo tanto se supondrá que para este grupo de estudiantes se sostiene WTT.

Los estudiantes del tipo 2 también tienen garantizada la asignación a una de las opciones rankeadas. En su caso, podrían incorporar una estrategia similar a la detallada para los estudiantes de tipo 1. Si su predeterminado es su primera o segunda opción preferida, podrían completar las opciones inferiores restantes con centros muy demandados para no quedar asignados en el paso 5, y ser asignados a su

⁷En el mecanismo de *Boston* (o de aceptación inmediata), por ejemplo, se considera primero la primera opción de cada estudiante. Si para un estudiante no hay cupo en la primera opción, se intentará asignar a los siguientes estudiantes a su primera opción. Esto reduce la probabilidad del estudiante de entrar en su segunda opción, por lo que tiene incentivos a no rankear una opción muy demandada en primer lugar.

predeterminado en el paso 6. También se asumirá que esta estrategia implica un conocimiento detallado del algoritmo y es riesgosa para el estudiante, por lo que también se asumirá WTT para los estudiantes de tipo 2.

El caso más problemático es el de los estudiantes de tipo 3. Si a un estudiante no le gusta su predeterminado, podría incorporar opciones que no estén entre sus tres más preferidas, pero que tengan una probabilidad más alta de entrar y sí sean mejores opciones que el predeterminado. Para este grupo no parecería sostenerse el supuesto de WTT, por lo que se asumirá solamente estrategias no dominadas contando al predeterminado como cuarta opción. Este supuesto sí puede mantenerse porque aunque el estudiante incluya una opción “segura” para evitar el predeterminado, el uso del dictador serial en el paso 5 hace que no tenga incentivos a colocar sus tres opciones rankeadas en un orden distinto al de su preferencia. Si bien el uso de este supuesto en lugar de WTT reduce la capacidad de inferir preferencias a partir de los rankings, los estudiantes que pertenecen a este grupo representan una porción menor del total, como se observa en el Cuadro 5.

Un detalle importante es que los estudiantes de tipo 1 pueden distinguirse en los datos porque se observa si tienen o no un liceo predeterminado. Sin embargo, para los estudiantes que tienen predeterminado lo que observamos es si lo rankea o no. Asumir que rankearlo implica que esté realmente entre sus preferidos requiere tomar un supuesto de comportamiento adicional. Los estudiantes que tienen un liceo predeterminado no tienen incentivos a rankear su predeterminado si este no está entre sus opciones preferidas, ya que esto solo les quitaría la posibilidad de rankear una opción más preferida sin ningún beneficio. Por esto, puede asumirse que los estudiantes que rankean su predeterminado son del tipo 2 y los que no lo rankean son del tipo 3.

En definitiva, el sistema de asignación uruguayo queda definido como un conjunto de I estudiantes que eligen centros educativos j dentro de su *choice set* J_i enviando un ranking de tres centros $L_i = ((j_i^1, j_i^2, j_i^3))$. El ranking que el estudiante asigna a cada centro es $rk_i(j)$, siendo $1 \leq rk_i(j) \leq 3$ si $j \in L_i$ y $rk_i(j) = \emptyset$ si j no está en la lista. Una parte de los estudiantes cuenta con un centro predeterminado $j_i^p \in J_i$, mientras el resto de los estudiantes no tiene ningún centro predeterminado, $j_i^p = \emptyset$. La utilidad del estudiante i por ser asignado al centro j tiene la forma $U_i(j) = \beta X_{ij} + \varepsilon_{ij}$, al igual que en el modelo presentado anteriormente. Si el estudiante i envía el ranking $L_i = (j_i^1, j_i^2, j_i^3)$, revela la siguiente relación de preferencias (se omite el subíndice i):

$$\begin{cases} U(j^1) > U(j^2) > U(j^3) > U(j^p) & \text{si } j^p \neq \emptyset \text{ \& } rk(j^p) = \emptyset \\ U(j^1) > U(j^2) > U(j^3) > U(j') & \text{en otro caso} \end{cases} \quad (2)$$

Donde $j' = j \in J_i$ tal que $(rk(j') = \emptyset)$, son todas las opciones del *choice set* no rankeadas. La ecuación (2) refleja el uso del supuesto de estrategias no dominadas para los estudiantes que tienen predeterminado pero no lo rankean, y WTT para el resto de los estudiantes.

La identificación en modelos de elección discreta requiere que se asuma cierta estructura en el término de error. Lo más usual en la literatura de organización industrial es utilizar variantes de modelos *Logit*, en los que se asume que ε_{ij} es *iid* y tiene una distribución *extreme value*, o *Probit*, en los que se asume una distribución conjunta de los inobservables $\varepsilon_i \sim N(0, \Omega)$. En general, los modelos Logit suponen una forma conveniente de estimación utilizando el método de máxima verosimilitud o métodos de momentos, aunque cuentan con la limitación de asumir la propiedad de independencia de alternativas irrelevantes. Este supuesto es particularmente problemático al usar rankings, porque implica asumir que la elección de cada opción rankeada es independiente entre sí (Train, 2009, cap. 2). Existen variantes como los modelos *Mixed Logit* que incorporan coeficientes aleatorios y de esa forma superan esta limitación, que pueden estimarse por simulación de la máxima verosimilitud. Sin embargo el elevado número de opciones lleva a sesgos en estos modelos (Train, 2009) e incluso suelen no tener una solución cuando se pretende asumir algún tipo de comportamiento estratégico en los rankings (Agarwal y Somaini, 2020). Es por esto que los modelos *Probit* se vuelven más atractivos, siendo que la estructura que se asume para el término de error permite cualquier tipo de correlación entre los inobservables de las alternativas, y de esa forma relaja la propiedad de independencia de alternativas irrelevantes (McCulloch y Rossi, 1994; Train, 2009). La estimación de esta clase de modelos se realiza usualmente mediante procedimientos bayesianos, que superan los sesgos de la estimación por simulación de la máxima verosimilitud aún con un número elevado de opciones.

4.3. Preferencias heterogéneas y el rol de la información

El modelo desarrollado asume que las familias conocen las características de los centros. Este es un supuesto importante, siendo que la falta de información sobre la calidad de los productos afecta el comportamiento de los consumidores (Akerlof, 1970). En el contexto de la elección escolar, el acceso a la información sobre la calidad de los centros podría ser más restringido para las familias de contextos socioeconómicos más vulnerables (Hastings y Weinstein, 2008). En el extremo, si las familias más vulnerables no tienen ninguna información sobre la calidad de los centros, sus decisiones estarán guiadas únicamente por características que sí conocen como la distancia y no le darán valor a la calidad. Por otro lado, aún cuando todas las familias tuvieran información precisa sobre las características de los centros, las preferencias de las familias más vulnerables podrían incorporar información incorrecta sobre

los retornos a la inversión en capital humano y sobre la importancia de la calidad de los centros (Jensen, 2010).

En definitiva, las diferencias entre grupos en las preferencias estimadas capturarán al mismo tiempo las diferencias en la información disponible sobre las características de los centros y las preferencias reales de las familias. A la vez, una parte de las preferencias estimadas estará asociada a la información de las familias sobre los retornos educativos y la relevancia de la calidad del centro educativo. En el marco de este trabajo, estos tres componentes no pueden distinguirse. Al imponer a las familias vulnerables las preferencias estimadas para el resto de las familias, se pretende capturar el efecto potencial que tendría una intervención en la información de las familias vulnerables si esta fuera la causa de la heterogeneidad en las preferencias.

5. Estrategia empírica

El objetivo de la estrategia empírica es medir la relación entre la heterogeneidad en las preferencias y la segregación de los estudiantes en los centros de distinta calidad. El primer paso para alcanzar este objetivo es estimar el set de parámetros β que representan los determinantes de la utilidad de los estudiantes por centros educativos en $U_i(j) = \beta X_{ij} + \varepsilon_{ij}$. Para esto se asumirá el modelo de comportamiento explicado en la sección anterior y se realizará un procedimiento de estimación bayesiano basado en McCulloch y Rossi (1994). En segundo lugar, se construyen escenarios contrafactuales con el objetivo de medir los cambios en la segregación y su relación con la heterogeneidad en las preferencias. A continuación se explica el procedimiento de estimación de las preferencias en detalle y se exponen los supuestos de identificación. Luego se detalla la construcción de escenarios contrafactuales y la medición de la segregación.

5.1. Procedimiento de estimación

En primer lugar, se considera a la utilidad como una variable latente de la cual depende la decisión binaria de elegir un centro en determinado *choice set*. McCulloch y Rossi (1994) establece que dada la estructura *Probit* del modelo (en particular, la distribución normal conjunta del error en la utilidad de cada opción), pueden tomarse muestras de la distribución posterior del conjunto de parámetros de interés, $p(\beta \mid \text{datos})$ utilizando las distribuciones posteriores condicionales $p(U \mid \beta, \text{datos})$ y $p(\beta \mid U, \text{datos})$, donde en el contexto de este trabajo los *datos* son los rankings enviados por las familias. A la vez, demuestran que tomando ciertos *prior* generales sobre los parámetros del modelo estas distribuciones posteriores tienen formas conocidas con las cuales se puede hacer inferencia bayesiana. El uso de la distribución posterior de la variable latente, en este caso la utilidad, es lo que se conoce como *Data Augmentation*.

Para la estimación de los parámetros se utiliza el algoritmo de *Gibbs Sampling*. El algoritmo comienza tomando valores arbitrarios de los parámetros β y luego sigue los siguientes pasos:

1. Para cada estudiante i , tomar una muestra $U_i(j)$ desde $p(U_i(j) \mid \beta, U_i(j'), \text{datos}) = N(X_{ij}\beta, 1)$ truncada $[\underline{U}_i(j), \bar{U}_i(j)]$.
2. Tomar una muestra de β desde $p(\beta \mid U, \text{datos}) = N(b, (XX)^{-1})$ con $b = (XX)^{-1}X'U$.
3. Repetir los pasos 1 y 2 n veces (donde n es el número de iteraciones).

Las muestras de los parámetros β tomadas en cada iteración forman Cadenas de Markov (una por

cada parámetro), y el promedio de cada cadena es asintóticamente equivalente al estimador de máxima verosimilitud (Van der Vaart, 2000). Es necesario definir la cantidad n de iteraciones del Gibbs sampler y la cantidad de iteraciones que se descartan para eliminar las iteraciones en las que las cadenas están convergiendo (*burn-in*). Adicionalmente, como las simulaciones sucesivas de cada parámetro están correlacionadas entre sí, suele guardarse 1 de cada t simulaciones para la estimación (*thining*).

La estimación requiere que se definan los truncamientos de la distribución posterior de utilidades. Los truncamientos reflejan los rankings de cada estudiante a la luz del supuesto de comportamiento estratégico tomado, tal como se describe en la sección anterior. Formalmente, para los estudiantes que se asume WTT los truncamientos son:

$$\bar{U}_i(j) = \begin{cases} +\infty & \text{si } rk_i(j) = 1 \\ \min_{j' \in L_i} \{U_i(j') : rk_i(j') < rk_i(j)\} & \text{si } rk_i(j) > 1 \\ \min_{j' \in L_i} \{U_i(j')\} & \text{si } rk_i(j) = \emptyset \end{cases}$$

$$\underline{U}_i(j) = \begin{cases} \max_{j' \in L_i} \{U_i(j') : rk_i(j') > rk_i(j)\} & \text{si } rk_i(j) < 3 \\ \max_{j' \notin L_i} \{U_i(j')\} & \text{si } rk_i(j) = 3 \\ -\infty & \text{si } rk_i(j) = \emptyset \end{cases}$$

Para los estudiantes que se asume estrategias no dominadas, los truncamientos son:

$$\bar{U}_i(j) = \begin{cases} +\infty & \text{si } rk_i(j) \in \{1, \emptyset\} \\ \min_{j' \in L_i} \{U_i(j') : rk_i(j') < rk_i(j)\} & \text{si } rk_i(j) > 1 \end{cases}$$

$$\underline{U}_i(j) = \begin{cases} -\infty & \text{si } rk_i(j) \in \{3, \emptyset\} \\ \max_{j' \in L_i} \{U_i(j') : rk_i(j') > rk_i(j)\} & \text{si } rk_i(j) < 3 \end{cases}$$

Estos truncamientos representan las relaciones de preferencias que pueden inferirse de los rankings según lo expuesto en la ecuación (2). Por ejemplo si consideramos el caso de los estudiantes para los que se asume WTT, si una opción está rankeada por un estudiante en primer lugar, el límite superior de la utilidad que le reporta es infinito porque no se fuerza a que esta utilidad sea menor que la de ninguna otra opción. Para las opciones rankeadas en segundo y tercer lugar, se limita el valor de la utilidad en la utilidad de la opción que esté rankeada en una posición más alta. Para las opciones no rankeadas, la utilidad puede ser como máximo la utilidad reportada por la tercera opción, pues se asume que todas las

no rankeadas son peores que las rankeadas. La interpretación del resto de los truncamientos es análoga.

El procedimiento se realizó utilizando una versión modificada del paquete *stabest* desarrollado por [Ortega et al. \(2020\)](#) para el software libre *R*. En el caso de [Ortega et al. \(2020\)](#), se estiman adicionalmente las preferencias de los centros educativos por los estudiantes. Se desarrolló una nueva versión del paquete tomando como base el código original en C++ y *R* de forma de estimar únicamente las preferencias de los estudiantes.

5.2. Especificación del modelo

La especificación del modelo requiere, a grandes rasgos, dos definiciones: qué variables incluir en la matriz X_{ij} y los *choice sets* de cada estudiante.

Características incluidas en el modelo. El modelo admite la inclusión de características de los centros y de interacciones entre las mismas y características de los estudiantes. Las características de los centros serán necesarias para estimar las elasticidades de los padres hacia ellas, y la inclusión de interacciones con características de los estudiantes permite introducir heterogeneidad en el modelo para comprender las diferencias en las preferencias de distintos tipos de estudiantes. Usualmente los trabajos que introducen este tipo de heterogeneidad interactúan las características de los centros con una variable que indica la pertenencia a algún grupo vulnerable. Se incluirán interacciones entre las características de los centros y una variable binaria que identifica a los estudiantes que reciben TUS ⁸.

La distancia y alguna medida de la calidad del centro son las características más comúnmente incluidas en este tipo de modelos. En este trabajo se incluye la distancia del centro al hogar del estudiante y además 4 grupos de características que se detallan a continuación.

En primer lugar, se incluyen características que intentan capturar la capacidad del centro de generar resultados educativos positivos en sus estudiantes. En el caso de Uruguay, es difícil obtener medidas del efecto de los centros sobre los resultados educativos. Sin embargo, [Abdulkadiroğlu et al. \(2020\)](#) encuentra que los padres no consideran medidas sofisticadas como el valor agregado docente, sino que dan más importancia a los resultados promedio de la escuela. En este sentido, se incluye el porcentaje de estudiantes aprobados y el porcentaje de desvinculación estudiantil. Se considera el porcentaje de aprobación de 2021 y el promedio del porcentaje de desvinculación entre 2021 y 2022 ⁹.

En segundo lugar, se incluye un conjunto de variables asociadas a los recursos del centro educativo:

⁸A modo de ejercicio de robustez, en el Anexo A.5 se incluyen estimaciones que toman valores de corte del índice de vulnerabilidad y la tenencia de doble TUS para identificar estudiantes vulnerables.

⁹La aprobación no está disponible para 2022, y en el caso de los liceos se dispone de datos desde 2020. Se decidió relegar los datos de 2020 por las particularidades del cursado en el marco de la pandemia por COVID-19

el porcentaje de docentes del centro que tienen un cargo efectivo allí, la cantidad de estudiantes por docente y la matrícula estudiantil. Como los docentes efectivos deben permanecer en el centro y deben concursar para trasladarse, se pretende que capture la estabilidad del plantel docente. La presencia de docentes efectivos también podría reflejar características no observadas que sí son tomadas en cuenta por los docentes cuando eligen un centro para tener su cargo efectivo. Adicionalmente, la cantidad de cargos efectivos refleja el presupuesto del centro. La cantidad de estudiantes por docente tiene una relación inversa con la calidad y es un reflejo del presupuesto por estudiante del centro, dado el peso del salario docente en el presupuesto de los centros educativos. La matrícula estudiantil busca controlar preferencias por centros más grandes. De este conjunto de variables se considera el promedio de 2021 y 2022, de forma de tomar una medida más estable que un solo año.

Resulta difícil pensar que las familias conocen con precisión estas variables, a pesar de que los datos de resultados educativos están disponibles públicamente en la web de ANEP. Su inclusión en el modelo no implica suponer que las familias las conocen, sino que se pretende capturar las expectativas de las familias sobre la calidad de los centros. La misma debe ser suficientemente estable en el tiempo para que las familias tengan una idea formada sobre ella. Como se pretende que las variables del modelo estén correlacionadas con esta característica, es necesario que también sean estables en el tiempo. En el Anexo A.3 se presenta un análisis de la estabilidad en el tiempo de las características.

Aún si las variables del modelo están correlacionadas con la calidad de los centros, se asume que las familias toman decisiones a partir de la información que tienen sobre la calidad, pero en la estimación se incluyen realizaciones de las características. Si la diferencia entre la información de las familias y los valores reales de estas características es aleatoria entre las familias, este problema puede interpretarse como error de medida en las variables del modelo y las estimaciones tendrán un sesgo de atenuación. Sin embargo, el nivel de información con el que cuentan las familias podría estar correlacionado con sus características, por ejemplo con su nivel de vulnerabilidad (como se ve en [Hastings y Weinstein \(2008\)](#), por ejemplo). En este sentido, las preferencias estimadas para el grupo de estudiantes vulnerables podrían resultar menores a las del resto por tres razones que no pueden distinguirse en este trabajo: por tener menos información sobre la calidad de los centros, y por lo tanto darle menor valor a las características asociadas; por tener información incorrecta y sobrestimar la calidad de los centros que eligen; y por efectivamente tener una valoración menor de la calidad, aun cuando tengan información correcta. Esta es una limitación importante del enfoque de estimación, que podría ser superado en futuros trabajos que intervengan de forma controlada la información de las familias.

En tercer lugar, se incluyen la cantidad de homicidios y hurtos denunciados en un entorno de 1

kilómetro para controlar la seguridad en el entorno de los centros. Por último, se incorpora la composición de los pares utilizando el porcentaje de estudiantes que reciben TUS entre los asignados al centro. Se espera que si existen preferencias por pares del mismo contexto socioeconómico, esta variable tenga un coeficiente negativo para los estudiantes que no reciben TUS y un coeficiente positivo para los que sí la reciben.

Distancias y definición del *Choice set*.

La estimación del modelo requiere la definición del *choice set* para cada estudiante. Si bien el mecanismo permite a todos los estudiantes elegir cualquier centro educativo en el país, definir *choice sets* tan amplios para todos los estudiantes es problemático en términos computacionales. Además hacerlo implicaría asumir que los padres procesan información de todos los centros educativos del país, y pueden rankearlos según sus preferencias. En la realidad, los agentes enfrentan costos de búsqueda y tienen una capacidad limitada de procesar información sobre todas las opciones posibles. Kapor *et al.* (2016), por ejemplo, realiza una encuesta a las familias que participan de la inscripción centralizada de liceos en New Haven, donde los estudiantes podían inscribirse a 12 centros, y encuentra que solo un quinto de los participantes consideró todos los centros, mientras que un tercio consideró menos de 6.

El cuadro 9 muestra la cantidad de centros que los estudiantes tienen disponibles en determinados entornos a su hogar. Se observa que los estudiantes de Montevideo tienen una oferta cercana mayor, siendo que en algunos casos los estudiantes del resto del país tienen muy pocas alternativas cercanas. Este punto es relevante ya que la identificación de las preferencias de los estudiantes que cuentan con pocas opciones es menor y es probable que sus preferencias no sean bien representadas por las estimaciones de preferencias en el promedio. Al considerar los estudiantes que reciben TUS, se observa que también suelen tener una oferta menor, particularmente en la cantidad de liceos.

Cuadro 10: Cantidad de centros cercanos

	Total	Montevideo	Resto del país	Sin TUS	Con TUS
<20km					
Centros	14 [5-99]	104 [99-109]	9 [4-15]	21 [7-102]	11 [5-98]
UTU	4 [2-31]	33 [31-35]	3 [2-5]	7 [2-32]	4 [2-31]
Liceos	10 [3-68]	71 [68-75]	6 [2-10]	13 [4-70]	8 [3-67]
<10km					
Centros	9 [4-48]	64 [52-69]	6 [3-10]	11 [5-52]	9 [4-45]
UTU	3 [1-14]	18 [15-20]	2 [1-3]	4 [2-16]	3 [1-14]
Liceos	6 [2-34]	45 [37-49]	3 [2-7]	7 [3-37]	6 [2-31]
<5km					
Centros	6 [3-12]	22 [15-28]	4 [2-7]	7 [4-15]	6 [3-11]
UTU	2 [1-4]	7 [4-8]	2 [1-2]	2 [1-5]	2 [1-4]
Liceos	4 [2-8]	16 [10-20]	3 [1-5]	5 [2-10]	4 [2-7]
Choice set					
Centros	13 [10-57]	64 [52-71]	10 [10-14]	15 [10-59]	12 [10-56]
UTU	5 [3-16]	18 [14-21]	4 [3-5]	5 [3-17]	5 [3-16]
Liceos	8 [7-40]	46 [37-50]	7 [6-9]	10 [7-41]	8 [7-40]

Mediana [25 %-75 %]. Elaboración propia con base en datos de ANEP.

Tomando esta información en cuenta, se definieron los *choice sets* en dos pasos. En primer lugar, se filtran los centros según el departamento. Se consideran todos los centros que estén en el mismo departamento que el estudiante o que alguna de las opciones que rankeó el estudiante. Si el estudiante vive en Canelones o rankeó algún centro en Canelones, también se incluyen los centros de Montevideo.

En el segundo paso, se calcula para cada estudiante el máximo entre 10 kilómetros y la distancia más lejana a un centro que haya rankeado. Se eliminan todos los centros que estén a una distancia mayor a la calculada. Se utiliza la distancia máxima a algún centro rankeado como un indicador de la disposición del estudiante a desplazarse, siendo que la lejanía no hizo que dejara de considerar un centro. En caso de que el *set* resultante tenga menos de 10 opciones, se agregan las opciones más cercanas sin considerar el límite de distancia (aunque sí deben cumplir el requisito de estar en el mismo departamento). Como ejercicio de robustez, en el Anexo A.5 se presentan estimaciones que definen los *choice sets* utilizando solamente el primer paso, y estimaciones que no consideran a los estudiantes que no tienen centros cerca disponibles.

Además de la definición de los *choice sets*, también es necesario definir la *outside option*. La *outside option* es la alternativa de no elegir ninguna de las opciones en el *choice set*, y los modelos de elección discreta requieren definir la utilidad que otorga esta opción. El mecanismo de perinscripción está

diseñado para aplicar a los centros de enseñanza públicos, por lo que no es posible incluir los centros privados en el modelo ni identificar las preferencias de los padres por estos centros. Sin embargo, es razonable pensar que su presencia puede jugar un rol en la elección de los padres. En particular, si se observa que las familias tienen diferentes posibilidades de acceder a un centro privado, tendrían *outside options* diferentes y la estimación se complejizaría. Como los resultados de la preinscripción se divulgan en diciembre y las inscripciones a centros privados suelen darse con anterioridad a esa fecha, se asumirá que la decisión de las familias ocurre en dos etapas. Se asume que las familias que envían su preinscripción a los centros públicos ya descartaron la opción de asistir a un privado y estos no juegan ningún rol en la estimación de las preferencias. En este sentido, los parámetros estimados por el modelo serán representativos de las familias que optan por el sistema público de educación media. Este camino hace más simple la definición de la *outside option*, ya que una vez que las familias descartaron la opción de ir a un centro privado, la alternativa a asistir a un centro público es no asistir a la educación media y se asumirá una utilidad de 0.

5.3. Identificación e interpretación de los parámetros.

El supuesto central de la estimación de las preferencias es que las familias reportan sus preferencias en los rankings de la forma en que se explicita en el modelo de comportamiento presentado. Si bien se mostró que los incentivos del algoritmo y los resultados de la asignación de años anteriores sustentan la hipótesis de WTT para la mayor parte de los estudiantes, el hecho de que no cuenten con ninguna información sobre el algoritmo es una limitación de este trabajo. Podría ocurrir que las familias formen creencias equivocadas sobre su probabilidad de asignación y descarten opciones preferidas aunque hubiesen tenido chances de entrar si las rankeaban. Esta limitación podrá ser superada en trabajos posteriores que incorporen de alguna forma la información con la que cuentan las familias (como se realiza por ejemplo en [Kapor et al. \(2016\)](#)). En caso de que este supuesto no se cumpla y efectivamente las familias incorporen su expectativa sobre la probabilidad de asignación, las estimaciones no capturarán únicamente las preferencias reales.

En segundo lugar, se supone que las familias tienen expectativas formadas sobre la calidad educativa de los centros, aunque la interpretación de los resultados no asume que todas las familias cuenten con información completa. Como se explicó, se supone que esas expectativas están correlacionadas con la calidad de los centros y esta última con las variables incluidas en el modelo. La evidencia que sustenta este supuesto es que las variables consideradas son estables en el tiempo, como se muestra en el [Anexo A.3](#).

También deben considerarse los supuestos usuales de los modelos de elección discreta. Train (2009) resume las características principales de la identificación en los modelos *Probit* en 2 puntos centrales. En primer lugar, *solo las diferencias en utilidad importan* dado que sumar una constante a la utilidad de todas las alternativas disponibles no altera las decisiones de los agentes ni la estimación de las preferencias. Por otro lado, las decisiones de los agentes tampoco cambian si se multiplica la utilidad de todas las alternativas por una constante, por lo cual *la escala de la utilidad es irrelevante*. Esto hace que no puedan identificarse los parámetros que afectan a la escala o el nivel de la utilidad, simplemente porque no influyen en la decisión que se observa. En términos prácticos, no pueden incluirse en el modelo características que no varíen entre las alternativas, como por ejemplo las características de los estudiantes sin interactuar con las características de los centros. Adicionalmente, se vuelve relevante que exista variación de las características en el *choice set* de los estudiantes ¹⁰.

Adicionalmente, la interpretación de los parámetros estimados requiere definir una normalización de la escala de la utilidad. Esto se consigue asumiendo cierta estructura en el término del error. En este trabajo se sigue a Ortega *et al.* (2020) y se asume que el error está centrado en 0 y tiene varianza unitaria.

En el contexto de la elección escolar, suele utilizarse el coeficiente de la distancia como numerario de la utilidad (Agarwal *et al.*, 2021). De esa forma, los parámetros pueden interpretarse como el aporte de cada característica a la utilidad en términos de *disposición a trasladarse*. Sin embargo, este enfoque implica tomar el supuesto de que la distancia es exógena una vez que se condiciona en el resto de las características incluidas en el modelo. La posibilidad de que exista *residential sorting* es un aspecto central y ocupa un lugar en la discusión de todos los trabajos que utilizan la distancia como numerario. Si las familias sistemáticamente eligen vivir cerca de los centros que prefieren por razones inobservables que no son controladas por las variables incluidas en el modelo, la distancia deja de ser exógena y probablemente se esté subestimando el valor de ser asignado a un centro cercano (Hastings *et al.*, 2005; Abdulkadiroğlu *et al.*, 2017). En este trabajo se asumirá la distancia es exógena utilizando el argumento de Neilson (2013), basado en el hecho de que la ubicación residencial no determina las opciones disponibles ni la prioridad en la asignación, de la misma forma que en Uruguay.

Es importante notar que la identificación de la utilidad depende del cumplimiento de los supuestos de comportamiento, y su interpretación en términos de bienestar medido en términos de distancia depende del supuesto de exogeneidad de la distancia. Sin embargo, no es necesario asumir que el resto de las características incluidas en el modelo sean exógenas si no se les da una interpretación causal. Agarwal *et al.* (2021) señala que la endogeneidad de estas variables podría generar sesgos en la construcción

¹⁰El Anexo A.3 presenta descriptivas al respecto.

de escenarios contrafactuales que alteren la asignación, si la asignación de las características depende de la asignación de estudiantes. Esta limitación debe ser tenida en cuenta en la interpretación de los escenarios contrafactuales que se construyen en este trabajo. Es de esperar que las tasas de aprobación y desvinculación de los centros cambiaren si se altera la asignación de estudiantes. Por otro lado, los docentes efectivos podrían reaccionar a los cambios de composición e intentar cambiar su cargo de centro si tienen preferencias sobre la vulnerabilidad de los estudiantes. En la sección siguiente se discute la interpretación de los escenarios contrafactuales a la luz de esta condición.

5.4. Escenarios contrafactuales

Para analizar los efectos de la heterogeneidad de las preferencias sobre la segregación, se comparan las asignaciones de estudiantes en tres escenarios alternativos. Como medida de segregación se utiliza la proporción de estudiantes de cada grupo de vulnerabilidad asignados a cada cuartil de calidad de las características consideradas. Esto refleja en qué medida cada grupo accede a los centros de distintos niveles de calidad. Se considera que no existiría segregación si la proporción de estudiantes de cada grupo de vulnerabilidad dentro de cada cuartil de calidad fuera idéntica a la proporción de los grupos de estudiantes en la población. Esta medida se analiza gráficamente y utilizando el *Mutual Information Index (M)* (Theil y Finizza, 1971). De forma simple, en el contexto de este trabajo M mide cuánto se desvía la composición por grupos de vulnerabilidad dentro de los cuartiles de calidad con respecto a la composición de la población total. Un nivel más alto del índice muestra un mayor nivel de segregación.

El primer escenario es la asignación de los estudiantes al centro más cercano con cupo sin tomar en cuenta sus preferencias. Para su construcción se simuló la asignación utilizando el algoritmo de ANEP pero sustituyendo las preferencias declaradas por un ranking de cercanía a los centros. Es decir, se supone que cada estudiante rankea los centros sólo en función de la distancia al centro. De esta forma, el primer centro en el ranking es el centro más cercano. Este escenario se toma como un *benchmark* en el cual la segregación está fundamentalmente asociada a la distribución de estudiantes y centros en el espacio. Como se utiliza el mismo algoritmo y la misma forma de priorización de los estudiantes según el índice de vulnerabilidad, el único cambio de este escenario con respecto a la asignación real es que no se toman en cuenta las preferencias de las familias.

El segundo escenario es la asignación que efectivamente sucedió en 2022, considerando las preferencias de las familias expresadas en los rankings y asignando con el mecanismo de ANEP.

El último escenario impone a los estudiantes del grupo vulnerable las preferencias por calidad del grupo no vulnerable. En particular, se imputa a los estudiantes que reciben TUS las preferencias por

porcentaje de docentes efectivos, estudiantes por docente, matrícula, tasa de aprobación y tasa de desvinculación de los estudiantes que no reciben TUS. El primer paso en la construcción de este escenario es obtener los rankings que enviarían las familias con TUS con las nuevas preferencias imputadas. Siguiendo a [Ortega et al. \(2020\)](#) se calcula la utilidad de cada estudiante por cada centro en su *choice set* utilizando las características de los centros y las nuevas preferencias, sumando una simulación en particular del término de error. La simulación del error incorporada representa el gusto particular de cada estudiante por cada centro, y garantiza el cumplimiento de las condiciones impuestas por la ecuación (2). Por ejemplo, este término podría capturar la presencia de hermanos o la cercanía con el lugar de trabajo de algún familiar. En este escenario contrafactual, se asume que esta valoración inobservable se mantiene incambiada. Así, la utilidad estimada para el estudiante del grupo vulnerable i de ser asignado al centro j es:

$$\hat{U}_i(j) = X_{ij} \times \hat{\beta}_{cf} + \hat{\varepsilon}_{ij}$$

Donde $\hat{\varepsilon}_{ij}$ es una simulación del término de error de la valoración del estudiante i por el centro j , y $\hat{\beta}_{cf}$ es un vector de parámetros que para las características relacionadas con la calidad toma el promedio estimado para los estudiantes que no reciben TUS, y para el resto de las características toma la estimación promedio para los estudiantes que reciben TUS. Una vez obtenida la valoración de cada estudiante por cada centro en su *choice set* con las nuevas preferencias impuestas, se generan los rankings que enviarían dado ese ordenamiento asumiendo que no tendrían ningún comportamiento estratégico y se simula la asignación corriendo el algoritmo utilizado por ANEP¹¹. Para simular la asignación contrafactual se incorporan nuevamente los estudiantes que se dejaron de considerar para la estimación de preferencias (estudiantes con 15 años o más, estudiantes de escuelas especiales, estudiantes que no rankearon ninguno de los centros considerados y estudiantes que fueron asignados a centros privados) de forma de no aumentar artificialmente la cantidad de cupos disponibles en el sistema. Para estos estudiantes se utilizaron sus rankings originales.

Un primer objetivo de la comparación de los escenarios es evaluar el efecto de la introducción de un mecanismo que tome en cuenta las preferencias sobre la segregación en comparación con el *benchmark* de la segregación que ya existe a causa de la segregación residencial. Para esto alcanza con comparar el primer y segundo escenario. En segundo lugar, la comparación del segundo y tercer escenario permite medir el efecto de la heterogeneidad de las preferencias sobre la segregación en el contexto del mecanis-

¹¹El algoritmo utilizado es una réplica en Matlab basada en el código utilizado por ANEP.

mo actualmente utilizado. Por último, interesa evaluar si la introducción del algoritmo acompañado de la homogeneización de las preferencias alcanza a revertir el efecto de la segregación residencial, para lo que se comparan el primer y tercer escenario.

El análisis del acceso a centros de calidad en las asignaciones contrafactuales requiere cierta cautela. Como se mencionó en la sección anterior, las tres variables consideradas para medir la calidad de los centros podrían depender de la asignación de estudiantes. Podría suceder que al homogeneizar las preferencias aumentara la cantidad de estudiantes vulnerables en un centro con buenos resultados de aprobación y desvinculación, disminuyendo la segregación en calidad. Sin embargo, el propio cambio en la composición podría afectar los resultados educativos del siguiente período, con lo que el efecto final sobre la segregación es incierto. El punto que se pretende mostrar en este trabajo es que en el siguiente período, los nuevos resultados serán conocidos para los estudiantes que ingresan a la educación media y sus preferencias determinarán el nivel de segregación en términos de esos nuevos resultados. En definitiva, se pretende medir el efecto de la heterogeneidad de las preferencias sobre la segregación que se genera al momento de la asignación en cada período. Más allá de este punto, los efectos estimados no deben ser tomados como resultados de equilibrio. Obtenerlos requeriría ampliar el modelo de este trabajo en al menos dos dimensiones. Por un lado, los cambios en la demanda podrían generar presiones competitivas que eleven la productividad de los centros educativos (Hoxby, 2003; Neilson, 2013; Hanushek y Rivkin, 2003). Para comprender estos efectos es necesario incorporar un modelo de la oferta de centros educativos. Por otro lado, debería incorporarse al modelo la elección de las familias entre centros educativos públicos y privados. Es posible que los resultados de equilibrio supongan que algunas familias opten por centros privados, tal como encuentra Epple *et al.* (2018).

6. Resultados

6.1. Heterogeneidad en las preferencias

El Cuadro 11 presenta los resultados de la estimación de preferencias dejando correr el Gibbs Sampler por 500.000 iteraciones, con un período de *burn-in* de 250.000 iteraciones y tomando una de cada 250 simulaciones para reducir la correlación serial¹². El coeficiente de UTU representa cuántos kilómetros están dispuestos a trasladarse en promedio para pasar de ser asignados de un liceo a una UTU. Para el resto de las características, los coeficientes representan la disposición a trasladarse para ser asignado a un centro que está en el percentil 90 de la distribución en vez de a uno que está en el percentil 10 de la característica asociada al coeficiente.

Cuadro 11: Preferencias estimadas por grupo

	Sin TUS (1)	Con TUS (2)	Diferencia TUS (3)	
UTU	0,88 [0,25 ; 1,45]	2,88 [2,04 ; 3,71]	2,00 [0,96 ; 3,06]	
Recursos	Efectivos	12,69 [11,92 ; 13,5]	7,50 [6,33 ; 8,77]	-5,19 [-6,62 ; -3,71]
	Est./Doc.	-5,54 [-6,13 ; -4,93]	-4,30 [-5,23 ; -3,43]	1,24 [0,13 ; 2,29]
	Matrícula	24,09 [23,24 ; 24,93]	24,97 [23,86 ; 26,08]	0,88 [-0,46 ; 2,26]
Resultados	Aprobación	9,42 [8,82 ; 9,98]	6,42 [5,63 ; 7,2]	-3,01 [-4 ; -2,05]
	Desvinc.	-6,28 [-6,83 ; -5,69]	0,56 [-0,12 ; 1,29]	6,84 [5,94 ; 7,71]
Pares		-3,47 [-4,18 ; -2,81]	11,36 [10,34 ; 12,29]	14,82 [13,66 ; 16,04]
	Homicidios	-2,94 [-3,35 ; -2,53]	-1,43 [-1,94 ; -0,96]	1,51 [0,87 ; 2,13]
	Hurtos	-4,03 [-4,46 ; -3,59]	-8,00 [-8,76 ; -7,29]	-3,98 [-4,8 ; -3,13]

Intervalos de credibilidad al 95% entre paréntesis rectos. *UTU* es una *dummy* que vale 1 para las UTU. *Efectivos* es el porcentaje de docentes efectivos. *Est./Doc.* es la cantidad de estudiantes por cargo docente. *Matrícula* es la matrícula de estudiantes. *Aprobación* es la tasa de aprobación y *Desvinc.* la tasa de desvinculación. Todas estas variables están medidas a nivel de Ciclo Básico tomando el promedio 2021-2022 (salvo para la tasa de aprobación que es la del 2022). *Pares* es el porcentaje de estudiantes asignados al centro que reciben TUS. *Homicidios* y *Hurtos* son la cantidad de homicidios y hurtos en 2022 en el radio de 1 kilómetro del centro. Los coeficientes estimados reportan la disposición a trasladarse por pasar de un centro en el percentil 10 al 90 de la característica asociada al coeficiente. Elaboración propia con base en datos de ANEP y Ministerio del Interior.

¹²En el Anexo A.4 presenta un análisis de convergencia de la estimación.

Los resultados muestran que, con una probabilidad de 95 %¹³, todas las variables incluidas son valoradas por ambos grupos de estudiantes, salvo la tasa de desvinculación que no es considerada por los estudiantes que reciben TUS. En primer lugar, las familias valoran negativamente los centros que están más alejados, siendo que el coeficiente de la distancia es negativo (el coeficiente no se presenta en la tabla porque es el numerario del resto de los coeficientes). Las familias prefieren en promedio las UTU, valoran positivamente la presencia de docentes efectivos, el tamaño de los centros en términos de matrícula y la tasa de aprobación. A la vez, valoran negativamente la cantidad de estudiantes por docente, la tasa de desvinculación (salvo los que reciben TUS, que no la consideran) y las medidas de criminalidad en el entorno. Ambos grupos valoran positivamente la presencia de estudiantes de su mismo tipo en el centro, siendo que el porcentaje de estudiantes con TUS asignados al centro es valorado negativamente por las familias sin TUS y positivamente por las familias con TUS. En línea con lo esperado, las características asociadas a la calidad de los centros son valoradas positivamente.

Adicionalmente, todas las variables salvo la matrícula son valoradas de forma distinta por los estudiantes de ambos grupos. En particular, los estudiantes que reciben TUS le dan menos valor a todas las características asociadas a la calidad. Esto está en línea con la hipótesis manejada en la sección 4.3 de que los estudiantes más vulnerables podrían no estar accediendo a la misma información sobre la calidad de los centros o sobre la importancia de esta para las trayectorias educativas y los retornos de la educación. De cualquier forma, la estimación no permite distinguir estos dos efectos. Adicionalmente, los estudiantes que reciben TUS valoran más la opción de UTU por sobre liceo.

En el Cuadro 19 del Anexo A.5 se presentan estimaciones de las preferencias con definiciones alternativas de los grupos de vulnerables y no vulnerables. Más allá de que estas estimaciones dan robustez al resultado de que las preferencias son heterogéneas entre grupos, también ofrecen algunos indicios sobre los mecanismos que explican estas diferencias. En línea con de Melo y Machado (2018), la valoración de la UTU es aún más grande cuando se agrupa a los estudiantes utilizando puntos de corte del índice de vulnerabilidad, que también es creciente en los resultados escolares negativos. Sin embargo, la diferencia no es mayor al considerar estudiantes que reciben doble TUS en vez de TUS, por lo que la preferencia diferencial por UTU parece estar más asociada a los peores resultados educativos previos que a la intensidad de la pobreza. Por otro lado, la brecha en la valoración de los resultados educativos (tasas de aprobación y desvinculación) aumenta con las tres definiciones de vulnerabilidad alternativas, pero la brecha más grande se verifica al considerar los estudiantes con niveles altos del índice de vulnerabilidad.

¹³Como la econometría bayesiana no adopta un enfoque frecuentista de la probabilidad, no se utiliza el valor p para definir la significación estadística de las estimaciones. Alternativamente, se construyen intervalos de credibilidad utilizando la distribución posterior estimada. En el Cuadro 11 se proporcionan intervalos que contienen el parámetro real con una probabilidad del 95 %.

Esto sugiere que la menor valoración está asociada a la intensidad de la pobreza pero en mayor medida a los peores resultados previos del estudiante. La diferencia en la valoración del porcentaje de docentes efectivos también es mayor cuando se considera el índice de vulnerabilidad o recibir doble TUS, pero sin distinción entre los distintos puntos de corte.

La heterogeneidad identificada en las preferencias también es robusta a modificar la definición de los *choice set*, incluir efectos fijos por centro, quitar de la consideración los estudiantes que no cuentan con centros cercanos o permitir heterogeneidad en la valoración de la distancia. Todos estos ejercicios de robustez se detallan en el Anexo A.5.

6.2. Segregación

El Cuadro 12 muestra el índice de segregación calculado en la asignación que sólo considera las distancias y su variación en los dos escenarios en que la asignación toma en cuenta las preferencias¹⁴. La columna 2 muestra que al introducirse el algoritmo de asignación de ANEP, la segregación aumenta en las tres características consideradas de manera significativa a un 95 % de confianza. En el caso de la segregación en cuartiles de tasa aprobados y desvinculación, la segregación se duplica. Este aumento de la segregación puede asociarse al hecho de que se tomen en cuenta las preferencias de las familias, en un contexto donde estas son heterogéneas. Este resultado supone que existe un conjunto de centros que tienen en su proximidad tanto estudiantes vulnerables como no vulnerables. Cuando la asignación considera únicamente la proximidad, estos centros reciben estudiantes vulnerables y no vulnerables, determinando un menor nivel de segregación. Parte de lo que explica el aumento de la segregación puede observarse en el Cuadro 13. Los estudiantes menos vulnerables aprovechan el algoritmo para expresar sus preferencias por calidad y moverse a los mejores centros, saliendo de su centro cercano si es de mala calidad. Su asignación aumenta en los centros del cuartil más alto en las tres características al tiempo que disminuye en los centros de peor calidad. Mientras tanto, los estudiantes vulnerables siguen eligiendo en promedio centros cercanos similares a los que obtendrían en la asignación por proximidad.

¹⁴En el Anexo A.5 se presentan gráficamente los cambios en la asignación.

Cuadro 12: *Benchmark* de segregación asignando por distancia y variación en escenarios que toman en cuenta las preferencias (índice *M*)

	Asignación por distancia (1)	Pref. heterogéneas (2)	Pref. homogéneas (3)
Efectivos	0,019 [0,017;0,021]	0,005 [0,002;0,008]	-0,001 [-0,004;0,001]
Aprobados	0,004 [0,003;0,005]	0,004 [0,002;0,005]	0,000 [-0,002;0,001]
Desvincul.	0,008 [0,007;0,009]	0,008 [0,006;0,01]	0,000 [-0,002;0,002]

Se reporta el índice *M* y el intervalo de confianza al 95% construido mediante *bootstrap* (5000 iteraciones) entre paréntesis rectos. Las columnas 2 y 3 presentan la variación del índice de segregación en cada escenario en comparación al valor de la columna 1. Elaboración propia con base en datos de ANEP.

Cuadro 13: Variación de estudiantes asignados contra *benchmark* de asignación por distancia

	Preferencias heterogéneas				Preferencias homogéneas			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Efectivos								
Doble TUS	-52	37	29	-14	-127	-144	69	202
TUS	7	38	-115	70	-39	-39	-80	158
AFAM-PE	-88	-45	-210	345	-85	-12	-262	361
Sin PS	-49	-378	-167	599	-4	-338	-171	519
Total	-182	-348	-463	1.000	-255	-533	-444	1.240
Aprobación								
Doble TUS	37	48	146	-231	-258	53	256	-51
TUS	-1	55	21	-75	-79	29	57	-7
AFAM-PE	-304	149	-25	182	-265	97	-8	178
Sin PS	-515	-171	196	495	-438	-96	171	369
Total	-783	81	338	371	-1.040	83	476	489
Desvinculación								
Doble TUS	-50	62	222	-234	-300	-7	233	74
TUS	-45	55	134	-144	-134	15	121	-2
AFAM-PE	-200	-168	279	91	-139	-143	224	60
Sin PS	-423	-303	-111	842	-334	-157	-152	649
Total	-718	-354	524	555	-907	-292	426	781

Los valores reportados son la cantidad de asignados en el escenario *benchmark* de asignación por distancia menos la cantidad de asignados en la asignación con preferencias heterogéneas u homogéneas según corresponda. Elaboración propia con base en datos de ANEP.

Por otro lado, la columna 3 del Cuadro 12 muestra que una vez que se imponen preferencias homogéneas por las características asociadas a la calidad educativa, la introducción del algoritmo no exa-

cerba ni disminuye la segregación asociada a la ubicación residencial de los estudiantes. La heterogeneidad en las preferencias por el resto de las características (presencia de estudiantes del mismo grupo de vulnerabilidad, criminalidad en el entorno y subsistema educativo) no parece estar jugando ningún papel en la segregación que se introduce al considerar las preferencias de las familias en el mecanismo de asignación. Es posible que estas características sean más salientes para las familias que las características asociadas a la calidad educativa. Este es un aspecto interesante, siendo que las preferencias que introducen segregación son las que podrían estar asociadas a la falta de información de las familias más vulnerables.

Más allá de que la consideración de las preferencias no altera la segregación si las preferencias por calidad son homogéneas entre familias, sí tiene un efecto sobre la proporción absoluta de estudiantes asignados a los centros de mayor calidad, como se observa en el Cuadro 13. Cuando se le imponen a los estudiantes vulnerables las preferencias por calidad del resto, tienen un comportamiento similar al descrito para los estudiantes no vulnerables y abandonan los centros de menor calidad para elegir centros mejores. En definitiva, si las preferencias por calidad son homogéneas entre grupos, la introducción de un mecanismo que las tenga en cuenta produce una mejora en la eficiencia del sistema, en el sentido de que los cupos disponibles en los centros de mayor calidad son aprovechados y se observa un traslado de estudiantes desde centros de peor calidad a ellos. Por otro lado, también supone costos para algunos estudiantes no vulnerables que son asignados a un centro de menor calidad.

Sin embargo, aún en el escenario en que las preferencias por calidad son homogéneas, la priorización de la voluntad de los estudiantes más vulnerables que existe en el algoritmo de ANEP no logra reducir la segregación asociada a dónde viven los estudiantes. Esto señala que las familias valoran la cercanía de los centros educativos y por lo tanto darles la libertad a las familias más vulnerables de elegir el centro que prefieran no es suficiente para garantizar el acceso igualitario a centros de calidad, aún cuando valoren esta característica en la misma medida que el resto de los estudiantes.

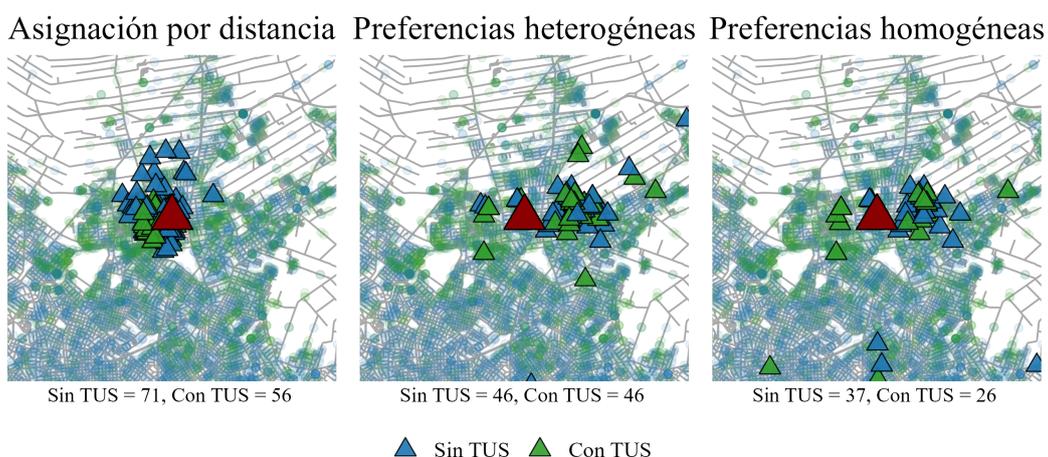
6.3. Dinámica de los cambios

El aumento de la segregación al considerar las preferencias heterogéneas de las familias señala que existen centros con distintos niveles de calidad que tienen en su proximidad estudiantes vulnerables y no vulnerables. Cuando la asignación se realiza considerando únicamente la proximidad, ambos grupos son asignados a estos centros en proporciones similares. Sin embargo, cuando los estudiantes pueden elegir según sus preferencias, los estudiantes menos vulnerables abandonan los centros de mala calidad y eligen los de buena calidad en mayor medida que los estudiantes vulnerables, aumentando la segregación en el

sistema. A continuación se presentan dos ejemplos reales que muestran esta dinámica, a los que se les llamará A y B.

Caso A: un centro con bajos valores en los indicadores de calidad. El liceo A de Montevideo se encuentra entre el 25 % de centros con menor porcentaje de docentes efectivos, menor tasa de aprobación y mayor tasa de desvinculación. Si se consideran 5 kilómetros a la redonda del liceo, hay 3690 estudiantes que participaron del sistema de preinscripción, de los cuales un 42 % recibe TUS. Cuando la asignación se realiza considerando únicamente la proximidad, un 44 % de los estudiantes asignados recibe TUS. Esta proporción aumenta al 50 % de los estudiantes cuando se toman en cuenta las preferencias al momento de la asignación, y además estas son heterogéneas. En este escenario, el centro tiene una sobre representación de estudiantes vulnerables aún en la zona en la que se encuentra. Como se observa en la Figura 6, el aumento en esta proporción se explica porque la cantidad de estudiantes no vulnerables que eligen el centro disminuye en mayor medida que la cantidad de estudiantes vulnerables. Cuando se impone que las preferencias por calidad sean homogéneas, la participación de los estudiantes que reciben TUS entre los asignados cae a un 41 %. Esta proporción es únicamente un reflejo de la participación de los estudiantes vulnerables en el entorno del centro, es decir, la introducción de las preferencias no supone un nivel adicional de segregación al que explica la ubicación residencial de los estudiantes. Por último, la imposición de preferencias homogéneas en lugar de heterogéneas supone una reducción de un 30 % en la matrícula de este centro. En total, un 17 % de los centros de todo el país muestran esta misma dinámica.

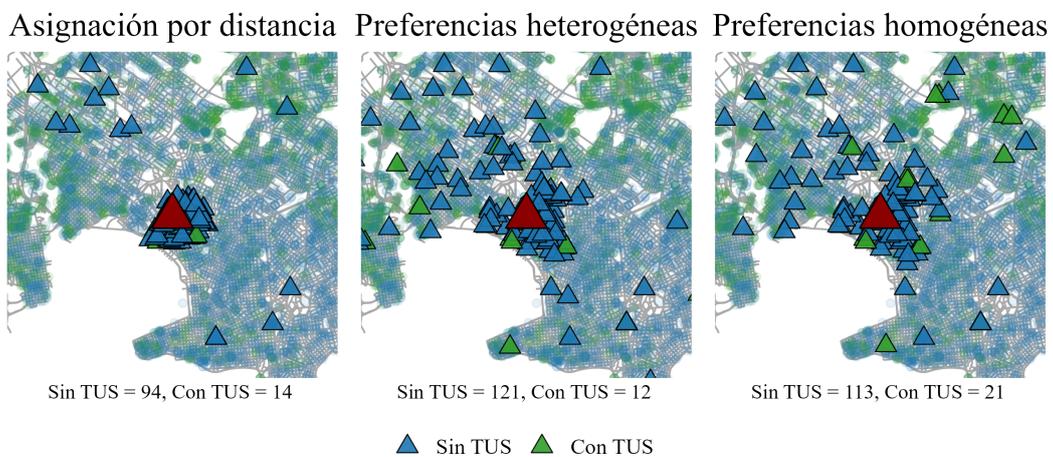
Figura 6: Asignación a un liceo con bajos valores en los indicadores de calidad



Los círculos de colores muestran el resto de los estudiantes que participan de la preinscripción según su grupo de vulnerabilidad. Elaboración propia con base en datos de ANEP.

Caso B: un centro con altos valores en los indicadores de calidad. El liceo B de Montevideo muestra la dinámica contraria. Se encuentra en el cuartil más alto de los tres indicadores de calidad considerados. Se encuentra en una zona menos vulnerable, siendo que de los 4183 estudiantes que viven a menos de 5 kilómetros, un 27% recibe TUS. La asignación por proximidad resulta en una sub representación de estos estudiantes, siendo que un 13% de los asignados reciben TUS. Cuando se toman en cuenta las preferencias para la asignación, la segregación se exagera siendo que el porcentaje de estudiantes con TUS asignados a este centro disminuye a un 9%. Como muestra la Figura 7, esto resulta principalmente de una mayor cantidad de estudiantes no vulnerables que eligen este centro por tener un nivel mayor de calidad. Cuando las preferencias por calidad son homogéneas, la cantidad de estudiantes con TUS que eligen el centro aumenta. Además, como el centro se encuentra con todos sus lugares utilizados, los estudiantes vulnerables que eligen este centro desplazan algunos estudiantes no vulnerables, dado que son priorizados por el mecanismo de asignación. Esto resulta en una mayor participación de los estudiantes vulnerables en el centro, alcanzando el 16%. En total, un 6% de los centros del país tienen el mismo comportamiento.

Figura 7: Asignación a un liceo con altos valores en los indicadores de calidad



Los círculos de colores muestran el resto de los estudiantes que participan de la preinscripción según su grupo de vulnerabilidad. Elaboración propia con base en datos de ANEP.

7. Comentarios finales

Este trabajo pretende ser una primera aproximación a comprender las preferencias de las familias por centros educativos en el sistema de educación media pública uruguayo y cómo se relacionan estas preferencias con el acceso desigual a centros de calidad y la segregación. Este objetivo es relevante porque, a pesar de la definición política de otorgar prioridad en la elección escolar a los estudiantes más vulnerables, su acceso a los centros de mayor calidad es restringido.

Con ese objetivo se construyó un modelo estructural de demanda que utiliza los rankings enviados por las familias al momento de la preinscripción. Los supuestos principales del modelo se apoyan en las propiedades teóricas del algoritmo utilizado por ANEP y en la información que se asume que tienen disponible las familias. La estimación de los parámetros se realizó utilizando un procedimiento bayesiano de aumento de datos.

Las preferencias estimadas muestran que las familias más vulnerables le dan un menor peso a las características de los centros educativos relacionadas con la calidad educativa. Este resultado es robusto a la definición del grupo vulnerable, la definición del *choice set*, la introducción de efectos fijos por centro educativo, la consideración de distintas muestras de estudiantes y a permitir que la valoración por distancia sea también heterogénea. Adicionalmente, se mostró que esta heterogeneidad lleva a que la introducción de un mecanismo de asignación que toma en cuenta las preferencias de los estudiantes exacerba la segregación existente por la ubicación espacial de los estudiantes y centros educativos. El mecanismo que está por detrás de este aumento en la segregación es que los estudiantes menos vulnerables expresan sus preferencias en el sistema de asignación y acceden a centros de más calidad, mientras que la asignación de los estudiantes vulnerables no cambia demasiado con respecto a la asignación que sólo toma en cuenta la distancia. Por último, se generó un escenario contrafactual en el que las familias vulnerables tienen la misma valoración de las características asociadas a la calidad educativa que las familias no vulnerables. La homogeneización de las preferencias resultaría en una mejora en el acceso de los estudiantes vulnerables a los centros de mayor calidad y en menores niveles de segregación en el sistema que cuando las preferencias son heterogéneas. Adicionalmente supondría un aumento de la matrícula agregada en centros de mayor calidad, y una reducción en los centros de peor calidad. Por otro lado, este cambio tendría un costo para algunos estudiantes no vulnerables que serían desplazados a centros de menor calidad que los que acceden actualmente.

Estos resultados son relevantes para la evaluación de la política implementada por ANEP al introducir un mecanismo centralizado de asignación que toma en cuenta las preferencias de las familias y

prioriza a los estudiantes más vulnerables. En particular, se mostró que esta política tiene un costo en términos de una mayor segregación en el acceso a centros educativos de calidad, que es consecuencia de la heterogeneidad en las preferencias de las familias. En caso de que las preferencias de las familias vulnerables estuvieran asociadas a diferencias en el acceso a la información, intervenir la información podría mejorar los resultados del mecanismo en términos de igualdad en el acceso a centros de calidad, aunque no sería suficiente para reducir las brechas existentes por la segregación residencial.

La principal limitación de este trabajo es que sólo alcanza a medir efectos parciales de eliminar la heterogeneidad en las preferencias. El modelo desarrollado no permite evaluar cómo cambiaría la calidad de los centros educativos ante el aumento de la demanda por calidad de los estudiantes y los cambios en la composición estudiantil. Tampoco permite evaluar si podría tener un efecto sobre la decisión de las familias entre elegir un centro público o uno privado. Estas cuestiones podrían abordarse en trabajos posteriores que incorporen un modelo de la oferta por parte de los centros educativos y amplíen el modelo de demanda para considerar también la decisión entre el sistema público y privado.

Además de este camino, los resultados de este trabajo dejan abiertas las siguientes líneas de investigación. Una línea que surge de forma inmediata es comprender qué explica la heterogeneidad existente en las preferencias de estudiantes según su nivel de vulnerabilidad. Este es un aspecto clave para las posibilidades de recomendación de política. La brecha en las preferencias de las familias por la calidad educativa puede reflejar restricciones en el acceso a la información de las familias más vulnerables. En particular, las familias más desfavorecidas podrían no contar con información sobre las características de los centros porque obtener esta información implica asumir costos de búsqueda. También podrían tener información incorrecta sobre la importancia de la calidad del centro sobre los resultados educativos, y la importancia de estos resultados en los ingresos futuros. Esta hipótesis podría contrastarse mediante experimentos que intervengan de forma controlada la información que tienen las familias con respecto a estos puntos (por ejemplo, ver [Allende *et al.* \(2019\)](#)). Este trabajo muestra que conocer la forma de homogeneizar las preferencias de los estudiantes podría habilitar un mayor beneficio del mecanismo de asignación utilizado por ANEP.

Referencias

- Abdulkadiroğlu, A., Agarwal, N., y Pathak, P. A. (2017). The welfare effects of coordinated assignment: Evidence from the new york city high school match. *American Economic Review*, 107(12):3635–3689.
- Abdulkadiroglu, A., Pathak, P. A., Roth, A. E., y Sönmez, T. (2006). Changing the boston school choice mechanism.
- Abdulkadiroğlu, A., Pathak, P. A., Schellenberg, J., y Walters, C. R. (2020). Do parents value school effectiveness? *American Economic Review*, 110(5):1502–39.
- Abdulkadiroğlu, A. y Sönmez, T. (2003). School choice: A mechanism design approach. *American economic review*, 93(3):729–747.
- Abdulkadiroglu, A. y Sönmez, T. (2013). Matching markets: Theory and practice. *Advances in Economics and Econometrics*, 1:3–47.
- Agarwal, N. y Somaini, P. (2018). Demand analysis using strategic reports: An application to a school choice mechanism. *Econometrica*, 86(2):391–444.
- Agarwal, N. y Somaini, P. (2020). Revealed preference analysis of school choice models. *Annual Review of Economics*, 12:471–501.
- Agarwal, Nikhil and Somaini, Paulo and others (2021). Empirical models of non-transferable utility matching. *Online and Matching-Based Market Design*.
- Akerlof, G. A. (1970). 4. the market for ‘lemons’: quality uncertainty and the market mechanism. *Market Failure or Success*, 66.
- Allende, C. (2019). Competition under social interactions and the design of education policies. *Job Market Paper*.
- Allende, Claudia and Gallego, Francisco and Neilson, Christopher and others (2019). Approximating the equilibrium effects of informed school choice. Technical report.
- Berry, S., Levinsohn, J., y Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pp. 841–890.
- Berry, S., Levinsohn, J., y Pakes, A. (2004). Differentiated products demand systems from a combination of micro and macro data: The new car market. *Journal of political Economy*, 112(1):68–105.

- Berry, S. T. y Haile, P. A. (2021). Foundations of demand estimation. En *Handbook of industrial organization*, volumen 4, pp. 1–62. Elsevier.
- Black, S. E. (1999). Do better schools matter? parental valuation of elementary education. *The quarterly journal of economics*, 114(2):577–599.
- Black, S. E. y Machin, S. (2011). Housing valuations of school performance. En *Handbook of the Economics of Education*, volumen 3, pp. 485–519. Elsevier.
- de Melo, G. y Machado, A. (2018). Educational trajectories. evidence from uruguay. *International Journal of Educational Research*, 92:110–134.
- Echenique, F. y Yenmez, M. B. (2015). How to control controlled school choice. *American Economic Review*, 105(8):2679–94.
- Epple, D., Jha, A., y Sieg, H. (2018). The superintendent’s dilemma: Managing school district capacity as parents vote with their feet. *Quantitative Economics*, 9(1):483–520.
- Fack, G., Grenet, J., y He, Y. (2019). Beyond truth-telling: Preference estimation with centralized school choice and college admissions. *American Economic Review*, 109(4):1486–1529.
- Failache, E., Salas, G., y Vigorito, A. (2018). Desarrollo en la infancia y trayectorias educativas de los adolescentes. un estudio con base en datos de panel para uruguay. *El trimestre económico*, 85(337):81–113.
- Gale, D. y Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.
- Gelman, A. y Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.
- Hanushek, E. A., Lavy, V., y Hitomi, K. (2008). Do students care about school quality? determinants of dropout behavior in developing countries. *Journal of Human Capital*, 2(1):69–105.
- Hanushek, E. A. y Rivkin, S. G. (2003). Does public school competition affect teacher quality? En *The economics of school choice*, pp. 23–48. University of Chicago Press.
- Hastings, J. S., Kane, T. J., y Staiger, D. O. (2005). Parental preferences and school competition: Evidence from a public school choice program.

- Hastings, J. S. y Weinstein, J. M. (2008). Information, school choice, and academic achievement: Evidence from two experiments. *The Quarterly journal of economics*, 123(4):1373–1414.
- Hoxby, C. M. (2003). School choice and school productivity. could school choice be a tide that lifts all boats? En *The economics of school choice*, pp. 287–342. University of Chicago Press.
- INEEd (2023). Informe sobre el estado de la educación en Uruguay 2021-2022.
- Jensen, R. (2010). The (perceived) returns to education and the demand for schooling. *The Quarterly Journal of Economics*, 125(2):515–548.
- Kapor, A., Neilson, C., y Zimmerman, S. (2016). Heterogeneous beliefs and school choice. *American Economic Review*.
- Kojima, F. (2012). School choice: Impossibilities for affirmative action. *Games and Economic Behavior*, 75(2):685–693.
- Laverde, M. (2022). Distance to schools and equal access in school choice systems. Technical report.
- McCulloch, R. y Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64(1-2):207–240.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.
- Méndez-Errico, L. y Ramos, X. (2022). Selection and educational attainment: why some children are left behind? evidence from a middle-income country. *Education Economics*, 30(6):624–643.
- Neilson, C. (2019). The rise of centralized choice and assignment mechanisms in education markets around the world. *Unpublished manuscript*.
- Neilson, C. e. a. (2013). Targeted vouchers, competition among schools, and the academic achievement of poor students. *Documento de trabajo. Yale University*.
- Oosterbeek, H., Sóvágó, S., y van der Klaauw, B. (2021). Preference heterogeneity and school segregation. *Journal of Public Economics*, 197:104400.
- Ortega, J., Klein, T., y Aue, R. (2020). What happens when separate and unequal school districts merge?
- Pathak, P. A. y Sönmez, T. (2008). Leveling the playing field: Sincere and sophisticated players in the Boston mechanism. *American Economic Review*, 98(4):1636–52.

Theil, H. y Finizza, A. J. (1971). A note on the measurement of racial integration of schools by means of informational concepts.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Van Buuren, S. y Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of statistical software*, 45:1–67.

Van der Vaart, A. W. (2000). *Asymptotic statistics*, volumen 3. Cambridge university press.

A. Anexo

A.1. Preparación de los datos

Características de los estudiantes. La información sobre los estudiantes proviene de los datos que completan al momento de la preinscripción. Estos incluyen los rankings, el liceo predeterminado, el centro asignado y el paso en el que se asignó. También cuenta con información sobre la escuela a la que asistieron: el quintil socioeconómico (según clasificación de ANEP) y el tipo de la escuela (especial, de tiempo completo o extendido, etc.). Sobre los estudiantes, está disponible el índice de vulnerabilidad y sus componentes (cantidad de faltas y nota final de los últimos tres años escolares, si recibe AFAM-PE, TUS y/o doble TUS), su edad, sexo y dirección.

La dirección de los estudiantes en conjunto con información georreferenciada de los centros educativos brindada por ANEP se utilizó para calcular la distancia de cada estudiante a cada opción. La georreferenciación de los estudiantes a partir de sus direcciones fue realizada por Matías Faggetti en el marco de su tesis de maestría. Agradezco a Matías por tener la amabilidad de poner estos datos a disposición de este proyecto. Adicionalmente, se imputaron las coordenadas de 197 estudiantes que no contaban con dirección en los datos originales. Para los casos en que había otros estudiantes de la misma escuela (184) se tomaron las coordenadas aleatoriamente de otro estudiante de la misma escuela. Para los 13 estudiantes restantes se tomaron aleatoriamente las coordenadas de otro estudiante del mismo departamento y que hubiera marcado la misma primera opción. Con las coordenadas se calcularon las distancia geodésicas en el software *R* usando el paquete *geosphere*.

Los rankings de los estudiantes se editaron de forma de eliminar los centros que no son considerados (centros privados, de educación no formal y planes extraedad). Nótese que los supuestos tomados en el modelo teórico de comportamiento hacen que esta edición no afecte la estimación de los parámetros. El procedimiento sólo asume que las opciones que se mantienen en el ranking editado son mejores que las opciones que no están rankeadas. Se supone que esta relación es independiente de haber rankeado opciones que no son consideradas en este análisis.

Del total de 37.259 estudiantes considerados, 115 estudiantes tienen algún dato faltante. El procedimiento utilizado no admite la inclusión de datos faltantes. En particular, 112 estudiantes no tienen información sobre los planes sociales que reciben, 3 estudiantes no tienen información de las inasistencias y 2 estudiantes no tienen información sobre sus notas. Se imputaron estos datos faltantes usando el paquete *mice* del software *R* usando *predictive mean matching* (Van Buuren y Groothuis-Oudshoorn, 2011). El método busca la observación más parecida en base a las características observadas y le imputa su valor en

el faltante. Se realizó el procedimiento utilizando todas las características observadas de los estudiantes, incluyendo edad, sexo, escuela, tipo de escuela, barrio, notas y faltas. El método realiza simulaciones de la imputación hasta que el valor imputado converja. Los detalles del procedimiento pueden verse en la documentación del paquete disponible online. Si bien esta estrategia introduce cierto error de medida, en acuerdo con [Ortega *et al.* \(2020\)](#) se asume que la ganancia de incorporar a estos estudiantes al análisis es mayor al costo asumido. En particular, contar con todos los estudiantes es relevante para la estimación de nuevas asignaciones contrafactuales.

Características de los centros. Los datos sobre características de los centros provienen principalmente de ANEP, aunque se registran por separado para el caso de secundaria (Dirección General de Educación Secundaria) y educación técnico-profesional (Dirección General de Educación Técnico-Profesional). Estos datos incluyen información sobre los resultados educativos y sobre características del plantel docente. En el caso de secundaria, la información original se encuentra a nivel de cada plan/liceo y está dividida entre la información sobre docentes e información sobre los resultados educativos. Para los docentes se tiene información de 2018 a 2022. Si un docente dicta más de una asignatura en el mismo liceo se lo cuenta una vez por cada asignatura que dicta con un grado distinto. La información incluye la cantidad de horas semanales en el centro en tramos de 10 horas, la tenencia de título, efectividad y grado, sexo y edad en tramos. Para los resultados educativos se tiene información de 2020, 2021 y 2022 (salvo porcentaje de aprobados en 2022). Los resultados pueden abrirse a nivel de cada año o ciclo básico. En el caso de UTU, los datos de docentes se encuentran a nivel de cada asignación horaria. Es decir, para un docente en un centro se tiene una línea por cada grupo que dicta y también por otras tareas de coordinación. Se eliminaron las horas referidas a tareas de coordinación y se tomó el mismo criterio que en secundaria, contando un docente por cada asignatura dictada con un grado distinto. Estos datos están disponibles desde 2015 a 2022 e incluyen la cantidad de horas, el carácter del cargo (efectivo o interino), la fecha de posesión del cargo, sexo y edad. Los datos originales de estudiantes están disponibles desde 2015 a 2021 y también están a nivel de cada estudiante, por lo que se agruparon a nivel de cada programa/centro.

Del total de 403 centros considerados, hay 15 para los que no se tiene información de aprobados y 7 para los que no se tiene ni aprobados ni desvinculación. En línea con lo explicado para los datos faltantes de estudiantes, se utilizó el mismo procedimiento de imputación utilizando como características observables la información sobre docentes y el barrio, considerando por separado liceos y UTU.

Para construir las variables que miden el delito en la cercanía de los centros se utilizaron datos georreferenciados de denuncias de hurtos y rapiñas y registros de homicidios del Ministerio del Interior,

disponibles a nivel de cada acto entre 2016 y 2022. Se utilizó la georreferenciación de los centros educativos para contar la cantidad de denuncias de cada delito en un entorno de 1 kilómetro.

A.2. Construcción del índice de vulnerabilidad

El índice de vulnerabilidad toma valores entre 0 y 1 y se construye a partir de la ponderación de un índice de 0 a 1 dentro de cada categoría, de la siguiente forma:

$$I = 0,3 \times PlanesSociales + 0,1 \times Notas + 0,3 \times Faltas + 0,3 \times Extraedad$$

Planes sociales.

$$PlanesSociales = \begin{cases} 0 & \text{si no recibe planes sociales} \\ 0,25 & \text{si recibe AFAM} \\ 0,5 & \text{si recibe AFAM y TUS} \\ 1 & \text{si recibe AFAM y doble TUS} \end{cases}$$

Notas. Se consideran las notas de los últimos 3 años escolares de la siguiente forma:

$$Notas = 0,5 \times (1(N_t < B) \times 0,5(N_t = B)) + 0,25 \times 1(N_{t-1} = B) + 0,25 \times 1(N_{t-2} = B)$$

Donde B es la nota mínima de aprobación y N_t es la nota de aprobación del año t (en el año de la preinscripción es la última nota disponible).

Inasistencias. Se consideran las inasistencias de los últimos 3 años escolares de la siguiente forma:

$$Faltas = \begin{cases} 0,1 \times \sum_{t-2}^t F_t & \text{si } \sum_{t-2}^t F_t < 100 \\ 1 & \text{si } \sum_{t-2}^t F_t > 100 \end{cases}$$

Donde F_t es la cantidad de inasistencias en el año t .

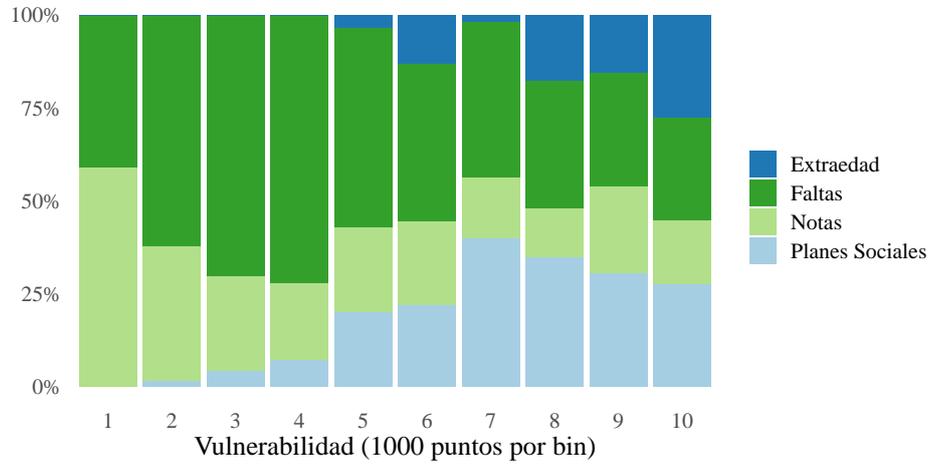
Extraedad. Se considera EE como la cantidad de años por encima de la edad esperada al 30 de abril del año de prescripción (11 años).

$$Extraedad = \begin{cases} 0 & \text{si } EE = 0 \\ 0,5 & \text{si } EE = 1 \\ 1 & \text{si } EE > 1 \end{cases}$$

La Figura 8 muestra el peso relativo de cada componente en el índice de vulnerabilidad, en tramos

de 1000 puntos sobre 10000. Se observa que para valores más bajos, la cantidad de inasistencias es el factor más relevante del índice, y va perdiendo peso frente a los planes sociales en niveles más altos.

Figura 8: Composición del índice de vulnerabilidad por tramo



Elaboración propia con base en datos de ANEP.

A.3. Características de los centros

Variación en los *choice sets*. Como se menciona en la [Subsección 5.3](#), la estimación necesita que exista variación de las características dentro de los *choice set* de los estudiantes. Para comprobarlo, se calculó el desvío estándar de cada característica dentro del *choice set* de cada estudiante.

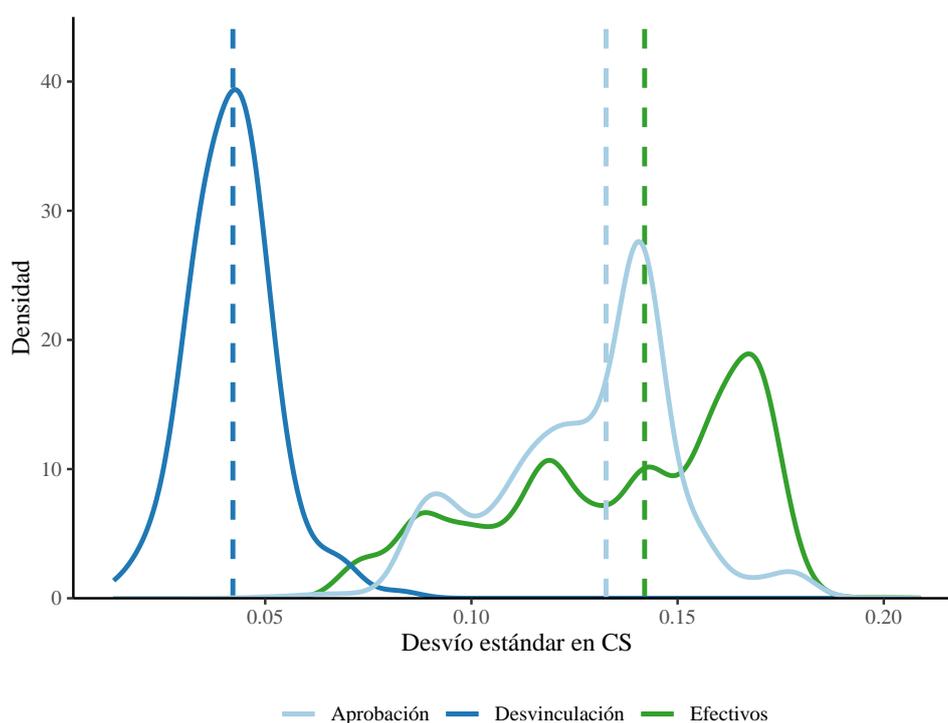
Cuadro 14: Desvío estándar de características dentro de CS

	Total	Montevideo	Resto del país
Efectivos	0,14 [0,06 - 0,14 - 0,21]	0,16 [0,08 - 0,17 - 0,18]	0,12 [0,06 - 0,12 - 0,21]
Est./Doc.	1,84 [0,61 - 1,84 - 3,28]	1,83 [1,13 - 1,84 - 2,19]	1,85 [0,61 - 1,84 - 3,28]
Matricula	217 [50 - 223 - 321]	224 [178 - 224 - 294]	214 [50 - 218 - 321]
Aprobación	0,13 [0,04 - 0,13 - 0,20]	0,14 [0,09 - 0,14 - 0,18]	0,12 [0,04 - 0,12 - 0,20]
Desvinc.	0,04 [0,01 - 0,04 - 0,08]	0,04 [0,02 - 0,04 - 0,06]	0,04 [0,01 - 0,04 - 0,08]
Pares	0,17 [0,07 - 0,17 - 0,28]	0,22 [0,08 - 0,22 - 0,26]	0,15 [0,07 - 0,15 - 0,28]
Homicidios	1,54 [0,00 - 1,25 - 3,80]	2,92 [0,70 - 2,92 - 3,80]	0,94 [0,00 - 0,70 - 3,49]
Hurtos	385 [17 - 247 - 908]	735 [90 - 772 - 908]	231 [17 - 212 - 864]

Promedio [Mínimo - Mediana - Máximo] Elaboración propia con base en datos de ANEP y Ministerio del Interior.

El Cuadro 14 muestra que en promedio y en la mediana todas las características tienen variación dentro del *choice set* de los estudiantes. Sin embargo, hay estudiantes para los cuales la proporción de docentes efectivos, las tasas de aprobación y desvinculación y la cantidad de homicidios varían poco dentro de su *choice set*. En la Figura 9 se muestra que la cantidad de estudiantes que no tienen variación en la proporción de docentes efectivos, tasa de aprobación y desvinculación es baja, por lo que no representaría una amenaza sustantiva a la identificación.

Figura 9: Variación de características en CS



Elaboración propia con base en datos de ANEP.

Estabilidad de las características. La interpretación que se hace en el modelo de las variables incluidas supone que las familias tienen expectativas relativamente precisas sobre la calidad educativa y los recursos de los centros, que a la vez están correlacionados con las variables del modelo. Para que este supuesto se cumpla, es necesario que las características incluidas sean relativamente estables en el tiempo. Como es esperable que la posición relativa de los centros en términos de la calidad esperada por las familias no cambie, importa que los niveles relativos entre los centros de las variables no cambien. El Cuadro 15 muestra que la cantidad de estudiantes por docente varía muy poco dentro de cada centro para el período en el que se cuenta con datos. La proporción de docentes efectivos y las tasas de aprobación y desvinculación sí muestran cambios en el período.

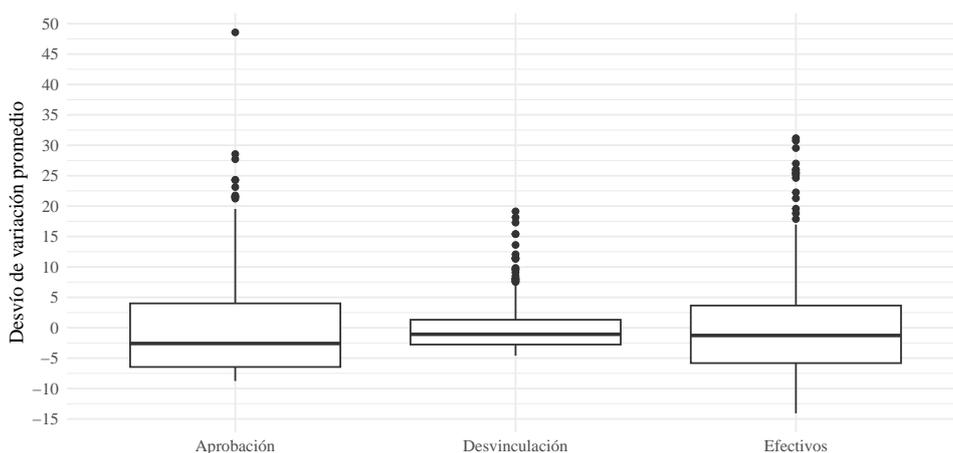
Cuadro 15: Variación de características dentro del centro (p.p.)

	Mín	p25	Mediana	Promedio	p75	Máx
Aprobación	0	2,3	6,2	8,8	12,8	57,3
Desvinc.	0	1,8	3,5	4,6	5,9	23,7
Efectivos	0	8,3	12,8	14,1	17,7	45,2
Est./Doc.	0	0,3	0,6	0,7	1,0	5,3

Para estudiantes por docente, aprobación y desvinculación se considera el período 2020-2022. Para efectivos se considera el período 2018-2022. Se calcula la variación como la diferencia entre el máximo y mínimo valor registrado en el período. Elaboración propia con base en datos de ANEP.

Sin embargo, la Figura 10 muestra que la mayor parte de esa variación estaría asociada a tendencias generales de todos los centros. La mayoría de los centros se desvía de la variación promedio en menos de 5 puntos porcentuales.

Figura 10: Variación en el tiempo de características por centro



Elaboración propia con base en datos de ANEP.

A.4. Análisis de convergencia

La convergencia de las MCMC se comprobó mediante el análisis gráfico y el análisis del Potential Scale Reduction Factor (PSRF) y Effective Sample Size (ESS). El PSRF (Gelman y Rubin, 1992) es una medida de cómo cambiaría la distribución posterior estimada si se incrementara infinitamente la cantidad de iteraciones. Se espera que si la cadena efectivamente convergió este estadístico sea cercano a 1. Aunque no exista una regla, en general se descarta la convergencia si es mayor a 1,2. El Cuadro 16 muestra que la estimación del PSRF es menor a 1,2 para todos los parámetros aún considerando la cota superior del intervalo de confianza al 95 %. Por otro lado, el ESS es una medida de la correlación serial de las simulaciones. Como las simulaciones sucesivas de los parámetros en las cadenas de Markov están correlacionadas, este estadístico representa la cantidad de simulaciones independientes equivalente a las simulaciones de la cadena. En definitiva, muestra cuántos valores independientes se están utilizando para estimar el parámetro de interés. Se observa en el Cuadro 16 que el ESS es bajo para una parte de los parámetros estimados, lo cual podría mejorarse aumentando la cantidad total de iteraciones. Por último, las gráficas de la Figura 11 muestran las 1000 simulaciones finales de cada cadena de Markov utilizadas para estimar los parámetros. No se observa ninguna tendencia, lo cual sustenta la hipótesis de convergencia.

Cuadro 16: Estadísticos de convergencia

	PSRF	Cota sup. I.C.	ESS
Distancia	1,00	1,00	556
Distancia al cuadrado	1,00	1,00	235
UTU	1,01	1,05	784
Efectivos	1,00	1,00	478
Estudiantes por docente	1,00	1,01	1001
Matrícula	1,00	1,01	1001
Aprobación	1,00	1,00	149
Desvinculación	1,00	1,00	476
Porc. estudiantes con TUS	1,00	1,00	391
Homicidios	1,00	1,01	908
Hurtos	1,00	1,00	1281
TUS * UTU	1,00	1,00	1001
TUS * Efectivos	1,00	1,01	475
TUS * Estudiantes por docente	1,00	1,01	882
TUS * Matrícula	1,00	1,00	1001
TUS * Aprobación	1,00	1,00	235
TUS * Desvinculación	1,00	1,00	653
TUS * Porc. estudiantes con TUS	1,00	1,00	396
TUS * Homicidios	1,01	1,05	1001
TUS * Hurtos	1,00	1,00	1160

El PSRF y ESS se estimaron utilizando el paquete *coda* en R.

Figura 11: Análisis gráfico de convergencia

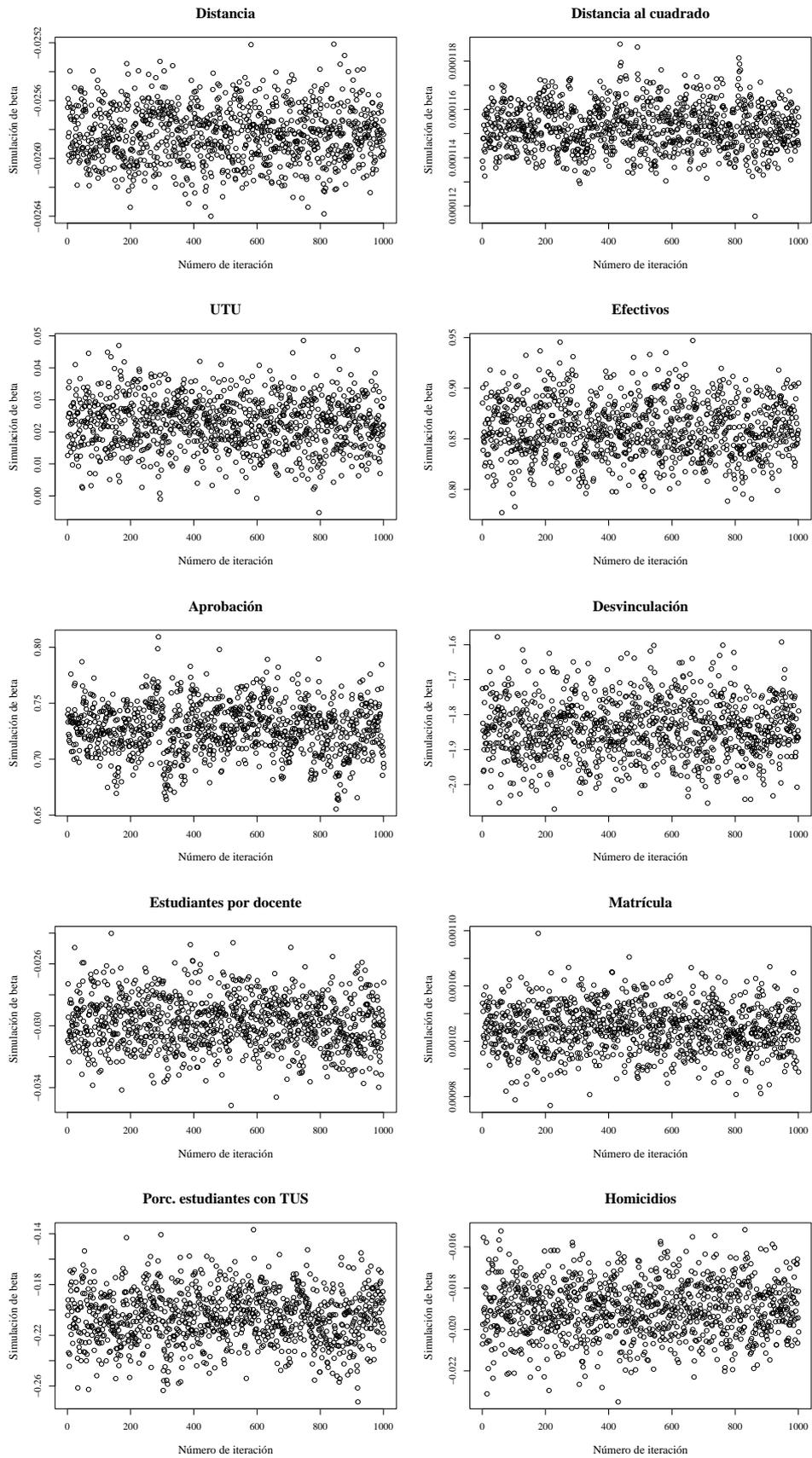
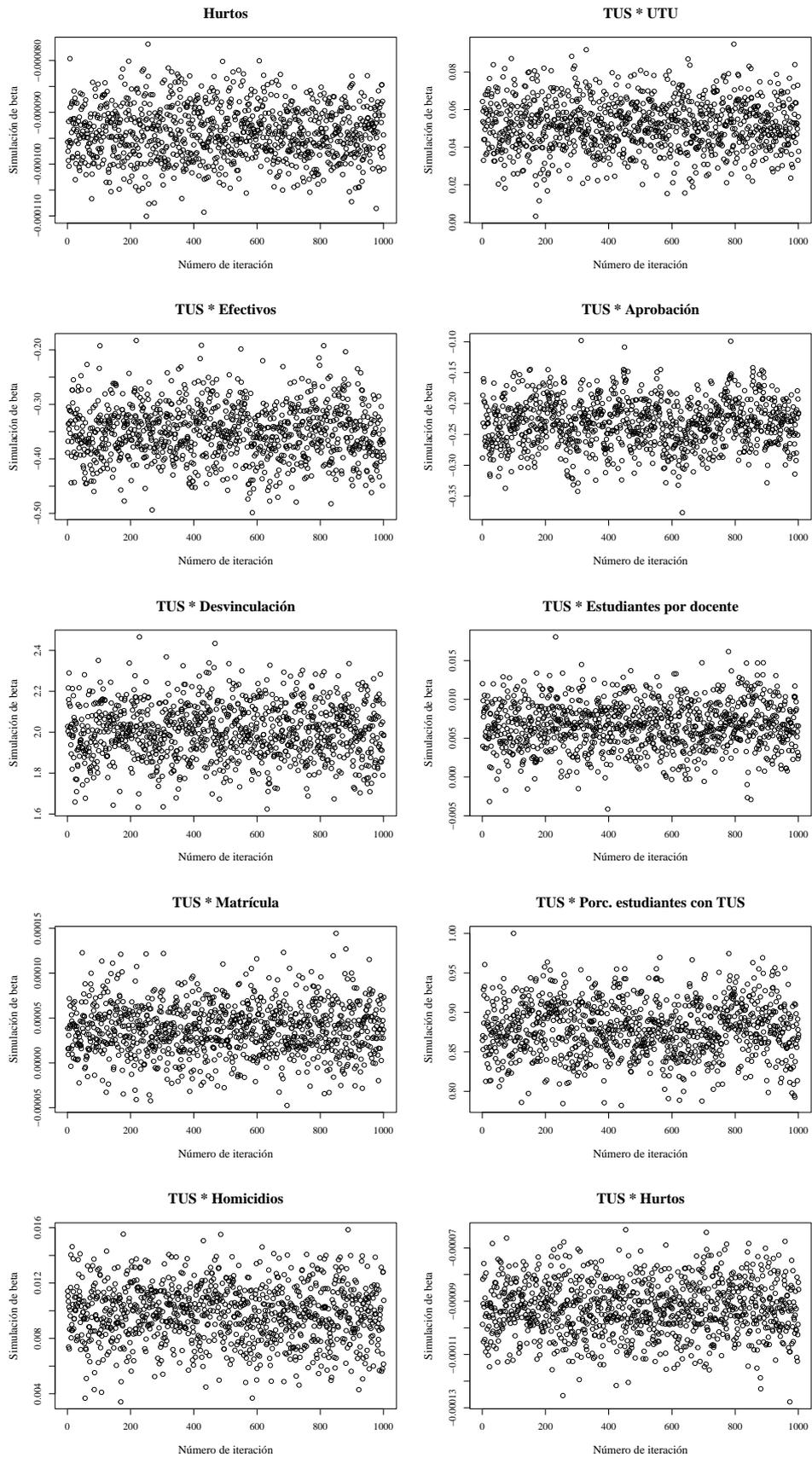


Figura 11: Análisis gráfico de convergencia (cont.)



A.5. Robustez

Se realizaron estimaciones alternativas de las preferencias para darle mayor robustez a los resultados presentados. La Figura 12 muestra la composición por grupos de cada cuartil de calidad en los distintos indicadores que surge de las estimaciones principales. En primer lugar, se realizaron estimaciones que amplían los *choice sets* de los estudiantes, incorporando todas las opciones que se encuentren en su departamento y en el departamento de alguna de sus opciones rankeadas (a esta opción se le llama *choice set B*). Los resultados se presentan en el Cuadro 18. Se observa que el signo de todos los coeficientes se mantiene, salvo para el caso de la valoración por la UTU de los estudiantes que no reciben TUS, cuyo intervalo de credibilidad incluye al 0. Las diferencias entre grupos también se sostienen. Por otro lado, la magnitud de casi todas las estimaciones es menor. Esto es consecuencia de que al agregar centros lejanos en los *choice set* que no son rankeados el modelo interpreta que la distancia es más relevante. Como la distancia es numerario de los coeficientes presentados, la magnitud disminuye. La Figura 13 muestra la composición contrafactual con preferencias homogéneas por cuartil de calidad con las estimaciones alternativas. La menor magnitud de la valoración de las características atenúa el efecto sobre la composición.

En segundo lugar, se estimó el modelo utilizando efectos fijos por centro para capturar aspectos inobservables que no estuvieran asociados a las características incluidas en el modelo. La inclusión de efectos fijos hace que no sea posible estimar los parámetros asociados a las características de los centros si no están interactuadas con alguna característica que varíe entre los estudiantes. De esta forma, se interactuaron las características de los centros con la *dummy* que identifica a los estudiantes que reciben TUS para obtener una nueva estimación de la diferencia en la valoración entre los dos grupos de estudiantes. El Cuadro 17 muestra que aún controlando por efectos fijos de los centros, las diferencias en la valoración de cada característica entre ambos grupos siguen existiendo con probabilidad 95 % con cualquiera de los dos *choice set*, salvo por la cantidad de estudiantes por docente, cuyo intervalo de probabilidad incluye al 0. Las estimaciones con efectos fijos no incluyen la *dummy* UTU porque sería multicolineal con los efectos fijos.

Cuadro 17: Diferencia en preferencias de estudiantes con y sin TUS

	Diferencia TUS (1)		Diferencia TUS (2)		Diferencia TUS (3)	
	$\hat{\beta}$	IC 95 %	$\hat{\beta}$	IC 95 %	$\hat{\beta}$	IC 95 %
UTU	2,00	[0,96;3,06]	-	-	-	-
Efectivos	-5,19	[-6,62;-3,71]	-6,14	[-7,66;-4,6]	-3,30	[-4,17;-2,37]
Est./Doc.	1,24	[0,13;2,29]	-1,07	[-2,29;0,08]	-0,51	[-1,24;0,23]
Matrícula	0,87	[-0,46;2,26]	2,19	[0,78;3,58]	1,41	[0,63;2,28]
Aprobación	-3,00	[-4,00;-2,05]	-2,41	[-3,41;-1,43]	-1,32	[-1,88;-0,70]
Desvinc.	6,84	[5,94;7,71]	6,05	[5,07;6,98]	3,21	[2,66;3,78]
Pares	14,82	[13,66;16,04]	18,47	[17,28;19,76]	11,28	[10,56;12,02]
Homicidios	1,51	[0,87;2,13]	1,66	[0,93;2,35]	1,13	[0,67;1,55]
Hurtos	-3,98	[-4,80;-3,13]	-4,22	[-5,17;-3,28]	-2,53	[-3,08;-2,03]
EF centro	No		Sí		Sí	
Choice set	A		A		B	

Elaboración propia con base en datos de ANEP.

Un tercer elemento que se modificó en especificaciones alternativas es el corte de los grupos de estudiantes. El Cuadro 19 presenta estimaciones que introducen heterogeneidad en las preferencias dividiendo a los estudiantes según si tienen un índice de vulnerabilidad mayor al percentil 75 (índice 4769), mayor al percentil 90 (índice 6477), o si reciben doble TUS. Las estimaciones muestran que también existen diferencias en la valoración de las características al considerar estos grupos. Al igual que en las estimaciones con efectos fijos, dejan de observarse diferencias en la valoración de la cantidad de estudiantes por docente. La magnitud de la diferencia en la valoración de las características asociadas a la calidad es mayor en estas especificaciones. Por un lado, las mayores diferencias al considerar los estudiantes con doble TUS en lugar de los estudiantes con TUS son evidencia sugerente de que existe una relación entre la intensidad de la pobreza y la menor valoración de la calidad de los centros. Por otro lado, la mayor magnitud al considerar los estudiantes por encima del percentil 75 del índice de vulnerabilidad sugiere que no solo la pobreza está relacionada con la menor valoración de la calidad, sino que también el mal desempeño educativo previo. La Figura 14 presenta la composición por grupo de ingresos dentro de cada cuartil de las distintas características consideradas como indicadores de calidad. Para la construcción de estos escenarios, se modificaron únicamente las preferencias y rankings de los estudiantes considerados vulnerables según el corte definido en cada especificación. Los gráficos muestran que los efectos de homogeneizar las preferencias son similares en las distintas definiciones del corte de vulnerabilidad.

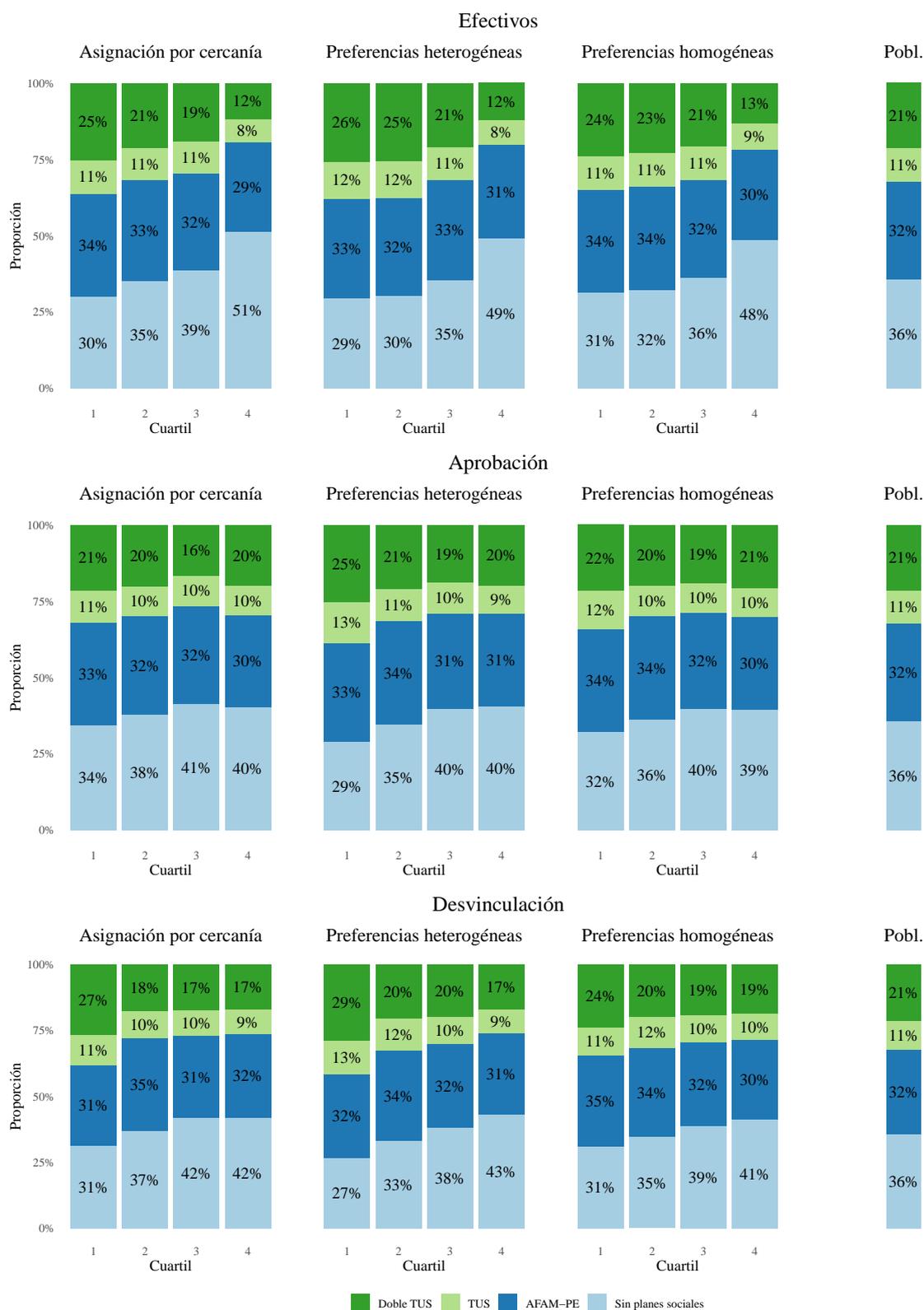
En cuarto lugar, se realizaron estimaciones eliminando de la muestra los estudiantes con pocos o

ningún centro cercano tomando distintos radios, considerando que podrían tener una forma distinta de formar sus preferencias. En particular, se estimó eliminando los estudiantes sin ningún centro y con menos de 3 centros a 20, 10 y 5 kilómetros. El Cuadro 20 muestra que la valoración por las características es igual en signo, similar en magnitud y en cuanto a la diferencia entre los estudiantes que reciben o no TUS. La magnitud de la valoración de las características es mayor cuanto más restrictiva es la muestra de estudiantes. Esto sugiere que los estudiantes de centros urbanos le dan más valor a la calidad de los centros. Es posible que esto suceda porque tienen que comparar más opciones y por lo tanto obtienen algo más de información sobre los centros, aunque no es posible determinar la validez de esta hipótesis en este trabajo. Las figuras 15 y 16 muestran que la asignación contrafactual que surge de estas estimaciones es similar a la que surge de la especificación principal del trabajo.

Por último, se realizaron estimaciones de las preferencias introduciendo heterogeneidad entre los grupos de vulnerabilidad en la valoración de la distancia. El Cuadro 21 presenta los coeficientes estimados bajo esta especificación. A diferencia de los resultados principales presentados, estos coeficientes no tienen una interpretación en términos de disposición a trasladarse, porque no están normalizados utilizando el coeficiente de distancia. La interpretación técnica de estos coeficientes es que representan el valor de la característica en relación al valor de un desvío estándar de los factores inobservables (Train, 2009, cap. 2). Se observa que si bien se identifica que la distancia afecta menos a los estudiantes más vulnerables, la diferencia es pequeña y no afecta el resultado de que las preferencias por las características asociadas a la calidad de los centros son heterogéneas. La Figura 17 muestra que la asignación resultante de imponer preferencias homogéneas por calidad es similar a la que se obtiene si se asume que la valoración de la distancia es homogénea.

El Cuadro 22 resume todos los ejercicios de robustez en términos de cómo afectan la medición del efecto de la heterogeneidad de las preferencias sobre la segregación. Se construyeron asignaciones contrafactuales imponiendo homogeneidad en las preferencias por características asociadas a la calidad utilizando cada una de las estimaciones alternativas de preferencias. Se encuentra que las conclusiones alcanzadas en este trabajo son robustas a estas variantes de estimación de las preferencias.

Figura 12: Proporción de estudiantes según grupo de vulnerabilidad, especificación principal



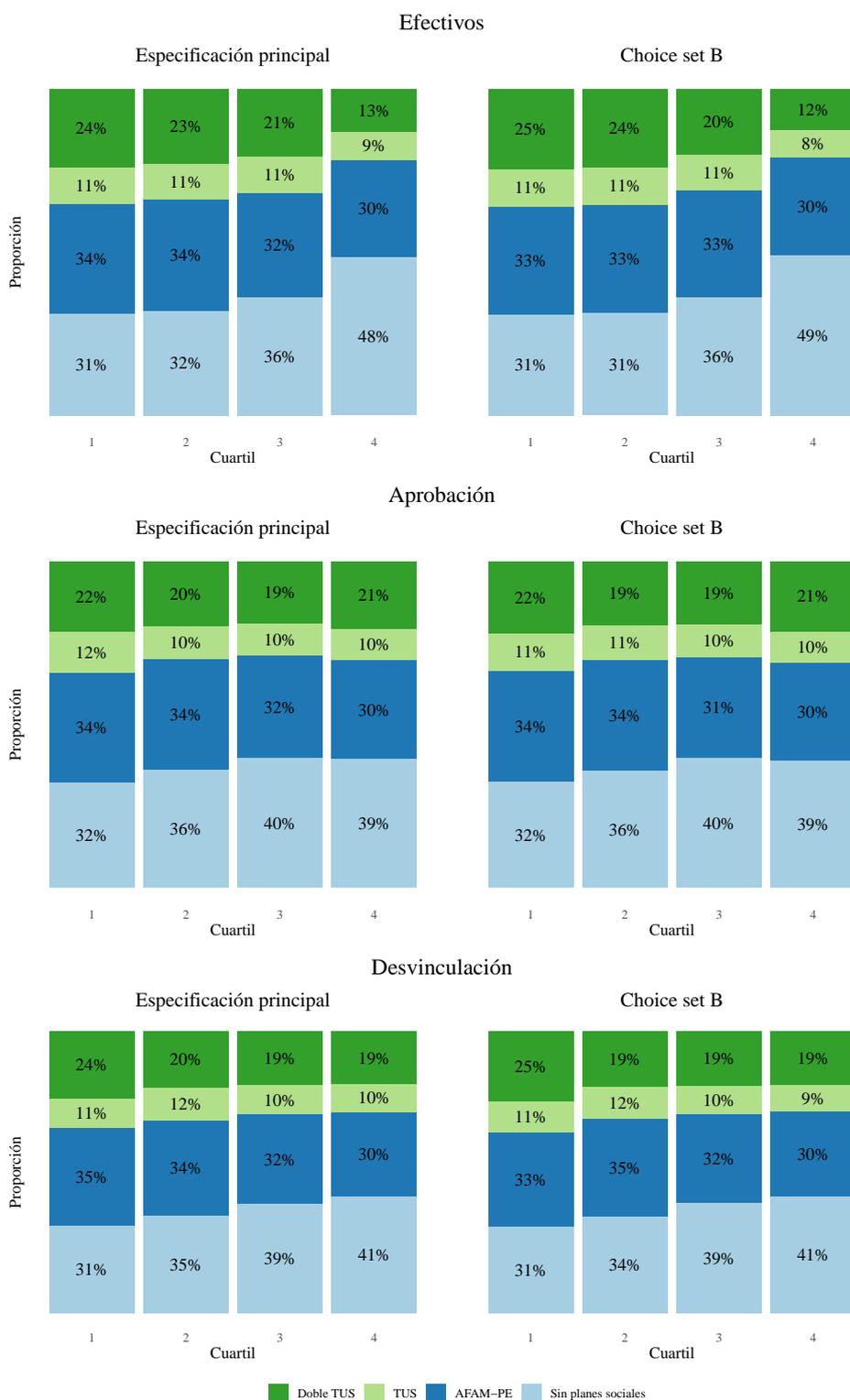
Pobl. representa la composición por grupo en la población de estudiantes. *Efectivos* es el porcentaje de docentes efectivos. *Aprobación* es la tasa de aprobación y *Desvinculación* la tasa de desvinculación. Todas las variables están medidas a nivel de Ciclo Básico tomando el promedio 2021-2022 (salvo para la tasa de aprobación). Los cuartiles se calcularon dentro de cada departamento. Elaboración propia con base en datos de ANEP.

Cuadro 18: Preferencias estimadas bajo ambas opciones de *choice set*

	Sin TUS		Con TUS		Diferencia TUS	
	$\hat{\beta}$	(1) IC 95 %	$\hat{\beta}$	(2) IC 95 %	$\hat{\beta}$	(3) IC 95 %
<i>Choice set A</i>						
UTU	0,88	[0,25;1,45]	2,88	[2,04;3,71]	2,00	[0,96;3,06]
Efectivos	12,69	[11,92;13,5]	7,50	[6,33;8,77]	-5,19	[-6,62;-3,71]
Est./Doc.	-5,54	[-6,13;-4,93]	-4,30	[-5,23;-3,43]	1,24	[0,13;2,29]
Matrícula	24,09	[23,24;24,93]	24,97	[23,86;26,08]	0,87	[-0,46;2,26]
Aprobación	9,42	[8,82;9,98]	6,41	[5,63;7,2]	-3,00	[-4;-2,05]
Desvinc.	-6,28	[-6,83;-5,69]	0,56	[-0,12;1,29]	6,84	[5,94;7,71]
Pares	-3,47	[-4,18;-2,81]	11,35	[10,34;12,29]	14,82	[13,66;16,04]
Homicidios	-2,94	[-3,35;-2,53]	-1,43	[-1,94;-0,96]	1,51	[0,87;2,13]
Hurtos	-4,03	[-4,46;-3,59]	-8,00	[-8,76;-7,29]	-3,98	[-4,8;-3,13]
<i>Choice set B</i>						
UTU	-0,20	[-0,59;0,2]	1,04	[0,52;1,56]	1,24	[0,61;1,9]
Efectivos	7,66	[7,15;8,12]	5,35	[4,6;6,06]	-2,31	[-3,23;-1,43]
Est./Doc.	-5,14	[-5,52;-4,77]	-4,10	[-4,69;-3,49]	1,04	[0,35;1,73]
Matricula	15,14	[14,64;15,62]	15,65	[14,96;16,37]	0,51	[-0,34;1,29]
Aprobación	7,17	[6,84;7,51]	5,16	[4,71;5,64]	-2,01	[-2,6;-1,39]
Desvinc.	-3,69	[-4,03;-3,35]	0,25	[-0,19;0,66]	3,94	[3,39;4,48]
Pares	-2,48	[-2,92;-2,07]	6,99	[6,48;7,54]	9,47	[8,79;10,24]
Homicidios	-1,79	[-2,04;-1,54]	-0,69	[-1,01;-0,35]	1,10	[0,7;1,52]
Hurtos	-4,65	[-4,89;-4,41]	-6,94	[-7,39;-6,51]	-2,29	[-2,8;-1,81]

Choice set A es el utilizado en la especificación principal presentada en el cuerpo del documento, *Choice set B* es la definición alternativa presentada en este anexo. Elaboración propia con base en datos de ANEP.

Figura 13: Asignación con preferencias homogéneas, *choice set* alternativo



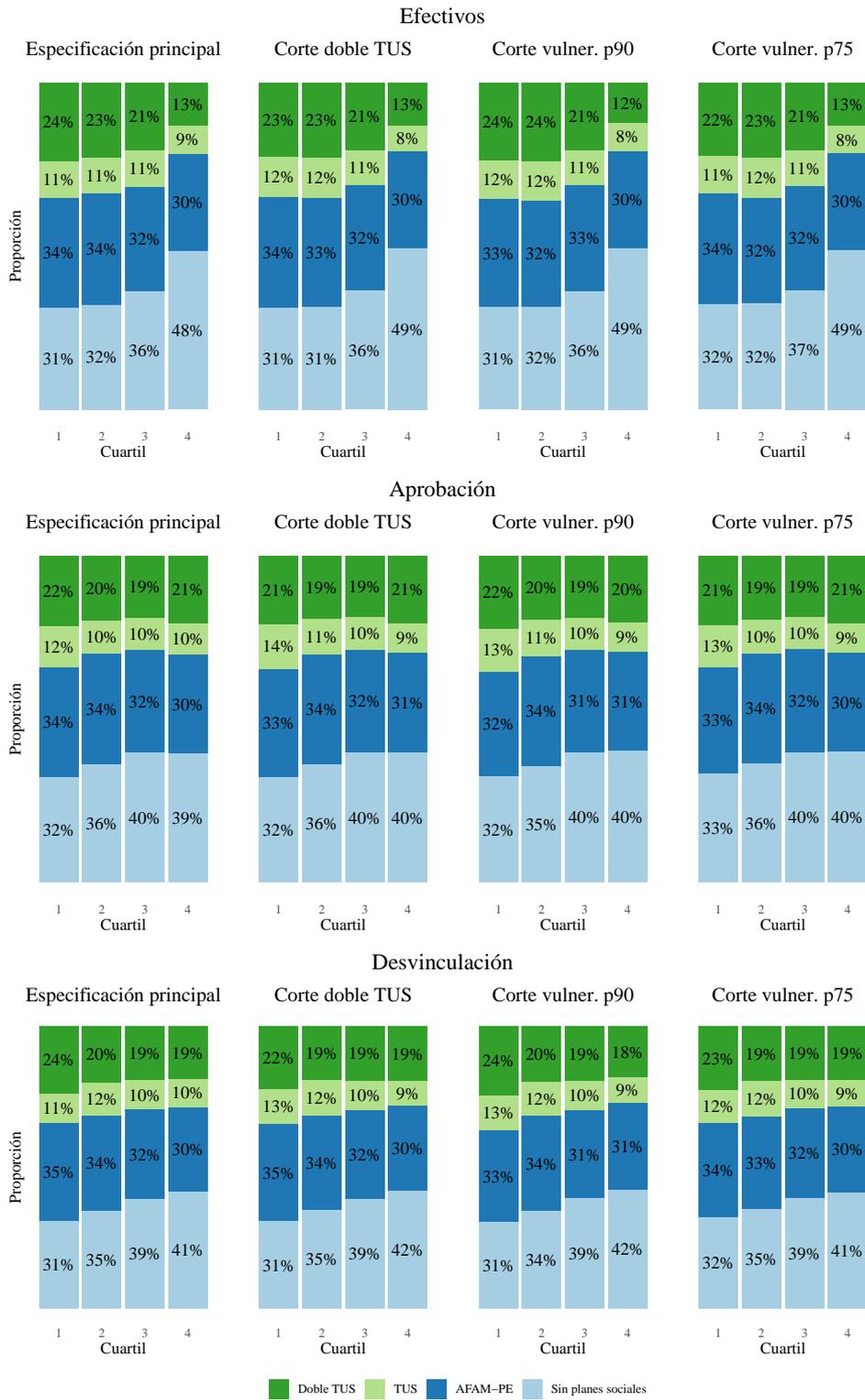
Elaboración propia con base en datos de ANEP.

Cuadro 19: Estimación de preferencias, cortes de vulnerabilidad alternativos

	Especificación principal (1)		V75 (2)		V90 (3)		Doble TUS (4)			
	$\hat{\beta}$	IC 95 %	$\hat{\beta}$	IC 95 %	$\hat{\beta}$	IC 95 %	$\hat{\beta}$	IC 95 %		
Estudiantes sin TUS										
UTU	0,88	[0,25;1,45]	6,04	[5,28;6,83]	0,71	[0,18;1,29]	0,64	[0,14;1,12]	1,08	[0,55;1,65]
Efectivos	12,69	[11,92;13,5]	-	-	12,96	[12,12;13,71]	11,66	[10,99;12,38]	12,80	[12,01;13,59]
Titulados	-	-	11,30	[10,43;12,15]	-	-	-	-	-	-
Est./Doc.	-5,54	[-6,13;-4,93]	-5,89	[-6,57;-5,3]	-4,95	[-5,6;-4,35]	-4,83	[-5,39;-4,28]	-5,19	[-5,78;-4,58]
Matricula	24,09	[23,24;24,93]	26,04	[25,22;26,93]	23,74	[22,87;24,56]	23,84	[23,17;24,59]	23,90	[23,1;24,63]
Aprobación	9,42	[8,82;9,98]	9,77	[9,21;10,33]	9,03	[8,45;9,58]	9,06	[8,57;9,55]	9,37	[8,87;9,89]
Desvinc.	-6,28	[-6,83;-5,69]	-6,35	[-6,89;-5,79]	-6,42	[-6,95;-5,87]	-4,85	[-5,31;-4,41]	-5,75	[-6,27;-5,3]
Pares	-3,47	[-4,18;-2,81]	-4,97	[-5,62;-4,33]	-2,57	[-3,13;-2,04]	-0,08	[-0,44;0,27]	-0,96	[-1,49;-0,44]
Homicidios	-2,94	[-3,35;-2,53]	-4,05	[-4,44;-3,65]	-2,65	[-3,03;-2,24]	-2,44	[-2,76;-2,07]	-2,88	[-3,26;-2,5]
Hurtos	-4,03	[-4,46;-3,59]	-2,84	[-3,27;-2,4]	-4,22	[-4,63;-3,77]	-4,84	[-5,23;-4,46]	-4,28	[-4,72;-3,88]
Diferencia estudiantes con TUS										
UTU	2,00	[0,96;3,06]	2,40	[1,08;3,71]	3,64	[2,61;4,67]	6,18	[4,7;7,67]	2,05	[0,9;3,22]
Efectivos	-5,19	[-6,62;-3,71]	-	-	-7,44	[-8,97;-5,97]	-7,38	[-9,52;-5,24]	-7,56	[-9,29;-5,97]
Titulados	-	-	-1,50	[-3,1;0,29]	-	-	-	-	-	-
Est./Doc.	1,24	[0,13;2,29]	2,10	[0,98;3,26]	0,49	[-0,62;1,59]	0,78	[-0,94;2,41]	0,62	[-0,6;1,9]
Matricula	0,87	[-0,46;2,26]	-0,47	[-1,75;0,82]	0,00	[0;0]	-0,19	[-2,11;1,64]	1,45	[-0,05;3,01]
Aprobación	-3,00	[-4;-2,05]	-3,46	[-4,53;-2,45]	-3,49	[-4,4;-2,56]	-5,38	[-6,65;-3,96]	-3,96	[-5,04;-2,87]
Desvinc.	6,84	[5,94;7,71]	6,63	[5,69;7,55]	8,20	[7,32;9,09]	8,43	[7,2;9,65]	7,62	[6,78;8,58]
Pares	14,82	[13,66;16,04]	16,00	[14,83;17,22]	9,05	[8,01;10,06]	5,38	[4,46;6,33]	9,88	[8,83;10,87]
Homicidios	1,51	[0,87;2,13]	2,27	[1,61;2,96]	1,21	[0,54;1,84]	1,29	[0,44;2,18]	1,55	[0,85;2,31]
Hurtos	-3,98	[-4,8;-3,13]	-4,73	[-5,54;-3,88]	-3,50	[-4,28;-2,66]	-3,37	[-4,57;-2,12]	-4,04	[-5,02;-3,09]

Elaboración propia con base en datos de ANEP.

Figura 14: Asignación con preferencias homogéneas, cortes de vulnerabilidad alternativos



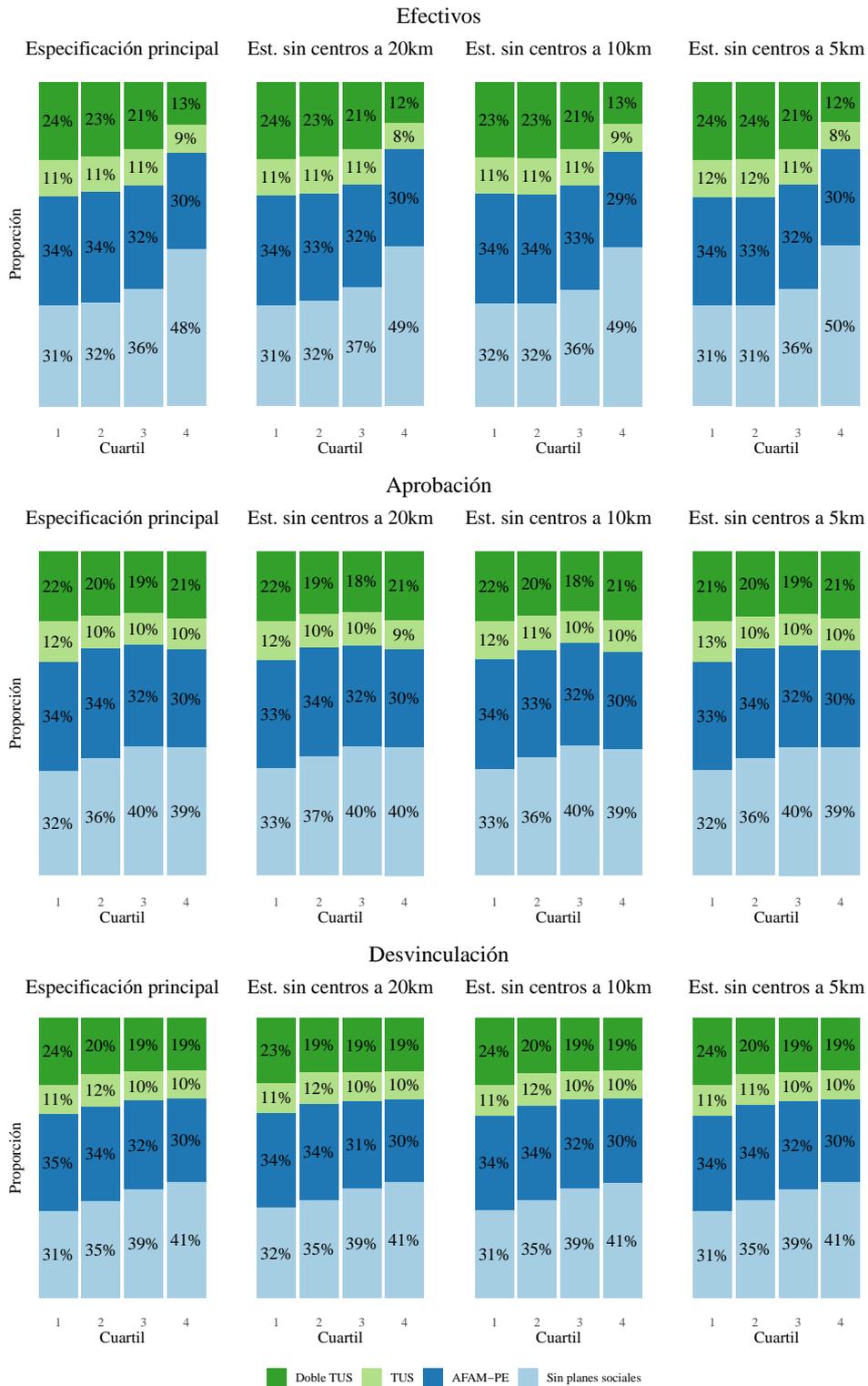
Elaboración propia con base en datos de ANEP.

Cuadro 20: Especificaciones alternativas filtrando de estudiantes sin centros cercanos

	Sin opc. a 20km (1)		Sin opc. a 10km (2)		Sin opc. a 5km (3)		Sin 3 opc. a 20km (4)		Sin 3 opc. a 10km (5)		Sin 3 opc. a 5km (6)	
	$\hat{\beta}$	IC 95%	$\hat{\beta}$	IC 95%	$\hat{\beta}$	IC 95%	$\hat{\beta}$	IC 95%	$\hat{\beta}$	IC 95%	$\hat{\beta}$	IC 95%
Estudiantes sin TUS												
UTU	0,80	[0,22 ; 1,37]	0,52	[-0,07 ; 1,12]	0,47	[-0,13 ; 1,13]	1,64	[0,97 ; 2,32]	1,02	[0,35 ; 1,7]	0,67	[-0,04 ; 1,42]
Efectivos	12,78	[11,98 ; 13,62]	12,40	[11,58 ; 13,25]	12,40	[11,54 ; 13,23]	15,85	[14,88 ; 16,87]	16,20	[15,27 ; 17,16]	16,13	[15,11 ; 17,32]
Est./Doc.	-5,65	[-6,26 ; -5]	-5,63	[-6,26 ; -5,05]	-5,67	[-6,27 ; -5,07]	-5,82	[-6,49 ; -5,13]	-5,99	[-6,66 ; -5,3]	-6,59	[-7,28 ; -5,85]
Matricula	24,24	[23,45 ; 25,12]	23,84	[22,99 ; 24,68]	24,05	[23,28 ; 24,91]	26,33	[25,4 ; 27,19]	27,81	[26,84 ; 28,69]	29,39	[28,41 ; 30,42]
Aprobación	9,51	[9 ; 10,04]	9,78	[9,2 ; 10,34]	9,88	[9,29 ; 10,49]	10,99	[10,26 ; 11,7]	12,17	[11,54 ; 12,79]	12,87	[12,22 ; 13,55]
Desvinc.	-6,29	[-6,83 ; -5,77]	-6,24	[-6,78 ; -5,73]	-6,34	[-6,89 ; -5,78]	-6,94	[-7,55 ; -6,35]	-7,12	[-7,77 ; -6,52]	-7,38	[-8,03 ; -6,69]
Pares	-3,38	[-4,11 ; -2,7]	-3,30	[-3,95 ; -2,61]	-3,12	[-3,83 ; -2,41]	-1,35	[-2,18 ; -0,57]	-0,85	[-1,55 ; -0,1]	0,03	[-0,79 ; 0,87]
Homicidios	-2,94	[-3,34 ; -2,53]	-2,86	[-3,28 ; -2,46]	-2,94	[-3,33 ; -2,54]	-2,71	[-3,13 ; -2,28]	-2,57	[-3 ; -2,16]	-2,37	[-2,79 ; -1,96]
Hurtos	-4,04	[-4,47 ; -3,61]	-3,99	[-4,42 ; -3,57]	-3,92	[-4,37 ; -3,47]	-4,10	[-4,55 ; -3,67]	-3,98	[-4,48 ; -3,54]	-3,57	[-4,06 ; -3,12]
Diferencia estudiantes con TUS												
UTU	1,99	[0,96 ; 3,01]	2,26	[1,22 ; 3,28]	2,46	[1,45 ; 3,46]	2,38	[1,3 ; 3,47]	3,05	[1,92 ; 4,23]	3,48	[2,21 ; 4,74]
Efectivos	-4,97	[-6,38 ; -3,52]	-4,69	[-6,2 ; -3,22]	-4,49	[-5,96 ; -3,05]	-4,68	[-6,42 ; -3,18]	-4,77	[-6,38 ; -3,16]	-5,20	[-6,9 ; -3,38]
Est./Doc.	1,33	[0,22 ; 2,47]	1,31	[0,18 ; 2,46]	1,42	[0,35 ; 2,55]	1,62	[0,49 ; 2,79]	1,69	[0,54 ; 2,98]	2,02	[0,69 ; 3,29]
Matricula	0,91	[-0,42 ; 2,21]	1,05	[-0,24 ; 2,33]	0,87	[-0,45 ; 2,14]	0,77	[-0,64 ; 2,24]	0,54	[-0,84 ; 1,91]	0,20	[-1,18 ; 1,59]
Aprobación	-2,89	[-3,83 ; -1,93]	-3,18	[-4,08 ; -2,21]	-3,23	[-4,17 ; -2,29]	-3,43	[-4,49 ; -2,39]	-4,07	[-5,14 ; -3,13]	-4,78	[-5,8 ; -3,65]
Desvinc.	6,92	[6,01 ; 7,8]	6,83	[5,97 ; 7,69]	6,87	[5,98 ; 7,76]	7,14	[6,22 ; 8,13]	7,51	[6,47 ; 8,5]	7,68	[6,61 ; 8,63]
Pares	15,03	[13,85 ; 16,21]	14,86	[13,62 ; 15,97]	14,78	[13,56 ; 15,95]	15,61	[14,42 ; 16,87]	15,59	[14,26 ; 16,88]	15,05	[13,72 ; 16,37]
Homicidios	1,50	[0,83 ; 2,16]	1,44	[0,75 ; 2,1]	1,54	[0,9 ; 2,19]	1,54	[0,83 ; 2,24]	1,52	[0,82 ; 2,27]	1,61	[0,91 ; 2,34]
Hurtos	-4,02	[-4,85 ; -3,17]	-3,96	[-4,78 ; -3,15]	-4,05	[-4,9 ; -3,2]	-4,12	[-4,99 ; -3,2]	-4,19	[-5,17 ; -3,24]	-4,18	[-5,09 ; -3,27]

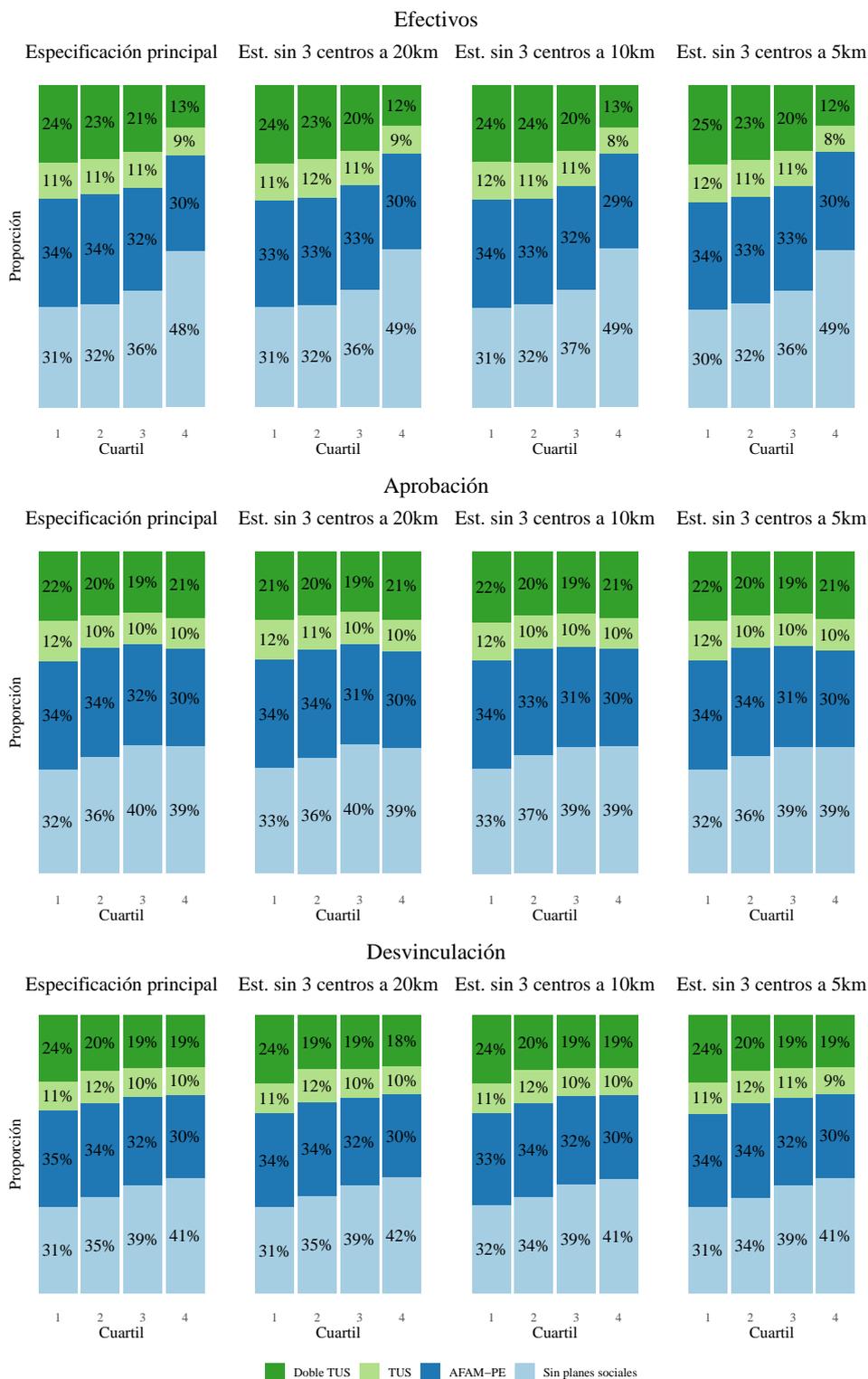
Elaboración propia con base en datos de ANEP.

Figura 15: Asignación con preferencias homogéneas, filtrando estudiantes sin centros cercanos (a)



Elaboración propia con base en datos de ANEP.

Figura 16: Asignación con preferencias homogéneas, filtrando estudiantes sin centros cercanos (b)



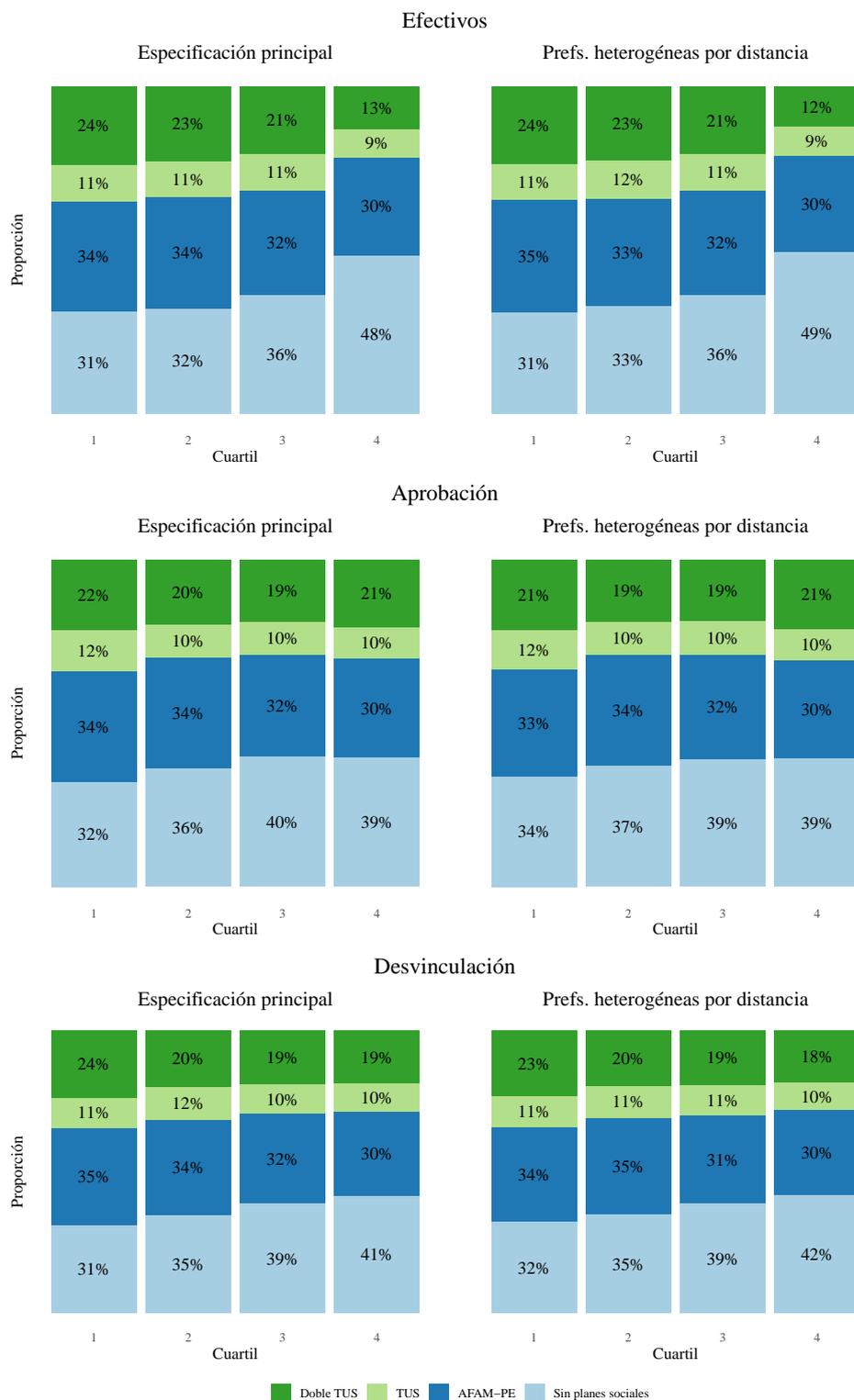
Elaboración propia con base en datos de ANEP.

Cuadro 21: Estimación de preferencias introduciendo heterogeneidad en la valoración de la distancia

	Sin TUS		Con TUS		Diferencia TUS	
	(1)		(2)		(3)	
	$\hat{\beta}$	IC 95 %	$\hat{\beta}$	IC 95 %	$\hat{\beta}$	IC 95 %
Preferencias por distancia homogéneas						
Distancia	-0,026	[-0,026;-0,025]	-0,026	[-0,026;-0,025]	-	-
Distancia ²	0,000	[0;0]	0,000	[0;0]	-	-
UTU	0,023	[0,007;0,037]	0,074	[0,053;0,095]	0,052	[0,025;0,079]
Efectivos	0,860	[0,808;0,913]	0,508	[0,427;0,593]	-0,351	[-0,446;-0,253]
Doc./Est.	-0,030	[-0,033;-0,027]	-0,023	[-0,028;-0,018]	0,007	[0,001;0,012]
Matrícula	0,001	[0,001;0,001]	0,001	[0,001;0,001]	0,000	[0;0]
Aprobación	0,727	[0,682;0,769]	0,495	[0,433;0,556]	-0,232	[-0,309;-0,158]
Desvinc.	-1,840	[-1,995;-1,672]	0,163	[-0,035;0,378]	2,003	[1,742;2,274]
Pares	-0,206	[-0,248;-0,167]	0,674	[0,614;0,73]	0,879	[0,813;0,946]
Homicidios	-0,019	[-0,022;-0,016]	-0,009	[-0,012;-0,006]	0,010	[0,006;0,014]
Hurtos	0,000	[0;0]	0,000	[0;0]	0,000	[0;0]
Preferencias por distancia heterogéneas						
Distancia	-0,027	[-0,027;-0,03]	-0,023	[-0,024;-0,02]	0,003	[0,002;0,004]
Distancia ²	0,000	[0;0]	0,000	[0;0]	0,000	[0;0]
UTU	0,082	[0,067;0,1]	0,126	[0,106;0,15]	0,045	[0,018;0,07]
Efectivos	0,862	[0,808;0,92]	0,535	[0,448;0,62]	-0,327	[-0,435;-0,22]
Doc./Est.	-0,030	[-0,033;-0,03]	-0,019	[-0,024;-0,01]	0,011	[0,005;0,02]
Matrícula	0,001	[0,001;0]	0,001	[0,001;0]	0,000	[0;0]
Aprobación	0,703	[0,664;0,74]	0,458	[0,396;0,53]	-0,246	[-0,327;-0,17]
Desvinc.	-1,746	[-1,901;-1,59]	0,207	[0,003;0,43]	1,953	[1,676;2,24]
Pares	-0,219	[-0,26;-0,18]	0,677	[0,619;0,73]	0,895	[0,821;0,96]
Homicidios	-0,021	[-0,023;-0,02]	-0,010	[-0,013;-0,01]	0,011	[0,007;0,02]
Hurtos	0,000	[0;0]	0,000	[0;0]	0,000	[0;0]

A diferencia del Cuadro 11 presentado en la sección de resultados, los coeficientes de esta tabla no tienen una interpretación en términos de disposición a trasladarse. Técnicamente, los coeficientes representan el efecto sobre la utilidad de la característica en relación al efecto de un desvío estándar de los factores inobservables. Elaboración propia con base en datos de ANEP.

Figura 17: Asignación con preferencias por calidad homogéneas y preferencias por distancia heterogéneas



Elaboración propia con base en datos de ANEP.

Cuadro 22: Variación de la segregación al homogeneizar las preferencias por calidad

	Efectivos (1)		Aprobación (2)		Desvinc. (3)	
	M	IC 95 %	M	IC 95 %	M	IC 95 %
Especificación principal	-0,006	[-0,009;-0,004]	-0,004	[-0,006;-0,003]	-0,008	[-0,011;-0,006]
Choice set A	-0,005	[-0,008;-0,002]	-0,005	[-0,006;-0,003]	-0,008	[-0,010;-0,006]
Prof. por distancia heterogéneas	-0,007	[-0,010;-0,004]	-0,005	[-0,006;-0,003]	-0,009	[-0,011;-0,007]
Vuln. p75	-0,009	[-0,012;-0,007]	-0,004	[-0,006;-0,003]	-0,010	[-0,012;-0,008]
Vuln. p90	-0,005	[-0,008;-0,002]	-0,003	[-0,004;-0,001]	-0,006	[-0,008;-0,004]
Doble TUS	-0,008	[-0,011;-0,004]	-0,004	[-0,005;-0,002]	-0,008	[-0,010;-0,006]
Sin opc. a 20km	-0,007	[-0,010;-0,004]	-0,005	[-0,006;-0,003]	-0,009	[-0,011;-0,007]
Sin opc. a 10km	-0,007	[-0,010;-0,005]	-0,004	[-0,006;-0,003]	-0,008	[-0,010;-0,006]
Sin opc. a 5km	-0,006	[-0,009;-0,004]	-0,004	[-0,005;-0,003]	-0,009	[-0,011;-0,006]
Sin 3 opc. a 20km	-0,006	[-0,009;-0,003]	-0,004	[-0,006;-0,003]	-0,008	[-0,010;-0,006]
Sin 3 opc. a 10km	-0,007	[-0,010;-0,004]	-0,004	[-0,006;-0,003]	-0,009	[-0,011;-0,007]
Sin 3 opc. a 5km	-0,007	[-0,009;-0,003]	-0,005	[-0,006;-0,003]	-0,009	[-0,011;-0,007]

Elaboración propia con base en datos de ANEP.