Peer Influence and College Major Choices in Male-Dominated Fields

Rocío Valdebenito*

Department of Agricultural and Consumer Economics
University of Illinois at Urbana-Champaign

[Link to latest version]

October 18, 2023

Abstract

This paper investigates the causal effect of high school peers' choices on individuals' college major choices and explores whether the gender of the individuals and their peers is a relevant mediating factor on these effects. The empirical approach employs a regression discontinuity design, using student-level data from 2005 to 2019 from the centralized admission system of Chilean universities, where admission cutoffs are unpredictable and defined each year for each program-university combination. This paper analyzes peers exposed to different fields of study, such as comparing individuals whose peers were enrolled in male-dominated fields with a counterfactual option in a female-dominated one. The main finding shows a positive and significant impact of having a peer enrolled in a technology and engineering program on an applicant's enrollment and application to the same field only when the counterfactual alternative of the peers' admission is classified in the field of humanities, social sciences, and education. The results suggest substantial heterogeneity by gender. While male students exhibit a higher inclination towards technology and engineering fields when influenced by peers who have opted for male-dominated disciplines, female students are less likely to pursue that field when exposed to such peers.

Keywords: Gender, Higher Education, Choice.

JEL: J16, I23, J24.

*E-mail address: riv2@illinois.edu

1 Introduction

One of the most significant transformations in developed and emerging economies is the increased educational attainment of women over the past few decades. These changes have had an impact on women's labor force participation and have partially narrowed the gender wage gap. However, despite improvements in college attendance rates, notable gender differences persist in choosing college majors. Recent empirical evidence highlights that these gender differences in majors explain a substantial portion of the gender wage gap, as jobs in male-dominated fields tend to offer higher salaries compared to those in female-dominated fields (Altonji et al., 2012; Sloane et al., 2021). Moreover, several studies suggest that while workers on STEM (Science, Technology, Engineering, and Mathematics) play a crucial role in driving economic growth, gender diversity within these fields can enhance innovation and overall performance (Bayer & Rouse, 2016; Peri et al., 2015). Despite some progress in reducing the gender gap in male-dominated majors over time, a persistent and significant gap continues to hinder labor market equality (Goldin et al., 2006).

College enrollment and major choices represent one of the most important decisions after high school graduation, with important implications for labor market outcomes. Unfortunately, it is difficult to understand what determines majors and field choices. Several studies have highlighted the significance of peers, individuals from classrooms, high schools, neighborhoods, and families, as they constitute a central aspect of teenagers' social environment (Aguirre & Matta, 2021; Altmejd, 2022; Anelli & Peri, 2019; Barrios-Fernández, 2022; Brenøe & Zölitz, 2020; Dustan, 2018; Lavy & Schlosser, 2011; Li, 2018). These studies, among others, have established that peers play a vital role in shaping individuals' decisions through various mechanisms, including social influence, role modeling, and aspirational effects. Recent experimental evidence further emphasizes the importance of role models for young female students. Given the gender imbalances in college majors, it becomes challenging for female students to interact with peers pursuing male-dominated fields (Porter & Serra, 2020).

This paper investigates the causal relationship between high school peers and individuals' choices of majors that are traditionally overrepresented by men. Particularly, it uses administrative student-level data from 2005 to 2019 from the applications and admissions to Chilean universities.

Chile provides an interesting educational setting with large gender gaps in college majors at enrollment and graduation. In this setting, students apply to specific majors (programs)-university combinations, a distinct feature in which students do not choose majors later in the program. The transition from high school to university is based on a centralized admission system in which students cannot manipulate admission cutoffs, and the system awards admission only based on test scores and measures of high school academic performance. Taking advantage of the selection process and employing a regression discontinuity design approach, this paper aims to provide causal evidence on the effect of high school peers. Specifically, this paper aims to answer the following questions: Does having an older peer enrolled in a male-dominated field affect applicants' choices? And, are these peer effects mediated by the gender of the student and his/her peers? In this context, high school peers are defined as students in the same high school as potential applicants, that are expose to their college enrollment decisions one year ahead of the applicants.

The extensive body of literature on peer effects in education has primarily concentrated on the analysis of the role of peers within various educational contexts and across different educational levels, namely elementary, secondary, and post-secondary education, with a focus on diverse educational outcomes. More than 30 years of research have tried to address the empirical challenges associated with identifying peer effects. However, there exists a broad consensus that peers play a crucial role in shaping students' outcomes and choices (Epple & Romano, 2011; Sacerdote, 2011).

This paper aims to contribute to three important gaps from this literature. The first gap corresponds to analyzing the transition from secondary to post-secondary education. Previous papers have explored, for instance, the influence of high-ability peers on individual test scores, or the impact of exposure to predominantly female or male cohorts on performance and major choice at the college or high school level (Briole, 2021; Calkins et al., 2023; Landaud et al., 2020; Mouganie & Wang, 2020). Utilizing the educational context and data from Chile, this paper examines the influence of peers at the secondary level on subsequent post-secondary education choices, encompassing program and field selections made by students during their application and enrollment processes.

A second gap involves the estimation of endogenous peer effects, which focuses on the impact of peers' outcomes rather than their background, commonly referred to as contextual effects. Within this body of literature, a common challenge relates to the independent identification of both these effects, which causes significant econometric difficulties, and only few studies have addressed this issue empirically (Aguirre & Matta, 2021; Barrios-Fernández, 2022; De Giorgi et al., 2010). This paper aims to contribute to the estimation of endogenous peer effects by investigating a quasi-random shock that exclusively impacts peers' major choices, as opposed to their test scores, parents' education, or any other background variable that could be correlated with major selection and college enrollment.

Finally, this paper aims to contribute to a third gap related to understanding peer effects in the context of major choices, particularly on fields traditionally considered male-oriented. Previous empirical research has identified peers' gender and ability as important variables, but there is limited knowledge about the impact of peers' choices in specific fields of study on individuals' decision-making processes. Furthermore, the studies that have analyzed the effects on major choices typically concentrate on peer effects within the same educational level, usually secondary or post-secondary levels.

The primary findings of this study indicate that, on average, applicants tend to apply and enroll in technology and engineering fields when they are exposed to high school peers who have also made that choice, as opposed to peers pursuing female-dominated fields. Nevertheless, significant gender differences emerge among applicants. Specifically, male students exhibit a higher inclination towards technology and engineering fields when influenced by peers who have opted for male-dominated disciplines, rather than female-dominated ones. In contrast, female students display a reduced propensity to pursue technology and engineering when exposed to such peers. These findings could align with recent empirical research, which suggests that competitive environments with high-achieving peers may discourage female students from pursuing scientific fields (Brenøe & Zölitz, 2020; Fischer, 2017; Landaud et al., 2020) as well as the literature indicating that the gender gap in major choice stems from highly heterogeneous preferences (Bordón et al., 2020; Calkins et al., 2023). Although this study does not investigate the underlying mechanisms of the results, it is worth noting that exposure to peers who have enrolled in male-dominated fields, known for their high competitiveness, could potentially contribute to the reluctance to pursue such majors.

This paper is organized as follows. Section 2 explains the main challenges in the peer effects literature, Section 3 briefly describes the Chilean institutional context, Section 4 presents the steps require in the sample construction, Section 5 presents the empirical strategy while Section 6 exhibits the empirical tests used to validate the main assumptions from the empirical strategy. Finally, Sections 7 and 8 present the results and final remarks.

2 Peer effects in education

A growing body of literature has focused on explaining the gender gap in higher education choices. From score gaps in math and science to gender stereotypes and role models, there are multiple factors that help explain these gaps. Recent literature suggests that, among other factors, high-school subject choices and achievement at earlier school stages are relevant for explaining gender gaps (Card & Payne, 2021). However, differences in ability, such as math achievement, do not fully explain these gaps (Cimpian et al., 2020; Riegle-Crumb et al., 2012), and beliefs that men are naturally more skilled in quantitative domains are empirically unfounded (Favara, 2012).

Peers, on the other hand, represent a central aspect of teenagers' social environment as they can impact a variety of important outcomes, including test scores, educational attainment, and trajectories (Balestra et al., 2021; Black et al., 2013; Carrell et al., 2018).[1] A key goal in this literature is to learn how the composition of peer groups influences different educational outcomes. However, several identification challenges exist in the peer effects literature, as noted by the contributions of Manski (1993) and Moffitt (2001).

Naturally, peer selection is not random, and correlated outcomes can be confounded by the factors that explained the group formation in the first place. A second aspect is that peer effects can be driven by different social interactions: contextual (or exogenous) effects, corresponding to changes in an individual's behavior due to peers' characteristics, endogenous effects that are changes in the individual's behavior due to the prevalence of that behavior among peers, and correlated effects that correspond to similarity of outcomes between peers due to having similar individual characteristics or experiencing similar shocks or institutional environment. Under the presence of peers effects, the challenge is to distinguish among exogenous, endogenous or correlated

[1] See Epple & Romano (2011) and Sacerdote (2011) for an extensive review.

effects. Finding any effect can be signal of, for example, the "birds of a feather flock together" phenomenon rather than any actual peer effect. In addition, if outcomes of peers simultaneously affect each other, then it becomes even harder to separate contextual and endogenous effects, which is the so-called reflection problem (Manski, 1993).

The literature on peer effects has tried different empirical approaches to develop credible identification strategies. One of these approaches is to estimate peer effects by leveraging the random assignment of students at the classroom level (Anelli & Peri, 2019; Duflo et al., 2011; Goulas et al., 2022) or roommates and classrooms at the college level (Elsner et al., 2021; Sacerdote, 2001; Zimmerman, 2003). Another approach includes Brenøe & Zölitz (2020) that explores the natural within-school variation in peer composition and peer characteristics across time. Several findings can be retrieved from these empirical approaches, for example, Anelli & Peri (2019) finds that male students exposed to high school cohorts composed of more than 80% male peers are more likely to choose male-dominated majors in Italy. Brenøe & Zölitz (2020), using Danish data, finds evidence that having a large proportion of female peers in class decreases the likelihood of enrolling in STEM fields. Another concept that arises in this literature is the impact of the "peer quality," measured for example, by either peers' performance on standardized tests or grades (Balestra et al., 2021; Card & Payne, 2021; Mouganie & Wang, 2020). The evidence on this matter is inconclusive, and two recent papers show somewhat different results. While Balestra et al. (2021) show that the presence of peers with high intellectual ability affects the likelihood of selecting STEM occupations for men only, Mouganie & Wang (2020) find that high-performing female peers in math increase the likelihood that women choose STEM tracks. Despite the useful conclusions from these analyses that inform us about the actual presence of peer effects, the identification strategy from these approaches is unable to separate contextual and endogenous effects.

A different empirical approach explores the variations generated by random shocks that affect peers' outcomes. A recent example of this approach can be found in Barrios-Fernández (2022), in which the author explores the quasi-random experiment generated by loan eligibility in the Chilean centralized admission system to analyze the influence of neighbors' enrollment on potential applicants' college choices. Other examples include Altmejd et al. (2021) and Aguirre & Matta (2021), where the authors explore the admission score cutoffs to identify the effects of older siblings' trajectories on younger siblings' college choices. All of these studies have the empirical advantage to independently identify endogenous peer effects by isolating them from correlated effects. As Section 5 explains, the identification strategy in this paper is closely aligned with the papers above.

3 Chilean institutional context

The Chilean post-secondary education system comprises 60 universities that offer bachelor's degrees, and 43 that participate in a centralized admission system. Universities that do not participate in this admission system are predominantly private and typically serve lower-scoring students.[2] The participating universities are all non-profit and can be public, private, or private-parochial. Although the institutions of the centralized admission system span a wide range of selectivity levels, it also includes the country's most prestigious and traditional universities.

During their senior year of high school, students sign up to take a series of standardized tests to apply to any of the academic programs offered in the centralized admission. The series of tests, called PSU (*Prueba de Selección Universitaria*, in Spanish) consists of two mandatory tests, mathematics and language, and one of two optional tests, science and history. Besides PSU scores, the students' performance measures of high school GPA and GPA ranking,[3] are the only other components of the weighted average score considered in the system. Each program-institution has specific weights that apply to each component of the weighted score, and the information about such weights is public and available to students before sending their applications.

After being informed of their scores, students submit a list from most to least preferred of up to 10 program-institution combinations, referred to in this paper as choices or alternatives.[4] After receiving the application list from the students alongside with the specific weighted scores computed for each program-institution alternative,[5] the system's algorithm implements a *deferred acceptance*

[2]The centralized admission system is called *Sistema Único de Admisíon*, or Unified System of Admission.

[3]Starting in the admission year of 2012, the GPA ranking is an average measure of relative performance–in terms of GPA, with respect the current and previous cohorts.

[4]An alternative can be, for example, Civil Engineering at the University of Chile. If the student is also interested in Economics at the same institution, she could include that as a second-best alternative in her preference list.

[5]For example, Civil Engineering at the University of Chile in 2021 assigned 10% to the GPA score, 20% to GPA ranking, 10% to language PSU score, 45% to math PSU score, and 15% to science PSU score. This particular program requires that applicants have to take the science test instead of the history test. Other

assignment mechanism, to determine which students are offered admission to each program.[6]

Chile is similar to many other countries with trends indicating that a greater number of female students have been enrolling in tertiary education compared to their male counterparts. Figure 1 depicts the trend where, starting from 2016, more women than men are participating in university enrollments through the centralized admission system. However, striking differences occur when analyzing patterns at the field level. Figure 2 shows that the lack of female graduates in the fields such as "Engineering, Manufacturing, and Construction", and "Information and Communication Technologies" is an issue in most countries, including Chile. The OECD average share of female graduates in these two fields are around 7% and 2%, respectively. In Chile, these shares reach 7.6% and 0.7%, respectively (OECD, 2017).

Enrollment data at universities from the centralized admission system in Chile show significant gender gaps by fields. Figure 3 shows the average freshmen enrollment by men and women across different fields of study. On average, there are almost 10,000 more men than women enrolled in the field of "Technology and Engineering" (TE) per year, and women represent around 25% of the freshmen enrollment in such field. In contrast, the fields of "Humanities, Social Sciences, Arts, and Education" (HASSE), and "Health" show opposite patterns, with almost 5,000 more women than men per year enrolled in such fields, respectively.

Enrollment patterns indicate that fields dominated by males have experienced an even greater male dominance compared to the past, while the same trend is observed in female-dominated fields. Figure 4 illustrates the annual evolution of the total freshmen enrollment by gender in four of the fields of study presented in Figure 3, specifically the ones with the highest and lowest difference between male and female enrollment. The most male-dominated field–TE–shows that the gender gap has increased over time. In contrast, the field of "Business" exhibits a gender-balance pattern without significant changes across time. Regarding the female-dominated fields, both HASSE and "Health" have shown an increase in the gap between female and male enrollment over time.

programs, in contrast, allow students to decide which test they want to take, and therefore the highest score between these two components is the one used to calculate the weighted score. At the end, regardless of the program, only one test (either science or history) is used in the final calculation of the weighted score.

[6]See Rios et al. (2021) for further details on the admission system.

4 Data and sample construction

This paper focuses on the transition from high school to universities over the period 2005–2019, using individual-level data where the information has been previously anonymized with a unique student *id* that allows the identification of educational trajectories. The main two sources of information are the Department of Evaluation, Measurement, and Educational Registry (DEMRE, in Spanish) and the Ministry of Education (*Centro de Estudios Mineduc*). DEMRE is the agency in charge of the standardized tests and the entire process of the centralized admission system to the universities from the Council of Rectors of Chilean Universities. The datasets provided by DEMRE are the scores of each of the subjects included in the college admission test, the ranking of preferences submitted by the applicants, self-reported socioeconomic information, admissions offered from the centralized system, and enrollment at the universities involved in the centralized system. From the Ministry of Education, the main datasets are student enrollment records for secondary education, schools' *id*, and academic performance with consolidated GPAs at the end of each academic year.[7]

The intuition behind the sample construction of peers data is based on a regression discontinuity design (RDD), where the main objective is to obtain from the application data, an exogenous variation around the cutoff that affects peers' admission to a male-dominated field. Peers and potential applicants are students who attended the same high school, but peers were exposed to the choice of enrollment one year before the potential applicants.

The construction of the peers data follows three important steps. The identification of the undersubscribed programs, the elimination of dominated alternatives, and the construction of target and counterfactual alternatives.[8].

First, undersubscribed programs are programs that could not fill completely its seats, and therefore cutoffs cannot be identified. I define the admission cutoff c_{jt} to a program-institution j at a given year t, as the minimum weighted score among students who were offered admission, in

[7]See Online Appendix A.1 for further details.

[8]See Online Appendix A.2 for further details and examples on the construction of the peers data.

programs with at least one not admitted applicant:

$$c_{jt} = \min\{s_{ijt}\} \quad \text{s.t. } i \text{ is offered admission to } j \text{ in the admission year t,} \tag{1}$$

where s_{ijt} is the average weighted score of applicant i obtained when applying to the alternative j at year t, calculated as:

$$s_{ijt} = \sum_{l} s_i^l \alpha_{jt}^l, \tag{2}$$

where α_{jt}^l is the weight that the program-institution (j) assigns to component l in the academic year t, and s_i^l is the score student i obtained in the specific component l–math, language, history/science or GPA scores.

The second aspect of data construction is the elimination of dominated alternatives (Abdulkadiroğlu et al., 2014; Aguirre & Matta, 2021; Aguirre et al., 2022). A dominated alternative happens when an applicant submits a highly selective alternative (i.e., relatively higher cutoffs) in a lower-ranked position. For example, assume the case of a student who ranks in the first place a program j with very low selectivity, followed by a program k with very high selectivity. If the student is above the cutoff of k, and by consequence, he is also above the cutoff of j, he would be admitted to program j because it is a preferred choice. So, being above the cutoff of k would have no effect on the assignment to k. Thus, in this stage I identify and eliminate dominated alternatives from the sample, because keeping dominated alternatives in the data reduces the statistical power of the first stage. In other words, for a given applicant, the resulting sample after this cleaning procedure contains ordered preferences in which any lower ranked choices are alternatives where the applicant could in fact be admitted, if she is below the cutoff in a higher-ranked choice.

An important aspect of the Chilean centralized admission system is that weights of each component of the final weighted score and cutoffs are program specific. This feature adds another layer of complexity to the admission system in which simply comparing programs' cutoffs is not enough to define high/low selectivity programs.[9] Aguirre & Matta (2021) and Aguirre et al. (2022)

[9] Note that in the specific case when two programs assign identical components' weights, then the simple comparison between their cutoffs is enough to define which program is more selective.

explain the concept of *relative selectivity*, which helps to identify when a lower-ranked program is relatively more selective than a higher-ranked program, from the applicant's perspective.[10] If that is true, as explained before, the relatively more selective program would not survive the elimination procedure.[11]

The third step consists of merging fields of study classifications to the program-institution alternatives. Since each preference consists of a specific program-institution, it is possible that students apply to the same field across their preference list. In this step, I collapse consecutive alternatives classified in the same field. For example, if an applicant submits five consecutive preferences in the same field,[12] then I keep the alternative in which the applicant is closer to the cutoff.

The fourth step in the data construction consists of creating observation pairs of a preferred field (j) and a counterfactual or also called fall-back (k) field. The main objective in this step is constructing a pair in which the counterfactual alternative serves as a plausible scenario in the case the student is not admitted in his target choice. For example, if a pair (j, k) is $j =$TE, and $k =$Health, then two outcomes can be observed, first, the student is above the cutoff of his preferred choice (j) and therefore he is admitted, and second, he is below the cutoff of his preferred choice and therefore, given the sample construction and cleaning, he would be admitted in his counterfactual alternative, Health. In other words, the counterfactual alternative is what would have happened if the student is below the cutoff of his preferred choice, and importantly, in a given neighborhood around the target's cutoff, the sample construction is able to "randomize" admission to Technology and Engineering, and Health. [13]

As expected, students have the option to apply to a wide variety of programs and fields, resulting in the existence of multiple field combinations to explore. The aim of this study is to investigate

[10] Following Aguirre & Matta (2021) and Aguirre et al. (2022), *Relative Selectivity* is calculated as $\phi_{ij} = \frac{s_{ij} - c_j}{\sqrt{\sum_l (\alpha_j^l)^2}}$, which represents the euclidean distance from applicant's i scores (components) to the admission frontier defined by the cutoff at program j.

[11] See Table A.1 with the results of the elimination procedure and the resulting number of observations after each iteration. According to this table, around 59% of the observations survived the elimination procedure.

[12] Applying to engineering programs at different universities or different engineering programs at the same university, or a combination of both.

[13] See Tables A.2 and A.3 for an example of a preference list and the resulting construction of target and counterfactual pairs.

whether having a peer enrolled in a male-dominated field influences the likelihood of potential applicants choosing the same field of study. However, the counterfactual alternative can encompass various possibilities. In this paper, three main samples are used, all with the same target field, Technology and Engineering,[14] Specifically, I consider two female-dominated fields (HASSE and Health)[15] and a gender-balanced field, Business. [16]

Finally, I connect peers and potential applicants. Potential applicants consist of the whole universe of high school graduates between 2005 and 2019, while peers data consists of the subsample of peers at the admission margins of j and k, that survived the elimination procedure, over the period 2004–2018. The merging process is performed using the same school id between potential applicants and peers, but combining lagged cohorts of peers being one year ahead of the applicants.[17] In other words, I perform a merging process by conditioning students at the same high school, where peers' admission process is observed at $t-1$, and potential applicants' admission process is observed at time t.[18] Thus, the merging process includes only one observation per potential applicant, and in a given year, all students from the same high school, are connected with only one peer at the admission margin of interest.

Table 1 shows descriptive statistics of the analytical sample for the applicants and their peers, over the period 2005–2019. There are in total 321,966 potential applicants with an older peer at the admission margin to either TE or HASS, that represents in total 2,945 unique peers. Given that peers are students who self-select themselves to participate in the centralized admission system after observing their scores,[19] it is expected that peers present higher PSU scores and GPA than potential applicants. Moreover, in this sample, peers are students who have at least one alternative submitted to a TE program; it is thus also expected that the proportion of female students is lower

[14]Examples of programs classified as TE include Civil Engineering, Engineering in Informatics and Computing, Civil Engineering in Construction, among others.

[15]Examples of programs classified as HASSE include Psychology, Sociology, Pedagogy, Graphic Design, Humanities, Linguistics, Philosophy, and Social Work, while examples of programs classified as Health include Nursing, Medicine, Kinesiology, Medical Technology, Nutrition, and Odontology.

[16]Table A.4 presents all possible combinations in the data and the frequency of each combination. As well as the main margins of interest.

[17]See Figure A.1 for an example of how peers and applicants are connected in a given school.

[18]In order to avoid duplicate potential applicants, if there are in the same high school multiple peers on the margin of applying to j and k that survived the elimination procedure, then I only keep the peer that is closer to the cutoff of j.

[19]Students with extremely low-score are less likely to submit their applications because weights and the previous year's cutoffs are public information; thus, they can expect that admission would not be awarded.

among the peers than among the potential applicants. As Table 1 indicates, the share of female students is 40% versus 52% at the sample of peers and potential applicants, respectively. Finally, peers also present better socio-economic status, as their parents' education and family income are higher than the potential applicants' parents' status.

5 Identification strategy

Peers' college choices are not random and could be affected by many confounding factors, but each admission cutoff can be used as a separate natural experiment that provides the exogenous variation needed to estimate the impacts of the peers' college admission. As explained in Section 4, the peers' sample consists on a subsample of students that are in the margins of admission to a TE program, and different fall-back programs in the fields of HASSE, Health, or Business. The empirical strategy in this paper employs a regression discontinuity (RD) approach that leverages the centralized admission system's unpredictable cutoffs, for which a subset of applicants, the cutoff effectively randomizes admission offers to a male-dominated field.

For each potential applicant i, I identify an older peer p in the same school (s) as i that applies to universities in the admission year $t-1$, to an alternative j and a counterfactual option k. Thus, I estimate a Fuzzy RD design that captures the effect of having a marginal peer admitted in a TE program. I start by defining the running variable for all peers p with a preferred and counterfactual fields (j, k), such that j is TE and k is HASSE, Business or Health:

$$r_{pjt} = s_{pjt} - c_{jt}, \tag{3}$$

which measures the distance between the peer's weighted score applying to field j and its cutoff. If $r_{pjt} \geq 0$, then the peer is admitted in his preferred field j–TE, and if $r_{pjt} < 0$, then he is admitted to his counterfactual alternative.

In this setting, an indicator for being above the cutoff of a TE program is used as an instrument for the actual enrollment of the peer in the same field. Therefore, the first stage is represented as:

$$\text{Peer enrolls in TE}_{pj,t-1} = \pi_1 Z_{pj,t-1} + h(r_{pj,t-1}) + \beta_s + \psi_t + \gamma_j + \nu_{ipjt}, \tag{4}$$

where $Z_{pj,t-1}$ is an indicator variable that captures whether the peer p crossed the admission cutoff of the TE program, then $Z_{pj,t-1} = \mathbb{1}[r_{pj,t-1} \geq 0]$. The outcome variable from this stage, Peer enrolls in $\text{TE}_{pj,t-1}$, is a binary indicator that equals one when the peer is enrolled in TE and zero otherwise. The function $h(\cdot)$ represents a polynomial of the running variable $r_{pj,t-1}$, that it can be a first-order or higher-order function. The terms β_s, ψ_t, and γ_j, are school and year fixed effects, respectively. And ν_{ipjt} is an error term.

The second-stage on this procedure includes, as a regressor, the predicted outcome of Equation (4). Equation (5) represents the second-stage as follows,

$$\text{Applicant chooses TE}_{ipjt} = \tau \text{Peer enrolls in TE}_{pj,t-1} + h(r_{pj,t-1}) + \mu_s + \alpha_t + \delta_j + \varepsilon_{ipjt}, \tag{5}$$

the outcome variable **Applicant chooses TE**$_{ipjt}$ captures whether the applicants follows, either by enrollment or application, the field of the peer that the applicant is exposed to at high school. In practice, the primary outcomes analyzed are three, first "applicant's first choice is TE", which a binary variable that takes value one when the potential applicant submits as their/his/her first choice a program in a TE field. The variable takes the value of zero if either the application was made to another field or there was no application at all. Second, "applicant has TE as any choice", that is binary variable that takes the value one when the potential applicant submits in any of her preference list alternatives a program in a TE field. The variable takes the value of zero if either the application was made to another field or there was no application at all. And finally, the variable "enrollment in TE", that is a binary outcome that takes the value one when the potential applicant enrolls in a TE program. The variable takes the value of zero if no enrollment is observed, or the applicant enrolled in a different field.

The parameter of interest τ, recovers the effect of peer's enrollment into a TE program on potential applicants' choices in the same field. Similarly to the first stage, μ_s, α_t, and δ_j are terms capturing school, admission year, and preferred alternative fixed effects, respectively. And $h(\cdot)$ represents the polynomial function of the running variable and ε_{ipjt} an error term.

I estimate Equation (5) using the RD robust approach proposed by Calonico et al. (2014a,b), which is a non-parametric approach for RD design that does not impose strong assumptions on the shape of the relationship between the running variable and the outcome. In this draft, I present local linear polynomial estimation to both sides of the threshold using triangular kernels and optimal bandwidths selected by the *rdrobust* package (Calonico et al., 2020), that are chosen to minimize the mean squared error.

6 Validation of RD assumptions

RD designs require that students whose weighted score is near the threshold are comparable in terms of observable and unobservable characteristics regardless of their actual admission status, and that the treatment assignment is not manipulable (Lee & Lemieux, 2010). This section aims to explore empirically these assumptions. A first falsification test explores the manipulation of the running variable, which in practice translates into testing whether the number of observations below the cutoff is substantially different from the number of observations above the cutoff. Figure 5 presents the density test suggested by Cattaneo et al. (2020), where the robust bias-corrected test of the null hypothesis of "no manipulation" shows a *p-value* of 0.7206, providing evidence in favor of the continuity of the running variable. Aside from the graphical and statistical evidence, as explained in section 3, the national entry exam is part of a complete centralized admission system, in which students are unable to manipulate programs' cutoff and thus, manipulation of the running variable is an unlikely scenario.

A second test examines whether, around the cutoff, treated and control individuals are similar in terms of baseline observable characteristics. If baseline covariates, that are expected to be correlated with the outcome, are not continuous at the cutoff, the continuity assumption of the potential outcome functions are likely to fail. Figure 6 shows the 95% confidence intervals of the treatment effect[20] on a set of socioeconomic and individual characteristics, such as parents' education, family income, gender, and GPA. Figures 6a and 6b show the continuity test for potential applicants and peers, respectively. As it is shown, all estimates are not statistically significant, providing evidence

[20] The coefficient associated with the indicator variable that captures whether the running variable is above zero.

that baseline covariates are continuous at the cutoff.

Finally, an additional assumption applicable to Fuzzy RD designs, is that the threshold-crossing indicator is a good instrument of the treatment assignment. A visual representation of that relationship is shown in Figure 7, that exhibits the first stage presented in Equation 4. The discontinuity of the peers' enrollment in TE at the cutoff supports the strong association between being above the admission cutoff and actual enrollment of the peers.

7 Results

The study analyzes three primary outcomes:"application in the first choice to TE", "application to TE in any choice", and "enrollment in TE". Table 2 presents the average treatment effect of having a peer enrolled in TE on these primary outcomes, along with the first stage estimates that indicate the effects of a peer crossing the admission threshold on their actual enrollment.

Across all panels in Table 2, the first stage estimates reveal a strong positive relationship between crossing the admission cutoff and the peer's enrollment, which aligns with the visual representation in Figure 7. Table 2 is divided in three panels, each one corresponds to an specific fall-back field. The RD estimates in panel A, show that on average there is positive and statistically significant effect of having a high-school peer that one year before the applicant was enrolled in a TE field, relative to being admitted in a HASSE program. The results from Table 2 indicate that having a peer enrolled in TE increases the probability of choosing a TE program as the first choice by 3.3 percentage points (26%), listing a TE program in any choice by 4.2 percentage points (21%), and enrolling in a TE program by 2.5 percentage points (25%). panel B shows that when the sample randomizes between peers admitted in TE versus Health, peers' enrollment in TE does not have a positive impact on applicants' TE application or enrollment. In fact, it makes applicants less likely to apply and enroll in TE programs by approximately 14 to 18 percentage points. Panel C focuses on Business as the fallback alternative. Although there is a slightly significant and positive effect on the application to TE in the second column, there is close to zero and no significant effect of peers' TE enrollment on applicants' admission and enrollment in the same field.

Table 3 analyzes the potential role of the applicant's gender in the transmission of peer effects.

The table consists of two columns, with the first column representing male potential applicants and the second column representing female potential applicants. The purpose is to examine whether gender influences the impact of peers on the outcomes. The results from Panel A reveal that, on average, the effects of having a peer enrolled in TE, compared to HASSE, on female applicants' submission and enrollment are small and not statistically significant. In contrast, male potential applicants demonstrate a higher likelihood of applying and enrolling in TE programs when they are exposed to peers enrolled in TE. All the estimates for male applicants are positive and statistically significant at the 95% confidence level. For instance, when exposed to a peer enrolled in TE, relative to HASSE, male potential applicants are 8.5 percentage points more likely to enroll in TE. These findings suggest that male potential applicants primarily capture the positive effect observed in Panel A of Table 2.

Table 3, Panel B, presents the results when the fall-back option is Health. These findings indicate that male applicants are not affected by peers who are at the margin of enrollment in TE versus Health. The likelihood of enrolling in TE remains unchanged for male applicants. In contrast to the results in Panel A, it is the female potential applicants who capture the overall negative effect presented in Panel B of Table 2. This result suggests that when a female potential applicant is exposed to a peer who is at the margins of TE and Health, their probability of enrolling in a TE program decreases. It would be interesting to further analyze whether, for example, female applicants are more likely to apply to Health programs. Moving to Panel C of Table 3, the results are presented by applicant's gender when the fall-back alternative is Business. On average, male applicants are more likely to apply to TE programs when their peers are at this margin. However, for both men and women, there are close to zero and non-significant estimates when it comes to TE enrollment.

Table 4 presents the results when the sample is divided by both the gender of the applicant and the peer. Panel A represents male potential applicants, while Panel B represents female potential applicants. In each panel, the first column shows the effect of exposure to a male peer, while the second column shows the effect of exposure to a female peer.

Panel A of Table 4 demonstrates that male potential applicants are positively influenced by both male and female peers. Although statistical tests should be used to analyze the statistically significant differences between these two coefficients, the effect size of male applicants being exposed to a male peer is noticeably larger than when exposed to a female peer. In the case of female potential applicants, the results show some inconsistencies when exposed to male and female peers. While it appears that male peers enrolled in TE negatively affect the likelihood of application, they also increase the likelihood of enrollment. A similar inconsistent pattern is observed in the exposure to female peers. Although these peers have a negative impact on the probability of first-choice application and enrollment in TE for female applicants, they also increase the probability of application as any choice.[21]

It is important to acknowledge that in the high school environment, students are exposed to multiple peers. The overall results presented in this paper demonstrate the impact of specific peers who, due to the institutional context, experience a random shock in their enrollment in TE. Therefore, an important factor that may mediate this context is the size of the high school cohorts. Figure 8 illustrates the coefficients by dividing the sample into terciles based on high school size, with the first tercile representing schools with the smallest cohorts. The figure reveals a U-shaped pattern in the impact of having a peer enrolled in TE relative to HASS. Negative and statistically significant effects are observed at both the smallest and largest extremes of the cohort size distribution, while a positive effect is observed for schools in the middle portion of the distribution.

8 Conclusions

This paper resides at the intersection of two significant areas within the economics of education literature: gender gaps in college major selection and peer effects. While extensive research has been conducted in these domains across various educational contexts, there are still notable limitations and opportunities for further advancements. Recent empirical studies have aimed to explore the causal impact of peers' outcomes and characteristics on college major choice. However, isolating the endogenous and exogenous effects that emerge due to the simultaneity problem, as described in the seminal works by Manski (1993) and Manski (1993), remains challenging.

[21] Results by gender of the applicant and the peer for Health and Business as fall-back alternatives are presented in the Online Appendix Tables A.5 and A.6, respectively.

An increasing body of literature on college choice and gender gaps in areas such as STEM has provided useful insights into understanding why women and men choose differently. However there still is a large knowledge gap to be filled. This paper aims to advance the current knowledge in this area by taking advantage of older peers' college application data. The Chilean centralized admission system provides a useful setting in which isolating endogenous effects is possible as the contributions of Altmejd et al. (2021), Barrios-Fernández (2022), and Aguirre et al. (2022) have noted.

This paper utilizes data from the admission system to identify samples that represent various and relevant margins of interest. Within these samples, crossing the admission threshold creates a random shock on peers' enrollment in the field of "Technology and Engineering" compared to other fields such as "Humanities, Social Sciences, Arts, Education," "Business," and "Health." In Chile, as in many other countries, there is a gender-based segregation in the choice of fields of study, with a predominantly male enrollment in Technology and Engineering and a significant female enrollment in areas like Health and Education. Therefore, the primary analysis in this manuscript aims to shed light on the importance of having a peer enrolled in Technology and Engineering compared to different counterfactual alternatives.

The main results of this study demonstrate a causal relationship between the educational paths of peers and potential applicants, who are students in the same high school and must make decisions about university applications one year after their peers. Specifically, it is found that having an older peer enrolled in a male-dominated program increases the likelihood of potential applicants choosing a male-dominated program, but only when the alternative field is "Humanities, Social Sciences, Arts, Education." On the other hand, when the counterfactual alternative for the peer is "Health," the probability of enrolling in Technology and Engineering decreases. However, it is important to note that these contrasting results are primarily influenced by the gender of the applicant. Male potential applicants are positively affected when the peer is at the Technology and Engineering versus Humanities, Social Sciences, Arts, Education (HASSE) margin, while female applicants are negatively affected when the peer is at the Technology and Engineering versus Health margin.

Furthermore, the results indicate that the gender of the peer plays a significant role in explaining these outcomes, particularly for male potential applicants. Male applicants are considerably more influenced by male peers in terms of their likelihood to apply and enroll in a Technology and Engineering program.

Future research should consider exploring other sources of heterogeneity, as focusing solely on the gender dimension from the potential applicants' perspective may not provide a comprehensive understanding. Factors such as students' abilities, socioeconomic status, and peers' admission to highly selective programs, among others, could contribute to a more comprehensive understanding of this form of social interaction among high school students.

References

Abdulkadiroğlu, A., Angrist, J., & Pathak, P. (2014). The Elite Illusion: Achievement Effects at Boston and New York Exam Schools. *Econometrica*, 82(1), 137–196.

Aguirre, J. & Matta, J. (2021). Walking in your footsteps: Sibling spillovers in higher education choices. *Economics of Education Review*, 80, 102062.

Aguirre, J., Matta, J., & Montoya, A. M. (2022). Joining the Old Boys' Club: Women's Returns to Majoring in Technology and Engineering. unpublished.

Altmejd, A. (2022). Inheritance of fields of study. unpublished.

Altmejd, A., Barrios-Fernández, A., Drlje, M., Goodman, J., Hurwitz, M., Kovac, D., Mulhern, C., Neilson, C., & Smith, J. (2021). O Brother, Where Start Thou? Sibling Spillovers on College and Major Choice in Four Countries. *The Quarterly Journal of Economics*, 136(3), 1831–1886.

Altonji, J. G., Blom, E., & Meghir, C. (2012). Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers. *Annual Review of Economics*, 4(1), 185–223.

Anelli, M. & Peri, G. (2019). The Effects of High School Peers' Gender on College Major, College Performance and Income. *The Economic Journal*, 129(618), 553–602.

Balestra, S., Sallin, A., & Wolter, S. C. (2021). High-Ability Influencers? The Heterogeneous Effects of Gifted Classmates. *Journal of Human Resources*, (pp. 0920–11170R1).

Barrios-Fernández, A. (2022). Neighbors' Effects on University Enrollment. *American Economic Journal: Applied Economics*, 14(3), 30–60.

Bayer, A. & Rouse, C. E. (2016). Diversity in the Economics Profession: A New Attack on an Old Problem. *Journal of Economic Perspectives*, 30(4), 221–242.

Black, S. E., Devereux, P. J., & Salvanes, K. G. (2013). Under Pressure? The Effect of Peers on Outcomes of Young Adults. *Journal of Labor Economics*, 31(1), 119–153. Publisher: The University of Chicago Press.

Bordón, P., Canals, C., & Mizala, A. (2020). The gender gap in college major choice in Chile. *Economics of Education Review*, 77, 102011.

Brenøe, A. A. & Zölitz, U. (2020). Exposure to More Female Peers Widens the Gender Gap in STEM Participation. *Journal of Labor Economics*, 38(4), 1009–1054.

Briole, S. (2021). Are Girls Always Good For Boys? Short and Long Term Effects of School Peers' Gender. *Economics of Education Review*, 84, 102150.

Calkins, A., Binder, A. J., Shaat, D., & Timpe, B. (2023). When Sarah Meets Lawrence: The Effects of Coeducation on Women's College Major Choices. *American Economic Journal: Applied Economics*, 15(3), 1–34.

Calonico, S., Cattaneo, M. D., & Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2), 192–210.

Calonico, S., Cattaneo, M. D., Farrell, M. H., & Titiunik, R. (2017). Rdrobust: Software for Regression-discontinuity Designs. *The Stata Journal: Promoting communications on statistics and Stata*, 17(2), 372–404.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014a). Robust Data-Driven Inference in the Regression-Discontinuity Design. *The Stata Journal*, 14(4), 909–946.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014b). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. *Econometrica*, 82(6), 2295–2326.

Card, D. & Payne, A. A. (2021). High School Choices and the Gender Gap In STEM. *Economic Inquiry*, 59(1), 9–28.

Carrell, S. E., Hoekstra, M., & Kuka, E. (2018). The Long-Run Effects of Disruptive Peers. *American Economic Review*, 108(11), 3377–3415.

Cattaneo, M. D., Idrobo, N., & Titiunik, R. (2019). A Practical Introduction to Regression Discontinuity Designs: Foundations. *Elements in Quantitative and Computational Methods for the Social Sciences*.

Cattaneo, M. D., Jansson, M., & Ma, X. (2020). Simple Local Polynomial Density Estimators. *Journal of the American Statistical Association*, 115(531), 1449–1455.

Cimpian, J. R., Kim, T. H., & McDermott, Z. T. (2020). Understanding persistent gender gaps in STEM. *Science*, 368(6497), 1317–1319.

De Giorgi, G., Pellizzari, M., & Redaelli, S. (2010). Identification of Social Interactions Through Partially Overlapping Peer Groups. *American Economic Journal: Applied Economics*, 2(2), 241–275.

Duflo, E., Dupas, P., & Kremer, M. (2011). Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya. *American Economic Review*, 101(5), 1739–1774.

Dustan, A. (2018). Family networks and school choice. *Journal of Development Economics*, 134, 372–391.

Elsner, B., Isphording, I. E., & Zölitz, U. (2021). Achievement Rank Affects Performance and Major Choices in College. *The Economic Journal*, 131(640), 3182–3206.

Epple, D. & Romano, R. E. (2011). Peer Effects in Education. In *Handbook of Social Economics*, volume 1 (pp. 1053–1163). Elsevier.

Favara, M. (2012). The Cost of Acting 'Girly': Gender Stereotypes and Educational Choices. unpublished.

Fischer, S. (2017). The Downside of Good Peers: How Classroom Composition Differentially Affects Men's and Women's STEM Persistence. *Labour Economics*, 46, 211–226.

Goldin, C., Katz, L. F., & Kuziemko, I. (2006). The Homecoming of American College Women: The Reversal of the College Gender Gap. *Journal of Economic Perspectives*, 20(4), 133–156.

Goulas, S., Griselda, S., & Megalokonomou, R. (2022). Comparative Advantage and Gender Gap in STEM. *Journal of Human Resources*.

Landaud, F., Ly, S. T., & Maurin, É. (2020). Competitive Schools and the Gender Gap in the Choice of Field of Study. *Journal of Human Resources*, 55(1), 278–308.

Lavy, V. & Schlosser, A. (2011). Mechanisms and Impacts of Gender Peer Effects at School. *American Economic Journal: Applied Economics*, 3(2), 1–33.

Lee, D. S. & Lemieux, T. (2010). Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2), 281–355.

Li, H.-H. (2018). Do mentoring, information, and nudge reduce the gender gap in economics majors? *Economics of Education Review*, 64, 165–183.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, 60(3), 531–542.

Moffitt, R. (2001). Policy Interventions, Low-Level Equilibria, and Social Interactions. In *Social Dynamics* (pp. 45–82). Booking Institutions Press.

Mouganie, P. & Wang, Y. (2020). High-Performing Peers and Female STEM Choices in School. *Journal of Labor Economics*, 38(3), 805–841.

OECD (2017). OECD Gender Data.

Peri, G., Shih, K., & Sparber, C. (2015). STEM Workers, H-1B Visas, and Productivity in US Cities. *Journal of Labor Economics*, 33(S1), S225–S255. Publisher: The University of Chicago Press.

Porter, C. & Serra, D. (2020). Gender Differences in the Choice of Major: The Importance of Female Role Models. *American Economic Journal: Applied Economics*, 12(3), 226–254.

Riegle-Crumb, C., King, B., Grodsky, E., & Muller, C. (2012). The More Things Change, the More They Stay the Same? Prior Achievement Fails to Explain Gender Inequality in Entry Into STEM College Majors Over Time. *American Educational Research Journal*, 49(6), 1048–1073.

Rios, I., Larroucau, T., Parra, G., & Cominetti, R. (2021). Improving the Chilean College Admissions System. *Operations Research*, 69(4), 1186–1205.

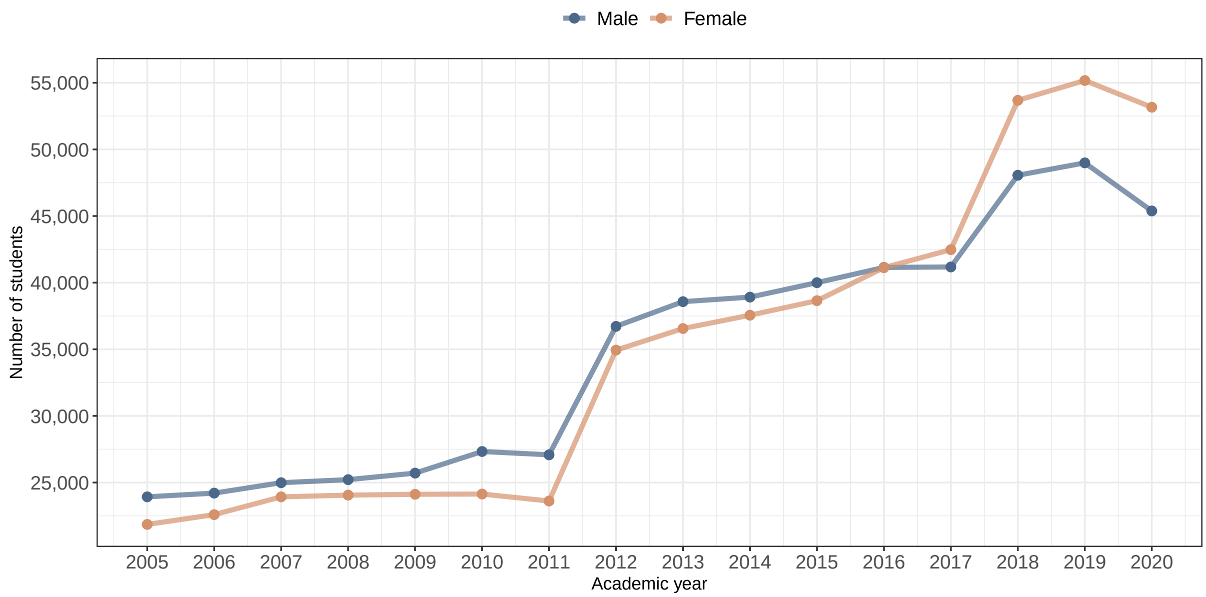
Sacerdote, B. (2001). Peer Effects with Random Assignment: Results for Dartmouth Roommates*. *The Quarterly Journal of Economics*, 116(2), 681–704.

Sacerdote, B. (2011). Chapter 4 - Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the Economics of Education*, volume 3 (pp. 249–277). Elsevier.

Sloane, C. M., Hurst, E. G., & Black, D. A. (2021). College Majors, Occupations, and the Gender Wage Gap. *Journal of Economic Perspectives*, 35(4), 223–248.

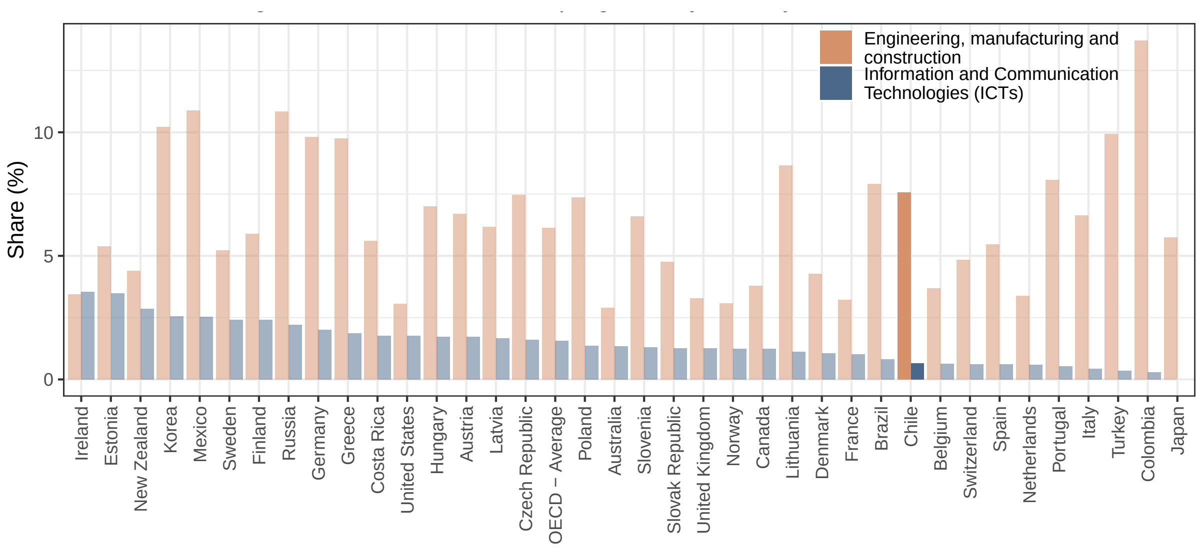
Zimmerman, D. J. (2003). Peer Effects in Academic Outcomes: Evidence from a Natural Experiment. *Review of Economics and Statistics*, 85(1), 9–23.

Figure 1: Freshmen enrollment by gender at universities from the centralized admission system



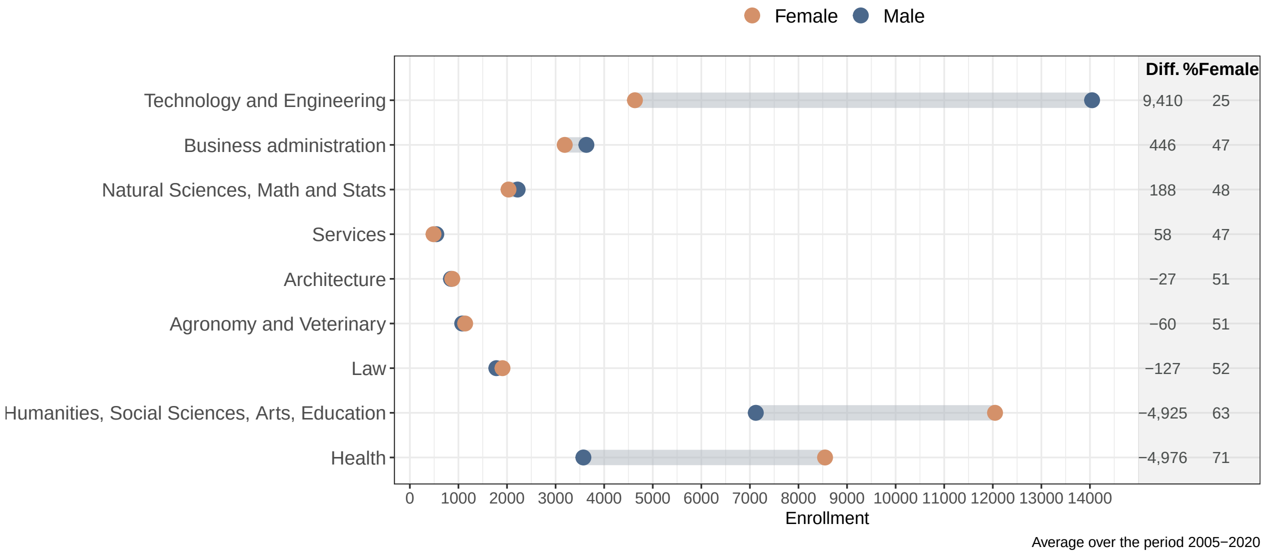
Note: This figure exhibits the total freshmen enrollment by gender at universities that are participants of the centralized admission system. Source: DEMRE administrative records.

Figure 2: Share of female graduates from bachelor's programs, by country and field



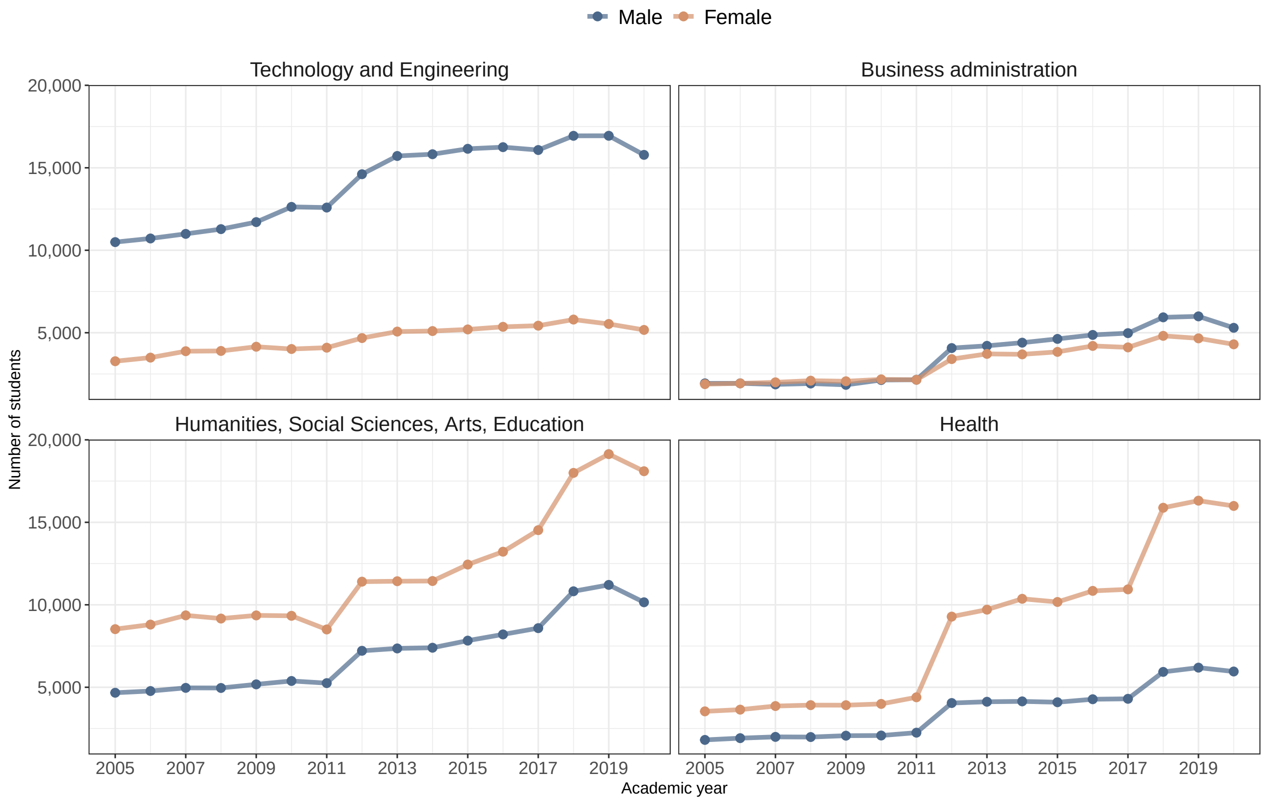
Note: This figure shows the share of female graduates by field across OECD countries. Source: Gender Data, OECD (2017).

Figure 3: Average freshmen enrollment by field and gender



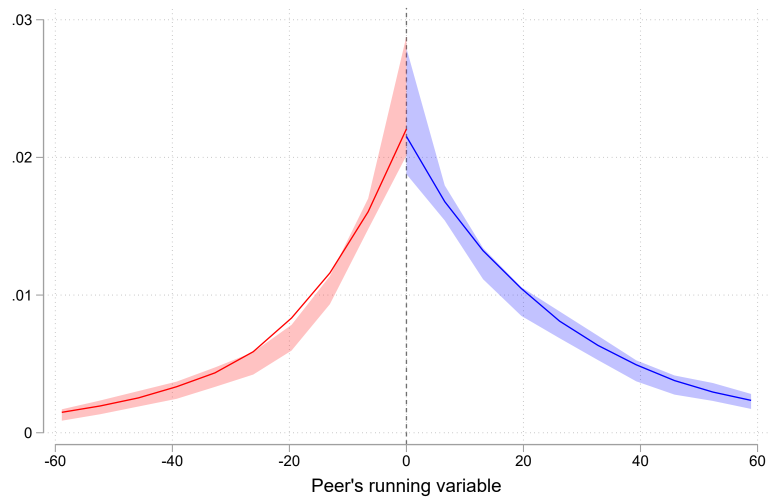
Note: This figure shows the average freshmen enrollment by gender at different fields of study. Grey bars represent the distance between men and female enrollment. The difference is depicted under the "Diff." column. The share of the female enrollment by field is despicted under the "%Female" column. For example, on average, there are 9,410 more men enrolled in Technology and Engineering than women. And women represent 25% of the enrollment at that field.

Figure 4: Enrollment by gender and field at programs from the centralized admission system



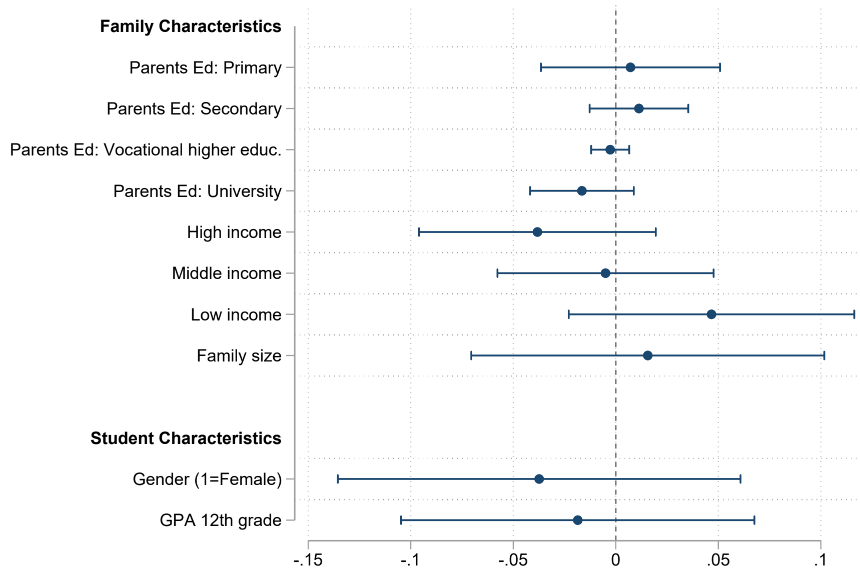
Note: This figure shows freshmen enrollment by year, gender and main fields of study.

Figure 5: Density plot of peer's running variable around the cutoff

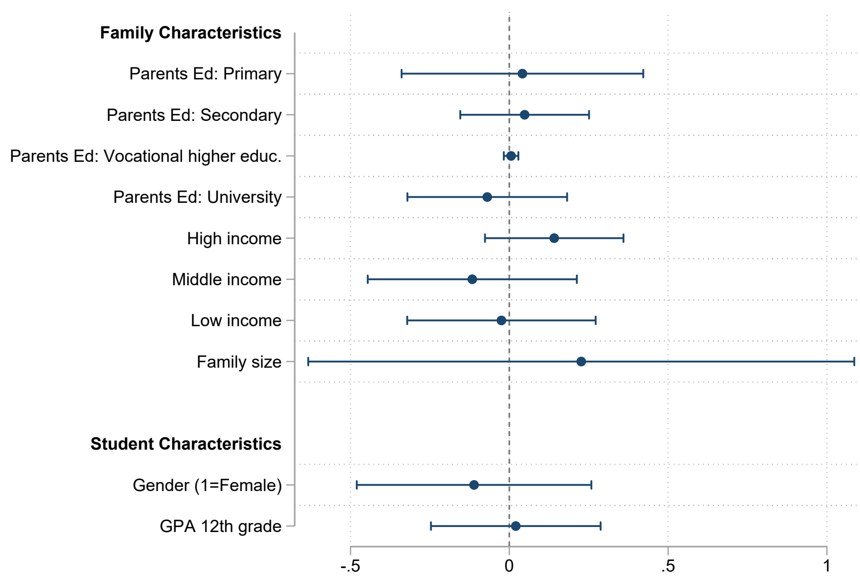


Note: This figure shows the density plot of the running variable around the cutoff, using the *rddensity* package that implements manipulation testing procedures using the local polynomial density estimators proposed in Cattaneo et al. (2020). The robust bias-corrected test proposed by the authors have a *p-value* of 0.7206, which provides empirical evidence in favor of the continuity of the running variable.

Figure 6: Potential applicants and peers' continuity test around the cutoff



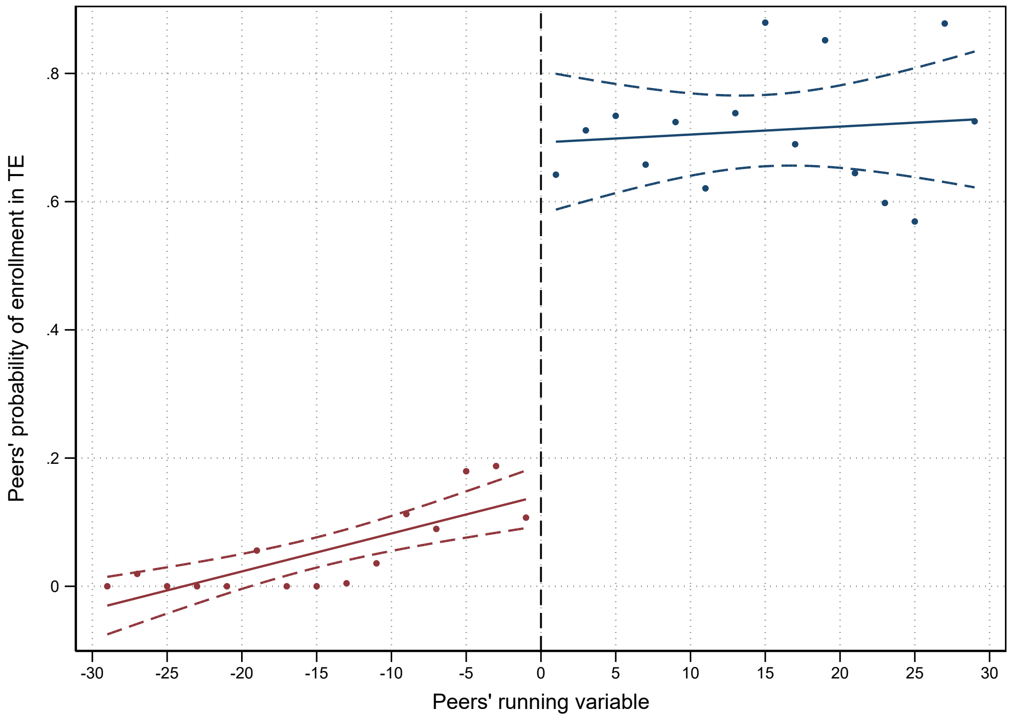
(a) Potential applicants' covariates



(b) Peers' covariates

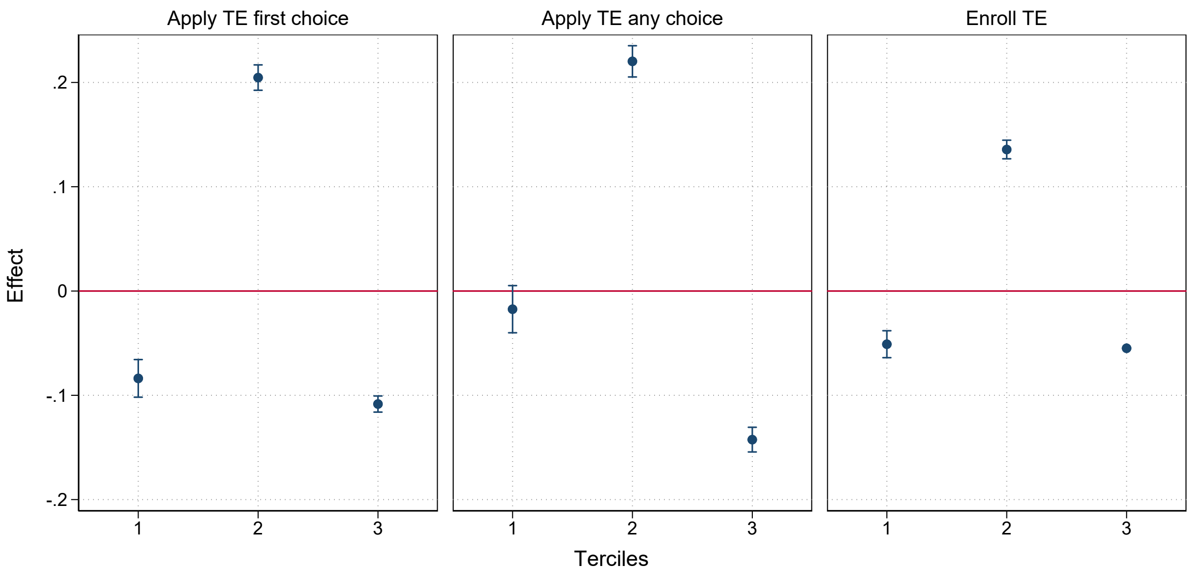
Note: This figure shows the bias-corrected treatment effects on baseline covariates, using confidence intervals obtained from robust standard errors following Calonico et al. (2014a,b). Each covariate was tested individually as the outcome, with optimal bandwidths chosen separately, as suggested in Cattaneo et al. (2019). Panel (a) and (b) present the point estimates and 95% CIs of potential applicant's covariates and peers, respectively.

Figure 7: First stage



Note: This figure shows the discontinuity at the cutoff of the first stage represented in Equation 4.

Figure 8: Heterogeneous treatment effect by school size terciles



Note: This figure shows the effect of peer's enrollment on Technology and Engineering on the main outcomes by terciles of the number of students in their last year of high school. Each coefficient and its 95% CI are obtained from separate regressions calculated for the corresponding tercile. The range of the number of students per school in each tercile is the following: in the first, second, and third tercile there are between 11–91, 92–191, and 192–694 students, respectively. Humanities, Social Sciences, Arts and Education is the fall-back alternative in this sample.

Table 1: Descriptive statistics for the analytic sample

	Mean (1)	SD (2)	Min (3)	Max (4)	N (5)
Potential applicants					
Parents Education - Primary	0.49	0.50	0.0	1.0	303,572
Parents Education - Secondary	0.11	0.32	0.0	1.0	303,572
Parents Education - Tertiary Vocational	0.02	0.15	0.0	1.0	321,801
Parents Education - Tertiary University	0.12	0.32	0.0	1.0	321,801
High Income	0.13	0.34	0.0	1.0	321,958
Mid Income	0.44	0.50	0.0	1.0	321,958
Low Income	0.43	0.49	0.0	1.0	321,958
Family size	4.46	1.48	1.0	39.0	309,450
Female	0.52	0.50	0.0	1.0	321,966
GPA 12th grade	5.79	0.53	1.0	7.0	319,667
PSU math score	533.23	107.05	150.0	850.0	321,966
PSU language score	528.72	105.01	164.0	850.0	321,777
Peers					
Parents Education - Primary	0.47	0.50	0.0	1.0	304,691
Parents Education - Secondary	0.15	0.35	0.0	1.0	304,691
Parents Education - Tertiary Vocational	0.02	0.15	0.0	1.0	321,917
Parents Education - Tertiary University	0.13	0.34	0.0	1.0	321,917
High Income	0.14	0.35	0.0	1.0	321,966
Mid Income	0.71	0.46	0.0	1.0	321,966
Low Income	0.15	0.36	0.0	1.0	321,966
Family size	4.36	1.45	1.0	20.0	310,184
Female	0.40	0.49	0.0	1.0	321,966
GPA 12th grade	5.97	0.43	4.5	7.0	320,771
PSU math score	584.31	69.17	355.0	850.0	321,966
PSU language score	573.90	65.46	395.0	809.0	321,966

Note: This table presents summary statistics for the estimation sample that corresponds 321,966 potential applicants and 2,945 unique peers. GPA is the final average score at the academic year that follows a scale from 1 to 7. PSU scores, are the score from the standardized test score with mean 500 and a standard deviation of 110. The minimum and maximum scores in the scale are 150 and 850 points, respectively.

Table 2: Average effect of high school peer's TE enrollment on potential applicant's outcomes

Outcome:	**Apply to TE in the first choice** (1)	**Apply to TE in any choice** (2)	**Enrollment in TE** (3)
Panel A: HASS as the fall-back field			
Peer enrolls in TE	0.033***	0.042***	0.025***
	(0.006)	(0.010)	(0.007)
First stage	0.485***	0.483***	0.503***
	(0.026)	(0.027)	(0.021)
BW Est. (h)	[9.299 ; 9.299]	[9.619 ; 9.619]	[8.361 ; 8.361]
Outcome mean	0.125	0.200	0.097
Number of applicants	62338	64410	55996
Panel B: Health as the fall-back field			
Peer enrolls in TE	-0.155***	-0.187***	-0.145***
	(0.010)	(0.015)	(0.008)
First stage	0.272***	0.324***	0.338***
	(0.013)	(0.012)	(0.012)
BW Est. (h)	[7.365 ; 7.365]	[6.937 ; 6.937]	[6.798 ; 6.798]
Outcome mean	0.146	0.234	0.123
Number of applicants	30165	28495	27604
Panel C: Business as the fall-back field			
Peer enrolls in TE	0.007	0.014*	-0.005
	(0.006)	(0.007)	(0.005)
First stage	0.399***	0.396***	0.393***
	(0.028)	(0.028)	(0.027)
BW Est. (h)	[10.056 ; 10.056]	[10.485 ; 10.485]	[10.771 ; 10.771]
Outcome mean	0.123	0.193	0.096
Number of applicants	100762	104092	106982
Year fixed effects	✓	✓	✓
School fixed effects	✓	✓	✓

Note: This table shows the effect of having a high-school peer enrolled in a TE program using a fuzzy RD approach. RD estimates are robust bias-corrected estimates computed using a linear local polynomial and a triangular kernel. Optimal bandwidths are chosen to be MSE optimal. All producedures are computed following the *rdrobust* package in Stata by Calonico et al. (2017, 2014b). Standard errors clustered at the school level. TE stands for Technology and Engineering, HASS stands for Humanities, Social Sciences, Arts, and Education. All regressions include school and time fixed effects.$^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$

Table 3: Average effect of high school peer's TE enrollment on potential applicant's outcomes by gender

Outcome:	**Apply to TE as first choice**		**Apply to TE in any choice**		**Enrollment in TE**	
	Male (1)	Female (2)	Male (3)	Female (4)	Male (5)	Female (6)
Panel A: HASSE as the fall-back field						
Peer enrolls in TE	0.095***	0.007	0.118***	-0.010	0.085***	-0.005
	(0.015)	(0.006)	(0.018)	(0.010)	(0.013)	(0.005)
First stage	0.475***	0.499***	0.475***	0.494***	0.485***	0.525***
	(0.026)	(0.026)	(0.027)	(0.027)	(0.021)	(0.021)
BW Est. (h)	[9.299 ; 9.299]	[9.299 ; 9.299]	[9.619 ; 9.619]	[9.619 ; 9.619]	[8.361 ; 8.361]	[8.361 ; 8.361]
Outcome mean	0.207	0.057	0.294	0.121	0.159	0.045
Number of applicants	28505	33833	29443	34967	25507	30489
Panel B: Health as the fall-back field						
Peer enrolls in TE	0.011	-0.193***	-0.140***	-0.137***	0.005	-0.179***
	(0.011)	(0.011)	(0.007)	(0.010)	(0.005)	(0.007)
First stage	0.253***	0.291***	0.323***	0.334***	0.342***	0.345***
	(0.012)	(0.014)	(0.011)	(0.013)	(0.010)	(0.013)
BW Est. (h)	[7.365 ; 7.365]	[7.365 ; 7.365]	[6.937 ; 6.937]	[6.937 ; 6.937]	[6.798 ; 6.798]	[6.798 ; 6.798]
Outcome mean	0.241	0.065	0.343	0.137	0.197	0.058
Number of applicants	14027	16138	13329	15166	12944	14660
Panel C: Business as the fall-back field						
Peer enrolls in TE	0.051***	-0.009	0.079***	-0.020***	0.009	0.007
	(0.014)	(0.007)	(0.016)	(0.008)	(0.013)	(0.005)
First stage	0.406***	0.389***	0.401***	0.387***	0.399***	0.385***
	(0.027)	(0.029)	(0.026)	(0.029)	(0.026)	(0.028)
BW Est. (h)	[10.056 ; 10.056]	[10.056 ; 10.056]	[10.485 ; 10.485]	[10.485 ; 10.485]	[10.771 ; 10.771]	[10.771 ; 10.771]
Outcome mean	0.205	0.056	0.287	0.115	0.159	0.044
Number of applicants	45687	55075	47216	56876	48570	58412

Note: This table shows the effect of having a high-school peer enrolled in a TE program using a fuzzy RD approach. RD estimates are robust bias-corrected estimates computed using a linear local polynomial and a triangular kernel. Optimal bandwidths are chosen to be MSE optimal. All producedures are computed following the *rdrobust* package in Stata by Calonico et al. (2017, 2014b). Standard errors clustered at the school level. TE stands for Technology and Engineering, HASS stands for Humanities, Social Sciences, Arts, and Education. All regressions include school and time fixed effects. $^{*}p < 0.1, ^{**}p < 0.05, ^{***}p < 0.01$

Table 4: Average effect of high school peer's TE enrollment on potential applicant's outcomes by gender, fall-back HASSE

Outcome:	**Apply to TE as first choice**		**Apply to TE in any choice**		**Enrollment in TE**	
	Male Peer (1)	Female Peer (2)	Male Peer (3)	Female Peer (4)	Male Peer (5)	Female Peer (6)
Panel A: Male Potential Applicant						
Peer enrolls in TE	0.228***	0.071***	0.442***	0.077***	0.900***	0.093***
	(0.013)	(0.004)	(0.025)	(0.006)	(0.127)	(0.003)
First stage	0.422***	0.659***	0.399***	0.642***	0.119***	0.668***
	(0.018)	(0.016)	(0.019)	(0.016)	(0.014)	(0.015)
BW Est. (h)	[9.299 ; 9.299]	[9.299 ; 9.299]	[9.619 ; 9.619]	[9.619 ; 9.619]	[8.361 ; 8.361]	[8.361 ; 8.361]
Outcome mean	0.208	0.204	0.299	0.288	0.162	0.155
Number of applicants	16565	11940	17151	12292	14613	10894
Panel B: Female Potential Applicant						
Peer enrolls in TE	-0.024***	-0.012***	-0.022***	0.033***	0.235***	-0.050***
	(0.006)	(0.002)	(0.007)	(0.004)	(0.027)	(0.002)
First stage	0.457***	0.746***	0.448***	0.738***	0.161***	0.740***
	(0.020)	(0.018)	(0.020)	(0.018)	(0.016)	(0.017)
BW Est. (h)	[9.299 ; 9.299]	[9.299 ; 9.299]	[9.619 ; 9.619]	[9.619 ; 9.619]	[8.361 ; 8.361]	[8.361 ; 8.361]
Outcome mean	0.057	0.056	0.122	0.121	0.044	0.046
Number of applicants	19212	14621	19876	15091	17081	13408

Note: This table shows the effect of having a high-school peer enrolled in a TE program using a fuzzy RD approach. RD estimates are robust bias-corrected estimates computed using a linear local polynomial and a triangular kernel. Optimal bandwidths are chosen to be MSE optimal. All producedures are computed following the *rdrobust* package in Stata by Calonico et al. (2017, 2014b). Standard errors clustered at the school level. TE stands for Technology and Engineering, HASSE stands for Humanities, Social Sciences, Arts, and Education. All regressions include school and time fixed effects. Standard errors clustered at the school level. The fall-back field in this table is Humanities, Social Sciences, Arts, Education. $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$

Online Appendix for

Peer Influence and College Major Choices in Male-Dominated Fields

Rocío Valdebenito

October 18, 2023

A Data construction

This section explains the data sources and steps needed to construct the final datasets used in the analysis. The section is divided in four parts, data sources, peers data, applicants data, and merging process.

A.1 Data sources

1. *Departamento de Evaluación, Medición y Registro Educacional* (DEMRE) 2004–2020. This organization is in charge of monitoring and implementing the admission to universities members of the centralized admission system. The datasets available from this agency contains student-level data of the entire centralized admission system. Particularly, the scores obtained from the whole universe of students taking the standardized college entry exam, self-reported socioeconomic information, applications to universities members of the system, the results of the submission process after the algorithm is implemented, and final enrollment.

2. DEMRE, programs' components weights, and number of seats per year and program over the period 2004–2020.

3. Student annual GPA, Ministry of Education. This datasets contains student-level variables of the annual GPA obtained by the students at the end of the academic year with the respective students and schools' identifiers.

A.2 Peers data

A.2.1 Identification of undersubscribed programs

A first step of the data cleaning consists of identifying programs that are undersubscribed, which are programs that were not able to fill their available seats completely in a given academic year. In these programs, I cannot identify their cutoff because any additional applicant would be admitted, regardless of the application score of the last admitted student.

Using data from the application process and the outcomes after the deferred acceptance mechanism, this step identifies, for each program-year, the list of admitted students and the list of non-admitted students. Therefore, after ordering the application weighted scores, I calculate the minimum score of the last admitted student, conditional on having at least one non-admitted student (i.e., below the admission cutoff) per program-year. Consequently, I can identify the cutoffs only for those programs where there is a 'waiting-list' of at least one student, as the available seats have been filled after the implementation of the algorithm.

A.2.2 Application data and cleaning procedures

First-time takers

It is important to note that students can take the college entry exam multiple times. The first step consists of combining all the years of the student-level datasets on scores and socioeconomic variables. If a student appears multiple times, this step keeps only the first time that the student took the test.

Eliminate undersubscribed programs from the list

A second step analyzes the datasets of the application list submitted by each student. This dataset is then merged with the information processed from the previous section where undersubscribed programs are identified. In this procedure, different cases can occur per student:

1. All the applications were made in oversubscribed programs: If that is the case, then all cutoffs across all alternatives are available.

2. If at least one alternative is made in an undersubscribed program:

   (a) And the admitted alternative is the undersubscribed program: In this case, the cutoff cannot be obtained for the alternative in which the student was admitted. Therefore, the running variable is impossible to calculate, and the entire list (student) is eliminated from the sample.

   (b) And the admitted alternative is not the undersubscribed program: In this case, the undersubscribed program does not affect the ability to calculate the running variable in the admitted choice. Therefore, if the undersubscribed program is just right below the admitted alternative, then the undersubscribed program would be used as the counterfactual for the admitted program. If, in contrast, the undersubscribed alternative is even further away (e.g., alternative 7th), then I eliminate the undersubscribed program from the list because its presence does not affect any target and counterfactual combination.

   (c) If the student was not admitted in any choice, then I eliminate the undersubscribed program from the list.

Identify relative selectivity and eliminate dominated alternatives

The main purpose of this step is to further clean the preference lists to have an ordered list where subsequent alternatives can be used as a counterfactual for a previous alternative. To accomplish this goal, this step borrows the concept of *relative selectivity* from Abdulkadiroğlu et al. (2014), Aguirre & Matta (2021), and Aguirre et al. (2022).

Relative Selectivity is calculated as $\phi_{ij} = \frac{s_{ij}-c_j}{\sqrt{\sum_l (\alpha_j^l)^2}}$, which represents the Euclidean distance between the applicant's scores (in math, language, science, history) and the admission frontier defined by the cutoff at program j. If the relative selectivity of a lower-ranked option is higher than that of a higher-ranked option, the relatively more selective option will be eliminated since it does not serve as a proper counterfactual. In other words, higher selective programs submitted in lower-ranked positions will not represent a real scenario of what would have happened if the student is below an admission cutoff in the higher-ranked option.

The table below shows the resulting number of observations after running the iterative process of dropping dominated alternatives:

Table A.1: Number of observations at each iterative process of elimination

Step	Number of observations
iteration 0:	3,586,477
iteration 1:	2,471,055
iteration 2:	2,198,677
iteration 3:	2,140,679
iteration 4:	2,130,907
iteration 5:	2,129,686
iteration 6:	2,129,569
iteration 7:	2,129,552

Constructing target and counterfactual

After cleaning the preferences in the previous step, I construct pairs of target and fallback/counterfactual alternatives. Each alternative corresponds to an specific program-university combination. Table A.2 exhibits an example of a preference list after the relative dominated alternatives are eliminated. There are 6 preferences that survived the elimination procedure. The student is admitted in the third choice, because it is the first one where the score is above the program-specific cutoff.

Table A.2: Illustration of a preference list

Preference	Program	University	Score	Cutoff	Status
1	Civil Engineering	University of Chile	685	710	No admitted
2	Business	University of Chile	675	695	No admitted
3	Civil Engineering	University of Santiago	685	675	Admitted
4	Business	University of Santiago	667	650	Admitted above
5	Mechanical Engineering	Diego Portales University	680	620	Admitted above
6	Electrical Engineering	Andres Bello University	678	610	Admitted above

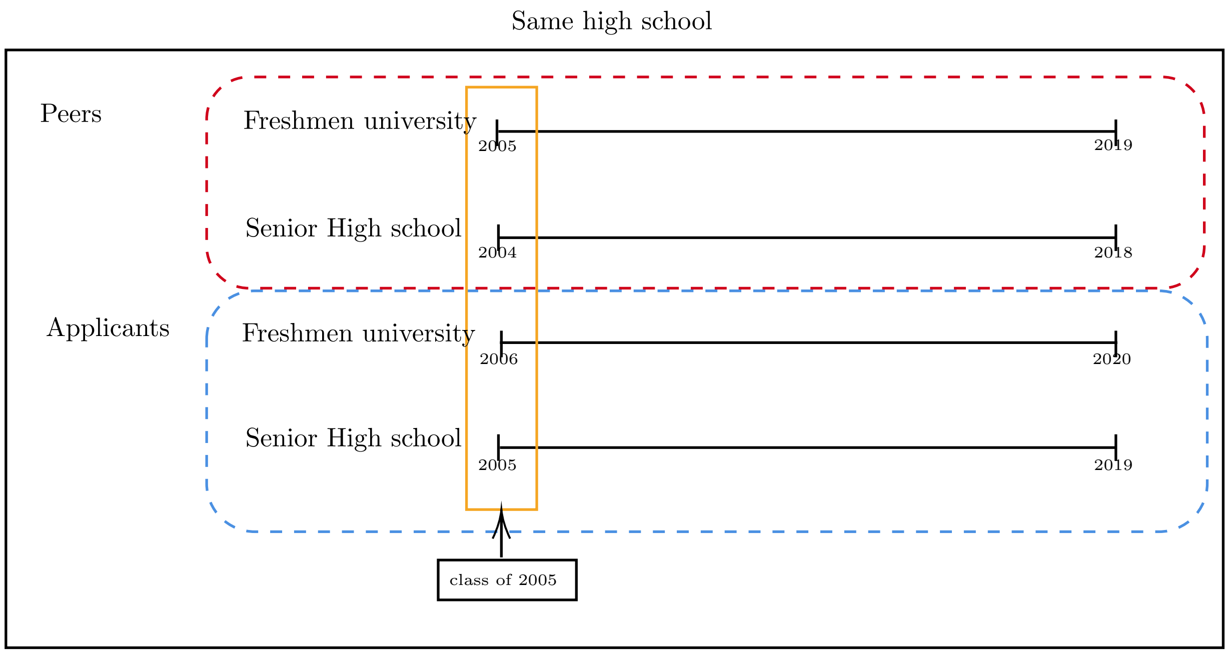
The main objective is to construct target and fall-back program pairs. The target program is the preferred program, while the fall-back or counterfactual represents the "what would have happened" scenario. However, if the student was not admitted to their target choice, then by construction, the fall-back choice is a representation of what actually occurred in practice. For example, in Table A.2, the first and second preferences are alternatives where the student was not admitted. Therefore, for each of those, the counterfactual scenario is the admitted choice (i.e., Civil Engineering at the University of Santiago). Moreover, for the admitted choice, the counterfactual would be the alternative just below, since the preference list is already cleaned from dominated choices and impossible choices are not present. In that case, the fourth preference represents what would have happened if the student's score was below the cutoff for their admitted choice.

Table A.3 presents the target and fall-back pairs obtained from the list presented in Table A.2.

Table A.3: Target and fall-back pairs

Preference	Target Program	Target University	Fall-back Program	Fall-back University
1	Civil Engineering	University of Chile	Civil Engineering	University of Santiago
2	Business	University of Chile	Civil Engineering	University of Santiago
3	Civil Engineering	University of Santiago	Business	University of Santiago

Figure A.1: Illustration of peers and potential applicants connection



Note: This figure shows how potential applicants and peers are connected. For example, the potential applicants' class of 2005, is connected with the sample of peers–in the margins of interests–that are in the same high school as the applicants and belong to the class of 2004.

Table A.4: Number of observations and representation of all target and fall-back combinations

Panel A: Total number of target fall back combinations

Target / Fallback	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	Nothing	Row total
Agronomy and Veterinary (1)	14,260	2,685	3,800	795	1,990	126	230	2,033	314	11,075	37,308
Natural Sciences, Math and Stats (2)	2,308	22,170	19,506	3,112	6,702	548	560	6,968	555	24,660	87,089
Technology and Engineering (3)	4,230	23,105	227,667	17,265	8,090	593	2,722	5,124	3,691	75,862	368,349
Business administration (4)	935	4,540	13,930	67,132	10,208	1,376	1,255	1,088	1,736	31,136	133,336
Humanities, Social Sciences, Arts, Education (5)	2,135	6,302	9,807	9,244	189,440	8,757	3,496	4,524	3,937	127,818	365,460
Law (6)	340	1,025	1,536	3,422	21,391	44,318	758	981	302	21,177	95,250
Architecture (7)	267	559	4,611	1,383	5,988	330	15,996	379	256	9,426	39,195
Health (8)	8,959	30,395	27,768	6,416	26,654	2,580	1,688	217,380	2,424	164,823	489,087
Services (9)	305	527	2,983	1,016	1,831	78	115	243	2,440	7,536	17,074

Panel B: Percent (%) by Target field

Target / Fallback	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	Nothing	Row total
Agronomy and Veterinary (1)	38.2	7.2	10.2	2.1	5.3	0.3	0.6	5.4	0.8	29.7	100
Natural Sciences, Math and Stats (2)	2.7	25.5	22.4	3.6	7.7	0.6	0.6	8.0	0.6	28.3	100
Technology and Engineering (3)	1.1	6.3	61.8	4.7	2.2	0.2	0.7	1.4	1.0	20.6	100
Business administration (4)	0.7	3.4	10.4	50.3	7.7	1.0	0.9	0.8	1.3	23.4	100
Humanities, Social Sciences, Arts, Education (5)	0.6	1.7	2.7	2.5	51.8	2.4	1.0	1.2	1.1	35.0	100
Law (6)	0.4	1.1	1.6	3.6	22.5	46.5	0.8	1.0	0.3	22.2	100
Architecture (7)	0.7	1.4	11.8	3.5	15.3	0.8	40.8	1.0	0.7	24.0	100
Health (8)	1.8	6.2	5.7	1.3	5.4	0.5	0.3	44.4	0.5	33.7	100
Services (9)	1.8	3.1	17.5	6.0	10.7	0.5	0.7	1.4	14.3	44.1	100

Panel C: Percent (%) over the total number combinations

Target / Fallback	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	Nothing	Row total
Agronomy and Veterinary (1)	0.9	0.2	0.2	0.0	0.1	0.0	0.0	0.1	0.0	0.7	2.3
Natural Sciences, Math and Stats (2)	0.1	1.4	1.2	0.2	0.4	0.0	0.0	0.4	0.0	1.5	5.3
Technology and Engineering (3)	0.3	1.4	13.9	1.1	0.5	0.0	0.2	0.3	0.2	4.6	22.6
Business administration (4)	0.1	0.3	0.9	4.1	0.6	0.1	0.1	0.1	0.1	1.9	8.2
Humanities, Social Sciences, Arts, Education (5)	0.1	0.4	0.6	0.6	11.6	0.5	0.2	0.3	0.2	7.8	22.4
Law (6)	0.0	0.1	0.1	0.2	1.3	2.7	0.0	0.1	0.0	1.3	5.8
Architecture (7)	0.0	0.0	0.3	0.1	0.4	0.0	1.0	0.0	0.0	0.6	2.4
Health (8)	0.5	1.9	1.7	0.4	1.6	0.2	0.1	13.3	0.1	10.1	30.0
Services (9)	0.0	0.0	0.2	0.1	0.1	0.0	0.0	0.0	0.1	0.5	1.0
Total	2.1	5.6	19.1	6.7	16.7	3.6	1.6	14.6	1.0	29.0	100.0

Note: This table shows the total number of observations of the entire dataset that contains all the possible combinations between target and fall-back fields. The cells highlighted in orange represent margins of target and fall-back combinations used in the estimations. Panel A shows the total number of pairs, Panel B shows the percent row by each target field, and Panel C shows the total representation over the entire possible combinations.

Table A.5: Average effect of high school peer's TE enrollment on potential applicant's outcomes by gender, fall-back Health

Outcome:	**Apply to TE as first choice**		**Apply to TE in any choice**		**Enrollment in TE**	
	Male Peer (1)	Female Peer (2)	Male Peer (3)	Female Peer (4)	Male Peer (5)	Female Peer (6)
Panel A: Male Potential Applicant						
Peer enrolls in TE	0.414***	-0.241***	0.026***	-0.317***	0.487***	-0.236***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
First stage	0.170***	0.692***	0.149***	0.752***	0.148***	0.887***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
BW Est. (h)	[7.365 ; 7.365]	[7.365 ; 7.365]	[6.937 ; 6.937]	[6.937 ; 6.937]	[6.798 ; 6.798]	[6.798 ; 6.798]
Outcome mean	0.245	0.236	0.349	0.337	0.201	0.193
Number of applicants	7049	6978	6746	6583	6594	6350
Panel B: Female Potential Applicant						
Peer enrolls in TE	-0.059***	-0.047***	0.136***	-0.060***	0.225***	-0.075***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
First stage	0.170***	0.692***	0.149***	0.752***	0.148***	0.887***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
BW Est. (h)	[7.365 ; 7.365]	[7.365 ; 7.365]	[6.937 ; 6.937]	[6.937 ; 6.937]	[6.798 ; 6.798]	[6.798 ; 6.798]
Outcome mean	0.064	0.065	0.137	0.138	0.060	0.057
Number of applicants	7825	8313	7327	7839	7138	7522

Note: This table shows the effect of having a high-school peer enrolled in a TE program using a fuzzy RD approach. RD estimates are robust bias-corrected estimates computed using a linear local polynomial and a triangular kernel. Optimal bandwidths are chosen to be MSE optimal. All producedures are computed following the *rdrobust* package in Stata by Calonico et al. (2017, 2014b). Standard errors clustered at the school level. TE stands for Technology and Engineering. All regressions include school and time fixed effects. Standard errors clustered at the school level. The fall-back field in this table is Health. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table A.6: Average effect of high school peer's TE enrollment on potential applicant's outcomes by gender, fall-back Business

Outcome:	**Apply to TE as first choice**		**Apply to TE in any choice**		**Enrollment in TE**	
	Male Peer (1)	Female Peer (2)	Male Peer (3)	Female Peer (4)	Male Peer (5)	Female Peer (6)
Panel A: Male Potential Applicant						
Peer enrolls in TE	0.116***	0.089***	0.133***	0.191***	0.080***	0.027***
	(0.016)	(0.004)	(0.022)	(0.005)	(0.015)	(0.003)
First stage	0.360***	0.703***	0.338***	0.713***	0.326***	0.715***
	(0.022)	(0.010)	(0.022)	(0.010)	(0.022)	(0.010)
BW Est. (h)	[10.056 ; 10.056]	[10.056 ; 10.056]	[10.485 ; 10.485]	[10.485 ; 10.485]	[10.771 ; 10.771]	[10.771 ; 10.771]
Outcome mean	0.202	0.211	0.282	0.296	0.157	0.164
Number of applicants	29637	16050	30615	16601	31353	17217
Panel B: Female Potential Applicant						
Peer enrolls in TE	-0.026***	-0.039***	-0.101***	-0.141***	0.030***	0.003
	(0.006)	(0.003)	(0.016)	(0.006)	(0.008)	(0.003)
First stage	0.320***	0.498***	0.296***	0.509***	0.283***	0.511***
	(0.025)	(0.011)	(0.026)	(0.012)	(0.026)	(0.012)
BW Est. (h)	[10.056 ; 10.056]	[10.056 ; 10.056]	[10.485 ; 10.485]	[10.485 ; 10.485]	[10.771 ; 10.771]	[10.771 ; 10.771]
Outcome mean	0.054	0.059	0.111	0.123	0.042	0.047
Number of applicants	35076	19999	36190	20686	37069	21343

Note: This table shows the effect of having a high-school peer enrolled in a TE program using a fuzzy RD approach. RD estimates are robust bias-corrected estimates computed using a linear local polynomial and a triangular kernel. Optimal bandwidths are chosen to be MSE optimal. All producedures are computed following the *rdrobust* package in Stata by Calonico et al. (2017, 2014b). Standard errors clustered at the school level. TE stands for Technology and Engineering. All regressions include school and time fixed effects. Standard errors clustered at the school level. The fall-back field in this table is Business. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$