INSTITUTO DE ECONOMÍA

MAGÍSTER en ECONOMÍA

2019

Contribution of social, spatial, and economic frictions to the socioeconomic school segregation: evidence from Chile

Anushka Baloian

www.economia.uc.cl

Contribution of social, spatial, and economic frictions to the socioeconomic school segregation: evidence from Chile

Anushka Baloian*

December 25, 2019

Abstract

A growing body of research has shown that increasing freedom of school choice may lead to higher socioeconomic school segregation. However, the evaluation of related educational policies requires a deeper understanding of the channels through which parental choices impact on socioeconomic school segregation, and a measure of the contribution of each of those channels. I look at rank-ordered Pre-Kindergarten preferences of high socioeconomic status (SES) and low-SES parents to explore this interrogation. Through the simulation of multiple counterfactual scenarios regarding school applications and the school admission process, I quantify the contribution of social, spatial, and economic frictions to socioeconomic school segregation in Chile. While removing social frictions leads to a significant decrease in socioeconomic school segregation, removing spatial frictions does not. Finally, removing economic frictions to all types of students leads to a significant increase in the levels of segregation due to highly congested school applications and heterogeneous preferences of high-SES and low-SES parents concerning school attributes.

*Agradezco especialmente a los profesores Kenzo Asahi y Nicolás Figueroa por su gran apoyo y disposición en el desarrollo de este trabajo. Agradecer también a Rodrigo Icarán, cuyo conocimiento sobre el algoritmo de selección del Sistema de Admisión Escolar fue imprescindible para llevarlo a cabo. Agradecer finalmente a mi familia por su imprescindible e incondicional apoyo durante mi paso por la universidad.

Contents

1 Introduction

Recognition of education as a fundamental element both for social and economic development has led to an unprecedented expansion in education access over the past century. In fact, more children today are enrolled in school than ever before (UNICEF, 2019). However, many challenges remain, and one of them is the socioeconomic segregation in schools. Recognizing that education is a key determinant of economic opportunities, differentiated educational experiences based on the socio-economic background of children may be a powerful mechanism of income inequality reproduction and exacerbation. Furthermore, concerns about socioeconomic school segregation (SES school segregation from now on) are not only a matter of equity. Several researchers have pointed out that such segregation has a significant impact on academic outcomes[1] and recent research suggests that economically diverse classrooms promote social empathy among high-SES students toward low-SES individuals (Rao, 2019). That being said, understanding the dynamics of SES school segregation is of high relevance for public policy.

The present work analyzes the contribution of social, spatial, and economic frictions to the SES school segregation in Chile.[2] Overall, I show that social frictions play a great role in explaining SES school segregation, which is attributable to an "outgroup avoidance" phenomenon, i.e., self-segregation behavior among upper-class families, who actively invest in order to enroll their children in schools with a low share of low-SES students. By contrast, in-group preferences are not observed among low-SES families, thus contradicting the hypothesis of "neutral ethnocentrism" that states that members of all social groups prefer to interact with individuals of their own group.
However, the outgroup avoidance we find could also be driven by unobservables such as school climate. This possibility cannot be completely discarded as I am not able to account for all elements affecting it. Nonetheless, the main specification controls for non-academic attributes that are summed up in a unique index, which likely captures a high fraction of elements that affect school climate.

Spatial frictions do not seem to increase SES school segregation. Furthermore, assuming no spatial frictions results in slightly higher levels of SES school segregation, which may be attributable to an intensification of the segregating behavior of high-SES families and a small rebound effect due to higher congestion in school applications.

Finally, setting economic frictions to zero results in a significant increase in the levels of SES school segregation, which is explained by two simultaneous phenomena. First, there is a strong rebound effect produced by highly concentrated school applications in a reduced amount of schools with high academic performance and a low share of low-SES students. In this scenario, high-SES applicants end up con-

[1] Van Ewijk & Sleegers (2010) provide a meta-analysis on this topic.

[2] Importantly, paid-private schools are excluded from the present analysis. They represent less than 8% of total school enrollment in Chile.

centrating in those schools, while low-SES applicants end up in schools with poorer academic performance and a higher share of low-SES students. Second, heterogeneous preferences of high-SES and low-SES parents regarding school attributes such as distance, academic performance, and share of low-SES students also explain a (smaller) fraction of these results.

The empirical approach to assess this study's research question is as follows. In the first stage, I analyze the choice problem of families using an exploded logit regression on the families' ranked school preferences, allowing for preference heterogeneity depending on the applicants' SES. I use information about the applications to Pre-Kindergarten (PK) during 2018 in the cities of Valparaíso, Viña del Mar and Concepción, considering schools academic and non-academic performance, monetary costs, proximity, and social distance between the applicant's SES and the socioeconomic composition of schools.

The second stage consists of replicating the school applications of students as well as the admission process, and simulating multiple counterfactual scenarios under specific assumptions; (1) the baseline scenario, (2) no social frictions scenario, (3) no spatial frictions scenario, and (4) no economic frictions scenario. Finally, I measure SES school segregation in each simulation. By comparing the levels of segregation between the baseline scenario and each counterfactual I assess the contribution of social frictions, spatial frictions, and economic frictions to SES school segregation in Chile.

This work relates to the literature on segregation, particularly to the one that focuses on its causes and the role played by private choices. One of the first authors that analyzed the impact of private choices on social segregation is Schelling (1971), who suggested that differences in preferences across social groups may play a first-order role in social segregation across multiple areas, such as housing, consumption, and education. This may be true even if individuals do not deliberately support social segregation. In this respect, a relevant question is whether private choices are inherently segregating, meaning that individuals want to reduce social interactions with the ones that are different to them, or if rather, residential segregation or taste differences across groups are the main factors behind socioeconomic segregation. This question has been mostly analyzed in the context of housing choices (Clark, 1986; Galster, 1988; Clark, 1989; Bobo & Zubrinsky, 1996; Timberlake, 2000; Card, Mas and Rothstein, 2008; Boustan, 2011). Also, Davis et al. (2019) develop a model of consumption segregation to understand the choices of restaurants by New York City residents belonging to different ethnic groups. This allows the authors to quantify the role of taste differences, spatial frictions, and social frictions.

Also, the present study is related to the school-choice literature, as it provides evidence on the relative importance of school characteristics, social frictions, spatial frictions, and economic frictions in families' school preferences. There is a growing literature that studies the implications of school choice for socioeconomic and ethnic segregation across schools. In this regard, a large amount of evidence suggests that the later is reinforced by parental school choices (Karsten et al., 2003; Denessen, Driessena, & Sleegers, 2005; Mickelson, Bottia, & Southworth, 2008; Schneider, Schuchart, Weishaupt, & Riedel, 2012; Brandén & Bygren, 2018).

However, less clear is the channel through which families 'preferences for schools may influence school segregation. The most direct one would be the explicit preferences of parents regarding the social composition of schools. This trend has been mainly observed among high-status groups. Saporito (2003) analyzes the choices for magnet schools of families in the US and finds that, while white families systematically avoid schools with higher shares of non-white students, wealthier families avoid the ones with higher shares of poor students. Also, Lankford and Wyckoff (2006) conclude that racial composition of school, as well as neighborhood characteristics, play a significant role in school choices of white families, thus contributing to racial segregation across the U.S. educational system. In line with this evidence, Havermans, Wouters and Groenez (2018) find that high-SES parents of Belgium have stronger preferences for schools with high proportions of members of their own social group, relative to low SES parents.

An additional channel through which freedom of school choice may increase educational segregation is proximity. In contexts of high residential segregation, school segregation may be a direct consequence of the former, as distance between home and school usually plays a relevant role in families preferences for schools (Hastings, Kane, & Staiger, 2005; Jacobs, 2013; Burgess et al., 2015; Havermans et al., 2018). Hence, if families systematically prefer to enroll their children in schools with higher proximity to their place of residence, high levels of school segregation may be just the direct consequence of the underlying residential segregation (Lindbom, 2010; Böhlmark, Holmlund, & Lindahl, 2015; Böhlmark, Holmlund & Lindahl, 2016).
Also, school segregation may be reinforced if there are systematic taste differences across groups regarding relevant school attributes. A common result of studies on school choices is that high-status families put greater weight on school academic achievement when deciding where to enroll their children (Kristen, 2008; Burgess et al., 2015; Havermans et al., 2018).

Finally, school segregation may be greater in educational settings where freedom of school choice coexists with the existence of tuition fees. Economic barriers may significantly reduce the choice set of low-SES families, while HIGH-SES families face lower financial constraints (Elacqua, 2012).

This work contributes to the literature on school choice and school segregation by narrowing the gap between both fields of research. Estimating families school preferences and comparing school segregation levels between the baseline scenario and the counterfactual scenarios mentioned in the first paragraph allows to quantify the contribution of social, spatial, and economic frictions to school SES segregation, which is of great importance in terms of public policy, as it provides direct insight on the potential effectiveness of policies aiming to reduce school segregation.
Despite the growing interest in the influence of families 'preferences on educational

segregation, few efforts have been made to properly identify the contribution of social, spatial, and economic frictions to the overall degree of SES school segregation. The existent literature has almost exclusively focused on the analysis of school preferences, without establishing a framework to connect their results to school SES segregation.

The structure of this work is as follows: Section 2 offers an overview of the Chilean educational system, its evolution, and the main features of the School Admission System (SAE). Section 3 describes the data set and presents summary statistics from the sample. Section 4 presents the empirical strategy, starting with the econometric specification of school preferences, then describing the methodology employed to simulate the school applications of students and the admission process, and finally specifying the counterfactual approach to quantify the contribution of social, spatial, and economic frictions to the SES school segregation. Section 5 presents the estimation results of the school preferences of families and their implications. Section 6 analyzes the model fit of the baseline simulation, and Section 7 presents the main results. Section 8 concludes. Supplementary material is displayed in the Appendix.

2 Chilean context

Since the establishment of the voucher system in 1981, the Chilean educational system is often perceived as an "extreme case of market-oriented education" (Bellei et al., 2014). Until 1981, the central government was directly in charge of public schools and provided them with funds. The private sector was composed of paid private schools that charged high tuition fees (unaffordable for the vast majority of Chilean families) and free private schools, which were partially financed with public funds. The educational reform of 1981 substantially altered the structure of primary and secondary education in Chile, by (i) transferring public education from the central to the local governments (municipalities), (ii) introducing a per-student subsidy or voucher for public and free private schools (voucher schools), (iii) allowing free entry for the school market, and (iv) allowing parents to freely choose among public schools and voucher schools (Gallego, 2006).

By directing the financial aid to families instead of doing it to schools, this system aims to provide parents with freedom of choice and encourage competition across schools to attract students, which is intended to promote a higher quality of education for all students. Since the 1981 reform, three types of schools make up the Chilean educational system; public schools, voucher schools, and private schools. The later are schools that do not receive any kind of public fund and are oriented to high-income families, attending around 8% of all Chilean students. These schools are excluded from the present analysis. Voucher schools serve around 55% of students and are privately-managed schools that receive public funds through the voucher program.

In the last decades, the Chilean educational system has experienced multiple reforms that aim, among others, to equalize educational opportunities across socioeconomic groups by reducing school segregation, providing additional financial aid to the most vulnerable students, and limiting the ability of schools to select their students arbitrarily. Until 2007, voucher schools were allowed to define their own admission policy and determine their tuition fees. In contrast, public schools are not able to charge any tuition and are required to admit all students as long as vacancies are available. This changed in 2008 when the Preferential School Subsidy Law (SEP) was passed, introducing important changes to the Chilean voucher system. Overall, this law aimed to equalize educational opportunities across students from different socioeconomic status and recognized that it is more expensive to educate students from disadvantaged families. The SEP reform established an additional subsidy for "Priority" students, i.e., those whose families are, approximately, in the bottom 40 percent of the income distribution. Also, concentration bonuses were introduced, meaning that schools with a higher concentration of low-SES students receive additional per-students funds. To be part of this program, schools are required to subscribe to the SEP program and meet its requirements; not to charge any tuition to Priority students, not to select students based on their academic achievement, and to take part in the SEP accountability system, which aims to monitor the proper use of the additional funds and to ensure the academic progress of students.

Despite the attempts to promote educational equity, SES school segregation is still high (Elacqua, 2012; Bellei, 2013; Valenzuela, Bellei, Ríos, 2014). In this regard, Valenzuela, Bellei De Los Ríos (2014) empirically assess the relationship between educational market at the municipal level and SES school segregation, concluding that educational market dynamics in Chile play an important role in the high levels of SES school segregation, at least at the municipal level. Also, Santos Elacqua (2016) show that parental choices and entry barriers established by schools, namely, tuition fees and selectivity, increase school segregation, as it cannot be entirely explained by residential segregation.

As a new attempt to reduce SES school segregation, the School Inclusion Law came into effect in 2016, substantially transforming the admission process of all public and private-subsidized schools (Sistema de Admission Escolar SAE, 2018). The new School Admission System (SAE) banns all types of selective practices of schools, aiming to prevent discrimination and provide the same educational opportunities to all students despite their academic and socioeconomic background. Under the new admission system, if the number of applicants does not exceed the vacancies available for a school, all students are automatically accepted in that school. If instead, the number of applicants is greater than the vacancies available, students are sorted in a list according to the priority criteria established by the Inclusion Law, which gives the highest priority to applicants whose sibling is already enrolled in that school. It also considers whether the student is identified as low-SES according to the SEP criteria (i.e., Priority student), whether the mother or father works in the school, and whether the student has previously been enrolled in that school. Finally, applicants with the same priority level are randomly selected.

The implementation of the new Admission System has been gradual. It was first implemented in the Magallanes region during the year 2016 and continued with the regions of Tarapacá, Coquimbo, O'Higgins, and Los Lagos in 2017. In 2018 it was expanded to the regions of Arica and Parinacota, Antofagasta, Atacama, Valparaíso, Maule, Bío Bío, Araucanía, Los Ríos, and Aysén. Finally, the new Admission System

was implemented in the Metropolitan region during the year 2019.
An important feature of the SAE is that applicants have to submit a rank-ordered and unrestricted list of school preferences. The school assignment mechanism is based on the "Deferred Acceptance" (DA) algorithm, proposed by Gale and Shapley (1962). A relevant characteristic of this mechanism is that it is strategy-proof, which means that the optimal strategy for students is to submit a list according to their true preferences (i.e., truth-telling is a weakly dominant strategy for students).This is particularly useful when trying to understand the school choices of families, allowing us to observe the rank-ordered preferences of all students applying to pre-kindergarten across the regions considered in this work.

3 Data and descriptive statistics

3.1 Data

Overall, I consider Pre-Kindergarten (PK) applications that took place during 2018 in the cities of Valparaiso, Viña del Mar, and Concepción. The reason why these cities are considered is that they are the most extensive urban settlements in which the SAE system was introduced during the year 2018. Regions where the SAE was introduced earlier are not considered in this analysis as an attempt to minimize the risk of strategic behavior of applicants that may modify their behavior as a response to the results of previous application processes.[3]

The SAE database contains the registry of all the schools that offered vacancies through the system during the year 2018 and their geographic coordinates, information on each applicant and their rank-ordered school choices, and their final enrollment. Specifically, the applicant's database provides information on student's SEP status (whether it is classified as a priority student or not), which I use as a measure of their socioeconomic status, household coordinates (with a random perturbation to protect privacy), and the respective georeferencing quality. I combine this information with the Simce dataset, which provides information on the average standardized test score of each school.
I also combine this information with the SEP dataset, which allows identifying the proportion of low-SES students (priority students) and high-SES students (non-beneficiaries students) in each educational establishment.
Also, the School-Directory database provided by MINEDUC allows identifying the co-payment fees of school, if existent. However, this information is displayed by intervals of payments, which implies that it is not possible to identify the exact amount that school i charges as copayment fee, but rather an interval in which it resides. In this regard, my approach is to take the median value of the interval as the copayment fee of each school.

[3] Even though the DA mechanism suggests that the best application-strategy is to reveal the real school preferences, applicants may develop alternative beliefs regarding the application process as they gather information on previous experiences.

Finally, I obtain the indicators on the non-academic performance of each school from the SNED (National System of Performance Evaluation) database.

Concerning the sample selection, the present analysis relies on two different data sets. First, the data set of applicants for whom I estimate school preferences (estimation sample). Not all applicants across Valparaíso, Viña del Mar, and Concepción are included in this sample, as for some of them (12%) the georeferencing information is not available, which is of great importance to quantify the contribution of spatial frictions to SES-school segregation, as explained in section 4. Also, rural schools and applicants are excluded from the present analysis, as school choices in the context of rural areas is probably a very distant phenomenon from school choices in urban spaces. School choices in the context of rural areas are significantly constrained by school availability and limited transportation, which makes it hard to conceive school “choices” of families located in rural areas as an expression of their preferences. Taking this into account, students who apply at least one rural school are left out of the sample (1.6 % of applicants). Finally, the current analysis is restricted to the regular process of application, as the applications during the complementary process may be influenced by the strategic behavior of families. This results in an estimation sample of 4452 applicants and 567 schools.
In contrast, the complete sample, which contains the applications of 5141 students, does not exclude applicants without georeferencing data or the ones applying to rural schools. However, I do not estimate school preferences for those applicants, but rather keep their real applications among the four simulated scenarios. Thus, the complete sample includes the simulated applications of the students considered in the estimation sample, as well as the real applications of the students excluded from it. This sample is used to simulate the school assignations determined by the SAE algorithm, in order to replicate the congestion levels of the real process and avoid general equilibrium effects that may arise if only considering the estimation sample to simulate the admission process.

3.2 Descriptive statistics

Table 1 contains descriptive statistics at the student level, separately for high-SES and low-SES students of the estimation sample, as well as the difference between both groups and its statistical significance, provided by a T-test. Overall, 1961 (38.1%) students are identified as low-SES, and the average Euclidean demographic distance (EDD) between them and the schools included in their choice set is almost 16 p.p. smaller than the one faced by high-SES students. Also, low-SES students face considerably lower tuition fees, which is mostly explained by the fact that they receive additional financial aid from the government. If we consider tuition fees without including the financial aid received by low-SES students, the annual cost of schools is only $10 USD lower for low-SES applicants. Besides, the distance between home and school is similar across both types of students, being slightly lower for low-SES students (0.12 km lower).
Concerning the type of schools in terms of tuition costs, the last three rows of Table 1

show that low-SES and high-SES students are surrounded by similar schools.

Table 1: Summary statistics of school's choice set by type of student

	(1) High-SES	(2) Low-SES	(3) Difference
Euclidean demographic distance	.578	.420	.159***
	[.024]	[.024]	(.000)
Distance	5.98	5.86	.119***
	[.952]	[.909]	(.028)
Average test score (std)	.039	-.021	.060***
	[.237]	[.211]	(.007)
Anual tuition fees (USD 2018)	231.5	121.0	110.54***
	[53.5]	[45.01]	(1.508)
Anual tuition fees (USD 2018) without financial aid	231.5	221.2	10.3***
	[53.5]	[45.01]	(1.508)
Non-Academic Index	1.93	1.89	.042***
	[.166]	[.145]	(.005)
Paid schools	.115	.112	.003***
	[.024]	[.021]	(.001)
Free-for-all schools	.740	.739	.001
	[.034]	[.033]	(.001)
Schools exclusively free for Low-SES students	.145	.149	-.005***
	[.022]	[.022]	(.001)
Observations	2,491 (61.9%)	1,961 (38.1%)	

Notes: This table presents summary statistics of all schools within an 11-km radius of student,s residence (i.e., Choice Set), at the student level. Standard errors in parentheses and standard deviations in brackets. Statistics are computed by considering all students of our sample and all the schools included in their Choice Set, without distinguishing between the ones they applied to and the ones they did not.*** indicates 1% statistical significance, ** indicates 5% statistical significance, and * indicates 10% statistical significance, determined by a T-test of differences.

Table 2 reports summary statistics of schools, differentiating between free-for-all schools (column 1), schools that are exclusively free for low-SES students (column 2), and paid schools (column 3). Free-for-all schools have the lowest avg. test (std) scores (-.053), while paid schools have the highest scores (1.08). Also, free-for-all schools exhibit the highest share of low-SES students (.655), while paid schools have the lowest (.233), as expected. Finally, the non-academic Index [4] reaches its highest value in paid schools. This is consistent with the existent literature, which shows that the highest levels of discipline, obedience, and social abilities are observed among high-SES students and, therefore, among paid schools[5]. Concerning the application of high-SES and low-SES students, Table 3 shows that all students, regardless of their SES, disproportionally apply to free schools. On average, 46% and 67% of all the schools applied by a high-SES student and a low-SES student are free schools, respectively. Concerning schools that are exclusively free for low-SES students, they represent only the 22% and 24% of applications of high-SES and low-SES students, respectively. Finally, the proportion of paid schools applied by a high-SES student is 30%, on average, and 11% for low-SES students. These facts suggest that, as expected, low-SES students disproportionally prefer free schools, relative to high-SES students. Also, they are less likely to apply to paid schools. Finally, there is no significant difference between the likelihood of high-SES and low-SES students to apply to schools that are exclusively free for low-SES students, as the difference between both groups is statistically insignificant (column 3 of Table 3).

Table 2: Summary statistics by type of school

	(1) Free-for-all	(2) Free for low-SES	(3) Paid
Avg. test score	-.053	.628	1.08
	[.891]	[.796]	[.784]
Share of low-SES students	.655	.423	.233
	[.137]	[.127]	[.124]
Non-academic Index	2.1	2.16	2.21
	[.981]	[.896]	[.978]
Observations	412	88	67

Notes: Standard deviations in brackets. Column 1 displays summary statistics of free-for-all schools. Column 2 displays summary statistics of schools that are exclusively free for low-SES students. Column 3 displays summary statistics of paid schools.

As expected, low-SES students face cheaper tuition fees (USD$445 less than high-SES students). Again, the fact that low-SES students receive additional financial aid from the government may be one explanation for the later. However, this trend is also reinforced by low-SES students systematically avoiding paid schools and concentrating their applications in free schools.

[4]The non-academic index includes some of the indicators considered by the SNED. It is explained with detail in Section 4.

[5]See, for example, Duncan, Brooks-Gunn,& Klebanov (1994), Duncan et al. (1998), Pinderhughes et al. (2000), Hosokawa & Katsura (2018).

Table 3: School applications, descriptive statistics by type of student

	(1) High-SES students	(2) Low-SES students	(3) Difference
Free-for-all schools	.462	.672	-.202***
	[.395]	[.386]	(.012)
Schools exclusively free for low-SES students	.222	.242	.012
	[.293]	[.320]	(.009)
Paid schools	.298	.105	.191***
	[.363]	[.234]	(.009)
Avg. Test score (std)	.718	.263	.455***
	[.682]	[.725]	(.021)
Anual Tuition fees (USD 2018)	554.4	109.1	445.3***
	[529.3]	[273.7]	(13.150)
Distance (km)	3.77	3.84	-.075
	[8.96]	[10.38]	(.290)
Euclidean demographic distance	.388	.486	-.098***
	[.166]	[.160]	(.005)
Share of low-SES students	.388	.514	-.126***
	[.166]	[.160]	(.005)
Non-academic index	2.18	2.15	.035*
	[.678]	[.716]	(.021)
Observations	2491	1961	

Notes: This table presents the means observed for the school applications at the student level, separately by the socioeconomic status of students. Only the schools to which the students applied are considered (irrespective of the ranking). Columns (1) and (2) contain the means observed for high-SES and low-SES students, respectively. Column (3) present the difference between both group and its statistical significance, determined by a T-test. Paid schools are schools that are not tuition-free and do not adhere to the SEP agreement. Free-for-all schools are schools that are tuition-free for both low and high-SES students.

With respect to the average test score and distance of schools applied by low-SES and high-SES students, the conclusions regarding the preferences of both types of students may be inaccurate if only looking at Table 3. First, the notable difference of academic performance between schools applied by high-SES and low-SES students exhibited in Table 3 (0.455 standard deviations) may suggest that low-SES students attribute substantially lower relevance to the academic performance of schools. However, this may be explained by the fact that low-SES students disproportionally apply to free schools and schools that are exclusively free for them (Table 3), both of which exhibit lower levels of academic performance (Table 2). In fact, if we only consider schools that charge the same tuition fees to low-SES and high-SES students, the gap between the average academic performance of schools applied by low-SES and high-SES students shrinks significantly, as exhibited in tables 13 and 14 in the Appendix (section A).

If only free schools are considered (table 13 of section A)[6] the earlier-mentioned gap reduces to less than a half (0.186 standard deviations). Still, this difference may be partially explained by low-SES students being surrounded by schools of lower academic performance, as shown in Table 1. Instead, if we focus on the schools that do charge tuition fees to students (table 14 of section A) [7], the gap in the avg. test score is reversed; the average test score of schools applied by low-SES students is 0.13 standard deviations higher and statistically significant.

In the second place, Table 3 may suggest that the distaste for distance (km) between home and school is lower for low-SES students. However, when considering schools where both types of students face the same tuition fees, this trend is reversed. On the one side, if only free schools are considered (table 13 of section A), the average distance between home and school is 0.4 Km lower for low-SES students than for high-SES students, but this difference is statistically insignificant. On the other side, if only paid school are considered (table 14 of section A), the average distance is 0.89 Km lower for low-SES students, and it is statistically significant. Finally, if we only consider schools that are exclusively free for low-SES students (table 15 of section A), the avg. distance is higher for the latter group.

Therefore, the fact that the avg. distance for low-SES applicants is higher is entirely explained by these families willing to enroll their children in schools that are exclusively free for them. This is not surprising, as these types of schools have, on average, higher academic performance (Table 2) relative to free-for-all schools. This suggests that low-SES parents put greater importance on the distance between home and schools, but they are willing to enroll in more distant schools to avoid paying tuition fees and enroll in schools with higher academic performance. In contrast, high-SES applicants are not willing to sacrifice proximity to enroll in this type of school, which is also reasonable, as these schools are not tuition-free for them and exhibit lower academic performance than paid schools.

Finally, low-SES families tend to apply to more socially distant schools than high-SES families (Table 3), as revealed by the average Euclidean demographic distance of their school applications.

Overall, statistics presented above suggest that (1) all types of parents face a high disutility for paying school tuition fees, but it is higher for low-SES families. (2) low-SES parents may have similar or only slightly lower preferences for school's academic performance, as a significant part of the observed gap is possibly explained by a stronger trade-off between academic performance and tuition cost faced by low-SES parents, as well as low-SES families being surrounded by schools with lower academic performance. Finally, (3) results suggest that low-SES students may put greater importance on the distance between home and schools, but they are willing to enroll in more distant schools in order to avoid paying tuition fees. Concerning to social preferences of high-SES and low-SES students, the data exhibited in this sections cannot give credible insight on students preferences regarding this attribute, as many covariates are probably at play, which is also true for the rest of the described

[6] For low-SES students, free schools include free-for-all schools and schools exclusively free for low-SES students. For high-SES students, free schools include only free-for-all schools

[7] For low-SES students, this includes paid schools. For high-SES students, paid schools and schools exclusively free for low-SES students are considered

features.
In what follows, I develop an econometric model that aims to identify the extent to which these stylized facts are related to the preferences of high-SES and low-SES applicants. Then, I analyze how social preferences (i.e., social frictions) distance (i.e., spatial frictions), and tuition costs (i.e., economic frictions) contribute to SES school segregation.

4 Empirical strategy

The empirical strategy of this work is made up of two parts, which are described in the following paragraphs. Section 4.1 describes the first stage and section 4.2 the second stage.

4.1 Preferences

In the first stage, I analyze the choice problem of families using an exploded logit regression on their ranked school preferences. Beggs, Cardell, and Hausman (1981) proposed this model, also known as rank-ordered logit regression, as a generalization of the conditional logit regression model developed by McFadden (1974). The main difference between both models is that the former allows modeling the rank-ordered preferences of the agents of interest while the latter limits the analysis to their most preferred alternative (first choice). This approach has, among others, been taken by Groenez, Havermans, and Wouters (2018), who analyze the school preferences of Belgian families using the information of their ranked school choices. Also, Tagle (2018) developed a similar approach to analyze school preferences of Chilean families.

Accordingly to the SAE, each applicant (i) ranks their school preferences in descending order (truth-telling assumption). In this model, I assume that families or students define their ranked preferences by considering spatial frictions, i.e., distance from home to school j (D_{ij}), social frictions, which are represented by the Euclidean Demographic Distance (EDD) between school socioeconomic composition and the applicant,s socioeconomic status (EDD_{ij}), economic frictions ($TuitionFees_{ij}$), average test score ($Score_j$) and a variable that controls for non-academic performance of schools ($NonAcademic_j$). Also, interaction terms for each of these variables are included for low-SES students to allow for heterogeneous preferences across groups. Finally, school preferences depend on an individual-specific preference shock e_{ij} for school j, which is exclusively known by the applicant and has an extreme value type 1 distribution. Hence, Applicant i has a utility equal to:

$$U_{ij} = V_{ij} + e_{ij} \tag{1}$$

when enrolling in school j., where V_{ij} is the observable component, specified by the following equation:

$$\begin{aligned} V_{ij} &= \gamma_1 D_{ij} + \gamma_2 EDD_{ij} + \beta_1 Score_j + \beta_2 TuitionFees_{ij} + \beta 3 NonAcademic_j \\ &+ \delta low-SES_i \times (D_{ij} + EDD_{ij} + Score_j + TuitionFees_{ij} + NonAcademic_j) \end{aligned} \tag{2}$$

Our measure of social frictions, represented by EDD_{ij}, is calculated as follows:

$$EDD_{ij} = \sqrt{(low-SES_i - Share_j)^2} \tag{3}$$

where $low-SES_i = 1$ if applicant i is identified as a low-SES student and $low-SES_i = 0$ otherwise. $Share_j$ is the share of low-SES students in school j.

Applicants can choose among all schools belonging to their choice set M_i, which I define as all the schools within a 10 km radius of their home address [8]. Applicant's i preferred set of schools is denoted by the subset Y_i. I assume that the choice set M_i is perfectly known by each applicant and that they choose the school that maximizes their total utility[9]. This is, the first-ranked alternative of their preferred school set is the one that provides the highest value for U_{ij} among the choice set M_i. The second one is the school that, considering the remaining schools belonging to the applicant's i choice set, provides the highest utility for her, and so on.[10] Taking this into account and under the assumption of an extreme value type 1 distribution of e_{ij}, the exploded logit specification can be interpreted as sequence of independent choices under a conditional logit specification in which each ranked option is chosen as the preferred one among the alternatives that have not yet been ranked (Train, 2003). If the choice set of applicant i is $M_i = [A, B, C, D, E]$, then the probability of choosing the school set $Y_i = [A, B, C]$, where A is ranked at first, B at the second and C at the third, is determined by:

$$Prob(Y_i|M_i) = \frac{e^{V_{iA}}}{\sum_{n=A,B,C,D,E} e^{V_{in}}} \times \frac{e^{V_{iB}}}{\sum_{n=A,B,C,D,E} e^{V_{in}}} \times \frac{e^{V_{iC}}}{\sum_{n=A,B,C,D,E} e^{V_{in}}}$$

Finally, this model assumes that non-ranked alternatives are equally valued by the applicant. Following the last example, if the school application of subject i is $Y_i = [A, B, C]$, the following preferences are assumed:

[8]The 10-km radius was chosen after simulating school applications under different assumptions with respect to the choice set (3km, 5km, 8km, 10km, 15km). When a 10-km choice set is assumed, simulated applications reach the highest model fit, suggesting that this alternative is the most convenient to identify the applicant's preference parameters and correctly predict their school choices.

[9]The perfect-knowledge assumption has been widely adopted in the recent literature, such as Gallego & Hernando, 2009; Cuesta et al., 2017; Groenez et al., 2018.

[10]In The SAE, there is no restriction on the number of schools that a student can include in her application. The simulated applications developed in this work consider the same amount of schools that each student applied for in the real application process.

$$(A \succsim B \succsim C \succsim D \sim E)$$

This is, A is weakly preferred over B, B is weakly preferred over C, C is weakly preferred over D, and D and E are equally preferred.

4.2 Simulation of application and admission process

In the second stage, I consider the estimated coefficients of Equation 2 to simulate the applications of all students under multiple counterfactual scenarios. After obtaining the simulated applications, the admission process is replicated by running the selection algorithm employed by the SAE[11]. Importantly, the simulation is executed over the same group of students that are considered to estimate the coefficients of Equation 2 (for the group of students over whom no preference parameters are estimated, I keep their real applications across all the simulations). Therefore, for what concerns this work, the estimated preference parameters correspond to the population coefficients.

For the baseline scenario and each counterfactual scenario, I run 250 simulations of the application process, according to the preference parameters exhibited in Section 5 (Table 4). For each school belonging to the choice set of each student, I compute the observed utility and add a random value with an extreme value type 1 distribution to compute the total utility of each student for each school. Simulated applications are obtained by ranking each school according to the total utility. For each scenario, this exercise is replicated 250 times.

The baseline scenario aims to replicate as close as possible the real application of students. Therefore, applications under this scenario are simulated by including all preference parameters defined in Equation 2.
In order to quantify the contributions of social, spatial, and economic frictions to SES school segregation, I develop three counterfactual scenarios that simulate the admission process while setting some of the estimated coefficients presented in Table 4 to zero. Table 17 (Appendix B) provides an overview of the baseline simulation and each counterfactual scenario.
The first counterfactual scenario aims to quantify the contribution of social frictions to SES school segregation. Under this scenario, the admission process is simulated by setting to zero the coefficient on EDD_{ij}, as well as the coefficient on the interaction term $low - SES_i \times EDD_{ij}$. This scenario can be interpreted as a hypothetical case in which applicants do not care about the socioeconomic composition of schools or the social distance between their socioeconomic status and the school socioeconomic composition. The second counterfactual scenario is designed to measure the contribution of spatial frictions to SES school segregation. Simulated applications

[11] Rodrigo Icarán (PUC) runs the selection algorithm employed by the SAE over the simulated applications of each counterfactual scenario.

under this scenario are obtained by setting to zero the coefficient on D_{ij}, as well as the coefficient on the interaction term $low - SES_i \times D_{ij}$. This means that I assume that distance between home and school is not a relevant factor when choosing schools. Alternatively, this scenario can be interpreted as a hypothetical case in which spatial frictions are null. Finally, the third counterfactual scenario captures the role of economic frictions on SES school segregation by setting to zero the coefficient on Tuition fees for both high-SES and low-SES students, which is equivalent to a universal gratuity policy.

4.3 SES school segregation

Finally, school segregation is measured in each of the previously described scenarios. Importantly, SES school segregation is computed by considering the reduced sample of applicants rather than all applicants, as my primary goal is to capture the variation of segregation levels under different assumptions. The students over whom I do not estimate preferences do not variate their school applications across the different scenarios, as I do not simulate alternative applications for them. Taking this into account, this group of applicants is not considered when computing segregation levels in any of the simulated scenarios.

In line with James & Taeuber (1985), this work considers segregation as a measure of dispersion or unevenness.[12] More specifically, I consider SES school segregation as the uneven distribution of social groups (low-SES and high-SES students) among schools. To measure SES school segregation, I rely on the Dissimilarity Index (D-Index), as it is commonly employed in the literature. In terms of the present work, the D-Index is interpreted as the proportion of minority members (low-SES students) that would have to be transferred between schools in order to achieve an even distribution among a given group of schools. The D-Index varies between 0 and 1, where 0 is the case of null segregation or even distribution, and 1 represents the case of maximum segregation. The D-Index is computed as:

$$DI = \frac{1}{2}\sum_{j=1}^{n} \left| \frac{low - SES_j}{low - SES} - \frac{high - SES_j}{high - SES} \right| \tag{4}$$

where $low - SES_j$ is the proportion of low-SES applicants enrolled in school j, $low - SES$ is the total amount of low-SES applicants, $high - SES_j$ is the share of high-SES applicants enrolled in school j, and $high - SES$ is the total amount of high-SES applicants.

5 Estimation results

This section reports the results of estimating an exploded logit regression of the form described by Equation 2 and provides some illustrative examples to facilitate its interpretation.

[12] See Massey & Denton (1988), and Gorard & Taylor (2002) for a review on the multiple dimensions of segregation, such as exposure, isolation, and clustering.

Table 4 shows the estimated coefficients of three alternative specifications to estimate the school preferences of families. Model 1 (column 1) omits EDD and Non-Academic Index. Model 2 (column 2) includes EDD while omitting the Non-Academic Index. Model 3 (column 3) includes both EDD and Non-Academic Index[13].

Across all specifications of Table 4, the estimated coefficients on distance from home to school and academic performance of school (avg. test score) suggest an important role for both attributes. Schools with higher test scores and schools located closer to home are preferred by all types of applicants. Also, columns 2 and 3 show that the Euclidean demographic distance is economically and statistically significant and thus relevant for understanding the school choices of families. Importantly, if EDD is omitted (column 1), the coefficient on avg. test score for high-SES applicants is around twice the value (from 0.641 to 0.329) than the coefficient when the Euclidean demographic distance is considered in the equation (column 2). For low-SES students, the coefficient on average test scores decrease from 0.531 to 0.234.[14] This suggests that the share of high-SES students is highly and positively correlated with the academic performance of schools, while the opposite is true for the share of low-SES students. In other words, if we observe school applications of low-SES and high-SES families, the strong tendency of both groups to prefer schools with higher academic performance hides something else: they are not only looking for schools with high academic performance, but they are also avoiding schools with high shares of low-SES students and looking for schools with high shares of high-SES students. Therefore, if social preferences are omitted, the preferences for academic performance are upward biased due to relevant omitted variables.

Also, a relevant result concerns the change in the coefficient on Log Distance when the Euclidean demographic distance is included (column 2) for high-SES and low-SES students. For the former, the coefficient on Log Distance passes from -1.135 (column 1) to -1.123 (column 2) when the Euclidean demographic distance is included. This could be interpreted as high-SES parents wanting their children to attend to closer schools not only because their value proximity but also because they prefer to enroll their children in schools with lower shares of low-SES students, which are the ones that are closer to them[15].

The opposite occurs for low-SES applicants, as the coefficient on Log distance for the later is more negative if the Euclidean demographic distance is considered (-1.165 in column 1 v/s -1.188 in column 2). This reveals that the desire of low-SES parents to enroll their children in schools with lower shares of low-SES students and higher shares of high-SES students push them to apply to schools that are farther

[13] Specifically, the Non-Academic Index includes (1) an indicator of community integration (parents and teachers communication and their involvement in school's educational project), (2) an indicator of working conditions evolution and overall proper functioning of school (administration, fulfillment of mandatory reporting and administrative transparency), and (3) a measure of educational initiative in term of school's disposition and ability to promote educational innovations and compromise the support of external agents in their pedagogical work.

[14] The coefficient on average test score for low-SES applicants is 0.641 – 0.110 = 0.531 in Model 1. In Model 2, it is equal to 0.329 – 0.095 = 0.234.

[15] See Figure 2 in Appendix A

from home, even though the long distance from home to schools implies a disutility[16].

With respect to the tuition fees charged by schools, again, ignoring the Euclidean demographic distance has a significant impact on our estimates. For high-SES applicants, the coefficient on Tuition fees is actually positive and statistically significant (0.006) if EDD is omitted (column 1). However, when the Euclidean demographic distance is included, it is negative and statistically significant (-0.007), which looks more reasonable. For low-SES applicants, the coefficient on Tuition fees moves from -0.006 to -0.016, more than doubling (in absolute terms) the coefficient when EDD is included and also doubling the coefficient for high-SES applicants. From this, the most evident conclusion is that the disutility of tuition cost is greater for low-SES families, which is expectable, as they probably face stronger financial constraints than high-SES families. A second conclusion that follows is that both high-SES and low-SES parents are willing to pay higher tuition fees in order to avoid enrolling their children in schools with higher shares of low-SES students.

Finally, column 3 of Table 4 shows that controlling for non-academic indicators of schools is relevant. For high-SES applicants, the coefficient on Log distance is slightly underestimated (in absolute terms) when I ignore this attribute, while the coefficients on the average test score, tuition fees, and EDD are overestimated (in absolute terms) when I don't control for non-academic aspects. Notably, the coefficient on average test scores falls from 0.329 to 0.310, suggesting that non-academic performance is positively related to academic performance and high-SES parents do also care about the former. The reduction (in absolute terms) in the coefficient on EDD for high-SES students is also notable when controlling for non-academic performance of schools (-4.804 in column 2 v/s -4.664 in column 3), suggesting that high-SES parents avoid schools with higher shares of low-SES students not only because an explicit social preference but also because schools with higher shares of low-SES students exhibit lower academic performance, on average[17].

In sum, Table 4 suggests that including both the Euclidean demographic distance (EDD) and the Non-Academic Index is relevant to identify the school preferences of families properly. That being said, the preferred specification corresponds to Model 3 (column 3), which includes EDD and controls for non-academic performance. In what follows, the main implications of this estimation are to be described.

[16] See Figure 2 in Appendix A
[17] See Figure 3 in Appendix A

Table 4: School preferences, exploded logit specification

	(1)	(2)	(3)
Log distance (km)	-1.135***	-1.123***	-1.121***
	(0.011)	(0.011)	(0.011)
Avg. test score (Std)	0.641***	0.329***	0.310***
	(0.015)	(0.017)	(0.018)
Tuition fees (USD 2018)	0.006***	-0.007***	-0.005***
	(0.000)	(0.000)	(0.000)
Euclidean Demographic Distance (EDD)		-4.804***	-4.664***
		(0.107)	(0.109)
Low-SES ×Log distance (km)	-0.030*	-0.064***	-0.065***
	(0.016)	(0.017)	(0.017)
Low-SES ×avg. test score 2016 (Std)	-0.110***	-0.095***	-0.065**
	(0.022)	(0.026)	(0.027)
Low-SES ×Tuition fees (USD)	-0.012***	-0.009***	-0.010***
	(0.001)	(0.001)	(0.001)
Low-SES ×EDD		8.339***	8.172***
		(0.151)	(0.153)
Non-Academic Index	NO	NO	YES
Observations	604954	604954	604954
Chi2	22683.4	26099.7	26652.2
AIC	107404.99	103992.71	103456.19
BIC	107472.87	104083.21	103637.20

Notes: This table presents results of an Exploded Logit estimation of school preferences. Standard errors in parentheses. Model 1 omitts EDD and Non-Academic Index. Model 2 includes EDD while omitting Non-Academic Index. Model 3 includes both EDD and Non-Academic Index.

5.1 Spatial frictions

The coefficient on Log distance (km) in column 3 of Table 4 suggests that spatial frictions play a relevant role in the school preferences of families. Despite their socioeconomic status, all applicants are less likely to apply to a school that is physically more distant from their home. However, distaste for long distances between home and school is slightly larger for low-SES students. For this group, a school that is located 1 km away from home is 3.27 times more likely to be included in their preferred school set compared to a school that has the same characteristics as the former but is located 2 km away from home. For high-SES students, the first school would be 3.07

times more likely to be included in their preferred school set than the second one.[18] In monetary terms, a low-SES student is willing to pay USD $54.8 in order to attend a school that is 1 km away from home rather than 2 km away.[19]. Instead, a high-SES would pay a maximum of USD $155.4 for the same change. This may seem contradictory, as we just said that high-SES students give less importance to the distance between home and school. That is, in fact, true, but the coefficient on tuition Fees and low-SES × tuition Fees also suggest that high-SES students are less reluctant to incur in higher monetary costs.

Finally, low-SES students face a stronger trade-off between the academic performance of schools and the distance. For them, changing from a school that is 2 km away to one that is 1 km away would be worth sacrificing 3.36 standard deviations of school's average test score. Instead, a high-SES student would sacrifice 2.5 standard deviations of school's avg. test score. This fact helps explain why the average test score of schools applied by low-SES students is so much lower than the ones applied by high-SES students.

In sum, the coefficient on Log distance suggests that the proximity of school is highly valued by types of families or students. However, the estimated parameters indicate that school proximity is an inferior feature, i.e., an increase in income decreases its demand, as low-SES families put more relevance on this attribute than high-SES families when choosing schools for their children.

5.2 Economic frictions

Concerning economic frictions, column 3 of Table 4 suggests that they play a relevant role in the school choices of both high-SES and low-SES families, as the coefficient across both groups is negative and statistically significant. However, the disutility of tuition fees for low-SES families triplicates the one for high-SES families (a coefficient of -0.015 for the former v/s a coefficient of -0-005 for the later). As expected, this suggests stronger financial constraints for low-SES families.

For high-SES families, an increment of USD $50 in annual tuition fees is equivalent to a 5.4 p.p. increment in the Euclidean demographic distance, a reduction of 0.81 in the avg. test score (std), or an increase of 1.5 km of distance from home to school for a student attending to a school that is 6 km away from home.[20]

To enable the comparison of economic frictions between high-SES and low-SES families in terms of the Euclidean demographic distance, avg. test score (std), and Log distance, let's assume for a moment that low-SES students have the same estimated coefficients on those attributes.[21] Under this scenario, an increase of USD $50 in

[18] Comparing two schools j and j' that are identical in every aspect $V_i = [EDD_i, Score, Size, TuitionFees_i]$ except D_{ij}, $P(d_{ij} = 1|V_i)/P(d_{ij'} = 1|V_i) = \exp(\gamma_1(D_{ij} - D_{ij'}))$, where $d_{ij} = 1$ if student i apply to school j, and $d_{ij} = 0$ otherwise.

[19] The utility increase of attending to a school that is 1 km away rather than 2 km away is $(-1.121 - 0.065) \times (ln1 - ln2) = 0.822$ for a low-SES student. The utility loss for a low-SES student of paying USD $ 54.8 additional tuition fees (annual) is $(-0.005 - 0.010) \times 54.8 = 0.822$

[20] 6 Km is the average distance from home to school in our sample.

[21] Otherwise, the differences between low-SES and high-SES families in the equivalent effect of an increase in annual tuition fees in terms of the above-mentioned attributes would not only reflect the

annual tuition fees for a low-SES family is equivalent to a 16 p.p. increase in the Euclidean demographic distance, a reduction of 2.42 in the avg. test score (std), or an increase of 5.7 km of distance from home to school for a low-SES student attending to a school that is 6 km away from home. However, if we consider the estimated coefficients for low-SES students, the same increase in annual tuition costs is equivalent to 21.4 p.p. decrease in the Euclidean demographic distance (opposite impact relative to high-SES students), a reduction of 3.06 in the avg. test score (std), and an increase of 5.3 km of distance from home to school for a student attending to a school that is 6 km away from home.
Finally, if we compare two schools, A and B, that are identical in all attributes excepting the annual tuition fees which are USD $50 higher in school B, a high-SES family would be 1.28 times more likely to apply to school A, while a low-SES family would be 2.12 times more likely to do so.

In sum, the consequences of economic frictions are statistically and economically significant for both high-SES and low-SES families, but its impact is considerably higher for low-SES families, as they are more financially constrained.

5.3 Social frictions

With respect to social frictions, the results suggest the existence of homophily among high-SES students only, since Table 4 reveals that the coefficient of our measure of social distance (EDD) is negative and statistically significant for this group (coefficient of -4.664, column 3 of Table 4). All else equal, high-SES students are less likely to apply to a school that has a larger share of low-SES students.
In contrast, low-SES students are more likely to apply to schools where the social composition is more distant to their own socioeconomic status. In other words, high-SES parents prefer to enroll their children in schools with a higher share of their own social group, while low-SES parents prefer to distance themselves from their own social group. Importantly, this fact cannot be attributed to differences in school academic quality, as the estimated equation includes an indicator for this aspect. Also, since I control for the distance between home and school, this result cannot be attributed to the joint impact of residential segregation and the disutility of distance. That being said, there is a segregating behavior of high-SES parents when choosing their children's school, while low-SES parents wish their children to increase contact with high-SES students.

To put it into perspective, consider two schools, A and B, which are identical except for their EDD. For a high-SES student, the Euclidean demographic distance is one standard deviation greater (0.213)[22] in school A than in school B. The estimated coefficient on EDD for this type of student indicates that a high-SES student is 37% more likely to apply to school B than to school A. To put it in terms of a trade-off between social friction and spatial frictions, to hold constant the utility of a high-SES student, a school that is 3.5 km away from home and becomes one standard deviation

difference in terms of economic frictions but also in the coefficients on the other attributes.

[22] See Table 16 in Appendix A.

more demographically distant would have to be 58.8% closer (1.44 km away from home) in terms of physical distance.
By contrast, low-SES students are less likely to apply to a school where the share of students from their same socioeconomic status is larger. If two schools, X and Y, are identical except for their EDD, which is one standard deviation greater in school X, a low-SES student would be twice as likely to apply to school X, relative to school Y.

However, the fact that the coefficient on the Euclidean demographic distance is negative for wealthier students could be due to omitted variables such as non-academic aspects of school quality like discipline, school life, etc. The preferred specification controls for non-academic aspects that are summed up in a unique index, but the data does not allow us to consider all non-academic aspects of schools, so I cannot completely discard this possibility. However, for a very reduced fraction of the schools included in my sample, there are additional non-academic indicators such as school life, promotion of healthy lifestyle habits, civic participation, etc. For this group of schools, the observed correlation between our measure of non-academic performance and the aforementioned attributes is greater than 0.5, which suggests that a significant part of non-academic features is captured by this index.

Importantly, the strong preferences of low-SES families for out-group members contradicts the hypothesis of "neutral ethnocentrism" developed by several social scientists (Bifulco & Ladd, 2006; Booker, Zimmer, & Buddin, 2005; Denessen, Driessena, & Sleegers, 2005; Weiher & Tedin, 2002). "Neutral ethnocentrism" states that members of all social groups prefer to interact with individuals of their own group, thus prevailing a behavior of self-segregation of all social groups.
In contrast, the present results suggest that this behavior prevails only among upper-class groups, which is consistent with the "outgroup avoidance," a model initially developed by Blumer (1958) in the context of US racial dynamics and further developed by Bobo (1999). This theory suggests that only upper-class members prefer to interact with members of their same status, as they actively seek to maintain a privileged social position[23].

6 Model Fit

In order to assess the model fit of the replication of the school admission process, this section reports the main results of the baseline simulation, based on the estimated coefficients of Equation 2 [24] and compares them with the real school application and

[23] In the context of school choices and racial segregation see Bifulco & Ladd (2009), Saporito (2003), and Saporito (2006).

[24] For the baseline scenario and each counterfactual scenario, I run 250 simulations of the application process, according to the preference parameters exhibited in Table 4. For each school belonging to the choice set of each student, I compute the observed utility and add a random value with an extreme value type 1 distribution to calculate the total utility of each student for each school. Simulated applications are obtained by ranking each school according to the overall utility. For each scenario, this exercise is replicated 250 times. Segregation levels are computed as the average segregation of these simulations. In

school assignation of applicants. In order to assess the model fit of this simulation, section 6.1 describes the main results regarding school applications in the baseline simulation and compares it to the real school applications of students. Section 6.2 does the equivalent for the school assignation of applicants.

6.1 Baseline simulation, descriptive statistics of applications

Table 5 presents summary statistics of real applications and the ones obtained in the baseline simulation, by type of applicant. Overall, the model shows a similar trend when comparing applications of low-SES and high-SES families. low-SES applicants disproportionately apply more to free schools (67.2% in real applications, 64% in simulated applications), and a very small amount applied to paid schools (10.6% in real applications, 12.6% in simulated applications). Also, low-SES students tend to apply more to schools with lower test scores, lower tuition fees, and higher Euclidean demographic distance. The only exception is distance, which my model predicts to be slightly longer for high-SES students. As described in Section 3.2, the average distance from home to school is larger for high-SES students, but this is possibly due to a stronger trade-off between tuition costs and distance for low-SES students.

In Table 5, if we compare the real applications of high-SES families (column 1) with the simulated applications in the baseline scenario (column 2), the later show slightly shorter distance between home and school (0.37 km less), a 5% higher share of paid schools among school applications of students, and a 5% lower share of free-for-all schools among their applications.

Besides this inaccuracy, simulated and real applications are highly similar for low-SES and high-SES students, and almost all the trends when comparing applications of both groups are preserved.

all scenarios, the real applications of students that are not considered in the estimation sample remain unchanged.

Table 5: Real applications and simulated application, descriptive statistics (mean)

	(1) Real: high-SES	(2) Simu: high-SES	(3) Real: low-SES	(4) Simu: low-SES
Distance (Km)	3.77	3.40	3.84	3.31
Euclidean Demog. Dist	.388	.387	.486	.499
Avg. test score (std)	.718	.721	.263	.309
Annual tuition fees (USD 2018)	554.4	560.9	109.1	116.9
Paid schools	.297	.351	.106	.126
Free schools (for all)	.461	.456	.672	.640
Schools exclusively free for low-SES	.242	.193	.222	.234
Non-Academic Index	2.18	2.49	2.15	1.96
Observations	2,491		1,961	

Notes: This table presents the means observed in the real application process and the 250 baseline simulations. In each case, only schools to which the students applied are considered (irrespective of the ranking). Columns (1) and (3) contain the means observed in the real application process for high-SES students and low-SES students, respectively. Columns (2) and (4) present the means observed in the simulated application, separately for high-SES students and low-SES students, respectively. Paid schools are schools that are not tuition-free and do not adhere to the SEP agreement. Free schools (for all) are schools that are tuition-free, either with or without the SEP agreement. Finally, Free schools (for low-SES only) are schools that are not tuition-free but adhere to SEP agreement, thus being tuition-free for low-SES students.

6.2 Baseline simulation: descriptive statistics of school assignation

In addition, Table 6 compares the main descriptive statistics of school assignation under the baseline simulation with the real output of the SAE 2018. Again, descriptive statistics of the former represent the average value of each attribute across 250 simulations of the School Admission Process. Overall, simulated and real school assignations are similar, and the replication preserves the trends when comparing high-SES and low-SES applicants. Again, the only exception is Distance, which is predicted to be slightly higher for high-SES students than low-SES students.
To assess the validity of my model, possibly the most relevant aspect is its ability to correctly replicate the levels of SES school segregation through the baseline simulation. Table 7 exhibits the levels of SES school segregation observed among the estimation sample [25]. The results show an accurate prediction of the D-Index by the baseline simulation, as the difference between the real D-index and the simulated one (average of 250 simulations) is less than 1 p.p. Moreover, the real D-Index falls inside the interval determined by the Pctile. 5 and Pctile.95 of the distribution of simulated levels of segregation, also suggesting an accurate prediction of the model (Figure 1).

[25] As previously mentioned, my measure of segregation excludes the school assignation of applicants for whom I don 't estimate preferences parameters.

Table 6: Real school assignations and simulated assignations (baseline), descriptive statistics

	(1) Real: high-SES	(2) Simu: high-SES	(3) Real: low-SES	(4) Simu: low-SES
Distance (km)	3.76	3.49	3.85	3.39
Euclidean demographic distance (EDD)	.428	.421	.454	.458
Avg. test score	.560	.601	.166	.156
Tuition Fees	450.7	485.7	88.14	83.25
Schools exclusively free for low-SES	.221	.204	.188	.202
Paid schools	.245	.283	.083	.081
Free schools	.535	.513	.729	.717

Notes: This table presents the means observed in the real application process and the simulated application process at the student level in each of the 250 simulations of the corresponding counterfactual, separately by the socioeconomic status of students. In each case, only the schools to which the students applied are considered (irrespective of the ranking). Columns (1) and (3) contain the means observed in the real application process for high-SES students and low-SES students, respectively. Columns (2) and (4) present the means observed in the simulated application, separately for high-SES students and low-SES students, respectively. Paid schools are schools that are not tuition-free and do not adhere to the SEP agreement. Free schools (for all) are schools that are tuition-free, either with or without the SEP agreement. Finally, Free schools (for low-SES only) are schools that are not tuition-free but adhere to the SEP agreement, thus being tuition-free for low-SES students.

Table 7: Dissimilarity Index (D-Index) of SES school segregation under baseline simulation and real admission process

	(1)
Baseline simulation (Avg)	0.581
Pctile. 5 & Pctile. 95	[0.561 ; 0.602]
Real admission process	0.572

Notes: This table presents the average level of SES school segregation for 250 simulations of the School Admission Process (SAE) of the year 2018 (baseline simulation), measured by the Dissimilarity Index of segregation (row 1), as well as a bracket-displayed interval containing the Pctile. 5 and Pctile. 95 of the distribution (row 2). Also, the Dissimilarity Index of segregation estimated for the real SAE is exhibited (row 3). Levels of segregation are computed by considering the school assignation of applicants for whom preference parameters are estimated (see Section 4.3 for more details). School assignations (real and simulated) are determined by the algorithm employed in the official School Admission Process (see Section 2 for more details).

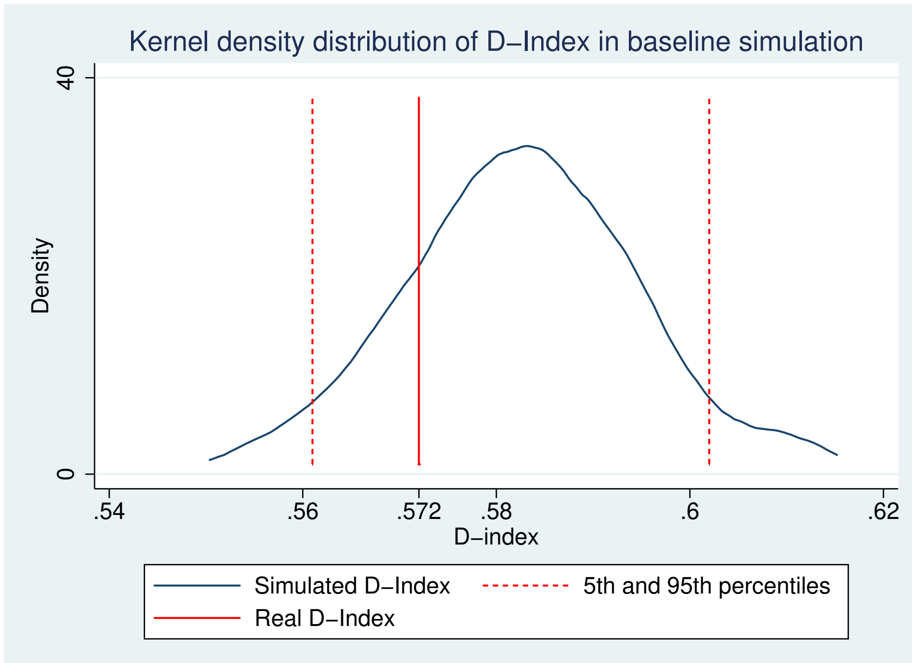

Figure 1: Kernel-density distribution of Dissimilarity Index (D-Index) under baseline simulation

Overall, the baseline simulation of school applications and school assignations is able to replicate the real School Admission Process accurately. Despite the existence of a small inaccuracy in the avg. distance simulated by the model, SES segregation levels are reproduced with high precision, which allows me to proceed with the comparative analysis of SES school segregation levels across the multiple counterfactual scenarios.

7 Socioeconomic school segregation

This section presents the resulting levels of segregation in each of the simulated scenarios, considering the school assignation determined by the selection algorithm of the School Admission System. Section 7.1 presents the main results of each simulation and its interpretation. In order to provide a better insight into the mechanism underlying the main results exhibited in section 7.1, section 7.2 presents descriptive statistics of school applications and school assignation under each counterfactual scenario.

7.1 Main results

Table 8 presents dissimilarity indices for low-SES students across the baseline scenario (column 1) and each counterfactual simulation, as well as the difference of the latter relative to the former. Column 2 shows the average D-Index that arises from setting the estimated coefficient on EDD and low-SES × EDD to zero. Column 3 exhibits the average D-Index that arises from setting the estimated coefficient on Log distance and low-SES × Log distance to zero. Finally, column 4 presents the average D-Index resulting from setting the estimated coefficient on tuition fees and low-SES × tuition fees to zero. Importantly, the set of schools, the school choice set of each student, and school characteristics are fixed across all scenarios. Therefore, the comparative analysis developed here does not capture the total effect of eliminating, social, spatial, or economic frictions, as general equilibrium effects may arise, such as supply responses or a change in applicant 's choice set.

Table 8: SES school segregation

	(1) Baseline	(2) No social	(3) No spatial	(4) No economic
D-Index	.581	.515	.587	.685
Change (p.p.)		-6.6***	0.6***	10.4***

Notes: This table reports Dissimilarity Index (row 1) for Low-SES students and High-SES students and its percentage-points reduction relative to the baseline simulation (Row 2). In all cases, the D-Index is computed by considering the school assigned to each student based on their simulated applications. Therefore, the D-Index reports SES school segregation levels across students that applied to Pre-Kindergarten through the SAE. Col.1 reports the D-Index for the Baseline simulation, which takes all the coefficients estimated in Table 7 to simulate students' applications. Col.2 sets the coefficients on EDD and $low-SES \times EDD$ to zero. Col. 3 sets the coefficients on $Logdistance$ and $low-SES \times Logdistance$ to zero. Col.4 sets the coefficient on $tuitionfees$ and $low-SES \times tuitionfees$ to zero. *** indicates statistical significance at the 1%, determined by a T-test of differences.

Comparing columns 1 and 2 suggest a first-order role of social frictions on SES school segregation of low-SES students, as the latter is reduced by 6.6 p.p. when social frictions are set to zero, implying a reduction of 11,4 % in SES school segregation. This significant reduction means that the segregating preferences of high-SES parents, i.e., their willingness to enroll their children in schools with higher shares of students of their own socioeconomic status and avoid schools with high shares of low-SES students, is so strong, that the integrating practices of low-SES families, who are willing to enroll their children in schools with higher shares of high-SES students, are not enough.

Comparing column 1 with column 3, eliminating spatial frictions slightly exacerbates SES school segregation for low-SES students. Despite the statistical significance of this result, its economic significance is reduced, as the increase of the D-Index is less than 1 p.p. and equivalent to a 1% reduction relative to the baseline scenario. However, it suggests a reduced role of residential segregation in SES school segrega-

tion, which may suggest that low-SES and high-SES families are integrated in terms of residence. However, the available evidence for the cities of Valparaíso, Viña del Mar, and Concepción contradicts this idea (Sabatini et al., 2010; Rasse, 2016). Therefore, a deeper exploration of these results is needed, which is developed in section 7.2.
Finally, comparing columns 1 and 4 indicates that setting economic frictions to zero increases SES school segregation of low-SES applicants. This result suggests that the current educational policy relative to tuition costs is successful in terms of reducing SES school segregation. The following section explores alternative mechanisms that may be underlying these results.

7.2 Descriptive statistics of school assignations under each counterfactual and exploration of mechanisms.

In order to understand the mechanism underlying the results mentioned above, tables 9 and 10 provides an overview of school assignation of low-SES and high-SES applicants under each counterfactual scenario, respectively.

Table 9: School assignation under each counterfactual scenario, low-SES students.

	(1) Baseline	(2) No social	(3) No spatial	(4) No economic
Distance (km)	3.39	3.30	5.27	4.51
Euclidean Demographic Distance (EDD)	.458	.414	.470	.561
Avg. test score (std)	.156	.031	.213	.492
Annual tuition Fees	83.3	51.5	85.9	511.2
Schools exclusively free for low-SES	.225	.185	.242	.158
Paid schools	.081	.057	.086	.343
Free-for-all schools	.694	.758	.671	.498

Notes: This table presents student-level descriptive statistics of school assignations under each counterfactual scenario, considering only low-SES applicants. Column 1 represents the Baseline scenario, column 2 the scenario without social frictions, column 3 the scenario without spatial frictions, and column 4 the scenario without economic frictions.

As expected, setting social frictions to zero implies that the average EDD for low-SES applicants reduces (column 2 of Table 9). Also, comparing columns 1 and 2 of Table 9 shows that low-SES applicants enroll in more closer schools when social frictions are omitted. This is consistent with low-SES parents willing to sacrifice proximity in order to enroll their children in schools with higher shares of high-SES students. Also, a higher share of low-SES students enrolls in free-for-all schools, which is explained by the fact that free-for-all schools have higher shares of low-SES students (Table 2), causing disutility for low-SES families. Therefore, when social frictions are assumed to be non-existent, the utility of attending to a free-for-all school increases.
For high-SES applicants (Table 10), setting social frictions to zero (column 2) implies

Table 10: School assignation under each counterfactual scenario, high-SES students.

	(1) Baseline	(2) No social	(3) No spatial	(4) No economic
Distance (km)	3.49	3.41	5.49	4.30
Euclidean Demographic Distance (EDD)	.421	.517	.413	.296
Avg. test score (std)	.601	.340	.638	.963
Annual tuition Fees	485.7	297.2	503.8	987.0
Schools exclusively free for low-SES	.201	.167	.216	.280
Paid schools	.283	.167	.294	.519
Free-for-all schools	.513	.667	.490	.201

Notes: This table presents student-level descriptive statistics of school assignations under each counterfactual scenario, considering only high-SES applicants. Column 1 represents the Baseline scenario, column 2 the scenario without social frictions, column 3 the scenario without spatial frictions, and column 4 the scenario without economic frictions.

that they enroll in schools that are slightly closer, more socially distant (higher the Euclidean demographic distance), have lower academic performance, and charge lower tuition fees. The increase in the Euclidean demographic distance is reasonable when social frictions are omitted, as high-SES students experience a disutility from higher levels of EDD. Also, the reduction in the avg. test score is consistent with the strong and negative correlation between avg.test score and share of low-SES students (see Figure 3 in Appendix A). Finally, ignoring social frictions results in a longer share of high-SES applicants enrolling in free-for-all schools, which plays a significant role in the great reduction of SES school segregation under this scenario.
It is worth remarking that SES school segregation, in the context of the present work, provides a measure of how disproportionately low-SES applicants do, relative to high-SES applicants, enroll in schools with high shares of low-SES students. If we compare the gap between the share of low-SES (Table 9) and high-SES applicants (Table 10) enrolled in free-for-all schools under the Baseline simulation (column 1) with the same gap under the simulation without social frictions, there is a notable reduction in the latter case (column 2 of Tables 9 and 10), as the gap decreases from 18.1 p.p to 9.1 p.p..[26] This indicates that both low-SES and high-SES parents are willing to pay higher tuition fees in order to avoid schools with high shares of low-SES students, but this willingness is higher for high-SES parents.

Concerning spatial frictions, the average distance increases by 55% and 57% for high-SES and low-SES applicants, respectively. Also, avg. test score (std) increases for both low-SES and high-SES applicants. Both facts suggest that eliminating spatial frictions allows applicants to enroll in schools that are farther but with greater academic performance. In other words, the additional utility of greater academic performance of those schools is no longer offset by the disutility of a larger distance. However, these facts alone do not explain why SES school segregation levels are higher

[26] The gap under the baseline simulation is $.694 - .513 = .18$. The gap under the simulation without social frictions is $.758 - .667 = 9.1$.

under this scenario. As already mentioned, the possibility of low-SES and high-SES families being residentially integrated does not seem feasible. Instead, an alternative hypothesis to explain the higher levels of segregation observed when spatial frictions are set to zero is based on a possible "rebound effect". A rebound effect would occur if congestion or over-demand increases when spatial frictions are set to zero. The reasoning is that in the absence of spatial frictions, there is a particular group of schools that experiment an increase in the number of applicants: the ones that have higher academic performance and the ones with lower shares of low-SES students[27]. This creates an over-demand of applicants in certain schools that may lead to higher SES school segregation if low-SES and high-SES applicants end up concentrating in different schools. To explore this possibility, Table 11 displays congestion measures that reflect the average number of schools that offer less vacancies than the number of students applying to them (Avg. congestion), the share of low-SES applicants who are assigned to highly congested schools relative to the total number of applicants assigned to those school (low-SES assigned high-congestion), the average number of applicants who are not assigned to any school included in their applications (Not assigned), the average test score of highly congested schools (High-congestion avg. test score), and the average share of low-SES students in highly congested schools (High-congestion avg. share of low-SES).

In line with the "rebound" hypothesis, congestion levels are higher when spatial frictions are set to zero (column 3), and those schools exhibit higher levels of academic performance. Importantly, the share of low-SES students in highly congested schools decreases by 5 p.p., suggesting that the elimination of spatial frictions promotes the segregating behavior of high-SES applicants.

However, the increase in the average number of schools that are congested is modest (3.3%). In addition, the share of low-SES applicants assigned to congested schools among all applicants assigned to those schools is only 1 p.p. lower relative to the baseline scenario and the share of applicants who are not assigned to any school included in their applications actually decreases by 1 p.p. Taking this into account, the evidence suggests the existence of a small rebound effect that is hardly enough to explain the higher or invariant levels of SES school segregation when spatial frictions are omitted. Furthermore, if we look at the segregation levels that result from assuming no congestion, i.e., taking the first school choice of applicants, the results are very similar (see appendix C, Table 19).[28]Therefore, the most likely mechanism that may be contributing to higher levels of SES school segregation is the segregating behavior of high-SES families that increases when spatial frictions are omitted.

[27]Academic performance and share of low-SES students are both relevant school attributes for applicants. "Better schools" are schools with higher academic performance and lower share of low-SES students. Also, schools that charge lower tuition fees are preferred, but this feature has student-level variation

[28]The change in the D-Index if we consider the first choice of applicants in both the baseline and counterfactual simulation is of 0.5 p.p., while the change when considering school assignments is of 0.6 p.p.

Table 11: Congestion measures

	(1) Baseline	(2) No spatial	(3) No economic
Avg. congestion	183.4	189.1	61.82
Avg. congestion (%)	33%	34%	11%
low-SES assigned high-congestion	43%	42%	.36 %
Not assigned, low-SES	3%	2%	53%
Not assigned, high-SES	5%	4%	48%
High-congestion avg. Test score (std)	.312	.441	.850
High-congestion avg. share of low-SES	49%	44%	.33%

Notes: "Avg. Congestion" reflects the average number of schools that offer fewer vacancies than the number of students applying to them, " low-SES assigned high-congestion" shows the share of low-SES applicants who are assigned to highly congested schools relative to the total number of applicants assigned to those school, "Not assigned" measures the average number of applicants who are not assigned to any school included in their applications, "High-congestion Avg. Test score" exhibits the average test score of highly congested schools, and "High-congestion avg. share of low-SES " shows the average share of low-SES students in highly congested schools. Column 1 contains statistics for the baseline simulation and column 2 for the counterfactual that sets spatial frictions to zero.

Finally, columns 4 of Tables 9 and 10 show that eliminating economic frictions results in a higher distance from home to school for both low-SES and high-SES applicants and a significant improvement in the academic performance of schools. Both facts suggest that eliminating economic barriers allows both high-SES and low-SES parents to enroll their children in schools with better academic performance while sacrificing proximity.

Also, there is an increase in avg. EDD for low-SES students, and a decrease in the avg. EDD of high-SES applicants. This implies that both low-SES and high-SES parents chose to enroll their children in schools with lower shares of low-SES applicants in the absence of economic restrictions. In addition, there is a 7 p.p. decrease in the percentage of low-SES applicants enrolled in schools that are exclusively free for them, the share of low-SES applicants enrolled in paid schools is quadrupled, and the proportion of them that enters in free-for-all schools decreases from 69% (column 1 of Table 9) to 50% (column 4 of Table 10). For high-SES students, the share of high-SES applicants (Table 10) that enrolls in schools that are exclusively free for low-SES students increases by 8 p.p., the proportion of applicants enrolled in paid schools increases from 28 % to 52 %, and the share of applicants that are accepted in free-for-all school decreases by 31 p.p. This means that low-SES applicants move from free-for-all schools and schools that are exclusively free for them to paid schools, while high-SES applicants also move to schools that are only free for low-SES students. This massive migration from free schools to paid schools is consistent with the latter exhibiting higher academic performance and lower share of low-SES, both attributes that are highly valued by all parents.

However, these facts do not explain why SES segregation levels are higher under this scenario. In this regard, Table 11 provides valuable information to understand the mechanism underlying this result. First, column 4 of Table 11 suggests that there is a high "rebound" effect in this scenario. The avg. number of congested schools is significantly lower in this simulation (11%), but there is a high proportion of low-SES ("Not assigned, low-SES ") and high-SES applicants ("Not assigned, high-SES ") that are not accepted in any of their preferred schools(53% and 48%, respectively). Both facts, together, suggest that eliminating economic frictions implies that a reduced amount of schools is preferred by a considerable proportion of applicants, resulting in highly concentrated school applications.

Nonetheless, these facts alone do not necessarily imply high levels of SES school segregation. If a small number of applicants is accepted in highly congested schools but there share of low-SES and high-SES applicants who are assigned is balanced, SES school segregation should not be particularly high. However, row 3 of Table 11 (column 3) shows that the share of low-SES applicants assigned to highly congested schools is particularly low in this scenario (36%). In other words, among all applicants that are accepted in a highly congested school, only 36% are low-SES students. This, combined with the low shares of low-SES students (not applicants) in highly congested schools (last row of Table 11) may explain the high levels of segregation produced in this scenario.

Also, Table 12 provides a more in-depth insight into school applications and assignations of low-SES and high-SES applicants when economic frictions are set to zero.

Table 12: No economic frictions, summary statistics of rebound effect

	(1) low-SES applicants	(2) high-SES applicants
Accepted in a highly congested school (relative to all same-SES applicants)	57.6%	71.5%
Highly congested school as first choice	77.6%	87.7%
Distance (km) to highly congested schools	5.4	5.3
Sibling priority in highly congested school	12.0 %	15.1 %
2nd choice is a highly congested school	99.7%	97.3%
Avg. test score if not accepted in a highly congested school	-.025	.455
Share of low-SES if not accepted in a highly congested school	59%	42.8%

Notes: This table presents additional summary statistics of school assignations and applications of low-SES and high-SES students when economic frictions are set to zero. Columns 1 and 2 display statistics for low-SES and high-SES applicants, respectively.

First, we confirm that low-SES applicants are less likely to be accepted in highly congested schools and that results revealed in Table 11 are not due to a higher share of high-SES applicants among all types of applicants. If only considering low-SES

applicants that are assigned to a school, the percentage of low-SES students accepted in a highly congested school is 57.6% (column 1, row 1) while the proportion of high-SES applicants that are accepted in a highly congested school, among all high-SES applicants that are assigned to a school, is 71.5% (column 2, row 1).
Therefore, a relevant question is why high-SES applicants are more likely to be accepted in highly congested schools than low-SES applicants. Table 15 shows that the proportion of low-SES applicants that put a highly congested school in their first preference (column 1, row 2) is lower (77.6%) than the share of high-SES applicants that does so (87.7%, column 2, row 2). This suggests that the utility of enrolling children in a highly congested school is higher for high-SES parents, which makes sense, as highly congested schools exhibit particularly high levels of academic performance (Table 11), an attribute that is more valued by high-SES families (Table 4).
Importantly, Table 12 shows that the average distance between home and highly congested schools is 2 km greater than the average distance of schools that are actually included in student's baseline applications (Table 6) for both low-SES and high-SES applicants. Again, this helps to explain why the value of highly congested schools is lower for low-SES parents, as the dislike for distance is greater for them (Table 4). Also, a 15.1 % of high-SES applicants have "sibling priority" in at least one highly congested school, while only a 12.0 % of low-SES applicants do, which may also explain the higher share of high-SES applicants assigned to highly congested school.
Importantly, the fact that the utility of enrolling children in a highly congested school is higher for high-SES parents does not imply that low-SES applicants do not apply to those schools. In fact, if we consider the second-ranked school, the share of low-SES applicants that apply to a highly congested school (99.7%) is higher (and almost 100%) than the share of high-SES applicants that does so (97.3 %), which confirms that congestion is driven by both low-SES and high-SES applicants, but the latter have weaker preferences for those schools. Also, Table 12 reveals that low-SES children that are not accepted in highly congested schools end up in schools with poorer academic performance and a higher share of low-SES students, relative to high-SES applicants.

Finally, Table 19 (Appendix C) shows that the increase in SES school segregation when assuming no economic frictions reduces from 10.4 p.p. to 4.2 p.p. if we assume no congestion. This suggests a first-order role of congestion in explaining the high levels of SES school segregation when economic frictions are set to zero. However, Table 19 also indicates that congestion is not enough to explain the increase in the D-Index, meaning that heterogeneous preferences of low-SES and high-SES parents concerning school attributes do also translate into higher SES school segregation when economic frictions are set to zero.
Overall, the present results suggest a high contribution of social frictions to SES school segregation, while spatial frictions do not seem to increase SES school segregation in our target population [29].
Furthermore, assuming no spatial frictions results in slightly higher levels of SES school segregation, which may be attributable to an intensification of the segregating

[29] It is important to remark that these results are limited to the public educational system of Chile, and they are not valid for the private school system and the population that participates in it.

behavior of high-SES families. Finally, setting economic frictions to zero results in a significant increase in the levels of SES school segregation, which is explained by two factors. First, there is a strong "rebound" effect produced by highly concentrated school applications in a reduced amount of schools with high academic performance and low share of low-SES students. In this scenario, high-SES applicants end up concentrating in those schools, while low-SES applicants end up in schools with poorer academic performance and higher share of low-SES students. Second, and in a smaller degree, heterogeneous preferences of high-SES and low-SES parents concerning school attributes such as distance, academic performance, and share of low-SES students also explain the increase in SES school segregation when economic frictions are set to zero.

8 Summary and conclusions

This work aimed to provide a reliable measure of the contribution of social, spatial, and economic frictions to SES school segregation. Using data on Pre-Kindergarten applications to estimate school preferences of low-SES and high-SES parents and developing a counterfactual analysis in which the school applications of students and the school admission process was simulated under no-social, no-spatial, and no-economic frictions assumptions (separately), three main conclusions were obtained. First, the social preferences of families play a first-order role in explaining the levels of SES school segregation in Pre-Kindergarten. Simulating the school admission process under an assumption of no social frictions translates into a 6.6 p.p. decrease in SES school segregation. This large reduction is explained by a self-segregating behavior of upper-class parents that is consistent with the "outgroup avoidance" hypothesis developed by Blumer (1958), which suggest that upper-class members actively seek to maintain their privileged social position, avoiding their interaction with out-group members.

Second, spatial frictions do not seem to contribute in a significant way to SES school segregation, suggesting a reduced capacity of residential segregation to explain it. These results may be attributable to an intensification of the segregating behavior of high-SES families when spatial frictions are set to zero.

Finally, setting economic frictions to zero results in a great increase in the levels of SES school segregation (10.4 p.p.), which is the result of two simultaneous phenomena. First, there is a strong "rebound" effect produced by highly concentrated school applications in a reduced amount of schools with high academic performance and low share of low-SES students. In this scenario, high-SES applicants end up concentrating in those schools, while low-SES applicants end up in schools with poorer academic performance and higher share of low-SES students. Second, and in a smaller degree, heterogeneous preferences of high-SES and low-SES parents concerning school attributes such as distance, academic performance, and share of low-SES students also explain the increase in SES school segregation when economic frictions are set to zero, as there is still an increase in SES school segregation when assuming no congestion in this scenario.

Although the present analysis does not provide a definitive answer, the findings described above have several key implications regarding educational policies aiming to reduce SES school segregation. First, the rebound effect observed when economic frictions are transversely set to zero provides relevant evidence concerning the effectiveness of a universal gratuity policy. In that scenario, reducing all tuition fees to zero to all types of students results in a high concentration of student's applications across a reduced number of schools, which are the ones with the highest levels of academic performance and the lowest shares of low-SES students. This dynamic results in a great increase in SES school segregation, thus being an ineffective tool to reduce SES school segregation. However, this does not mean that reducing tuition fees always increase SES school segregation, but it suggests that targeted policies that aim to alleviate the financial constraints of low-SES parents when choosing a school for their children may be more desirable than a universal gratuity initiative, as the rebound effect would probably be lower due to less congested applications. In this line, education programs such as the SEP point the right direction to reduce SES school segregation, but further research is needed in order to reveal how reducing economic barriers exclusively for poor students may impact SES school segregation.
In addition, the high relevance of social frictions in explaining SES school segregation suggests that the latter is mainly a consequence of self-selection and not school-side selection. This is particularly important in the context of the Chilean educational system, as one of the main reasons to implement the SAE was to reduce SES school segregation by banning schools from arbitrarily selecting their students. The evidence presented in this work suggest that the latter is not enough for that purpose, as self-segregating practices of upper-class parents prevail.
Inevitably, these results bring the voucher system into the discussion. The idea that free school choice could increase school segregation has been widely argued in the literature (Hansen and Gustafsson, 2016; Karsten et al., 2006; Schneider et al., 2012.). In this line, the evidence presented here suggests that this type of policy may promote educational stratification if not complemented with integrative efforts. These efforts should point in two directions. First, targeted financial aids aiming to reduce financial restrictions for the poor may have a great impact on reducing school segregation. low-SES parents, such as high-SES parents, wish to enroll their children in schools with low shares of low-SES students, but the high tuition fees charged by those schools prevent them from doing so. Second, promoting or even ensuring a minimum share of low-SES students enrolled in all schools could lessen the socioeconomic stratification of schools that arises from free school choice systems.

9 References

Bellei, C. (2013). El estudio de la segregación socioeconómica y académica de la educación chilena. *Estudios pedagógicos (Valdivia)*, 39(1), 325–345.

Bifulco, R., & Ladd, H. (2006). School choice, racial segregation, and test-score gaps: Evidence from North Carolina's Charter School Program. *Journal of Policy Analysis* *and Management* , 26(1), 31–56.

Billings, S., Deming, D., & Rockoff, J. (2014). School Segregation, Educational Attainment, and Crime: Evidence from the End of Busing in Charlotte-Mecklenburg. *The Quarterly Journal of Economics*, Volume 129, Issue 1

Blumer, H. (1958). Race prejudice as a sense of group position. *Pacific Sociological Review* , 1(1), 3–7.

Bobo, L., & Zubrinsky, C. (1996). Attitudes on residential integration: Perceived status differences, mere in-group preference, or racial prejudice? *Social forces*, 74(3), 883-909.

Bobo, L. (1999). Prejudice as group position: Microfoundations of a sociological approach to race and race relations.*Journal of Social Issues*, 55(3), 445–472.

Boustan, L. (2011). Racial Residential Segregation in American Cities. *The Oxford Handbook of Urban Economics and Planning,* Oxford University Press.

Böhlmark, A., Holmlund, H., & Lindahl, M. (2015). School choice and segregation: Evidence from Sweden (No. 2015: 8). Working Paper, IFAU-Institute for Evaluation of Labour Market and Education Policy.

Böhlmark, A., Holmlund, H. & Lindahl, M. (2016). Parental choice, neighborhood segregation or cream skimming? An analysis of school segregation after a generalized choice reform. *Journal of Population Economics*, 29:1155-1190.

Brandén, M., Bygren, M. (2018). School Choice and School Segregation: Lessons from Sweden's School Voucher System.

Burgess, S., Greaves, E., Vignoles, A., Wilson, D. (2015). What parents want: school preferences and school choice. The Economic Journal, 125(587), 1262-1289.

Card, D., Mas, A. & Rothstein, J. (2008). Tipping and the Dynamics of Segregation. *The Quarterly Journal of Economics*, Volume 123 (1), 177–218.

Chumacero, R., Gomez, D., & R. Paredes (2011). I Would Walk 500 Miles (if it paid).*Economics of Education Review*, Vol.30(5), pp.1103-1114.

Clark, W. (1986). Residential segregation in American cities: A review and interpretation. *Population research and Policy review*, 5(2), 95-127.

Clark, W. (1989). Residential segregation in American cities: Common ground and differences in interpretation. *Population Research and Policy Review*, 8(2), 193-197.

Coleman, J. (1990). Equality and achievement in education. San Francisco: Westview Press.

Davis, D., Dingel, J., Monras, J., & Morales, E. (2019). How segregated is urban consumption? *Journal of Political Economy*, 127(4), 000-000.

Denessen, E., Driessena, G., & Sleegers, P. (2005). Segregation by choice? A study of group-specific reasons for school choice. *Journal of education policy*, 20(3), 347-368.

Duncan, G. J., Brooks-Gunn, J., & Klebanov, P. K. (1994). Economic deprivation and early childhood development. *Child development*, 65(2), 296-318.

Duncan, G., Yeung, W., Brooks-Gunn, J., & Smith, J. (1998). How Much Does Childhood Poverty Affect the Life Chances of Children?*American Sociological Review* , 63(3), 406-423. Elacqua, G., Schneider, M. & Buckley, J. (2006). School choice in Chile: Is it class or the classroom? *Journal of Policy Analysis and Management*, Vol. 25 (3).

Elacqua, G.(2012). The impact of school choice and public policy on segregation: evidence from Chile. *International Journal of Educational Development*, vol. 32 (3), Amsterdam, Elsevier.

Gallego, F. & Hernando, A. (2008). On the Determinants and Implications of School Choice: Semi-Structural Simulations for Chile. *The Journal of the Latin American and the Caribbean Economic Association*, 9 (1): 197-244.

Gallego, F. & Hernando, A. (2009). School Choice in Chile: Looking at the Demand Side. *Documento de Trabajo*, 356.

Galster, G. (1988). Residential segregation in American cities: A contrary review. *Population research and policy review*, 7(2), 93-112.

Hastings, J., Kane, T., & Staiger, D. (2005). Parental preferences and school competition: Evidence from a public school choice program (No. w11805). National Bureau of Economic Research.

Hosokawa, R., & Katsura, T. (2018). Effect of socioeconomic status on behavioral problems from preschool to early elementary school–A Japanese longitudinal study.*PloS one*, 13(5), e0197961.

Imberman, S. A., Kugler, A. D., & Sacerdote, B. (2012). Katrina,s children: Evidence on the structure of peer effects from hurricane evacuees. *American Economic Review*, 102(5), 2048-82.

James, D. R., Taeuber, K. E. (1985). Measures of segregation. Sociological methodology, 15, 1-32.

Karsten et al. (2003). School choice and ethnic segregation. *Educational policy*, 17(4), 452-477.

Kristen, C. (2008). Primary school choice and ethnic school segregation in German elementary schools. *European Sociological Review*, 24(4), 495–510.

Lankford, H., & Wyckoff, J. (2006). The effect of school choice and residential location on the racial segregation of students. In Improving school accountability (pp. 185-239). *Emerald Group Publishing Limited* .

Lindbom, A., 2010. School Choice in Sweden: Effects on Student Performance, School Costs, and Segregation. *Scandinavian Journal of Educational Research*, 54:615–630.

Logan, J., Minca, E., & Adar, S. (2012). The geography of inequality: Why separate means unequal in American public schools. *Sociology of education*, 85(3), 287-301.

Mickelson, R., Bottia, M., & Southworth, S. (2008). School choice and segregation by race, class, and achievement.

Pinderhughes et al. (2000). Discipline responses: Influences of parents' socioeconomic status, ethnicity, beliefs about parenting, stress, and cognitive-emotional processes. *Journal of family psychology*, 14(3), 380.

Rao, G. (2019). Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools. *American Economic Review*, 109 (3): 774-809.

Rumberger, R., & Palardy, G. (2005). Does segregation still matter? The impact of student composition on academic achievement in high school. *Teachers college record*, 107(9), 1999.

Santos, H., & Elacqua, G. (2016). Segregación socioeconómica escolar en Chile: elección de la escuela por los padres y un análisis contrafactual teórico.

Saporito, S. (2003). Private choices and public consequences: Magnet school choice and segregation by race and poverty. *Social Problems*, 50(2), 181–203.

Saporito, S., & Sohoni, D. (2006). Coloring outside the lines: Racial segregation in public schools and their attendance boundaries. *Sociology of Education*, 9(2), 81–105.

Schelling, T. (1971). Dynamic Models of Segregation. *Journal of Mathematical Sociology*, 1:143–86.

Schneider, K., Schuchart, C., Weishaupt, H., & Riedel, A. (2012). The effect of free primary school choice on ethnic groups - Evidence from a policy reform. *European Journal of Political Economy*, 28(4), 430–444.

Tagle, F. (2018). ¿Condicionados por la Información?: Investigando el conocimiento que disponen los apoderados al momento de elegir establecimientos educacionales. *Tesis de Magíster en Economía UC*.

Timberlake, J. (2000). Still Life in Black and White: Effects of Racial and Class Attitudes on Prospects for Residential Integration in Atlanta. *Sociological Inquiry*, 70:420–45.

Valenzuela, J., Bellei, C., & Ríos, D. (2014). Socioeconomic school segregation in a market-oriented educational system. The case of Chile. *Journal of Education Policy*, 29(2), 217-241.

Van Ewijk, Reyn, & Peter Sleegers. (2010). The Effect of Peer Socioeconomic Status on Student Achievement: A Meta-Analysis. *Educational Research Review*, 5 (2): 134–50.

Appendices

A Descriptive statistics

Table 13: Free-School applications, descriptive statistics by type of student

	(1) High-SES students	(2) Low-SES students	(3) Difference
Avg. test score (std)	.371	.185	.186***
	[.700]	[.684]	(.023)
Distance (km)	3.94	3.90	.040
	[10.67]	[10.57]	(.356)
Share of low-SES students	.520	.538	-.019***
	[.127]	[.139]	(.004)
Non-Academic Index	2.26	2.17	.086***
	[.759]	[.734]	(.025)
Observations	1667	1901	

Notes: The present table shows the means observed for the school applications at the student level, considering only free schools, and separately by the socioeconomic status of students. For low-SES students, both Free-for-all schools and schools that are exclusively free for them are considered. For high-SES, only Free-for-all schools are considered. Only the schools to which the students applied are included (irrespective of the ranking). Columns (1) and (2) contain the means observed for high-SES students and low-SES students, respectively. Column (3) present the difference between both group and its statistical significance, determined by a T-test.

Table 14: Paid-School applications, descriptive statistics by type of student

	(1) High-SES students	(2) Low-SES students	(3) Difference
Avg. test score (std)	1.03	1.15	-.130***
	[.581]	[.651]	(.032)
Annual Tuition fees (USD 2018)	944.8	982.0	-37.18*
	[421.2]	[417.0]	(22.51)
Distance (km)	3.33	2.43	.893***
	[5.41]	[2.38]	(.267)
Share of low-SES students	.276	.222	.054***
	[.116]	[.093]	(.006)
Non-academic index	2.16	2.25	-.086*
	[.797]	[.960]	(.044)
Observations	1884	428	

Notes: This table presents the means observed for the school applications at the student level, considering only paid schools, and separately by the socioeconomic status of students. Only the schools to which the students applied are considered (irrespective of the ranking). Columns (1) and (2) contain the means observed for high-SES students and low-SES students, respectively. Column (3) present the difference between both group and its statistical significance, determined by a T-test.

Table 15: Applications to schools that are exclusively free for low-SES students, descriptive statistics by type of student

	(1) High-SES students	(2) Low-SES students	(3) Difference
Avg. test score (std)	.784	.644	.140***
	[.505]	[.555]	(.023)
Distance (km)	3.18	3.33	-.152
	[5.149]	[7.912]	(.282)
Annual Tuition fees (USD 2018)	745.5	0	745.5***
	[309.8]	[417.0]	(10.74)
Share of low-SES students	.349	0.379	-.271***
	[.091]	[.106]	(.004)
Non-Academic Index	2.40	2.32	.083**
	[.847]	[.803]	(.037)
Observations	1305	833	

Notes: The present table shows the means observed for the school applications at the student level, considering only schools that are exclusively free for low-SES students, separately by the socioeconomic status of students. Only the schools to which the students applied are included (irrespective of the ranking). Columns (1) and (2) contain the means observed for high-SES students and low-SES students, respectively. Column (3) present the difference between both group and its statistical significance, determined by a T-test.

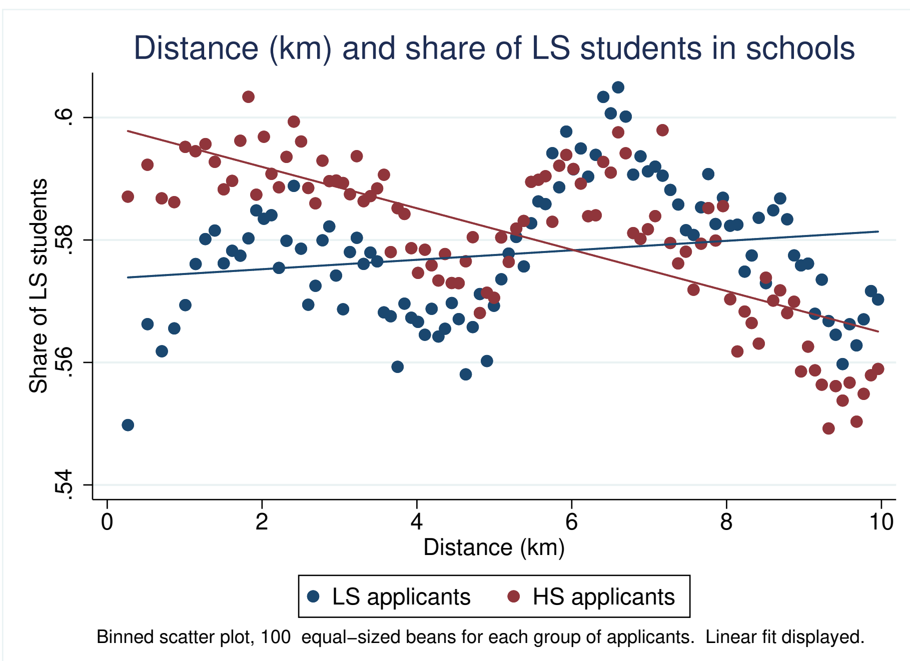

Figure 2: Correlation between distance (km) and share of low-SES students in schools

Table 16: Sample means of independent variables of the preference equation

	(1)
Avg. Test score (std)	.179
	[.956]
Annual Tuition fees (USD 2018)	179.7
	[427.9]
Distance (km)	5.89
	[3.24]
EDD	.508
	[.213]
Non-Academic Index	2.12
	[.963]
Observations	604954

Notes: This table presents the sample means and standard deviations of the variables included in the main estimating equation (Eq. 2). Standard deviations in brackets. Avg. Test score (std) and Non-Academic Index are computed by taking the mean across all schools included in our sample. Annual tuition fees, Distance, and EDD are calculated by considering all values for each student-rbd pair included in the choice set of each student.

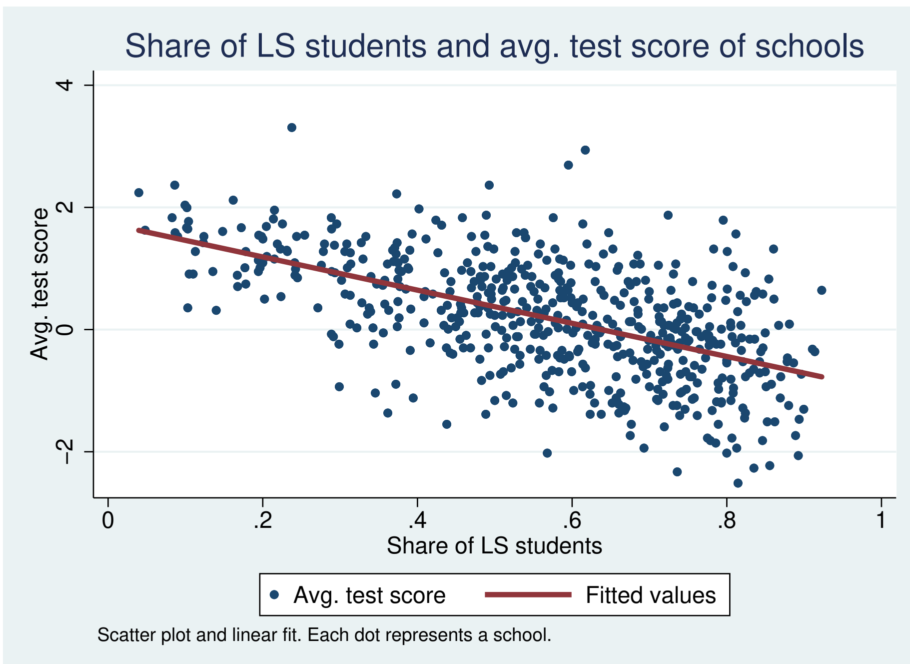

Figure 3: Correlation between share of low-SES students and avg. test score of schools

B Counterfactual design

Table 17: School preferences under the Baseline scenario and each counterfactual scenario

	Baseline scenario	Counterfactual 1 (No social frictions)	Counterfactual 2 (No spatial frictions)	Counterfactual 3 (No economic frictions)
Observable utility	$V^b_{ij} = \gamma_1 D_{ij} + \gamma_2 EDD_{ij} + \beta_1 Score_j + \beta_2 TuitionFees_{ij} + \beta 3 NonAcademic_j + \delta low-SES_i \times (D_{ij} + EDD_{ij} + Score_j + TuitionFees_{ij} + NonAcademic_j)$	$V^1_{ij} = \gamma_1 D_{ij} + \beta_1 Score_j + \beta_2 TuitionFees_{ij} + \beta 3 NonAcademic_j + \delta low-SES_i \times (D_{ij} + Score_j + TuitionFees_{ij} + NonAcademic_j)$	$V^2_{ij} = \gamma_2 EDD_{ij} + \beta_1 Score_j + \beta_2 TuitionFees_{ij} + \beta 3 NonAcademic_j + \delta low-SES_i \times (EDD_{ij} + Score_j + TuitionFees_{ij} + NonAcademic_j)$	$V^3_{ij} = \gamma_1 D_{ij} + \gamma_2 EDD_{ij} + \beta_1 Score_j + \beta 3 NonAcademic_j + \delta low-SES_i \times (D_{ij} + EDD_{ij} + Score_j + NonAcademic_j)$
Total utility	$V^b_{ij} + e_{ij}$	$V^1_{ij} + e_{ij}$	$V^2_{ij} + e_{ij}$	$V^3_{ij} + e_{ij}$

Notes: This table presents the Observable-utility equation employed for Baseline simulation (Equation 2) and for each counterfactual scenario, as well as the total utility that determines the school preferences of applicants in each case. In the Baseline scenario, preferences are computed, including all relevant variables considered in Equation 2. In Counterfactual 1, social frictions (EDD) are omitted. In Counterfactual 2 spatial frictions (Distance) are omitted, and Counterfactual 3 omits economic frictions.

C Simulated applications under each counterfactual scenario

Table 18: Counterfactual simulations, summary statistics of applications

	No EDD		No Distance		No economic	
	high-SES	low-SES	high-SES	low-SES	high-SES	low-SES
Distance (km)	3.41	3.24	5.73	5.70	4.59	4.92
Euclidean demo. dist.	.540	.412	.379	.509	.170	.827
Test score	.31	.07	.77	.37	1.2	1.1
Tuition fees (USD)	247.7	46.3	589.6	114.1	1500.4	1397.7
Paid schools	.139	.055	.361	.120	.762	.864
Free schools (for all)	.724	.787	.443	.631	.054	.108
Schools exclusively free for low-SES	.137	.158	.196	.249	.184	.028
Non-academic Index	2.16	2.10	2.19	2.14	1.70	1.57

Notes: This table presents the observed means of multiple variables for the simulated applications of each counterfactual scenario, separately by the socioeconomic status of students. Only the schools to which the student applied are considered (irrespective of the ranking). Column 1 and 2 (No EDD) refers to the counterfactual scenario where spatial frictions are omitted. Column 3 and 4 (No Distance) refers to the one that omits spatial frictions. Finally, columns 5 and 6 (No economic) considers the simulated applications of the counterfactual scenario that sets economic friction to zero. Each mean considers all of the schools included in the application of each student in each of the 250 simulations of the corresponding counterfactual scenario.

Table 19: SES school segregation using first choice

	(1)	(2)	(3)	(4)
	Baseline	No social	No spatial	No economic
D-Index	.602	.513	.607	.644
Change (p.p)		-8.9***	0.5***	4.2***

Notes: This table reports the Dissimilarity Index (row 1) for Low-SES students and High-SES students considering the first choice of applicants. Col.1 reports the D-Index for the Baseline simulation, which takes all the coefficients estimated in Table 7 to simulate students' applications. Col.2 sets the coefficients on EDD and $Low-SES \times EDD$ to zero. Col. 3 sets the coefficients on $Logdistance$ and $Low-SES \times Logdistance$ to zero. Col.4 sets the coefficient on $tuitionfees$ and $Low-SES \times tuitionfees$ to zero.