

Universidad de la República
Facultad de Ciencias Sociales
Departamento de Economía

Tesis de Maestría en Economía Internacional

Segregación educativa y elección escolar: El caso de la educación media básica en Uruguay¹

Matías Faggetti

Tutor: Juan S. Pereyra

Montevideo, Uruguay

2023

¹ Agradezco a Juan S. Pereyra por la guía y apoyo permanente durante las distintas etapas de elaboración del presente trabajo. Agradezco también por sus aportes y comentarios a Ignacio Cabrera. A Milton Silveira y Luana Mendez por compartirme los datos necesarios para realizar el trabajo, así como también a la ANII y el proyecto FMV_1_2021_1_166576. A Thilo Klein por su ayuda con el paquete de R “stabest” y a Michelle Cukerman y Mariana Laverde por sus comentarios. Finalmente agradezco a Andrea Barón y Victoria Landaberry por su ayuda con el software R.

RESUMEN

En este trabajo, se estudian los factores que contribuyen a la segregación en la educación media básica en Uruguay, específicamente entre la educación secundaria (liceos) y la educación técnica (UTU). A partir de los datos del nuevo sistema de asignación de estudiantes implementado en 2019, se muestra que la participación relativa de estudiantes vulnerables en UTU duplica o incluso triplica la cantidad de estudiantes vulnerables en liceos. Con base al análisis de escenarios contrafactuales, se concluye, en primer lugar, que la segregación residencial no explica, al menos en su totalidad la segregación educativa. Es decir, este fenómeno estaría mediado por las preferencias de los estudiantes o por características distintas a la distancia. En segundo lugar, se utilizan métodos bayesianos para la estimación de las preferencias cardinales de los estudiantes. Se encuentra que los estudiantes no vulnerables derivan una mayor utilidad por los liceos que por concurrir a una UTU, mientras que para los estudiantes vulnerables dicha variable no es relevante o incluso pueden tener preferencia por concurrir a una UTU (una vez que se controló por la variable distancia). Por lo tanto, el uso del índice de vulnerabilidad como única prioridad de los centros y el componente diferencial en las preferencias de estudiantes vulnerables y no vulnerables, estaría contribuyendo a la segregación del sistema.

Contenido

1. Introducción.....	5
2. Revisión de la literatura.....	8
3. Marco teórico y definiciones.....	10
4. Mecanismo de asignación en Uruguay, base de datos y estadísticas descriptivas	13
4.1. Oferta: Centros de educación	13
4.2. Demanda: Preferencias de estudiantes	13
4.3. Prioridades: Índice de vulnerabilidad social	15
4.4. Mecanismo de asignación	16
4.5. Distancia: Estadísticas descriptivas.....	17
5. Segregación	22
5.1. Distribución del índice de vulnerabilidad entre estudiantes de liceo y UTU	22
5.2. Participación relativa de estudiantes vulnerables o no vulnerables.....	24
5.3. Índice de Disimilitud: Duncan y Duncan	24
6. Metodología empírica	26
6.1. Ecuación a estimar	26
6.2. Metodología de estimación	27
6.3. Supuesto de identificación	28
7. Principales resultados.....	29
7.1. Simulaciones contrafactuales	29
7.2. Estimación empírica.....	32
7.3. Análisis de bienestar.....	35
8. Conclusiones	36
Referencias bibliográficas	38
Apéndice.....	40

1. Introducción

Disminuir la segregación en el sistema educativo es uno de los principales desafíos que enfrentan los gobiernos en materia educativa. Existe segregación cuando la distribución de alguna característica de los alumnos en un conjunto de centros educativos es distinta de la distribución de dicha característica en la población en general. Esto sucede, por ejemplo, cuando los estudiantes socioeconómicamente más vulnerables están sobrerrepresentados en algunos centros. Existe consenso de que la segregación educativa tiene consecuencias negativas en la trayectoria de los estudiantes. Por ejemplo, Card y Rothstein (2006) y Billings et al. (2012), muestran que la segregación educativa incrementa las diferencias en puntaje de pruebas y en logros educativos según diferentes grupos de estudiantes. Uruguay no es ajeno a este fenómeno, como se sostiene en INEEd (2019), existe segregación en todos los niveles educativos, donde la distribución más desigual se encuentra en el nivel inicial y primaria, seguido por la educación secundaria y finalmente la educación técnica. Por lo tanto, diseñar políticas públicas que disminuyan la segregación resulta fundamental para mejorar el sistema educativo.

Un problema para el diseño de políticas es la dificultad para determinar las causas de la segregación. Por ejemplo, consideremos un sistema de elección escolar en el cual los estudiantes pueden rankear los centros educativos de acuerdo a sus preferencias. Luego, a partir de las preferencias, se asignan los estudiantes usando las prioridades de los centros y un algoritmo centralizado. En este caso, la segregación podría ser efecto de las preferencias o de la oferta (o de ambas). Se estaría en el primer caso si el grupo de estudiantes vulnerables tiene preferencias homogéneas entre ellos, las cuales son distintas a las preferencias del grupo de estudiantes no vulnerables. Por otro lado, se estaría en el segundo caso cuando los dos grupos tienen preferencias similares, la cual depende de la distancia a los centros, y las prioridades en los centros están determinadas por la cercanía. Entonces, en este caso, la segregación educativa reflejará la segregación residencial. De esta forma, el conocimiento del escenario es relevante para el diseño de políticas para enfrentar la segregación.

En este trabajo, en primer lugar, se muestra evidencia de la segregación educativa que existe entre dos subsistemas de educación media básica en Uruguay: secundaria (liceos) y educación técnica. Luego, se estudian los posibles determinantes de la segregación mediante la estimación de las preferencias de los estudiantes por los centros educativos. El análisis es factible gracias a las particularidades del sistema de asignación a centros educativos de educación media básica en Uruguay.

En Uruguay, hasta 2019 los estudiantes eran asignados a centros de educación media básica por parte de la autoridad educativa (Administración de Educación Pública, ANEP), sin considerar las preferencias de los estudiantes y de acuerdo a un criterio principalmente geográfico. A partir de 2019, se cambió el sistema de asignación de estudiantes que finalizan su último año enseñanza primaria y comienza su primer año de educación media básica. El sistema se basa en tres elementos: 1) número de lugares disponibles de cada centro (cupó), 2) preferencias de los estudiantes por los centros y 3) orden de prioridad de los estudiantes.

En cuanto a los centros, existen dos subsistemas. Uno puede asociarse a una trayectoria más académica o tradicional denominado “Liceo”, y otro a un trayecto más técnico denominado

“UTU” (Dirección General de Educación Técnico Profesional). En general, se asocia al liceo como la trayectoria previa al ingreso a los estudios universitarios de grado, mientras que se asocia a las UTU con el aprendizaje técnico y de oficios, que permiten una salida laboral más temprana.

Cada estudiante que participa en el sistema debe ingresar en una aplicación tres opciones ordenadas según su preferencia, y la ANEP utiliza un algoritmo, denominado “Dictador Serial”, donde la prioridad de los centros sobre los estudiantes se basa únicamente en un índice de vulnerabilidad social construido por este organismo (la prioridad es la misma para todas las instituciones). A un mayor valor del índice, se considera que la vulnerabilidad del estudiante es mayor y tiene mayor prioridad. El algoritmo parte de este orden de prioridad y asigna al primer estudiante a su primera opción, luego el segundo estudiante en el orden es asignado a su opción preferida, dentro de aquellas en las cuales aún queda cupo, y así sucesivamente hasta completar toda la lista ordenada de estudiantes (o que los centros llenen sus cupos y queden estudiantes sin asignar).

Al analizar este índice de vulnerabilidad social por subsistema, se observa que la distribución acumulada del índice entre los asignados a liceos está dominada en sentido estocástico de primer orden por la distribución del índice de los asignados a UTU. Lo anterior también se cumple si se analiza en función de la primera opción rankeada por el estudiante y no su asignación efectiva. Esto implica que los estudiantes más vulnerables tienden a elegir y a ser asignados en mayor proporción a una UTU que a un liceo. También se observa que la participación relativa de estudiantes vulnerables en UTU duplica o incluso triplica la cantidad de estudiantes vulnerables en liceos. Adicionalmente también se encuentra segregación si es medida por el índice de disimilitud (Duncan-Duncan). Esta evidencia es el punto de partida del trabajo, y principal motivación para entender las preferencias de los estudiantes.

El **principal objetivo** del presente trabajo es investigar si la segregación encontrada se debe a la ubicación geográfica de los estudiantes o los centros, a temas relacionados con las preferencias de los estudiantes, o si tiene origen en el sistema de asignación, en cuanto a las prioridades o cupos de los centros.

Para ello, en primer lugar, se realiza de un ejercicio de simulación con asignaciones contrafactuales siguiendo a Greaves (2023). En él, se concluye que la distancia no explica, al menos en su totalidad, la dominancia estocástica que se observa al considerar ambas distribuciones en función de la asignación a la primera opción. Es decir que la segregación no sería sólo un efecto de la segregación residencial y estaría mediada por las preferencias de los estudiantes o por características distintas a la distancia. Incluso, si se asigna a los estudiantes a su opción más cercana, sin tener en cuenta los cupos, la dominancia desaparece.

En segundo lugar, se estiman las preferencias cardinales de los estudiantes de acuerdo a la metodología de Aue et al. (2020) a través de una estimación bayesiana. La hipótesis parte de que un centro con un lugar más alto en el ranking de un estudiante supone una utilidad mayor para dicho estudiante. Las variables que se consideran para la estimación refieren a: la distancia del domicilio del estudiante a cada centro educativo, una dummy que indica si el centro es un liceo o una UTU, el índice de vulnerabilidad social de los estudiantes y un índice de vulnerabilidad por centro (que se construye a partir del promedio del índice para los

estudiantes asignados a dicho centro). También se incluye un efecto fijo por centro y una variable de interacción entre la vulnerabilidad del estudiante y el centro educativo.

En las estimaciones, se encontró que los estudiantes tienen preferencias por centros educativos más cercanos (lo cual es un resultado estándar en la literatura) y que los estudiantes no vulnerables derivan una mayor utilidad por los liceos que los estudiantes vulnerables una vez que se controló por la variable distancia. Específicamente, se encuentra que el hecho de ser un liceo tiene un impacto negativo o no significativo (en comparación con pertenecer al subsistema UTU) para los estudiantes vulnerables, mientras que el efecto es positivo y significativo para los no vulnerables. Este hallazgo es importante para un posible rediseño del sistema. Si los estudiantes vulnerables tienen preferencias similares entre ellos, entonces al darle prioridad a estos estudiantes frente al grupo de no vulnerables, se genera segregación educativa. En otras palabras, si se cumple la hipótesis de que hay un componente diferencial en las preferencias de estudiantes vulnerables y no vulnerables, al usar como prioridad de los centros el índice de vulnerabilidad, se convalidan dichas preferencias, lo que impactaría en la segregación del sistema.

Finalmente se utilizan los modelos estimados para realizar análisis de bienestar, con resultados ambiguos, aunque no contradictorios con los supuestos de la metodología utilizada. En particular, se encuentra que la utilidad de concurrir al centro educativo más cercano es en promedio mayor a la utilidad de concurrir al primer centro rankeado. Es importante aclarar que estos resultados deben ser tomados con precaución por la limitación en los datos con los que se contó. Como línea de trabajo futuro se sugiere la incorporación de variables relevantes que podrían considerar las familias a la hora de rankear los centros (como, por ejemplo, indicadores de ambiente educativo por centro) y que pueden mejorar este tipo de estimaciones. El modelo ampliado será una herramienta fundamental para la evaluación de propuestas de política.

Los hallazgos de este trabajo sustentarían el diseño de distintas políticas públicas. En primer lugar, es clave mejorar el nivel de conocimiento de las familias acerca del mecanismo de asignación implementado, como forma de que, a partir de un mayor nivel de información, los participantes se encuentren en una mejor posición para la toma de decisiones. En el transcurso de esta investigación, se constató que no existe información pública clara y de fácil acceso sobre el funcionamiento del sistema de asignación: algoritmo, prioridades, etc. En particular, es importante mejorar el conocimiento de las familias sobre las consecuencias en las trayectorias (por ejemplo, en términos del ingreso futuro) de la elección de un centro en secundaria versus UTU. En segundo lugar, el diseño actual de las prioridades podría generar segregación como reflejo de las preferencias de los estudiantes. Por lo tanto, se podrían analizar distintos diseños de prioridades, combinados con una política de discriminación positiva (por ejemplo, cuotas). Sin embargo, también se debe considerar que quizás el mecanismo puede tener un impacto positivo no analizado, donde estudiantes vulnerables que en otro escenario no concurrirían al sistema educativo, en la actualidad sí lo hacen, concurriendo a una UTU. En tercer lugar, se podría realizar un análisis en mayor profundidad para el caso de estudiantes que concurren a UTU, donde existen dos tipos de planes (Ciclo Básico Tecnológico-CBT y Formación Profesional Básica-FPB) y de esta forma enriquecer la discusión. Por último, sería interesante la producción de indicadores de calidad o desempeño educativo por centro y la difusión de estos resultados a las familias,

como forma de dar más información acerca de las diferentes posibilidades educativas que puede impactar en el reporte de preferencias sobre los centros (como se sugiere en Allende et al., 2019).

La tesis se estructura de la siguiente manera: en el capítulo siguiente se realiza una breve revisión de la literatura de *school choice* y su impacto en la segregación escolar. En el capítulo tres se introducen algunas definiciones y el marco teórico. En el capítulo cuatro, se describe el sistema de asignación en Uruguay, así como también las principales estadísticas descriptivas. El capítulo cinco se muestra la segregación en el sistema bajo distintos indicadores. En el seis, se presenta la metodología utilizada mientras que en el siete se detallan los resultados de dicha estimación y de las simulaciones contrafactuales. Finalmente, en el capítulo ocho contiene las principales conclusiones.

2. Revisión de la literatura

El presente trabajo se enmarca en la literatura de la teoría del matching y, en particular, en los problemas de *school choice* y su relación con la segregación escolar.

La implementación del mecanismo de aceptación diferida en Estados Unidos a partir de los años 2000 promovió el desarrollo y aplicación de mecanismos de asignación escolar alrededor del mundo. Con ellos, también han aparecido en los últimos años trabajos que buscan cuantificar el impacto en la segregación escolar de dichos mecanismos.

El modelo teórico de *school choice* es original de Abdulkadiroğlu y Sönmez (2003). En este trabajo, los autores fueron pioneros en analizar los problemas de *school choice* como un problema de diseño de mecanismos. En el trabajo, se analizan los inconvenientes que arrojaban los mecanismos que existían hasta el momento en Boston, Columbia, Minneapolis y Seattle. Además, plantean dos mecanismos alternativos, a prueba de estrategias, como forma de solucionar algunos de los problemas que se generaban con los mecanismos preexistentes. Primero, proponen un mecanismo basado en el algoritmo de aceptación diferida de Gale y Shapley (1962) y, en segundo lugar, plantean el denominado *Top Trading Cycles Mechanism* (TTC por sus siglas en inglés), basado en el algoritmo de Scarf y Shapley (1974).

En los últimos años se han publicado trabajos que estiman preferencias cardinales de los estudiantes bajo diferentes supuestos y metodologías. En Agarwal y Somaini (2020) se presenta una revisión que describe el marco metodológico que se ha desarrollado para realizar estas estimaciones a partir de la información proveniente de los mecanismos de asignación escolar. Algunos ejemplos de este tipo de trabajos son Agarwal y Somaini (2018), Fack et al. (2019), Calsamiglia et al. (2020) o Kapor et al. (2020).

En el trabajo de Aue et al. (2020), que se usará en el presente documento como base para la metodología empírica, si bien se no estudia particularmente la segregación, se analizan cuáles serían los potenciales beneficios en términos de bienestar social, de consolidar o integrar distritos escolares en Hungría. Se concluye que la consolidación del sistema generaría ganancias de bienestar en los estudiantes y dicha ganancia es cuantificada en términos de distancia, donde el beneficio promedio sería equivalente a asistir a una institución 5 km más cercana al domicilio del estudiante. Abdulkadiroğlu et al. (2017)

analizan las ganancias en términos de bienestar de pasar de un sistema de asignación descentralizado a un mecanismo de asignación centralizado y coordinado en la ciudad de Nueva York. Los autores concluyen que el uso del mecanismo coordinado alcanza un 80% de las posibles ganancias que se obtendrían del pasaje de un modelo *benchmark* a otro (pasar de un escenario sin coordinación, donde los estudiantes son asignados al centro más cercano sujeto a las restricciones de capacidad, a un sistema centralizado maximizador de utilidad).

En cuanto al impacto de los sistemas en la segregación, Oosterbeek et al. (2021) estudian la heterogeneidad de las preferencias de los estudiantes en Ámsterdam. Con datos del sistema de asignación de la ciudad, muestran que diferencias en las preferencias de los estudiantes explican el 40% de la segregación por grupos étnicos y 25% de la segregación según ingresos de los hogares. Con base a simulaciones, los autores encuentran que la introducción de cuotas para minorías reduce la segregación, pero tiene como costo que aumenta el número de estudiantes que quedan asignados a opciones menos preferidas.

En Kurscher et al. (2020), se investiga si la adopción de un sistema centralizado de asignación puede tener impacto en la segregación de estudiantes con relación a un sistema descentralizado. Para esto los autores utilizan la metodología de diferencias en diferencias para el caso chileno, que pasó de tener un sistema descentralizado a utilizar un algoritmo de aceptación diferida en 2016. Los autores encuentran que la implementación del sistema centralizado aumentó los niveles de segregación. También aplicado a la experiencia chilena, en Escobar y Huerta (2022) se propone un mecanismo teórico alternativo para el sistema de asignación escolar en las principales ciudades de Chile. Con base a un análisis contrafactual y utilizando los datos del sistema escolar para contrastar las implicancias de dicho modelo, se concluye que con la introducción de “reservas para minorías” se podría disminuir la segregación escolar. En el trabajo también se explicita una relación de beneficio-costos, donde a mayor cantidad de estudiantes asignados a su primera opción, se tiene como contracara un número superior de estudiantes que quedan sin ser asignados.

Lee y Son (2022) usan datos de secundaria de Nueva York para evaluar el impacto del sistema centralizado en la segregación y buscan descomponer dicho impacto en relación con: la ubicación residencial de los estudiantes, sus preferencias, problemas de información y políticas de admisión de los centros. Usando el modelo estimado, cuantifican el impacto de variaciones exógenas sobre la estabilidad del matching y su efecto en la distribución de los estudiantes. Se analizan los resultados en términos de: 1-distribución étnica y 2-proporción de estudiantes asignados a una de sus primeras cinco secundarias rankeadas según grupos. Los autores encuentran que, si bien el sistema utilizado en la ciudad puede llegar a disminuir la segregación y mejorar el bienestar de los estudiantes, esto se ve seriamente comprometido frente a desviaciones de un marco de información completa. A su vez, encuentran que las políticas de admisión y búsqueda de los centros contribuyen a aumentar la segregación entre estudiantes de diferentes ascendencias.

Laverde (2022) estudia las limitaciones del mecanismo utilizado en la ciudad de Boston para la asignación de preescolares en un contexto de segregación residencial. La autora encuentra que los estudiantes mayoritarios son asignados más frecuentemente a centros de mejor rendimiento que los estudiantes minoritarios. A su vez, sostiene que la diferencia en la brecha de rendimiento entre estudiantes de diferentes grupos étnicos no es menor a las que

se generaría por un mecanismo que asigna en función de la ubicación residencial. Al realizar simulaciones en base a estimaciones de preferencias sobre las escuelas (que a su vez se basan en rankings brindados por los padres), se concluye que la mitad de la diferencia de la brecha de rendimiento escolar entre estudiantes mayoritarios y minoritarios se debe a la mayor distancia que deben recorrer los estudiantes minoritarios para acudir a mejores centros.

En Greaves (2023) se busca responder si las causas de la segregación escolar se deben a diferencias en las preferencias de los padres o restricciones para acceder a las escuelas más preferidas. Con base a simulaciones de asignación, se encuentra que, al levantar todas las restricciones en el sistema y asignar cada estudiante a su opción preferida, se tienen similares niveles de segregación a las que se alcanzan bajo la asignación efectiva. Sin embargo, al realizar una segunda simulación donde los estudiantes son asignados al centro más cercano a su domicilio, se logra reducir considerablemente la segregación. Estos resultados irían en línea con la idea de que las preferencias de los padres exacerbaban la segregación escolar causada por segregación residencial.

En Idoux (2023), se analiza la contribución de los criterios de admisión o prioridades de los centros a la segregación en las escuelas secundarias de Nueva York. En base a un modelo de demanda escolar, se concluye que la mitad de la segregación encontrada se debe a los criterios de admisión, mientras que las preferencias de las familias y la segregación residencial explican el resto.

En cuanto a trabajos sobre segregación educativa en Uruguay, en INEE (2021) se estudia la evolución de la segregación socioeconómica en la educación pública de Uruguay para el período 2013-2019. Para esto, se construyen índices de igualdad y de exposición para los niveles inicial, primaria y media básica (liceo y UTU) y se encuentra una distribución más desigual en los niveles de educación inicial y primaria. A su vez, la segregación en dichos niveles aumenta en el periodo analizado y podría sustentarse en fenómenos territoriales. Por otro lado, respecto a la dimensión exposición, que refiere a la probabilidad de que un estudiante vulnerable se encuentre con otro estudiante de similares características, se concluye que la educación media técnica (UTU) es donde se tiene mayores niveles de exposición. Esto podría deberse a la alta proporción de estudiantes vulnerables en este subsistema.

3. Marco teórico y definiciones

En el presente capítulo se introduce el modelo de *school choice* y se presentan algunas propiedades de las asignaciones y mecanismos.

Un *school choice problem* se define como una tupla $(T, S, \succ_t, \succ_s, q)$ donde:

- T es un conjunto de estudiantes. $T = \{t_1, t_2, \dots, t_n\}$.
- S es un conjunto de centros educativos. $S = \{s_1, s_2, \dots, s_i\}$. Donde $Q = \{q_{s_1}, q_{s_2}, \dots, q_{s_i}\}$ es un vector con la capacidad máxima de estudiantes que cada centro puede aceptar.
- \succ_t es la preferencia estricta del estudiante t sobre el conjunto de centros. De esta forma, $s \succ_t s'$ implica que el estudiante t prefiere al centro s sobre el centro s' . Asimismo, $s \succsim_t s'$ implica que: $s \succ_t s'$ o $s \sim_t s'$. Si se denota como \emptyset la opción

“externa” (*outside option*) a todas las escuelas S para cada estudiante, entonces el centro s es aceptable para el estudiante t si: $s \succ_t \emptyset$.

- \succ_s es la prioridad estricta del centro s sobre el conjunto de estudiantes. De esta forma, $t \succ_s t'$ implica que el estudiante t tiene una prioridad mayor que el estudiante t' en el centro s . Asimismo, $t \succcurlyeq_s t'$ implica que: $t \succ_s t'$ o $t \sim_s t'$. De la misma forma que antes, si se denota como \emptyset la opción externa a todos los estudiantes T para cada centro, entonces el estudiante t es aceptable para el centro s si: $t \succ_s \emptyset$.

Dado un conjunto T de estudiantes y un conjunto S de centros, se define un *matching* μ como una función $\mu: T \rightarrow T \cup S$. La notación $\mu(t) = s$ significa que el estudiante t queda asignado al centro s en el *matching* μ . Puede suceder que $\mu(t) = t$, es decir que el estudiante t queda sin asignar bajo el *matching* μ . Un *matching* es factible si la cantidad de estudiantes asignados al centro s bajo el *matching* μ es menor o igual a la capacidad máxima del centro, esto es $|\mu^{-1}(s)| \leq q_s$. Se trabajará en lo que sigue con *matching* factibles.

Un *matching* μ es estable si no puede ser bloqueado individualmente por un estudiante o por un centro, ni tampoco puede ser bloqueado por un par (t, s) con $t \in T$ y $s \in S$, es decir:

1. Condición de no bloqueo individual. $\mu(t)$ es aceptable para cada estudiante t y cada $t \in \mu^{-1}(s)$ es aceptable por cada centro s .
2. Condición de no bloqueo de a pares. No existe un par (t, s) con $t \in T$ y $s \in S$ tal que:
 - i. $s \succ_t \mu(t)$ y
 - ii. $|\mu^{-1}(s)| < q_s$ y $t \succ_s \emptyset$ o $t \succ_s t'$ para algún $t' \in \mu^{-1}(s)$.

Un *matching* μ es Pareto óptimo si no existe otro *matching* μ' tal que: $\mu'(t) \succcurlyeq_t \mu(t)$ para todo $t \in T$ y $\mu'(j) \succ_t \mu(j)$ para algún $j \in T$. Es decir, que no existe un *matching* alternativo a μ en el cual todos están por lo menos tan bien como en μ' , y un estudiante esté estrictamente mejor. Notar que en este modelo sólo se considera a los estudiantes cuando se habla de “bienestar”, los lugares en los centros educativos son objetos para asignar.

Un mecanismo ϕ es una función o procedimiento sistemático que mapea desde perfiles de preferencias de estudiantes, prioridades y capacidades de centros educativos a *matchings*. Es decir, si se fija las capacidades de los centros, y se denomina como φ al conjunto de todas las preferencias de estudiantes, ψ al conjunto de las prioridades de los centros y ω al conjunto de todos los *matching*, entonces se tiene que: $\phi: \varphi x \psi \rightarrow \omega$.

Un mecanismo ϕ es *strategy-proof* a prueba de estrategias si dado un perfil de preferencias de los estudiantes (\succ) , revelar las verdaderas preferencias (\succ_t) es al menos tan bueno como cualquier otra estrategia (\succ'_t) . Es decir que: $\phi_t(\succ) \succcurlyeq_t \phi_t(\succ'_t, \succ_{-t})$, donde $\phi_t(\succ)$ es la asignación del estudiante t en el *matching* $\phi(\succ)$.

Siguiendo a Abdulkadiroglu y Sonmez (1998) se denomina Dictador Serial a aquel mecanismo donde se considera un orden de prioridad de elección entre los estudiantes y, a partir de este orden, el primer estudiante es asignado a su primera opción, luego el estudiante que se encuentra en segundo lugar es asignado a su opción preferida, dentro de aquellas en las cuales aún queda cupo, y así sucesivamente hasta completar toda la lista ordenada de estudiantes (o que los centros llenen sus cupos y queden estudiantes sin asignar). Cuando las

prioridades de todos los centros educativos es la misma, entonces se puede correr el mecanismo de Dictador Serial usando ese orden. En este caso, el resultado coincide con el del mecanismo de aceptación diferida (Gale y Shapley, 1962).

El mecanismo Dictador Serial es *strategy-proof* (o a prueba de estrategias), estable y eficiente en el sentido de Pareto.

En muchos casos, los estudiantes se encuentran restringidos en el número de centros que pueden rankear, esto se conoce como “*Constrained school choice*” (Haeringer y Klijn, 2009), lo que destruye todas las propiedades anteriores. El mecanismo de Dictador Serial deja de ser *strategy-proof*, eficiente y estable. Este es el caso de análisis del presente documento, en el cual los estudiantes pueden rankear como máximo tres opciones.

El siguiente ejemplo ilustra el efecto de restringir el número de opciones que cada estudiante puede listar, en los incentivos del mecanismo. Se supone un mercado con tres estudiantes ($t_{1,2,3}$) y tres centros ($s_{1,2,3}$) donde cada uno tiene un cupo de un estudiante. Las preferencias de los estudiantes y la prioridad de los centros están dadas por:

- $t_1: s_2 \succ_{t_1} s_1 \succ_{t_1} s_3 \succ_{t_1} \emptyset$
- $t_2: s_1 \succ_{t_2} s_2 \succ_{t_2} s_3 \succ_{t_2} \emptyset$
- $t_3: s_1 \succ_{t_3} s_2 \succ_{t_3} s_3 \succ_{t_3} \emptyset$
- $s_{1,2,3}: t_2 \succ_{s_{1,2,3}} t_3 \succ_{s_{1,2,3}} t_1 \succ_{s_{1,2,3}} \emptyset$

De esta forma, si se corre el algoritmo conocido como Dictador Serial usando como orden la prioridad de los centros, el *matching* resultante es el siguiente:

$$\mu = \begin{pmatrix} t_1 & t_2 & t_3 \\ s_3 & s_1 & s_2 \end{pmatrix}$$

En cambio, si se restringe el número de centros que pueden rankear los estudiantes solamente a un centro y a su vez estos revelan sus verdaderas preferencias, el nuevo *matching* está dado por:

$$\mu' = \begin{pmatrix} t_1 & t_2 & t_3 \\ s_2 & s_1 & \emptyset \end{pmatrix}$$

Notar que el *matching* que resulta no es estable ni eficiente. Como se puede apreciar, al restringir la cantidad de opciones a rankear, t_3 tiene incentivos a no revelar su verdadera preferencia. Si bien su centro preferido es el primero, al verse obligado a declarar un solo centro, es una mejor estrategia declarar en el ranking a s_2 y de este modo el nuevo *matching* sería:

$$\mu'' = \begin{pmatrix} t_1 & t_2 & t_3 \\ \emptyset & s_1 & s_2 \end{pmatrix}$$

Donde $\mu'' \succ_{t_3} \mu'$ y que igualmente no sería un *matching* eficiente ni estable. Por lo cual el mecanismo Dictador Serial restringiendo las opciones que pueden rankear los estudiantes no es *strategy-proof*. Nótese además que si t_1 incorpora el comportamiento estratégico de t_3 , entonces debería revelar como su única opción s_3 y de esta forma se alcanza nuevamente el *matching* μ , donde $\mu \succ_{t_1} \mu''$.

4. Mecanismo de asignación en Uruguay, base de datos y estadísticas descriptivas

En este capítulo se describe el sistema de asignación de estudiantes que ingresan a la educación media pública. Se detalla la base de datos utilizada y se realiza una descripción de las estadísticas respecto al índice de vulnerabilidad social y la caracterización geoespacial.

El sistema de asignación se basa en tres elementos: 1) el número de lugares disponibles de cada centro (también llamado cupo), 2) las preferencias de los estudiantes por los centros y 3) un orden de prioridad de los estudiantes. Con estos tres elementos, se corre un algoritmo para hallar la asignación de cada estudiante. A continuación, se describe en detalle cada uno de estos componentes.

4.1. Oferta: Centros de educación

Los centros que participan en el sistema de asignación pertenecen a dos subsistemas distintos. Un subsistema puede asociarse a una trayectoria más académica o tradicional cuyos centros se denominan “Liceos”, y un trayecto más técnico denominado “UTU” (Dirección General de Educación Técnico Profesional – UTU). Si bien las características de estas instituciones han cambiado en los últimos años, se asocia al liceo como la trayectoria necesaria a realizar previo al ingreso a los estudios universitarios de grado, mientras que se asocia a las UTU con el aprendizaje técnico y de oficios, que permiten una salida laboral más temprana².

Al observar la distribución de la oferta educativa para 2023, se observan 381 opciones para liceos y 241 opciones para UTU en todo el país (en el caso de UTU las opciones corresponden a un centro y plan, por ejemplo, robótica). Aproximadamente un cuarto de los cupos y de las instituciones educativas corresponden a Montevideo.

Tabla N°1: Distribución y cupos de las opciones educativas

		Liceo	UTU	Total
N° de opciones	Montevideo	83	71	154
	Resto del país	298	170	468
	Total	381	241	622
Capacidad total	Montevideo	9.430	3.269	12.699
	Resto del país	23.781	7.416	31.197
	Total	33.211	10.685	43.896

4.2. Demanda: Preferencias de estudiantes

Cada estudiante que participa en el sistema –esto es, que se encuentra cursando sexto año de primaria en una escuela pública– debe ingresar tres opciones ordenadas según su preferencia en el sistema. Esto se hace alrededor de septiembre/octubre del año anterior al cual se inicia el ciclo de educación media básica. En 2022, 42.131 estudiantes reportaron preferencias, de

² Por más información ver: <https://www.anep.edu.uy/>. Ver opción “Subsistemas educativos”.

los cuales el 93% (39.198) declararon como primera opción un liceo o una UTU. Existen otras alternativas elegidas por los estudiantes, como la elección de un centro de educación privada, cursar un año extra de primaria para el caso de ciertos alumnos, o elegir otro tipo de institución diferente a las detalladas.

En cuanto a la demanda por las instituciones, se rankearon en total 512 centros (centros que fueron elegidos al menos una vez en una de las tres opciones de los estudiantes), casi la totalidad de la oferta disponible. A su vez, 274 corresponden a liceos y 238 a UTU. Al comparar la cantidad de veces que cada centro fue demandado como primera opción con la capacidad de dicha institución, se obtiene un indicador de sobredemanda. En particular, un centro se dice que está sobredemandado si esta diferencia es mayor a cero. En este sentido, 118 centros fueron sobredemandados, de los cuales 67 correspondió a liceos y 51 a UTU. Si se observa la sobredemanda por cantidad de estudiantes, se tiene que alcanzó a 1.971 alumnos para liceos y 767 para UTU.

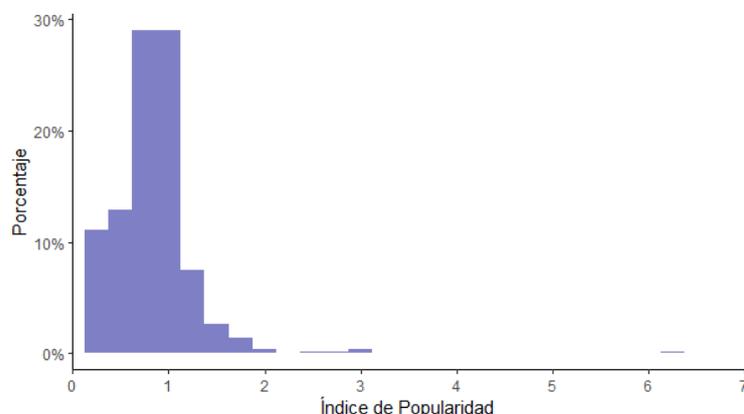
Tabla N°2: Centros sobredemandados como primera opción

	Centros	Part. % sobre centros	Estudiantes	Part. % sobre total
Liceo	67	18%	1.971	7%
UTU	51	21%	767	9%
Total	118	19%	2.738	7%

La participación de centros es sobre el total disponible por tipo de institución (liceo o UTU). La participación de los estudiantes es sobre el total de aquellos que optaron por un liceo o por una UTU como primera opción.

Para tener otra idea de la diferencia que existe entre la demanda y oferta, se calculó un índice de popularidad como el cociente entre el total de estudiantes que rankean cada centro como primera opción sobre la capacidad total de dicho centro. De esta forma, un centro será popular si su índice es mayor o igual a 1. En el gráfico N°1 se presenta un histograma del índice de popularidad y, tal como se puede apreciar, la mayoría de los centros tienen un índice cercano a 1.

Gráfico N°1: Histograma del índice de popularidad de los centros



Para trabajar con los datos de demanda de los estudiantes y realizar las estimaciones se realizaron algunos filtros a la base total. Se consideraron:

- 1) Estudiantes que fueron asignados a un liceo o una UTU³;
- 2) Estudiantes que fueron geolocalizados⁴;
- 3) Estudiantes que cumplieron con un criterio de distancia desde su domicilio a la institución asignada⁵.

De esta forma, se trabajó con 36.001 estudiantes (85% de la base original) de los cuales el 91% fue asignado a su primera opción. El 79% rankeó como primera opción un liceo mientras que el restante 21% rankeó en primer lugar una UTU.

Además del identificador para cada alumno que participó del mecanismo, en la base se cuenta con las siguientes variables: edad, escuela a la que concurrió el alumno, índice de vulnerabilidad social, las tres opciones elegidas, la opción a la cual fue derivado en el mecanismo y el domicilio del estudiante.

Tabla N°3: Estadísticas de estudiantes que participaron en 2022

	Media	Desvío Estándar	Min	Max	Mediana	N
Año de nacimiento	2010	0,772	1980	2012	2010	36.001
Edad	12	0,652	10	42	12	36.001
Índice de vulnerabilidad	0,378	0,213	0	1	0,345	36.001
Índice de vulnerabilidad - asignado a Liceo	0,349	0,198	0	1	0,301	28.417
Índice de vulnerabilidad - asignado a UTU	0,489	0,232	0	1	0,475	7.584
Distancia a Liceo asignado - km	5,110	9,680	0	60	1,552	28.417
Distancia a UTU asignada - km	6,000	10,767	0	60	1,857	7.584

4.3. Prioridades: Índice de vulnerabilidad social

El índice de vulnerabilidad es construido para determinar las prioridades que los alumnos tienen para ser asignados a los centros. En general, las prioridades se utilizan para decidir qué estudiantes asignar en el caso que la demanda por un centro sea mayor a la oferta de cupos. Centros distintos podrían tener prioridades distintas, aunque, en el caso de Uruguay, existe un único orden de prioridad para todos los centros. El índice toma valores entre 0 y 1 y, a un mayor valor del índice, se considera que la vulnerabilidad del estudiante es mayor y tiene mayor prioridad. En caso de empate entre estudiantes con el mismo valor del índice, se los ordena en función de la fecha y hora en la que se completó el registro.

³ Se excluyeron 2.875 estudiantes geolocalizados que fueron asignados a otras instituciones, como instituciones privadas, séptimo de primario u otras instituciones diferentes a un liceo u UTU. A su vez, de estos estudiantes, el 97% tuvo como primera opción una institución que no corresponde a un liceo o UTU.

⁴ Por más información sobre la geolocalización ver apéndice.

⁵ Por las características geográficas y demográficas de Uruguay se consideró como criterio para incluir en la base a los estudiantes cuyo domicilio se encuentra a menos de 25 km de la institución asignada para el caso de Montevideo, y a menos de 60 km para el resto del país. Unos 2.500 alumnos que cumplían con el criterio 1 y 2 no cumplían con el requisito de distancia. Se decidió excluir a estos estudiantes por considerar que pueden estar mal georreferenciados por diferentes motivos, como error en el procedimiento de geolocalización o inconvenientes en la declaración del domicilio.

Para la elaboración del índice se tienen en cuenta características del estudiante tales como: cantidad de inasistencias a la escuela, calificaciones escolares, acceso a prestaciones sociales del estado y nivel de extra edad (en referencia a la edad correspondiente según el año que cursa). En el gráfico N°2 se detalla el histograma del índice para la base de estudiantes considerada.

Gráfico N°2: Histograma: Índice de vulnerabilidad de estudiantes

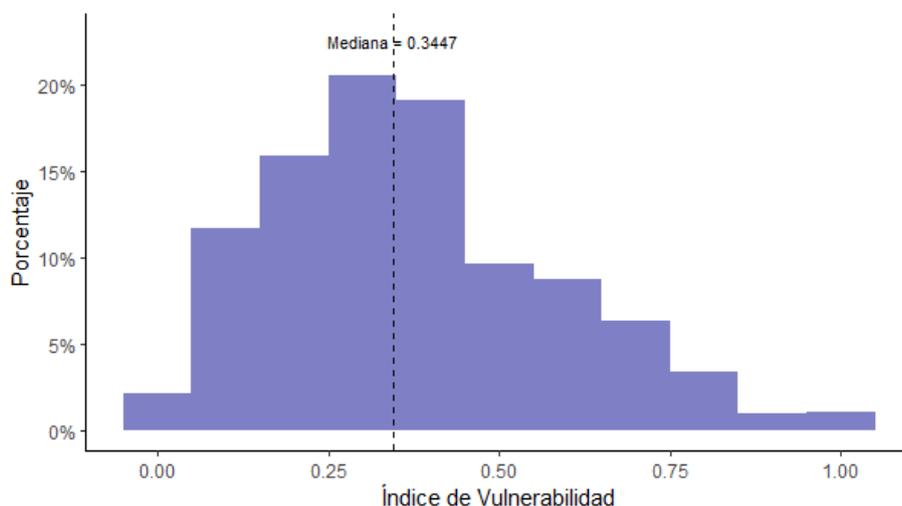


Tabla N°4: Descriptivos básicos del índice de vulnerabilidad de estudiantes

Medida	Valor
Mínimo	0
1er cuartil	0,2166
Mediana	0,3447
Media	0,3780
3er cuartil	0,5208
Máximo	1

4.4. Mecanismo de asignación

La ANEP utiliza el siguiente algoritmo que consta de seis pasos. En los primeros cuatro pasos se asignan a los estudiantes que: tienen un centro educativo asociado (CEA) que declararon en primer lugar, estudiantes de escuelas especiales, estudiantes de programas priorizados y estudiantes de escuelas rurales. El 13% del total de estudiantes fue asignado en estos pasos en 2022. El quinto paso es el principal, ya que en este se asignó al 87% de los estudiantes. Para ello, se aplica el algoritmo denominado “Dictador Serial”, donde la prioridad de los centros sobre los estudiantes se basa en el índice de vulnerabilidad social y dicha prioridad es la misma para todas las instituciones. Este ranking se construye ordenando los estudiantes por el índice de vulnerabilidad. En el último paso se asigna a los estudiantes sin centro de forma de minimizar el sobrecupo del sistema.

De esta forma, primero se considera al estudiante con mayor valor del índice de vulnerabilidad social y se lo asigna a su primera opción. Luego, se considera al segundo estudiante y se lo asigna a su opción más preferida, entre aquellas que aún tienen cupo

disponible, y así sucesivamente hasta que se consideran todos los estudiantes. En caso de que, para algún estudiante, sus tres opciones rankeadas no tengan cupo, se lo deja sin asignación en esta etapa. Luego se los considera en un último paso.

El 91% de los estudiantes que participaron del mecanismo en 2022 fueron asignados a su primera opción.

4.5. Distancia: Estadísticas descriptivas

A continuación, se detallan algunas descripciones geoespaciales respecto a los alumnos que participaron en el sistema de asignación.

En las tablas N°5 a N°7 se presentan características en cuanto a la cantidad de centros en determinados radios de distancia. En la Tabla N°5 se observa que en promedio los estudiantes de Montevideo cuentan con unos 16 centros en un radio de hasta 3 km, mientras que dicho valor es de 5 centros para estudiantes en el resto del país.

Tabla N°5: Cantidad de centros (promedio) según tipo liceo/UTU y según Montevideo o Interior

	Liceo	UTU	Total
Estudiantes de Montevideo			
Instituciones a 1 km	1,08	1,10	2,18
Instituciones a 3 km	8,15	8,00	16,16
Instituciones a 5 km	19,31	18,00	37,31
Instituciones a 10 km	52,48	49,58	102,05
Instituciones a 20 km	87,76	89,73	177,49
Estudiantes del Interior			
Instituciones a 1 km	0,88	0,73	1,61
Instituciones a 3 km	2,67	2,39	5,05
Instituciones a 5 km	3,70	3,37	7,06
Instituciones a 10 km	5,89	5,25	11,14
Instituciones a 20 km	16,56	15,15	31,70
Total			
Instituciones a 1 km	0,95	0,84	1,79
Instituciones a 3 km	4,39	4,15	8,53
Instituciones a 5 km	8,59	7,95	16,55
Instituciones a 10 km	20,50	19,15	39,64
Instituciones a 20 km	38,88	38,53	77,41

Si bien la oferta de liceos es levemente mayor a la de UTU para el total del país, para el caso de los estudiantes vulnerables (índice mayor a 0,6999) se tienen más UTU que liceos en un radio de hasta 3 km. Esto no sucede para el caso de los estudiantes no vulnerables (ver tabla N°6). De todas formas, las diferencias no parecen ser significativas.

Tabla N°6: Cantidad de centros (promedio) según liceo/UTU e índice de vulnerabilidad social

Vulnerables sí: índice > 0,6999	Liceo	UTU	Total
Estudiantes Vulnerables			
Instituciones a 1 km	0,89	0,91	1,80
Instituciones a 3 km	4,90	5,17	10,07
Instituciones a 5 km	10,47	10,36	20,83
Instituciones a 10 km	27,79	26,39	54,18
Instituciones a 20 km	53,80	53,84	107,64
Estudiantes No Vulnerables			
Instituciones a 1 km	0,95	0,84	1,79
Instituciones a 3 km	4,33	4,03	8,35
Instituciones a 5 km	8,37	7,67	16,03
Instituciones a 10 km	19,62	18,28	37,90
Instituciones a 20 km	37,09	36,69	73,78

Finalmente, en la Tabla N°7 se aprecia que en un radio de hasta 3 km los estudiantes vulnerables de Montevideo tienen menor cantidad de liceos disponibles que los no vulnerables. A su vez se mantiene la afirmación de que en dicho radio la oferta de liceos es menor a la de UTU para estudiantes vulnerables. No obstante, dichas diferencias son relativamente pequeñas.

Tabla N°7: Cantidad de centros (promedio) para estudiantes de Montevideo según tipo de centro e índice de vulnerabilidad

Vulnerables sí: índice > 0,6999	Liceo	UTU	Total
Vulnerables - Montevideo			
Instituciones a 1 km	0,98	1,12	2,09
Instituciones a 3 km	7,48	8,13	15,61
Instituciones a 5 km	17,88	17,86	35,73
Instituciones a 10 km	51,40	49,17	100,57
Instituciones a 20 km	88,89	90,78	179,67
No Vulnerables - Montevideo			
Instituciones a 1 km	1,10	1,09	2,20
Instituciones a 3 km	8,28	7,98	16,26
Instituciones a 5 km	19,58	18,03	37,61
Instituciones a 10 km	52,68	49,66	102,33
Instituciones a 20 km	87,54	89,54	177,08

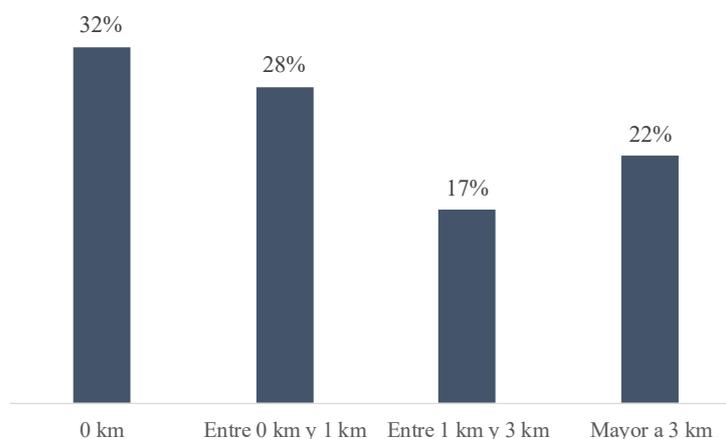
En cuanto a las opciones rankeadas, sobresale que el 32% de los alumnos tiene como primera opción el centro más cercano a su domicilio, mientras que el 12% elige en primer lugar a su segundo centro más cercano. En total, el 63% de los alumnos rankea en primer lugar a un centro que se encuentra en entre los cinco más cercanos a su domicilio.

Tabla N°8: Cercanía al domicilio de la primera opción rankeada. Part. % sobre el total de estudiantes

Centro más cercano a su domicilio	Rankeado como primer opción
Primero	32%
Segundo	12%
Tercero	9%
Cuarto	6%
Quinto	4%
Fuera de los 5 más cercanos	37%
Total	100%

A su vez, en el gráfico N°3 se muestra cómo se distribuye la diferencia entre la distancia a la primera opción de cada estudiante y su centro más cercano. Como se puede apreciar, para el 78% de los estudiantes dicha diferencia es menor a los 3km.

Gráfico N°3: Distancia entre primera opción y centro más cercano. Distribución sobre el total de estudiantes



Por otro lado, se aprecia que, para el total de la muestra, los estudiantes tienen un liceo más cerca de su domicilio que una UTU y esto también se cumple al observar estudiantes vulnerables y no vulnerables. Esta comparación refiere específicamente al tipo de centro más cercano al domicilio de cada estudiante ya que, como se comentó anteriormente, en un radio de hasta 3 km, los estudiantes vulnerables tienen una mayor oferta de UTU que de liceos.

Tabla N°9: Distancia en km del domicilio de los estudiantes al Liceo/UTU más cercano/a

Vulnerable si: índice >0,6999		1er cuartil	Mediana	Media	3er cuartil	Max
Total muestra	Liceo	0,53	0,90	1,71	1,43	46
	UTU	0,73	1,26	3,24	2,10	72
Vulnerable	Liceo	0,59	0,97	1,60	1,53	36
	UTU	0,74	1,25	2,41	1,98	71
No Vulnerable	Liceo	0,53	0,89	1,73	1,42	46
	UTU	0,72	1,26	3,34	2,10	72
Montevideo	Liceo	0,58	0,90	1,05	1,26	8
	UTU	0,74	1,19	1,36	1,75	9
Interior	Liceo	0,52	0,90	2,01	1,52	46
	UTU	0,71	1,31	4,10	2,44	72

En la Tabla N°10 se presentan algunas estadísticas descriptivas respecto a la distancia del hogar de los estudiantes a su institución asignada, así como también a cada una de las tres opciones que rankearon. El 50% de los alumnos es asignado a una institución que se encuentra a menos de 1,61 km de distancia de su hogar. Por otro lado, la distancia promedio en Montevideo (3,59 km) es considerablemente menor a la del interior del país (6,05 km), pero la mediana y el tercer cuartil es menor para el interior del país.

La distancia promedio (y la mediana) crece a medida que se pasa de la opción preferida (opción 1) a una menos preferida (opción 2 y opción 3). Además, en general se tiene que, para el caso de la mediana y el tercer cuartil, los estudiantes vulnerables rankean y son asignados a instituciones algo más lejanas que los estudiantes no vulnerables. Para el caso de Montevideo este fenómeno también ocurre para el promedio.

Finalmente, se calculó la distancia desde cada centro educativo al domicilio de todos los estudiantes asignados a dicha centro, para luego obtener medidas que podrían definirse como “área de captura” de los centros. Para cada opción se determinaron la distancia promedio a los estudiantes asignados y los cuartiles. En la Tabla N°11 se detalla el promedio de dichas medidas para los datos agregados, diferenciando entre Montevideo e Interior, así como entre liceo y UTU. Por ejemplo, para el total de casos, el “área de captura” de cada centro tiene una mediana de 3,8 km en promedio. Dicha área cae a 2,7 km para el caso de los centros de Montevideo. Asimismo, el “área de captura” es mayor en UTU que en liceos.

Tabla N°10: Distancia en km del domicilio de los estudiantes a instituciones

Distancia en Km	Medida	Asignación	Opción 1	Opción 2	Opción 3
Total datos	Media	5,30	5,28	7,85	9,83
	Mediana	1,61	1,57	2,32	2,88
	3er cuartil	3,92	3,88	6,20	8,32
Montevideo	Media	3,69	3,59	4,20	4,63
	Mediana	2,12	2,03	2,60	2,90
	3er cuartil	4,57	4,41	5,05	5,39
Interior	Media	6,04	6,05	9,54	12,34
	Mediana	1,40	1,39	2,17	2,88
	3er cuartil	3,50	3,48	8,02	13,67
Vulnerable	Media	5,16	5,15	6,92	8,57
	Mediana	2,03	2,02	2,52	3,16
	3er cuartil	4,96	4,94	6,36	7,55
No vulnerable	Media	5,32	5,30	7,96	9,98
	Mediana	1,57	1,52	2,29	2,85
	3er cuartil	3,79	3,75	6,17	8,43
Vulnerable Mdeo	Media	4,32	4,32	4,73	5,10
	Mediana	2,54	2,54	2,80	3,29
	3er cuartil	5,58	5,58	6,07	6,37
No vulnerable Mdeo	Media	3,57	3,46	4,10	4,55
	Mediana	2,05	1,96	2,55	2,84
	3er cuartil	4,34	4,19	4,89	5,20
Vulnerable Interior	Media	5,89	5,88	8,86	11,71
	Mediana	1,64	1,63	2,33	3,02
	3er cuartil	3,93	3,91	6,83	11,36
No vulnerable Interior	Media	6,05	6,07	9,60	12,40
	Mediana	1,39	1,37	2,15	2,86
	3er cuartil	3,45	3,44	8,13	14,00

Se consideraron vulnerables aquellos estudiantes con un índice de vulnerabilidad mayor a 0,6999.

Tabla N°11: Área de captura de los centros (en km)

	Mediana	Promedio	3er cuartil
Total	3,8	7,05	8,9
Montevideo	2,7	4,1	4,9
Interior	4,2	8,1	10,2
Liceo	2,9	6,8	8,5
UTU	4,9	7,4	9,3

En conclusión, si bien existe mayor oferta de liceos que de UTU, los estudiantes vulnerables (índice mayor a 0,6999) tienen mayor cantidad de UTU que de liceos en un radio de hasta 3 km, aunque dicha diferencia es pequeña. El 32% de los estudiantes rankea como primera opción el centro más cercano a su domicilio y el 75% vive a menos de 3,92 km del centro al cual es asignado. En general, los estudiantes vulnerables rankean y son asignados a instituciones algo más lejanas (para mediana y tercer cuartil). Para el caso de Montevideo este fenómeno también ocurre para el promedio.

5. Segregación

En esta sección, se investiga la existencia de segregación entre los estudiantes asignados a liceo y UTU. Para esto, es usual utilizar algún tipo de indicador que mida la segregación escolar como forma de analizar dicha distribución. Por definición, se podría afirmar que existe segregación escolar (o ausencia de ella) cuando la distribución de alguna característica de los alumnos en un grupo escolar particular (una escuela o una clase) difiere (o no difiere) de la distribución de dicha característica para la población en general. Existe consenso de que la segregación educativa tiene consecuencias negativas en los estudiantes. Por ejemplo, tiene efectos negativos al incrementar las diferencias en puntaje de pruebas y en logros educativos según diferentes grupos de estudiantes (Card y Rothstein (2006) y Billings et al., (2012)).

A continuación, se presentan algunas medidas de segregación escolar para el caso de los estudiantes que participaron del sistema en 2022 y fueron asignados a un liceo o UTU.

5.1. Distribución del índice de vulnerabilidad entre estudiantes de liceo y UTU

Como primera evidencia de segregación, en el Gráfico N°4 se observa que la distribución del índice de vulnerabilidad entre los asignados a liceos está dominada en sentido estocástico de primer orden por la distribución del índice de los asignados a UTU. También se cumple dicha dominancia si se considera la distribución del índice entre aquellos estudiantes que eligieron en primer lugar un liceo o una UTU (y no la asignación efectiva). Esto es esperable, ya que, como se mencionó anteriormente, el 91% de los alumnos fue asignado a su primera opción (ver Gráfico N°5). Esto implica que los estudiantes más vulnerables tienden a elegir y a ser asignados en mayor proporción a un centro de UTU que a un liceo.

Gráfico N°4: Distribución acumulada: Índice de vulnerabilidad de estudiantes según asignación

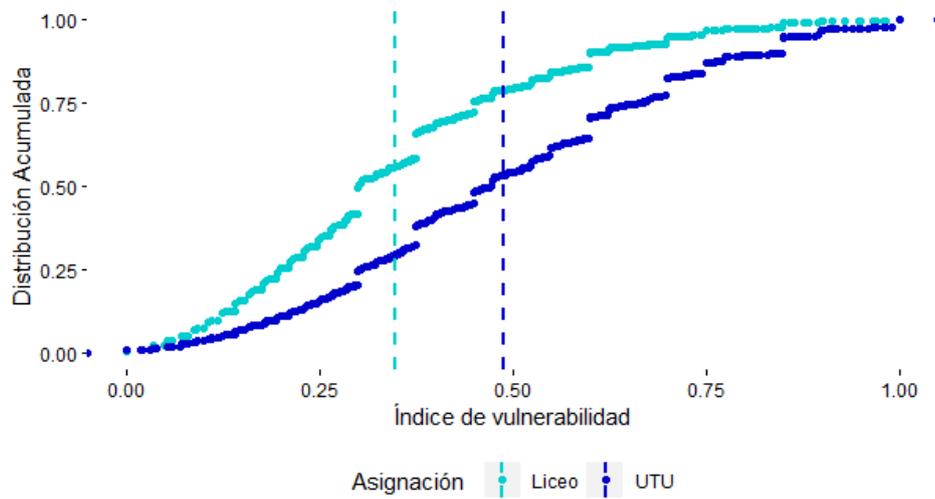
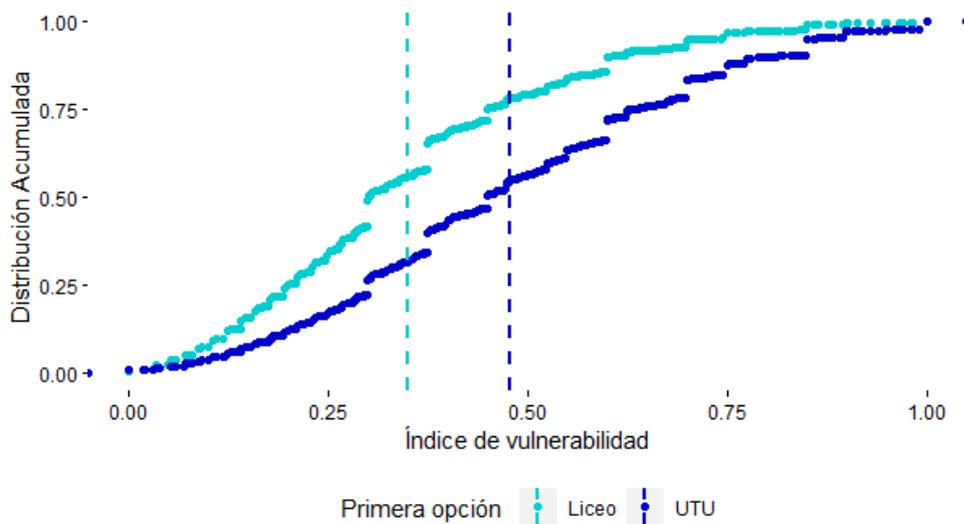


Gráfico N°5: Distribución acumulada: Índice de vulnerabilidad de estudiantes según primera opción



A su vez, se realizó un test de Kolmogorov-Smirnov para chequear las distribuciones acumuladas según la asignación a liceo y UTU. Los resultados van en la dirección de rechazar la hipótesis nula igualdad de distribuciones (se rechaza que ambas distribuciones pueden generarse a través de un proceso aleatorio) y por lo cual existirían diferencias entre liceo y UTU en la distribución de estudiantes según el índice de vulnerabilidad⁶.

⁶ El test de Kolmogorov-Smirnov se realizó en R con la función “*ks.boot*” del paquete “*Matching*”. Dicha función realiza una versión del test que utiliza *bootstrapping* ya que la distribución acumulada tiene saltos y no es continua. Por más información ver: <https://search.r-project.org/CRAN/refmans/Matching/html/ks.boot.html>

5.2. Participación relativa de estudiantes vulnerables o no vulnerables

Como segunda evidencia de segregación, y para observar la distribución de estudiantes según vulnerabilidad, se creó una variable dicotómica que separa a los estudiantes en vulnerables o no vulnerables. Para esto, se especificaron los resultados para tres casos:

- 1) Índice mayor a 0,5208 (tercer cuartil);
- 2) Índice mayor a 0,5999;
- 3) Índice mayor a 0,6999 (corte utilizado por ANEP para determinar a los estudiantes vulnerables priorizados).

En la Tabla N°12 se detallan los resultados para los tres cortes elegidos para el índice, donde se muestra la participación de estudiantes vulnerables/no vulnerables sobre el total dentro del tipo de institución (liceo o UTU). Se observa que la participación de estudiantes vulnerables en UTU duplica o incluso triplica (para el caso del corte mayor a 0,6999) la cantidad de estudiantes vulnerables en liceos.

Tabla N°12: Participación de estudiantes vulnerables y no vulnerables según tipo de institución y según corte

Muestra	N	Inst.	Vulnerables > 0,5208		Vulnerables > 0,5999		Vulnerables > 0,6999	
			No vulnerables	Vulnerables	No vulnerables	Vulnerables	No vulnerables	Vulnerables
Total muestra	28.417	Liceo	80,2%	19,8%	85,6%	14,4%	92,5%	7,5%
	7.584	UTU	55,7%	44,3%	64,4%	35,6%	77,2%	22,8%
Mdeo.	8.813	Liceo	73,1%	26,9%	79,6%	20,4%	88,3%	11,7%
	2.364	UTU	42,6%	57,4%	52,3%	47,7%	68,4%	31,6%

5.3. Índice de Disimilitud: Duncan y Duncan

Como forma de profundizar el análisis de las diferencias en la asignación según tipo de centro, se detalla el índice de disimilitud de Duncan, indicador utilizado para describir segregación entre grupos. Para el caso de estudiantes vulnerables y no vulnerables en la educación media pública de Uruguay, dicho índice se calcula de la siguiente forma:

$$D = \frac{1}{2} \sum_{i=1}^S \left| \frac{p_i}{P} - \frac{r_i}{R} \right|$$

Dónde p_i refiere a la cantidad de estudiantes vulnerables en el centro i y P es la cantidad total de estudiantes vulnerables. A su vez, r_i refiere a la cantidad de estudiantes no vulnerables en el centro i y R es la cantidad total de estudiantes no vulnerables.

El índice oscila entre cero (ausencia de segregación) y uno (máxima segregación). Un índice cercano a cero refiere a que la participación de estudiantes vulnerables y no vulnerables para cada centro se aproxima a la participación que tienen para el total de la muestra. En caso contrario, un índice cercano a uno refiere a que todos los estudiantes de un centro tienen iguales características (son todos vulnerables o son todos no vulnerables). El valor del índice puede interpretarse como la proporción de estudiantes que deberían cambiar de centro

educativo, sin ser reemplazados, para que la distribución de estudiantes sea homogénea (INEEd (2021)).

La Tabla N°13 muestra la presencia de segregación en la educación media pública para diferentes definiciones de estudiantes vulnerables. Si se consideran vulnerables a aquellos estudiantes con índice de vulnerabilidad mayor a 0,5999, se obtiene un índice de disimilitud de 37%. Esto implica que, para lograr una distribución homogénea de los estudiantes en cuanto la característica de vulnerables o no vulnerables, deberíamos reubicar al 37% de dichos estudiantes en los centros educativos.

El índice de disimilitud calculado para el total de la muestra (37%) es mayor que el valor del índice calculado entre aquellos asignados a liceo (casi 32%) o aquellos que fueron asignados a UTU (30%).

Tabla N°13: Índice de Duncan-Duncan: Total país

Índice mayor a:	Índice Disimilitud	Índice Disimilitud - Liceos	Índice Disimilitud - UTU
0,5208	36,04%	30,43%	30,61%
0,5999	37,06%	31,83%	30,16%
0,6999	40,69%	35,13%	31,24%

En la Tabla N°14 se detalla el índice de disimilitud para la muestra según la asignación a educación media (liceo y UTU) y, por otro lado, se evidencia el índice de disimilitud en función de la escuela a la que concurrieron los alumnos en su último año escolar⁷. Se aprecia que la segregación es mayor en primaria que en secundaria, en línea con INEEd (2021).

Tabla N°14: Índice de disimilitud de Duncan-Duncan: Educación Media Pública vs Escuela Pública

Índice mayor a:	Educación media	Escuela
0,5208	36,04%	38,58%
0,5999	37,06%	40,13%
0,6999	40,69%	42,56%

En el Apéndice se presenta otro indicador muy utilizado para la medición de segregación, denominado Índice de Información Mutua (*Mutual Information Index*), donde también se encuentra la existencia de segregación educativa y se descompone dicha segregación entre diferencias intragrupo e intergrupo (donde los grupos refieren a liceo o UTU). Según este indicador, casi dos tercios de la segregación se deben a diferencias a la interna de los tipos de instituciones (liceo o UTU) y un tercio se debe a diferencias entre liceo o UTU.

⁷ Para el cálculo del índice de disimilitud en función de la escuela a la que concurrían los alumnos, se descartaron 227 casos por no contarse con el identificador de escuela para el alumno.

6. Metodología empírica

En este capítulo se describe la metodología utilizada para la estimación de las preferencias cardinales de los estudiantes. Se siguió la estrategia detallada en Aue et al. (2020) y, además, se implementó la metodología con el paquete “stabest” del software R propuesto por dichos autores⁸.

6.1. Ecuación a estimar

Se quiere investigar si estudiantes con diferentes valores del índice de vulnerabilidad social tienen distintas preferencias por los centros. Para estimar estas preferencias, se cuenta con el ranking de hasta los tres centros preferidos por cada estudiante. La prioridad de los centros se basa en el índice de vulnerabilidad social y es igual para todos los centros, donde a mayor índice de vulnerabilidad mayor es la prioridad. También se cuenta con la distancia euclidiana desde hogar de cada estudiante a cada uno de los centros que componen el sistema. La distancia es un factor determinante a la hora de rankear centros por parte de los estudiantes, pero no se tiene en cuenta por parte de los centros para definir la prioridad, que sólo se basa en el índice de vulnerabilidad.

La hipótesis básica de la estimación es que un lugar más alto en el ranking supone una utilidad mayor para el estudiante de concurrir a dicho centro. Al seguir la nomenclatura utilizada por Aue et al. (2020), se define la utilidad que le brinda al estudiante t la escuela s (denominado $U_t(s)$) de la siguiente forma:

$$U_t(s) = U_{t0} + X_{ts}\beta + \epsilon_{ts} \quad (1)$$

donde $U_t(s)$ refiere a la utilidad latente a estimar junto con el parámetro β . A su vez, X_{ts} refiere a características observables de los centros e incluye (dependiendo del modelo) las siguientes variables: la distancia del hogar del estudiante t al centro s , un efecto fijo por centro⁹, una variable dummy que indica si el centro es un liceo o una UTU, la vulnerabilidad promedio de los estudiantes asignados al centro y una variable de interacción que relaciona la anterior variable con la vulnerabilidad del estudiante. U_{t0} refiere a la utilidad para el estudiante t de no quedar asignado y se puede suponer que dicha utilidad es cero (normalización, Train (2009)). Además, se supone que el error ϵ_{ts} es independiente entre las alternativas y se distribuye normal con varianza unitaria (darle una estructura a los errores es una condición necesaria en modelos de elección discreta para asegurar la identificación y facilitar la interpretación de los coeficientes estimados (Train, 2009))¹⁰.

En el caso del presente trabajo, cada estudiante ingresa al modelo con tres observaciones que refieren a las tres opciones de centros declaradas en su ranking, de los cuales será asignado a una de ellos.

⁸ Véase: <https://github.com/robertaue/stabest>

⁹ Se debe aclarar que estrictamente no se trata de un efecto fijo (no es una variable dummy por cada centro educativo) sino una variable que enumera los centros educativos (siguiendo la sugerencia de Aue et al., 2020).

¹⁰ Siguiendo el paquete de R propuesto en Aue et al. (2020) se asume la distribución normal pero también podría ser otra como la logística estándar, dando lugar a un modelo logit y no probit.

6.2. Metodología de estimación

Para estimar el parámetro β y la variable latente $U_t(s)$ se utiliza el procedimiento bayesiano, donde se considera una probabilidad a priori para el parámetro y se estima la función de densidad a posteriori del parámetro estructural $p(\beta/\text{datos})$ considerando $p(U/\beta, \text{datos})$ y $p(\beta/U, \text{datos})$.

Para generar las muestras de distribución a posteriori de las utilidades latentes se utiliza el algoritmo de Gibbs, donde el componente condicional utilizado para el muestreo está dado por:

$$p(U_t(s) / \beta, U_{-ts}, \text{datos}) \propto \exp \left\{ -\frac{(U_t(s) - X_{ts}\beta)^2}{2} \right\} \mathbf{1}(U_t(s) \in [\underline{U}_t(s), \overline{U}_t(s)])$$

Donde se supone que $\epsilon_{ts} \sim N(0,1)$ como condición necesaria en la literatura de elección discreta. A pesar de que los errores son independientes e incorrelacionados entre las alternativas, no sucede lo mismo con las utilidades, ya que los intervalos de truncamiento se determinan endógenamente.

Además, se define U_t como el vector de utilidades del estudiante t sobre el total de centros. Por su parte, U_{-ts} hace referencia a un vector que tiene el total de utilidades para el match entre estudiantes y centros menos las que incluyen al emparejamiento entre el estudiante t y el centro s .

De esta forma, el algoritmo de Gibbs se utiliza de la siguiente forma:

- 1) Para todo t y s , se generan muestras aleatorias de $U_t(s)$ a partir de $p(U_t(s) / \beta, U_{-ts}, \text{datos}) = N(X_{ts}\beta, 1)$, truncado en $[\underline{U}_t(s), \overline{U}_t(s)]$
- 2) Se generan muestras de β a partir de $p(\beta/U, \text{datos}) = N(b, (X'X)^{-1})$, con $b = (X'X)^{-1}X'U$
- 3) Se repiten n veces los pasos 1 y 2.

Para la determinación del truncamiento del intervalo de $U_t(s)$ se tienen distintas restricciones según diferentes supuestos de identificación del modelo. A su vez, dado que los límites del truncamiento dependen del ranking declarado por los estudiantes, pero también de la valoración de la utilidad latente de los estudiantes, estos deben ser calculados nuevamente en cada iteración del algoritmo de Gibbs. De esta forma, el vector de utilidades latentes en la iteración k se construye de la siguiente forma:

$$U_{ij}^{(k)} = \begin{cases} U_{ij}^{(k)} & \text{si el emparejamiento } ij \text{ fue calculado en la iteración } k \\ U_{ij}^{(k-1)} & \text{en otro caso} \end{cases}$$

6.3. Supuesto de identificación

En Aue et al. (2020) se sugieren tres supuestos de identificación para determinar el truncamiento de $U_t(s)$:

- 1) **Revelar las verdaderas preferencias es débilmente dominante para los estudiantes.** Esto implica que los estudiantes revelan sus tres opciones favoritas.
- 2) **Los estudiantes usan estrategias no dominadas.** Este caso es más débil que el anterior, pero algo más general. Implica solamente a las opciones rankeadas y no se puede concluir algo respecto a las opciones no rankeadas. Puede que algún centro no rankeado sea preferido a alguno rankeado. Pero el centro rankeado como primera opción es preferido al rankeado como segunda opción, y este a su vez es preferido al rankeado como tercera opción.
- 3) **El matching alcanzado por el mecanismo es estable.** La definición de estabilidad implica que no hay un par formado por un estudiante t y un centro s que cumpla que:
 - No hay un centro que prefiera al estudiante t frente a cualquiera de los asignados a él en el matching μ y;
 - $U_t(s) \geq U_t(\mu(t))$ (el estudiante t prefiere ser asignado a s sobre el centro asignado en el matching μ).

Lo anterior implica que se puede restringir la probabilidad condicional de la utilidad latente $U_t(s)$ al matching μ y a la valuación de U_{-ts} .

Como se discute en el próximo capítulo, en el presente trabajo se considera como supuesto de identificación el caso **de estrategias no dominadas**¹¹. De esta forma se es más flexible y se permite el supuesto de que pueden existir opciones no rankeadas para algún estudiante que son preferidas a alguna de las tres si rankeadas.

Añadiendo algo más de notación, si rk representa el vector con el total de rankings de hasta tres centros elegidos por los estudiantes y pr representa el ranking que tienen los centros sobre los estudiantes. Las preferencias ordinales del estudiante t están representadas por $L_t = (s_t^1, s_t^2, s_t^3)$ con $s_t^k \in S$ y $k = 1, 2$ o 3 . El ranking individual que el estudiante t asigna a la escuela s es $rk_t(s)$, donde $1 \leq rk_t(s) \leq 3$ si $s \in L_t$ o $rk_t(s) = \emptyset$ de lo contrario.

De esta forma, con el supuesto de identificación de estrategias no dominadas se cumple que $U_t(s) \geq U_t(s')$ si $rk_t(s) \geq rk_t(s')$ y $s, s' \in L_t$. Los límites superiores e inferiores para el truncamiento de $U_t(s)$ serían entonces:

$$\begin{aligned} \bar{U}_t(s) &= \begin{cases} +\infty & \text{si } rk_t(s) \in \{1, \emptyset\} \\ \min_{s' \in L_t} \{U_t(s') : rk_t(s') < rk_t(s)\} & \text{si } rk_t(s) > 1 \end{cases} \\ \underline{U}_t(s) &= \begin{cases} -\infty & \text{si } rk_t(s) \in \{L_t, \emptyset\} \\ \max_{s' \in L_t} \{U_t(s') : rk_t(s') > rk_t(s)\} & \text{si } rk_t(s) < L_t \end{cases} \end{aligned}$$

¹¹ Por mayor detalle sobre los límites de truncamiento para el caso de los otros supuestos de identificación ver Aue et al. (2020).

7. Principales resultados

En este capítulo se presentan los resultados de las estimaciones siguiendo la metodología comentada en el capítulo anterior. En forma previa y como una primera aproximación a la cuantificación del impacto de las preferencias de los estudiantes sobre la segregación, se realizan simulaciones contrafactuales siguiendo a Greaves (2023).

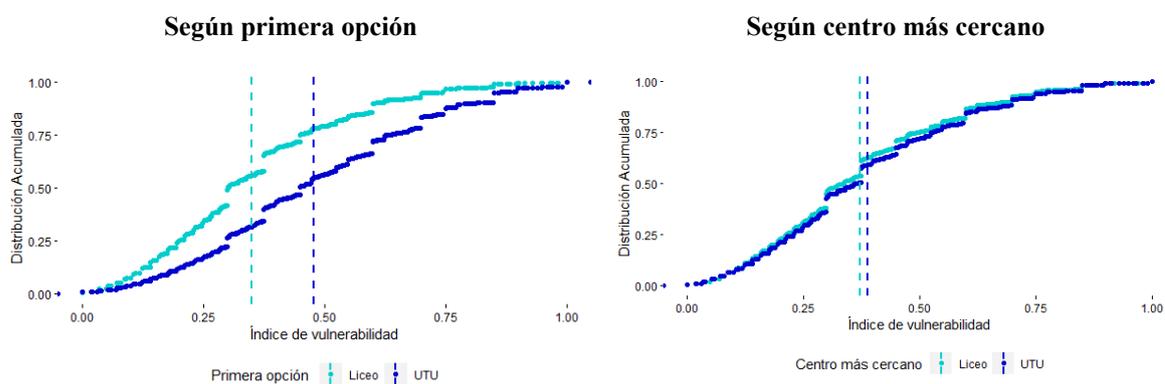
7.1. Simulaciones contrafactuales

En esta sección, se presentan las estadísticas descriptivas de segregación utilizadas en el capítulo 5 para dos escenarios contrafactuales:

Escenario contrafactual 1: Se compara la asignación de los estudiantes de acuerdo a su primera opción rankeada y la asignación a su centro más cercano. Para realizar esta simulación es necesario relajar la condición de los cupos de los centros. De esta forma, todos los estudiantes quedarán asignados a su primera opción en el primer caso, o a su centro más cercano en el segundo caso, independientemente de si el centro es un liceo o una UTU y sin tener en cuenta el cupo de cada centro como restricción.

En el Gráfico N°6, se presenta la distribución del índice de vulnerabilidad para ambos casos. Como se puede apreciar, la dominancia estocástica de primer orden entre liceo y UTU prácticamente desaparece al asignar a los estudiantes a su centro más cercano. Esto indicaría que la distancia no explica, al menos en su totalidad, la dominancia estocástica que se observa al considerar ambas distribuciones en función de la asignación a la primera opción. En otras palabras, la segregación que se observa en la asignación no sería sólo un efecto de la segregación residencial, esto es: los estudiantes más vulnerables tienden a vivir en las mismas zonas e ir a los centros más cercanos a su domicilio. La evidencia indicaría que la segregación está mediada por las preferencias de los estudiantes o por características distintas a la distancia¹².

Gráfico N°6: Distribución acumulada: Índice de vulnerabilidad de estudiantes



¹² Se realizó otra simulación en base a este escenario contrafactual 1, donde también se consideraron los cupos de cada centro como restricción a la hora de asignar a los estudiantes. En este caso, la dominancia estocástica no desaparece al asignar a los estudiantes según su centro más cercano, pero dicha dominancia es menor a la observada en la asignación según su primera opción. Igualmente, en este caso, surge el problema de que al tener en cuenta los cupos, quedan 3.588 estudiantes sin asignar bajo las preferencias en función del centro más cercano.

Adicionalmente, se presenta la comparación del cálculo de índice de Disimilitud (Duncan-Duncan) para ambos casos. Al asignar a los estudiantes en función de su centro más cercano, la segregación cae aproximadamente 10% en comparación a la asignación en función de la primera opción. Si se toma como corte para la definición de vulnerable/no vulnerable un índice mayor a 0,6999, se tiene que la segregación por características geoespaciales o residenciales arrojaría un índice de disimilitud de 27,8%, mientras que las preferencias de los estudiantes incrementaría el nivel de dicho índice en 11,1%, hasta alcanzar el 38,9%.

Tabla N°15: Índice de disimilitud de Duncan-Duncan: Primera opción vs Centro más cercano

Asignación	Índice mayor a:	Índice Disimilitud	Índice Disimilitud - Liceos	Índice Disimilitud - UTU
Primera opción	0,5208	34,2%	29,7%	27,6%
	0,5999	35,2%	31,2%	27,9%
	0,6999	38,9%	34,4%	29,0%
Centro más cercano	0,5208	24,3%	24,3%	23,7%
	0,5999	25,2%	24,7%	25,4%
	0,6999	27,8%	28,7%	25,4%

Escenario contrafactual 2:

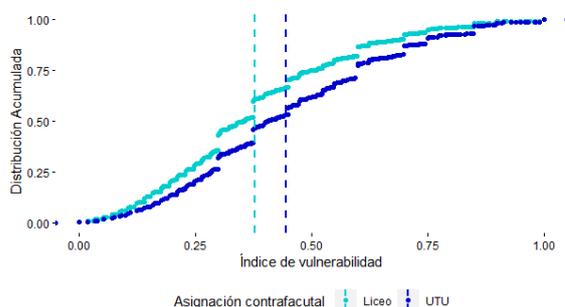
En este escenario, se compara la distribución del índice de vulnerabilidad entre dos alternativas: en un primer escenario, se corre el mecanismo de asignación donde las preferencias de los estudiantes son sus tres opciones más cercanas y la prioridad se basa en el índice de vulnerabilidad, y se compara con la alternativa de asignar a los estudiantes manteniendo las preferencias como sus tres opciones más cercanas, pero donde la prioridad pasa a ser también función de la cercanía al centro (y no por el índice de vulnerabilidad). De esta forma, el cambio entre las dos situaciones refiere a las prioridades de los centros¹³.

Como se observa en el Gráfico N°7, cuando la prioridad se basa en cercanía al centro la dominancia estocástica de primer orden deje de ser tan clara con respecto a cuando la prioridad se basa en el índice de vulnerabilidad.

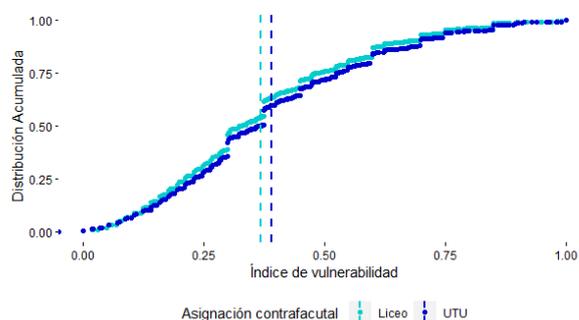
¹³ Al correr el mecanismo para estas dos situaciones contrafactuales, quedaron sin asignar 3.588 estudiantes para el caso donde la prioridad es el índice de vulnerabilidad (donde más del 90% serían estudiantes no vulnerables bajo cualquier corte del índice) y 3.808 estudiantes cuando la prioridad es la cercanía al centro educativo. Notar además que cuando se construye la prioridad es la misma para todos los centros cuando se considera el índice de vulnerabilidad pero que cuando se considera la cercanía al centro la prioridad varía entre éstos.

Gráfico N°7: Distribución acumulada: Índice de vulnerabilidad de estudiantes

Según prioridad: índice de vulnerabilidad



Según prioridad: cercanía al centro



Adicionalmente, como se desprende de la Tabla N°16, el índice de segregación es menor cuando se asigna a los estudiantes considerando como prioridad la cercanía al centro y no el índice de vulnerabilidad. A modo de ejemplo, si se considera como corte para determinar que un estudiante es vulnerable un índice mayor a 0,6999, se tiene que el índice de segregación arroja un valor de 42,4% en base a prioridad por índice de vulnerabilidad, mientras que, para el caso de prioridad por cercanía al centro, el índice de segregación disminuye hasta 29,8%.

Tabla N°16: Índice de disimilitud de Duncan-Duncan: Prioridad por índice de vulnerabilidad vs prioridad por cercanía al centro

Asignación	Índice mayor a:	Índice Disimilitud	Índice Disimilitud - Liceos	Índice Disimilitud - UTU
Top3-Centro cercano Prioridad índice	0,5208	41,1%	33,4%	57,6%
	0,5999	41,5%	34,1%	56,1%
	0,6999	42,4%	35,8%	52,8%
Top3-Centro cercano Prioridad cercanía	0,5208	25,3%	24,3%	26,4%
	0,5999	26,2%	24,8%	28,1%
	0,6999	29,8%	29,1%	28,9%

El fenómeno que podría estar por detrás de estos resultados es que, si se cumple la hipótesis de que hay un componente diferente en las preferencias entre estudiantes vulnerables y no vulnerables, el usar como prioridad el índice de vulnerabilidad, convalida dichas preferencias que tienen impacto en la segregación del sistema.

7.2. Estimación empírica

A continuación, se muestran los resultados de aplicar la metodología descrita en el capítulo 6. Como se comentó anteriormente, para cada estudiante se consideran tres observaciones que refieren a sus tres opciones rankeadas, donde una de ellas es el centro al que fue efectivamente asignado. Además de estas tres opciones, se considera un set de alternativas de centros individualizado por estudiante, según criterios que detallarán más adelante en el presente capítulo.

De esta forma, se estimaron 5 especificaciones para las preferencias de los estudiantes, donde se fueron agregando variables de la siguiente forma:

$$\text{Modelo 1: } U_t(s) = \beta_1(dist_{ts}) + \beta_2(dist_{ts}^2) + \epsilon_{ts}$$

$$\text{Modelo 2: } U_t(s) = \beta_1(dist_{ts}) + \beta_2(dist_{ts}^2) + \beta_3(centro_s) + \epsilon_{ts}$$

$$\text{Modelo 3: } U_t(s) = \beta_1(dist_{ts}) + \beta_2(dist_{ts}^2) + \beta_3(centro_s) + \beta_4(liceo_s) + \epsilon_{ts}$$

$$\text{Modelo 4: } U_t(s) = \beta_1(dist_{ts}) + \beta_2(dist_{ts}^2) + \beta_3(centro_s) + \beta_4(liceo_s) + \beta_5(vuln.prom.centro_s) + \epsilon_{ts}$$

$$\text{Modelo 5: } U_t(s) = \beta_1(dist_{ts}) + \beta_2(dist_{ts}^2) + \beta_3(centro_s) + \beta_4(liceo_s) + \beta_5(vuln.prom.centro_s) + \beta_6(interacción vuln._ts) + \epsilon_{ts}$$

Donde $U_t(s)$ refiere a la utilidad latente que le brinda al estudiante t asistir al centro s y se supone que el error se distribuye normal con varianza unitaria. Se asume que la utilidad depende de: la distancia desde el domicilio del estudiante al centro ($dist_{ts}$), dicha distancia al cuadrado, un efecto fijo por centro ($centro_s$), para tener en cuenta la heterogeneidad que puede existir sobre los centros, una dummy que indica si es el centro es un liceo o una UTU ($liceo_s$), la vulnerabilidad promedio de los estudiantes asignados al centro ($vuln.prom.centro_s$)¹⁴ y una variable de interacción que relaciona la anterior variable con la vulnerabilidad del estudiante ($interacción vuln._ts$).

A su vez, la estimación de los parámetros se realiza separando la muestra total de estudiantes en vulnerables y no vulnerables. Para esto, se consideraron tres cortes del índice de vulnerabilidad, donde el estudiante es vulnerable si el índice es mayor a: 0,5999, 0,6999 o 0,7999 respectivamente.

Por otro lado, si bien no se tiene información acerca de la cantidad de centros que las familias tienen en cuenta a la hora de reportar sus preferencias, parece razonable suponer que las familias consideran un conjunto de centros que es menor a la oferta total (que supera los 500 centros). De esta forma, se realizaron las estimaciones bajo el supuesto de que las tres opciones rankeadas surgen de un conjunto de elección por estudiante que incluye a todos los centros que se encuentran en un radio a su domicilio igual al valor máximo que surge de

¹⁴ La inclusión de la vulnerabilidad promedio por centro puede ser causa de endogeneidad. En Allende et al. (2019) se propone una solución para resolver este problema que se considerará en trabajos posteriores.

comparar la opción rankeada más alejada al domicilio, con tres alternativas¹⁵: 5 km, 10 km o 15 km.

Además, se consideró el supuesto de que las familias usan estrategias no dominadas como el más razonable. Dado el bajo nivel de información acerca del funcionamiento del sistema, parece razonable suponer que las familias no rankeen ciertas opciones por pensar que sus hijos no serán asignados a dichos centros. Es decir que se mostrarán los resultados bajo el supuesto de que las familias juegan estrategias no dominadas, donde es creíble el ranking sobre las opciones efectivamente rankeadas, pero no se pueden realizar conclusiones sobre las opciones no rankeadas. La estimación permite que algún centro no rankeado sea preferido a alguno rankeado. Dado que más del 90% de los estudiantes son asignados a su primera opción parece razonable suponer lo anterior, donde los estudiantes seleccionan y rankean aquellas opciones en las que creen razonable (según ellos) que puedan ser asignados.

En la Tabla N°17 se muestran los resultados del valor medio de los parámetros que surge de la distribución a posteriori, bajo el supuesto de identificación de estrategias no dominadas, donde se separa a los estudiantes en vulnerables o no vulnerables según el corte del índice en 0,6999 y se utilizó un set de elección de centros por estudiante con el tamaño de 10km. (en el apéndice como robustez se presentan los modelos alternativos, con resultados similares). En estimaciones bayesianas, es usual que se presente el valor medio de la distribución a posteriori del parámetro y un intervalo de confianza (y no el p-valor)¹⁶. Como se puede apreciar los estudiantes tienen preferencia por centros que se encuentran más cercanos a su domicilio, ya que el coeficiente asociado al parámetro distancia es negativo en todos los modelos. Este es un resultado estándar en la literatura para este tipo de estimaciones, donde además el parámetro asociado a la distancia al cuadrado es positivo, lo que implica desutilidad marginal decreciente.

El valor estimado del parámetro asociado a la dummy “Liceo” para los estudiantes no vulnerables (a partir del modelo 3) es siempre positivo y significativo, mientras que para el caso de los estudiantes vulnerables es en varios casos negativo o no significativo. Este resultado es robusto a diferentes estimaciones, variando el tamaño del set de elección de centros o con diferentes cortes en el índice de vulnerabilidad (ver apéndice). Lo anterior iría en línea con que estudiantes no vulnerables derivan una mayor utilidad por concurrir a liceos que los estudiantes vulnerables una vez que se controló por la variable distancia.

Respecto a la vulnerabilidad promedio por centro, se observa en el modelo 4 que, si bien el parámetro estimado es positivo en ambos casos, es mayor en valor para el caso de los estudiantes no vulnerables. Por lo cual, les brindaría mayor utilidad a los estudiantes vulnerables concurrir a centros con una vulnerabilidad promedio más alta. Adicionalmente,

¹⁵ A modo de ejemplo, para el caso de 5 km: si un estudiante tiene como opción rankeada más alejada un centro que se encuentra a 2 km, se considerarán en su set de alternativas todos los centros que se encuentren a menos de 5 km de su domicilio. Por otro lado, si la opción rankeada más alejada se encuentra a 7 km, se considerarán en su set de alternativas todos los centros que se encuentren a menos de 7 km de su domicilio. De esta forma, el set de elección de centros por estudiante es creciente en función de las tres alternativas: 5 km, 10km o 15km. Vale recordar que cómo se señala en el capítulo 4, el 75% de los estudiantes rankea como tercera opción (en promedio la más alejada del domicilio) alternativas que se encuentran como máximo a 8,32 km de su domicilio (para Montevideo dicha cifra es 5,39 km, mientras que para el resto del país asciende a 13,67 km).

¹⁶ Se realizaron también test de convergencia a través del paquete de R denominado “coda”. En el apéndice se muestran los resultados de la estimación y análisis de convergencia para el caso del modelo 5 con un millón de iteraciones. Las conclusiones respecto al valor estimado de los parámetros no cambian.

al incorporar la interacción con la vulnerabilidad del estudiante en el modelo 5, el valor estimado del parámetro asociado a la vulnerabilidad promedio por centro deja de ser significativa para estudiantes vulnerables (el intervalo de confianza contiene al cero) mientras que para el caso de los estudiantes no vulnerables pasa a ser significativa y negativa. Esto se interpretaría como una desutilidad respecto a la vulnerabilidad promedio por centro para los estudiantes no vulnerables.

Tabla N°17: Valor medio de la distribución a posteriori de los parámetros bajo el supuesto de identificación de estrategias no dominadas

Vulnerable>0,6999	Parámetro	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Vulnerables	dist_km	-0,02826 [-0,03189;-0,02463]	-0,03065 [-0,03411;-0,02711]	-0,03003 [-0,03353;-0,02583]	-0,02823 [-0,03151;-0,02472]	-0,02914 [-0,03336;-0,02464]
	dist_km2	0,00005 [0,00004;0,00006]	0,00005 [0,00004;0,00006]	0,00005 [0,00004;0,00006]	0,00005 [0,00003;0,00006]	0,00005 [0,00004;0,00006]
	centro		0,00134 [0,00113;0,00157]	0,00255 [0,00226;0,00284]	-0,00077 [-0,00127;-0,00022]	-0,00034 [-0,00088;0,00029]
	liceo			0,29896 [0,2136;0,3846]	-0,24115 [-0,35196;-0,09817]	-0,13302 [-0,25471;-0,01341]
	vuln_prom_centro				2,15169 [1,84244;2,53612]	0,78338 [-0,73969;2,3505]
	inter_vuln					1,69548 [-0,1503;3,36794]
No Vulnerables	dist_km	-0,03814 [-0,03936;-0,03677]	-0,03792 [-0,03928;-0,03643]	-0,03797 [-0,03946;-0,03666]	-0,03769 [-0,03900;-0,03630]	-0,03773 [-0,03906;-0,0365]
	dist_km2	0,00009 [0,00008;0,00009]	0,00009 [0,00008;0,00009]	0,00009 [0,00008;0,00009]	0,00009 [0,00008;0,00009]	0,00009 [0,00008;0,00009]
	centro		-0,00076 [-0,00085;-0,00067]	0,00213 [0,00191;0,00227]	0,00136 [0,00094;0,00167]	0,00145 [0,00097;0,00181]
	liceo			0,74349 [0,69048;0,7861]	0,61097 [0,52264;0,6729]	0,62226 [0,50205;0,70787]
	vuln_prom_centro				0,4527 [0,32679;0,61811]	-2,50396 [-2,72682;-2,31592]
	inter_vuln					8,47385 [8,12889;9,1405]

(*): Se consideran dentro del set de elección del estudiante todos los centros que se encuentran en un radio igual al máximo que surge entre la opción máxima rankeada y un valor de 10 km. Intervalo de confianza al 95% de la distribución a posteriori que se genera por medio del algoritmo de Gibbs. Número de iteraciones: 20.000. Adelgazamiento: 2. Número de descartes: 5.000.

En resumen, ambos grupos de estudiantes tienen preferencia por centros más cercanos a su domicilio. Para los estudiantes no vulnerables parecería que derivan una mayor utilidad en promedio de concurrir a liceos en relación a concurrir a UTU, mientras que dicha variable parece tener un efecto negativo o no significativo para los estudiantes vulnerables. A su vez, para el caso de los estudiantes vulnerables parecería no ser tan relevante la vulnerabilidad promedio de los centros ni la interacción de esta variable con la vulnerabilidad propia de los estudiantes vulnerables (la estimación en el modelo 5, de los parámetros asociados a estas variables contienen al cero en el intervalo de confianza). Para el caso de los estudiantes no vulnerables, preferirían centros que cuentan con un índice de vulnerabilidad promedio menor (ver estimación de dicho parámetro en modelo 5).

7.3. Análisis de bienestar

Usualmente, una de las aplicaciones posteriores a la estimación es la realización de análisis de bienestar. Una posibilidad es la comparación en términos de bienestar promedio para los estudiantes de ser asignados a su opción preferida o de ser asignado a su centro más cercano¹⁷. Es decir, se comparan dos escenarios, uno donde se tienen en cuenta las preferencias de los estudiantes y otro donde simplemente se aplica una regla y no se tienen en cuenta sus preferencias.

En línea con Aue et al. (2020), se calcula la diferencia de la utilidad de la primera opción elegida por el estudiante t respecto a la utilidad de concurrir a su centro más cercano. A su vez, como el valor en nivel de la utilidad no tiene una interpretación en sí mismo, se lo divide por la derivada de la utilidad respecto a la variable distancia, para tener una aproximación a la diferencia de utilidades (o bienestar) medida en términos de distancia que se traslada el estudiante t . La diferencia de utilidades sería entonces:

$$\Delta U_t^{km} = \frac{\widehat{U}_t(\text{primera opción})}{\left| \frac{\partial \widehat{U}_t(\text{primera opción})}{\partial \text{dist}_{ts}} \right|} - \frac{\widehat{U}_t(\text{centro más cercano})}{\left| \frac{\partial \widehat{U}_t(\text{centro más cercano})}{\partial \text{dist}_{ts}} \right|}$$

donde: $\frac{\partial \widehat{U}_t(\text{primera opción})}{\partial \text{dist}_{ts}} = \beta_1 + 2\beta_2 \text{dist}_{ts}$. Como el valor del coeficiente asociado a la distancia al cuadro (β_2) es muy pequeño, se usará la aproximación lineal, por lo que: $\frac{\partial \widehat{U}_t(\text{primera opción})}{\partial \text{dist}_{ts}} = \frac{\partial \widehat{U}_t(\text{centro más cercano})}{\partial \text{dist}_{ts}} = \beta_1$ y la expresión anterior queda:

$$\Delta U_t^{km} = \frac{\widehat{U}_t(\text{primera opción}) - \widehat{U}_t(\text{centro más cercano})}{|\beta_1|}$$

De esta forma, $\Delta U_t^{km} > 0$ nos podría informar cuántos km adicionales está dispuesto a trasladarse el estudiante t para concurrir a su centro preferido en relación a su centro más cercano.

En la Tabla N°18, los resultados muestran el promedio de la diferencia de utilidades antes mencionada, donde se calculan ambas utilidades a partir de los parámetros estimados para cada uno de los estudiantes, teniendo en cuenta las características de su centro más cercano y su opción preferida (distancia, efecto fijo, si es liceo o UTU, la vulnerabilidad promedio de los estudiantes que acuden al centro y la interacción de esta variable con el índice de vulnerabilidad del estudiante).

¹⁷ En ambos casos no se tienen en cuenta los cupos de los centros como restricción.

Tabla N°18: Cambio en las utilidades (Utilidad de concurrir a primera opción menos utilidad de concurrir al centro más cercano)

Modelo*	Vulnerable_6999	Promedio	Mediana
Modelo 3	Vulnerables	-3,1585	-0,1079
	No vulnerables	-2,9308	-0,0506
Modelo 4	Vulnerables	-1,3159	-0,0104
	No vulnerables	-2,9548	-0,0423
Modelo 5	Vulnerables	-1,3057	-0,0118
	No vulnerables	-2,3024	-0,0260

(*): Refieren a los modelos estimados en la Tabla N°17.

Como se desprende de la Tabla N°18, les brindaría a ambos grupos de estudiantes una mayor utilidad concurrir a su centro más cercano que a su primera opción (la diferencia de utilidades es negativa en promedio y muy cercana a cero para la mitad de la muestra). Si bien los resultados no son contradictorios con el modelo econométrico utilizado, ya que este permite que los estudiantes no rankeen opciones preferidas a las efectivamente rankeadas, si deberían ser tomados con precaución por varias razones. Se debe tener en cuenta que, en este análisis de bienestar, no se consideran los cupos en los centros como restricción. Además, faltaría incluir variables relevantes que las familias tienen en cuenta a la hora de rankear los centros (como, por ejemplo, indicadores de ambiente educativo por centro). En este sentido, una posible línea de trabajo futuro podría ser la reestimación de estos modelos con mejores indicadores acerca de los centros educativos.

8. Conclusiones

Usando información de la educación media básica pública en Uruguay, se buscó investigar si estudiantes con diferentes características tienen distintas preferencias por los centros de educación (Liceos o UTUs) y como esto impacta en la segregación educativa. Esto resulta importante ya que distintos estudios se ha encontrado que la segregación educativa tiene consecuencias negativas en los estudiantes (Card y Rothstein (2006) y Billings et al., (2012)).

En el presente documento se detallaron varios indicadores que muestran la presencia de segregación educativa. Se observó que: la participación relativa de estudiantes vulnerables en UTU duplica o incluso triplica la cantidad de estudiantes vulnerables en liceos. Además, la distribución del índice de vulnerabilidad entre los estudiantes asignados a liceos está dominada en sentido estocástico de primer orden por la distribución del índice de los asignados a UTU, lo que muestra la mayor vulnerabilidad (medida por el índice) de estudiantes que son asignados a UTU. Finalmente, se mostró la presencia de segregación educativa a través del índice de disimilitud (Duncan-Duncan), para diferentes definiciones de estudiantes vulnerables.

Para estudiar la relación entre las preferencias y la segregación educativa, se siguió la estrategia detallada en Aue et al. (2020) como forma de estimar las preferencias cardinales de los estudiantes, donde se utiliza un modelo de estimación bayesiana. Además, siguiendo a Greaves (2023) se realizaron simulaciones de asignaciones contrafactuales. Se encontró

que los estudiantes tienen preferencias por centros educativos más cercanos (resultado estándar) y que dicha variable no explica, al menos en su totalidad, la segregación encontrada en la asignación. Es decir que la segregación educativa no es sólo un efecto de la segregación residencial, la evidencia indicaría que la segregación está mediada por las preferencias de los estudiantes o por características distintas a la distancia.

Con base a la estimación de las preferencias cardinales, se encontró que dado todo lo demás constante, si un centro es liceo tiene un impacto negativo o nulo (en relación con pertenecer al subsistema UTU) en el caso de los estudiantes vulnerables, mientras que este efecto es positivo para los estudiantes no vulnerables.

A su vez, si bien es importante el cambio que se implementó en Uruguay desde el 2019 en la forma de asignación de los estudiantes (teniendo en cuenta sus preferencias), se encontró un efecto no buscado: si se cumple la hipótesis de que hay un componente diferencial en las preferencias de estudiantes vulnerables y no vulnerables, el usar como prioridad de los centros al índice de vulnerabilidad, convalida estas preferencias, lo que impactaría en la segregación del sistema. Como forma de disminuir la segregación, resulta clave mejorar el nivel de conocimiento de las familias acerca del mecanismo de asignación implementado, como forma de que, a partir de un mayor nivel de información, los participantes se encuentren en una mejor posición para la toma de decisiones. Brindar charlas informativas a las familias, en las propias escuelas donde concurren los estudiantes podría ser una posibilidad.

Un comentario adicional respecto a una posibilidad no analizada en el presente trabajo es que, si bien disminuir la segregación educativa podría ser un objetivo, se debe tener en cuenta que estudiantes vulnerables que en el actual sistema concurren a una UTU, podrían directamente no entrar al sistema educativo en otro tipo de sistema, lo que tendría consecuencias aún más negativas que la propia segregación. Además, para este caso, y como forma de enriquecer la investigación, se podría abrir el análisis en función de los dos planes que existen en UTU: Ciclo Básico Tecnológico-CBT y Formación Profesional Básica-FPB.

Finalmente, como línea de trabajo futuro, se sugiere la incorporación de indicadores de evaluación que brinden más información sobre las características de los centros de enseñanza (calidad de profesores, rendimiento de alumnos que concurren a dicho centro o infraestructura entre otros), como forma de dar más herramientas a las familias a la hora de reportar preferencias.

Referencias bibliográficas

- Abdulkadiroğlu, A. & Sönmez, T. (2003). "School choice: A mechanism design approach". *American Economic Review*, 93(3), 729-747.
- Abdulkadiroğlu, A. (2013). "School Choice". *Prepared for the Handbook of Market Design*, edited by Neeman, Z., Niederle, M., Roth, A. E. and Vulkan, N. Duke University.
- Abdulkadiroğlu, A., Agarwal, N. & Pathak, P. (2017). "The Welfare Effects of Coordinated Assignment: Evidence from the New York City High School Match". *American Economic Review*, 107(12): 3635–3689.
- Abdulkadiroğlu, A., Pathak, P., Schellenberg, J. & Walters, C. R. (2020). "Do parents value school effectiveness?". *American Economic Review*, 110(5): 1502–1539.
- Agarwal, N. & Somaini, P. (2018). "Demand analysis using strategic reports: an application to a school choice mechanism". *Econometrica*, 86(2): 391-444
- Agarwal, N. & Somaini, P. (2020). "Revealed Preference Analysis of School Choice Models". *Annual Reviews of Economics*, 12(1):471-501.
- Allende, C., Neilson, C., Gallego, F., (2019). "Approximating the Equilibrium Effects of Informed School Choice". Princeton University. *Working Paper* No. 618.
- Aue, R., Klein, T. & Ortega, J. (2020). "What happens when separate and unequal school districts merge?". *ZEW Discussion Papers*, No. 20-032.
- Billings, S. B., Deming, D. J. & Rockoff, J. E. (2012). "School segregation, educational attainment and crime: evidence from the end of busing in Charlotte-Mecklenburg". *NBER Working Paper* No. 18487.
- Burgess, S., Greaves, E., Vignoles, A. & Wilson, D. (2015). "What parents want: school preferences and school choice". *The Economic Journal*, 125 (September), 1262–1289.
- Cabrera, I. & Pereyra, J. (2022). "Informe para la Dirección Sectorial de Integración Educativa. Área Gestión de Sistemas de Información". *Anep*.
- Calsamiglia, C., Fu, C. & Güell, M. (2020). "Structural estimation of a model of school choices: The Boston Mechanism versus its alternatives," *Journal of Political Economy*, University of Chicago Press, 128(2):642-680.
- Card, D. & Rothstein, J. (2006). "Racial segregation and the black-white test score gap". *NBER Working Paper* No. 12078.
- Elbers, B. (2021). "A Method for Studying Differences in Segregation Across Time and Space". *Sociological Methods & Research*.
- Escobar, J. & Huerta, L. (2022). "Segregation and affirmative action in school choice"
- Fack, G., Grenet, J. & He, Y. (2019). "Beyond truth-telling: preference estimation with centralized school choice and college admissions". *American Economic Review*, 109(4): 1486-1529.

- Gale, D. & Shapley, L.S. (1962). “College admissions and the stability of marriage”. *The American Mathematical Monthly* Vol. 69 (1): 9-15.
- Greaves, E. (2023): “Segregation by choice? School choice and segregation in England”. *Education Economics*.
- Haeringer, G. & Klijn, F. (2009). “Constrained school choice”, *Journal of Economic theory*, 144(5), 1921-1947.
- Hastings, J., Kane, T. J. & Staiger, D. O. (2005). “Parental Preferences and School Competition: Evidence from a Public School Choice Program”. *NBER Working Paper No. 11805*.
- Idoux, C. (2023). “Integrating New York City Schools: The Role of Admission Criteria and Family Preferences”.
- INEEd (2021). “Evolución de la segregación socioeconómica en la educación pública de Uruguay 2013-2019”.
- Kapor, A.J., Neilson C.A. & Zimmerman S.D. (2020). “Heterogeneous beliefs and school choice mechanisms”. *American Economic Review*, 110(5):1274–1315.
- Kutscher, M., Nath, S. & Urzua, S. (2020): “Centralized Admission Systems and School Segregation: Evidence from a National Reform” Working paper, University of Maryland.
- Laverde, M. (2022). “Distance to schools and equal access in school choice systems”. Boston College.
- Lee, J. & Soon, S.J. (2022). “Distributional Impacts of Centralized School Choice”.
- Oosterbeek, H., Sóvágó, S. & van der Klaauw, B. (2021). “Preference heterogeneity and school segregation”. *Journal of Public Economics*, 197 104400.
- Shapley, L.S. & Scarf, H. (1974). “On cores and indivisibility”. *Journal of Mathematical Economics* vol.1 (1): 23-37.
- Taboga, M. (2021): "Markov Chain Monte Carlo (MCMC) diagnostics". *Lectures on probability theory and mathematical statistics*. Kindle Direct Publishing.
Online appendix: <https://www.statlect.com/fundamentals-of-statistics/Markov-Chain-Monte-Carlo-diagnostics>.
- Train, K. E. (2009). “Discrete Choice Methods with Simulation”. Cambridge University Press.

Apéndice

A.1 - Cálculo de la distancia del domicilio de los estudiantes a las instituciones.

Para calcular la distancia en kilómetros del hogar de cada uno de los estudiantes a las instituciones educativas, se procedió, en primer lugar, a georreferenciar el domicilio de los estudiantes (para aquellos con los que se contaba con la dirección) y de las instituciones educativas. Para cada observación, se buscó la latitud y longitud correspondiente para luego calcular una matriz de distancias en el *software* R a partir de las coordenadas.

A.1.1. - Geolocalización de instituciones educativas

Se obtuvieron los datos de 2.797 instituciones públicas¹⁸ que conforman la oferta educativa por número de identificación (RUEE) de: <https://sig.anep.edu.uy/siganep>.

Del total, 2.267 refieren a instituciones de educación de nivel inicial y primaria; 497 a educación secundaria (liceos) y formación técnica (UTU); y 33 son Centros de Formación Docente. En cuanto a los campos de la base de datos, se cuenta con información para cada institución de: nombre, departamento, localidad, paraje (si corresponde) y dirección.

Utilizando esta información, se procedió a obtener dos nuevos campos referidos a la latitud y longitud de cada institución en función de los resultados que arroja *Google Maps*. Para ello se utilizó un complemento de la aplicación *Google Sheets* que realiza *scraping web* en *Google Maps*. Luego se realizaron controles manuales a nivel de departamento en función de los resultados obtenidos, verificando que las instituciones se encontraran georreferenciadas dentro de los límites geográficos de los departamentos. Asimismo, se buscó manualmente aquellos resultados que no habían sido encontrados mediante la técnica anterior. En particular, se hizo una búsqueda más exhaustiva de 151 escuelas, en su mayoría escuelas rurales como forma de obtener una mejor aproximación de georreferenciación¹⁹.

La ubicación exacta de los liceos y UTU tiene especial relevancia para este trabajo de investigación, por lo que dichas instituciones fueron buscadas mediante la técnica anterior utilizando dos sets de información diferentes:

- 1) “Dirección, Paraje (si corresponde), Localidad, Departamento, Uruguay”
- 2) “Nombre y numeración de la institución, Paraje (si corresponde), Localidad, Departamento, Uruguay”

Luego se compararon los resultados arrojados por ambos procedimientos y 19 instituciones arrojaron resultados diferentes. De esta forma, se buscó en forma individual en la web la geolocalización de estos 19 casos (de los 497) para determinar cuál era su geolocalización correcta.

¹⁸ Se eliminaron de la base de datos las siguientes instituciones:

idRUEE: 1403575 por estar duplicada en la base de datos

idRUEE: 1216014 por estar duplicada con diferente dirección. Refiere al Liceo de Delta del Tigre, San José. Se consideró la dirección actual de dicha institución.

¹⁹ Para obtener una búsqueda más exacta en *Google Maps* para el caso de las escuelas rurales, se utilizó como fuente alternativa la siguiente página web: <https://laseguridad.social/uy/>

A.1.2.- Geolocalización de estudiantes

Para el caso de los estudiantes, se trabajó con la base de datos de Anep utilizada para la asignación de los estudiantes, que en total alcanzaba a 42.132 alumnos que ingresaban a liceos y UTU en el año 2023. Además del número identificador de cada alumno, se contaba con los siguientes campos²⁰: “Dirección completa”, “Localidad” y “Departamento”. Esta información refiere al domicilio que aparece en la base de datos de la matrícula de primaria.

Al igual que en el caso de los estudiantes, se realizó la búsqueda de las coordenadas (latitud y longitud) a imputar a cada estudiante mediante *scarping* en *Google Maps*, a través de un complemento de la aplicación *Google Sheets*. En este caso, la búsqueda por estudiante se definió de la siguiente forma: “Dirección completa, Localidad, Departamento, Uruguay”. De esta forma, se lograron georreferenciar 41.397 estudiantes de los 42.132 que componían la base (98,26% del total). De los 735 alumnos que no se logró georreferenciar su domicilio, para unos 500 de ellos fue porque no se contaba con información en el campo dirección.

Al igual que en el caso de las instituciones, se realizó un control a nivel de departamento, de forma que el domicilio de todos los estudiantes quedara georreferenciado dentro de los límites de dicho departamento²¹. Buscando en forma individualizada cuando dicha condición no se cumplía.

De esta forma del total de 42.132 alumnos, se pudo georreferenciar el domicilio para 41.397 alumnos mientras que no se pudo lograr dicho objetivo para 735 alumnos por falta de información.

A.2. - Segregación

Otro indicador muy utilizado para la medición de segregación es el Índice de Información Mutua (*Mutual Information Index*) que permite desagregar la medición de segregación total en: segregación entre tipo de instituciones (liceo y UTU) y, por otro lado, segregación a la interna de cada tipo de institución.

Siguiendo a Elbers (2021) se detalla el cálculo del *Mutual Information Index* (M) y una versión normalizada de este (H). El *Mutual Information Index* se define entonces como:

$$M(T) = \sum_{u=1}^U \sum_{g=1}^G p_{ug} \log \frac{p_{ug}}{p_u \cdot p_g}$$

donde U refiere al número de instituciones; G refiere a la cantidad de grupos (en este caso se tienen 2 grupos, estudiantes vulnerables y no vulnerables). A su vez, p_{ug} es la probabilidad de que el estudiante pertenezca a la escuela u y el grupo g . De igual modo, p_u es la probabilidad de que el estudiante pertenezca a la escuela u ; y p_g es la probabilidad de que el estudiante pertenezca a determinado grupo (vulnerable o no vulnerable en este caso).

²⁰ Se consideró excluyente contar con información en el campo “Dirección”. Para el resto de los campos puede que no se cuente con toda la información para todos los alumnos.

²¹ Algunos estudiantes de la frontera seca (por ejemplo: Chuy o Rivera) tienen su domicilio del lado brasileño.

La versión normalizada del índice se denomina Índice de Theil y se define de la siguiente forma:

$$H(T) = \frac{M(T)}{E(T)}$$

A su vez, $E(T)$ es la entropía de la distribución marginal de los grupos de T:

$$E(T) = - \sum_{g=1}^G p_g \log p_g$$

La versión normalizada del índice tiene la ventaja de restringir al índice entre cero y uno, donde a mayor valor del índice, mayor es el nivel de segregación. Asimismo, la segregación total se puede descomponer de la siguiente forma:

$$M(T) = M(S) + \sum_{s=1}^S p_s M(T_s)$$

$M(T)$ refiere al indicador de segregación antes mencionado, considerando las U instituciones según estudiantes vulnerables y no vulnerables. En forma análoga, $M(S)$ es el índice de segregación entre tipo de instituciones y se calcula considerando los S tipo de instituciones (liceo o UTU) según estudiantes vulnerables y no vulnerables. Finalmente, $M(T_s)$ refiere al índice de segregación a la interna de cada tipo de institución (liceo o UTU) y p_s es la participación de estudiantes del tipo de institución s (liceo o UTU).

Tabla N°19: Mutual Information Index

Vulnerable si: índice > 6999	M(T)	H(T)
Mutual Information	5,5%	16,0%
Entre tipos de institución (Liceo y UTU)	1,7%	5,1%
Dentro del tipo de institución:	3,7%	10,9%
Liceo	2,9%	10,9%
UTU	6,8%	12,7%
Part. %	79%	21%

Si se observa el indicador normalizado, $H(T)$, se aprecia que tiene un valor de 16%. Al descomponerlo, se tiene que casi dos tercios de dicha segregación (10,9%) se deben a diferencias a la interna de los tipos de instituciones (liceo o UTU) y un tercio se debe a diferencias entre tipo de instituciones (5,7%).

A.3. - Resultados

Tabla N°20: Valor medio de la distribución a posteriori de los parámetros para modelo 5 y radio de 5 km

Estudiantes	Parámetro	Vulnerable si índice mayor a:		
		0,5999	0,6999	0,7999
Vulnerables	dist_km	-0,0307 [-0,03355;-0,02761]	-0,02834 [-0,03284;-0,02429]	-0,03827 [-0,04597;-0,03003]
	dist_km2	0,00006 [0,00005;0,00007]	0,00005 [0,00004;0,00006]	0,0001 [0,00007;0,00013]
	centro	-0,00068 [-0,0012;-0,0002]	0,00039 [-0,00029;0,00122]	-0,00028 [-0,00174;0,00098]
	liceo	-0,18825 [-0,29636;-0,09972]	0,05607 [-0,08925;0,25617]	-0,13202 [-0,46965;0,18518]
	vuln_prom_centro	0,40154 [-0,40761;1,52161]	2,73681 [1,21405;4,36289]	1,19092 [-5,19704;7,52014]
	inter_vuln	2,2313 [0,72068;3,35511]	-0,66672 [-2,56905;1,22869]	1,17873 [-5,84972;7,87717]
	No Vulnerables	dist_km	-0,03801 [-0,03943;-0,03641]	-0,03769 [-0,03885;-0,03626]
dist_km2		0,00009 [0,00008;0,00009]	0,00008 [0,00008;0,00009]	0,00008 [0,00007;0,00008]
centro		0,00185 [0,00165;0,00214]	0,00142 [0,0011;0,00168]	0,00128 [0,00105;0,00151]
liceo		0,75766 [0,70585;0,81826]	0,61954 [0,5359;0,68433]	0,55634 [0,50584;0,60422]
vuln_prom_centro		-2,62648 [-2,87667;-2,3751]	-2,4715 [-2,72631;-2,24574]	-2,38385 [-2,67875;-2,16529]
inter_vuln		9,31972 [8,63266;10,068]	8,44995 [7,84277;9,03117]	7,7941 [7,28824;8,40937]

(*): Se consideran dentro del set de elección del estudiante todos los centros que se encuentran en un radio igual al máximo que surge entre la opción máxima rankeada y un valor de 5 km. Intervalo de confianza al 95% de la distribución a posteriori que se genera por medio del algoritmo de Gibbs. Número de iteraciones: 20.000. Adelgazamiento: 2. Número de descartes: 5.000.

Tabla N°21: Valor medio de la distribución a posteriori de los parámetros para modelo 5 y radio de 10 km

Estudiantes	Parámetro	Vulnerable si índice mayor a:		
		0,5999	0,6999	0,7999
Vulnerables	dist_km	-0,03018 [-0,0334;-0,02725]	-0,02914 [-0,03336;-0,02464]	-0,0409 [-0,04918;-0,03274]
	dist_km2	0,00006 [0,00005;0,00007]	0,00005 [0,00004;0,00006]	0,00011 [0,00008;0,00014]
	centro	-0,00048 [-0,00122;0,00027]	-0,00034 [-0,00088;0,00029]	0,00016 [-0,00066;0,00097]
	liceo	-0,13152 [-0,29267;0,03637]	-0,13302 [-0,25471;-0,01341]	-0,03355 [-0,18935;0,11873]
	vuln_prom_centro	-0,11104 [-1,08452;1,7024]	0,78338 [-0,73969;2,3505]	3,57143 [0,84996;6,85287]
	inter_vuln	2,92912 [0,62144;4,17264]	1,69548 [-0,1503;3,36794]	-1,618 [-5,43372;1,41177]
	No Vulnerables	dist_km	-0,03819 [-0,03948;-0,03683]	-0,03773 [-0,03906;-0,0365]
dist_km2		0,00009 [0,00008;0,00009]	0,00009 [0,00008;0,00009]	0,00008 [0,00007;0,00008]
centro		0,00189 [0,00146;0,0022]	0,00145 [0,00097;0,00181]	0,00115 [0,00083;0,00153]
liceo		0,7637 [0,66204;0,84027]	0,62226 [0,50205;0,70787]	0,5223 [0,45367;0,60904]
vuln_prom_centro		-2,55964 [-2,8148;-2,24599]	-2,50396 [-2,72682;-2,31592]	-2,43341 [-2,6738;-2,19674]
inter_vuln		9,07496 [8,20686;9,74574]	8,47385 [8,12889;9,1405]	7,92028 [7,40901;8,3967]

(*): Se consideran dentro del set de elección del estudiante todos los centros que se encuentran en un radio igual al máximo que surge entre la opción máxima rankeada y un valor de 10 km. Intervalo de confianza al 95% de la distribución a posteriori que se genera por medio del algoritmo de Gibbs. Número de iteraciones: 20.000. Adelgazamiento: 2. Número de descartes: 5.000.

Tabla N°22: Valor medio de la distribución a posteriori de los parámetros para modelo 5 y radio de 15 km

Estudiantes	Parámetro	Vulnerable si índice mayor a:		
		0,5999	0,6999	0,7999
Vulnerables	dist_km	-0,0304 [-0,03314;-0,02687]	-0,02838 [-0,03296;-0,02385]	-0,04037 [-0,04868;-0,03314]
	dist_km2	0,00006 [0,00005;0,00007]	0,00005 [0,00004;0,00006]	0,00011 [0,00008;0,00014]
	centro	-0,00022 [-0,00056;0,00013]	-0,00096 [-0,00153;-0,0004]	-0,00078 [-0,00147;0,00013]
	liceo	-0,06586 [-0,13804;0,00923]	-0,28721 [-0,41185;-0,14726]	-0,27007 [-0,43579;-0,09624]
	vuln_prom_centro	0,44786 [-0,35155;1,3174]	1,4967 [0,07357;2,88749]	2,68181 [-1,7951;6,30471]
	inter_vuln	2,21416 [1,17144;3,26255]	0,80756 [-0,97576;2,50464]	-0,50376 [-4,55392;4,50733]
No Vulnerables	dist_km	-0,03791 [-0,03937;-0,03666]	-0,03807 [-0,03926;-0,0369]	-0,03649 [-0,03782;-0,03523]
	dist_km2	0,00008 [0,00008;0,00009]	0,00009 [0,00008;0,00009]	0,00008 [0,00007;0,00008]
	centro	0,00165 [0,00116;0,00195]	0,00068 [0,00043;0,001]	0,00101 [0,00072;0,00134]
	liceo	0,70734 [0,5885;0,77121]	0,42563 [0,36788;0,4913]	0,48771 [0,42342;0,55976]
	vuln_prom_centro	-2,59187 [-2,85585;-2,31522]	-2,60596 [-2,82241;-2,3217]	-2,3828 [-2,57593;-2,21102]
	inter_vuln	9,20178 [8,40023;9,90842]	8,58864 [8,04586;9,00252]	7,73725 [7,3729;7,98729]

(*): Se consideran dentro del set de elección del estudiante todos los centros que se encuentran en un radio igual al máximo que surge entre la opción máxima rankeada y un valor de 15 km. Intervalo de confianza al 95% de la distribución a posteriori que se genera por medio del algoritmo de Gibbs. Número de iteraciones: 20.000. Adelgazamiento: 2. Número de descartes: 5.000.

A.4. – Análisis de convergencia

En metodologías que utilizan cadenas Markov Montecarlo como es el caso del muestreo de Gibbs, es usual la realización de análisis de convergencia de los parámetros estimados, donde se busca explicar si el muestro generado por el algoritmo provee una aproximación precisa de la distribución del parámetro a estimar. En concreto, mediante estos análisis se busca chequear dos tipos de problemas: 1-si gran parte del muestro generado mediante las cadenas Markov Montecarlo se han extraído de distribuciones que son significativamente diferentes de la función objetivo, y de esta forma la convergencia de la cadena será lenta; 2-verificar si el tamaño de la muestra es demasiado pequeño, por lo cual muchos términos de la cadena no son independientes entre sí. Como forma de mejorar la convergencia, dependiendo del problema que se presenta, se puede aumentar el número de iteraciones, así como también el número de descartes (Taboga (2021)).

En este sentido, en la Tabla N°23 se detallan los resultados de una estimación para el modelo 5, donde se realizaron un millón de iteraciones, con un valor de adelgazamiento de 250 y número de descartes de 400 mil. Como puede observarse, el valor medio de la distribución a posteriori no presenta mayores diferencias a lo mostrado en el capítulo de resultados. La distancia es significativa y negativa para ambos grupos de estudiantes. La dummy liceo no es significativa para el caso de los estudiantes vulnerables, mientras es positiva y significativa para los estudiantes no vulnerables.

Tabla N°23: Valor medio de la distribución a posteriori de los parámetros para modelo 5 con un millón de iteraciones

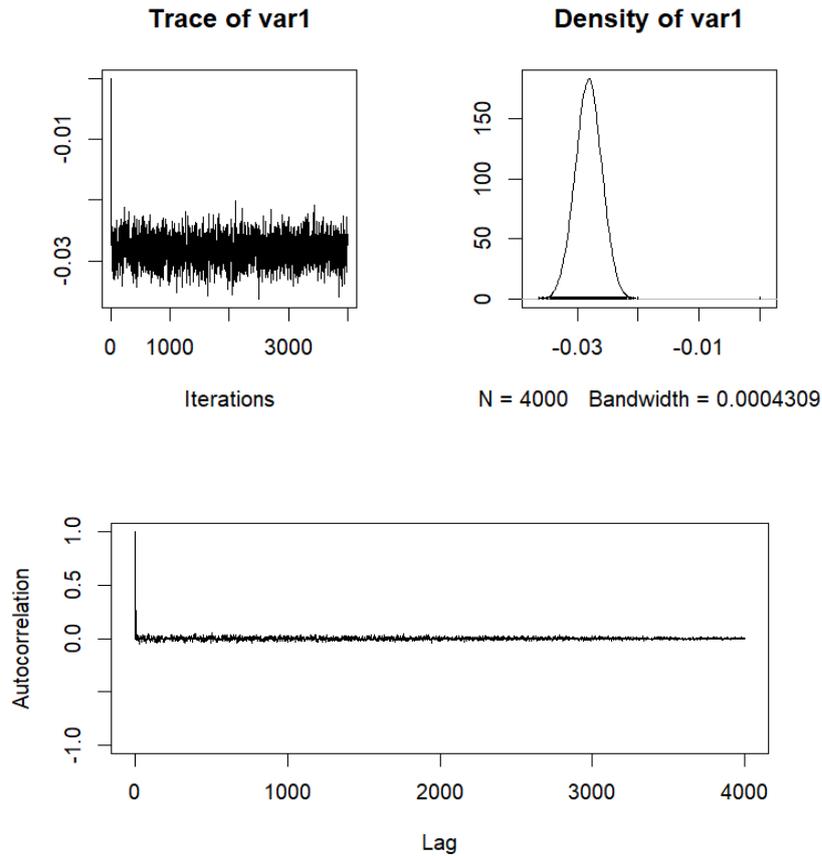
Estudiantes	Parámetro	Vulnerable si: índice > 0,6999
Vulnerables	dist_km	-0,02815 [-0,03232;-0,02401]
	dist_km2	0,00005 [0,00003;0,00006]
	centro	0,00091 [-0,00006;0,00205]
	liceo	0,18527 [-0,18578;0,46526]
	vuln_prom_centro	1,53857 [-0,80299;3,8299]
	inter_vuln	0,81255 [-1,8995;3,57378]
	No Vulnerables	dist_km
dist_km2		0,00009 [0,00008;0,00009]
centro		0,00205 [0,0016;0,0025]
liceo		0,77947 [0,66856;0,89052]
vuln_prom_centro		-2,4782 [-2,73546;-2,22103]
inter_vuln		8,51809 [7,88464;9,1415]

(*): Se consideran dentro del set de elección del estudiante todos los centros que se encuentran en un radio igual al máximo que surge entre la opción máxima rankeada y un valor de 10 km. Intervalo de confianza al 95% de la distribución a posteriori que se genera por medio del algoritmo de Gibbs. Número de iteraciones: 1.000.000. Adelgazamiento: 250. Número de descartes: 400.000.

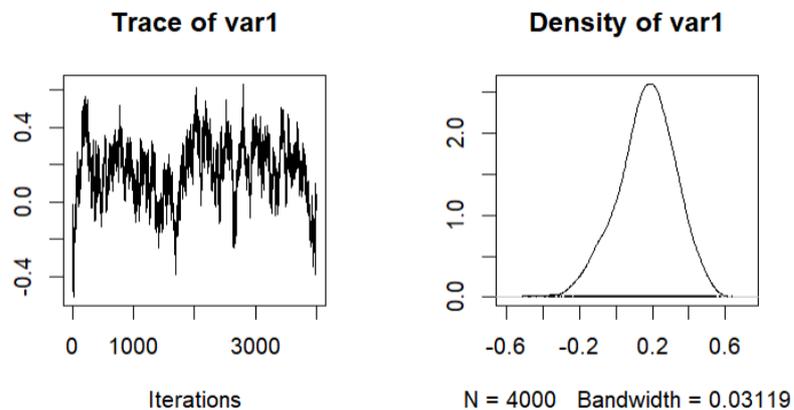
A continuación, se presenta un análisis gráfico para el estudio de la convergencia de la distribución a posteriori para dos parámetros asociados a las variables distancia y la dummy liceo. En general, cuando no se tienen los problemas 1 y 2 mencionados anteriormente, gran parte de la cadena calculada presenta una baja correlación. Además, mediante gráficos se puede observar como decrece la correlación en función de los retardos.

Gráfico N°23: Correlación de la cadena y autocorrelación - Estudiantes vulnerables

Variable “distancia”



Variable “liceo”



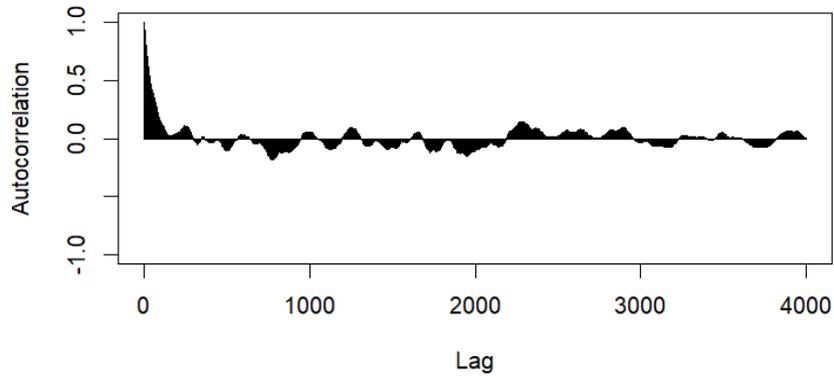
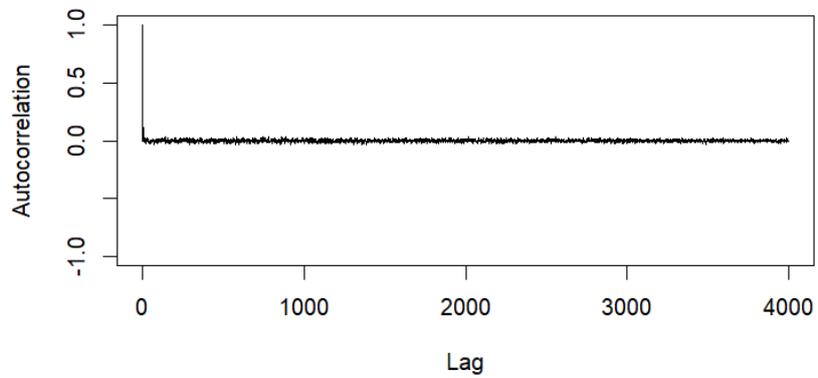
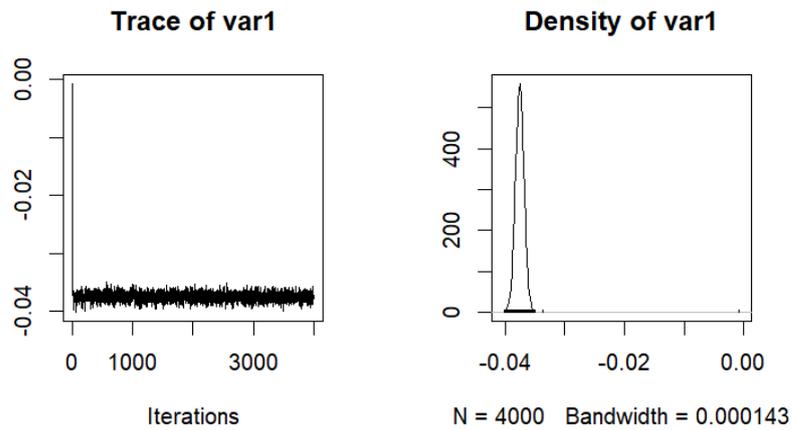


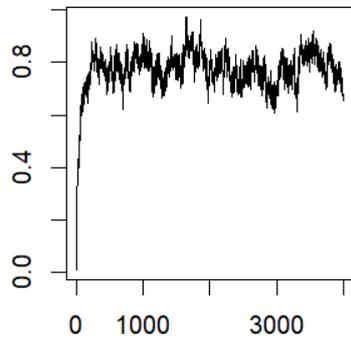
Gráfico N°23: Correlación de la cadena y autocorrelación - Estudiantes no vulnerables

Variable “distancia”



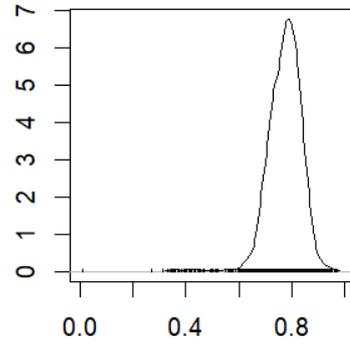
Variable "liceo"

Trace of var1



Iterations

Density of var1



N = 4000 Bandwidth = 0.01223

