CESifo WORKING PAPERS

11075 2024

April 2024

Keeping It in the Family: Student to Degree Match

Richard Murphy, Pedro Luís Silva

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 11075*

Keeping It in the Family: Student to Degree Match

Abstract

This paper examines systematic inequalities in the match between students and the university degree they apply to, and enroll in. Using linked administrative data on the population of Portuguese applicants we create a transparent and continuous measure of student-to-degree match employing minimal assumptions. We find that students who are the first in the family to attend post-secondary education consistently match to lower quality degrees across the entire achievement distribution. In contrast, only the highest achieving female students relatively undermatch. These gaps are larger at the application stage. We explore the role of student preferences and the consequences for intergenerational mobility.

JEL-Codes: I220, I230, I280.

Keywords: higher education, educational economics, college choice, mismatch, undermatch.

Richard Murphy
University of Texas at Austin / USA
richard.murphy@austin.utexas.edu

Pedro Luís Silva
Centre for Research in Higher Education Policies, University of Porto / Portugal
pedro.luis.silva@cipes.up.pt

April 2024
We thank Ana Tur Prats, Ana Rute Cardoso, Alberto Amaral and Miguel Urquiola for their helpful comments. Funding for this research was provided by FCT – Foundation for Science and Technology (Portugal) under the projects Ref. http://doi.org/10.54499/PTDC/CED-EDG/5530/2020, and http://doi.org/10.54499/UIDP/00757/2020 . Any remaining errors or omissions are the authors' responsibility. We thank the General-Directorate of Higher Education (DGES, Direção-Geral do Ensino Superior) and the General-Directorate of Education and Science Statistics (DGEEC, Direção-Geral de Estatísticas da Educação e Ciência) for the access to the two data sets used in this paper. The data sets were merged at the premises of DGEEC by the second author and cannot be disclosed to anyone who has not signed a confidentiality agreement with DGEEC.

I. Introduction

Increasing enrollments in higher education (HE) is a perennial concern of governments across the globe. A large literature has built up exploring various strategies to promote participation such as relaxing credit constraints (Carneiro & Heckman 2002, Lochner & Monge-Naranjo 2011, Murphy & Wyness 2023), providing better information (Hoxby & Turner 2015), or reducing administrative burdens (Bettinger et al 2012, Dynarski 2021). However, considerably less attention has been given to the university and major combination, which we will refer to as a degree, that students choose once they decide to pursue higher education. This is despite many of the same market failures leading students to make sub-optimal enrollment decisions on this intensive margin.

The choice of which degree course to apply to is important given the heterogeneous returns to subjects (Kirkeboen et al 2016) and institutions (Zimmerman 2019, Mountjoy 2022). This coupled with heterogeneity in the abilities of students means that a degree which is beneficial for one student may provide little utility to another. Sallee et. al. (2008) set out the assumptions required for academic assortative matching to improve the efficiency of human capital production (e.g. student and course complementarities). This has been complemented by recent empirical research which has shown the gains from students being well-matched to their degrees (Arcidiacono and Lovenheim, 2016; Dillon and Smith, 2020). The existence of these complementarities between degree and student quality means that matching students to degrees has large potential impacts on the aggregate returns to higher education for society.[1] In reality, exact assortative matching may not be optimal for all students. However, we assume that on average, an efficient higher education system would allocate students to degrees in which they would be the most well-matched. A system with market failures, such as credit constraints, informational constraints, and administrative barriers, may lead to students making suboptimal decisions.

In this paper, we construct comprehensive and transparent measures of student-to-degree match. This is constructed from detailed administrative data on the population of applicants to the Portuguese higher educational system. We use this information on student applications and matriculation to ascertain whether certain types of students are systematically mismatching and at

[1] Even in the absence of penalties to overmatching (Bleemer 2022), the existence of capacity constraints and undermatch penalties would deliver the same conclusion.

which stage in the process. The paper then examines the potential drivers of mismatch and makes comparisons to other higher education markets.

The key difficulty in studying student match is the role of preferences and separating them from market failures. Consider students from different groups (e.g., male/female) who are otherwise observationally equivalent, and who consistently enroll in different degrees. A researcher will find it hard to discern if this is driven by preferences, or market failures which one group is more exposed to. For example, high ability low-income female students may be qualified to attend high-ranking STEM degrees but may choose to enroll in a lower-ranking humanities degree instead. This could be due to a lack of information about the net benefits, or due to preferences. The existence of market failures may cause students revealed preferences to not match true normative preferences (Beshears et al 2008).

To make progress in understanding whether a student is truly mismatched requires either; i) perfect knowledge about student preferences; or ii) assumptions about the circumstances for when a student is mismatched. We are going to build on the structure set out by Campbell et al (2022) and make two assumptions: students have no degree preferences, and students gain the most from enrolling in a degree that matches their own abilities (Arcidiacono and Lovenheim, 2016; Dillon and Smith, 2017). These assumptions allow us to state that any student not enrolled in a degree that matches their ability is mismatched due to a market failure. In reality, degree preferences do exist, and so we regard this measure of mismatch as an upper bound.

We create continuous mismatch measures by combining administrative data on high school, application, and enrollment for all students who applied to a Portuguese public university between 2013/14 and 2019/20. We rank students nationally based on their academic attainment during high school, and rank degrees nationally on the median entry qualification of students who applied (application match) and among students who ultimately matriculated (enrollment match), weighting by course size. Each is converted into quality percentiles such that the degree n-th quality percentile is the degree that a student in the n-th ability percentile would attend under perfect assortative matching of students and colleges. The disparity between the student's percentile ranking and the degree's ranking forms our measures of mismatch.

The Portuguese system is especially suitable for exploring the role of student preferences in the matching to degrees as the application system is one-sided. After receiving their high school grades, students rank their choices of degrees, and admission is determined via a Gale–Shapley algorithm on the basis of their academic achievements. This setting allows for the isolation of student choice without the complication of potential action by universities, as they do not get to choose students. Moreover, as the financial aid students receive is the same for students regardless of high-school achievement, potential complications driven by university financial aid offers to high-achieving students are not present.

We examine the systematic differences in application and enrollment patterns for two key student characteristics; gender, and whether the student is the first in family (FIF) to attend college. For each, we employ two distinct approaches. First, are non-parametric plots of the student percentile against the degree percentile by these binary characteristics. This visual representation offers a comprehensive understanding of match throughout the attainment distribution without imposing functional form assumptions. Second, we estimate the average difference in mismatch, and through conditioning on individual covariates explore the determinates of any match inequalities. Our analysis is descriptive, rather than causal, and we interpret the conditional match gaps as being informative to the choices made by students.

We document substantial gaps in academic match across the attainment distribution for FIF compared to non-FIF students of around 6 percentage points. FIF students consistently undermatch, applying to and enrolling in degrees with lower-achieving peers compared to their more connected counterparts, a gap that exacerbates with their level of qualifications. These gaps persist even after controlling for individual demographics and other measures of achievement. The average magnitude of this disparity is 2 percentiles, the difference between studying Aerospace Engineering at the University of Lisbon versus Languages and Culture at the University of Porto.

The situation with gender mismatch is more complex. Unconditional on other achievement measures, women are overmatched relative to males (have worse test scores than male peers on the same degree), but accounting for school GPA females undermatch in terms of applications and match on enrollment. This undermatch is driven by females in the middle of the attainment distribution choosing health degrees compared to males who choose engineering degrees. Accounting for applicants' preferred major we find that the most qualified female students are undermatching and provide evidence that this is a result of women being more constrained in terms of geographic choice than men.

Our paper makes several key contributions to the student-to-degree match literature. Most existing papers on mismatch, have focused on high-achieving low-income students, using a binary measure of undermatch (Hoxby and Avery, 2012; Black et al. 2015, Dillon and Smith, 2017), or examined mismatch at different points in the distribution. We use continuous measures of mismatch, which allow us to examine the extent of mismatch, rather than its existence according to an arbitrary cutoff. Using binary measures may overrepresent the extent of mismatch if the thresholds are set too low, or vice versa, having a continuous measure provides transparency in the magnitude of the situation. Our approach allows us to present estimates from across the distribution of achievement, in doing so we highlight that simply focusing on high achievers obscures systematic undermatch throughout the skills distribution. We build on Dillon and Smith (2017), who use a survey of 2,406 NLSY97 individuals to document mismatch at each quartile of student achievement, by using administrative data of the full population of applicants to all public higher educations in Portugal. We consider all applicants between 2013 and 2019, consisting of 1.6 million applications from 333,460 applicants. The detailed nature of the data allows us to calculate student quality percentiles, and to explore the role of major choice in student mismatch. This is crucial as we find that field of study plays a major role in mismatch across genders.

Campbell et al (2022) use a similar approach applied to the English higher educational system, but they are limited in their ability to isolate student choice as they do not have application data, and in the English application system universities play a key role by accepting or declining applications. We calculate application and enrollment mismatch measures, and we find that not observing applications obscures key differences in behavior by gender. This differs from Dillon and Smith (2017), which focuses on enrollment to compute mismatch. In comparison to these papers set in the UK and the US, we observe that the Portuguese system with centralized admissions through a Gale–Shapley algorithm has a much lower degree of student mismatch.

This paper is the first to study mismatch based on FIF enrollment. Previous studies have focused on the race of the student (Cotton et al. 2022) or a composite measure of socio-economic status (Campbell et al, 2022). Having a direct measure of the student's family exposure to the higher education system is critical in order to study the role that matching to degrees plays in intergenerational mobility. Our finding that talented FIF students are enrolling in degrees with less qualified students, and which command lower returns, undermines the potential for higher education to have a positive impact on social mobility.

The remainder of this paper is structured as follows. Section 2 provides an overview of the institutional setting, dataset, and methodology employed to construct our undermatch indices. Section 3 presents the results obtained through our approaches and discusses the results, while Section 4 concludes.

2. Data and Methods

2.1 Institutional setting

Portugal has a centralized application system for public higher education institutions (HEIs), with no application fees, and students are provided with standardized information on all the degrees including typical grade requirements. Students can rank up to six degrees in their application and are matched using a Gale Shapley algorithm. Students can apply for more than one degree within the same institution. Applications are based on students' high school GPA and final exams. These exams are compulsory and need to be passed to graduate high school. There are around 50k vacancies, 45k applications and 40k enrollments each year across 13 Universities and 16 Polytechnics.

2.2 Data

We use individual-level administrative data on the population of state-school students who applied to HE in Portugal between 2013/14 and 2019/20. This comprises around 50,000 students per year. We then exclude those from the vocational high-school track for whom national exams were not mandatory (10.5%). From those, we consider students who were admitted in the first round of application (79.5%) and then exclude students with no defined field of study (2.2%). These restrictions are necessary for our analysis but could result in underestimating the extent of the match gap.

The applications data comes from Direção-Geral do Ensino Superior (DGES, Directorate General for Higher Education) which includes all listed degree preferences, application portfolios, and student demographics including gender and FIF. Student enrollment data comes from the *Registry of student enrollment and graduation in higher education* (RAIDES) from Direção-Geral de Estatísticas da Educação e Ciência (DGGEC). Appendix Table A1 presents some descriptive statistics of the application patterns and enrollments. We categorize majors into 25 subjects according to the International Standard Classification of Education by field (ISCED-F narrow). The three most popular subjects in terms of first preferences are Health (15.7%), Business and Administration (15.6%), and Engineering and engineering trades (15%). For these subjects, women are 77%, 59% and 25% of all enrollments respectively. For FIF students, they are 56%, 64% and 46% of all enrollments.

2.3 Student-to-degree match

For each student-degree, we derive an application match and enrollment match. Both metrics are the difference between degree quality and student quality. Each measure is calculated in three steps:

(1) **Student quality:** Calculate the percentile rank of individuals determined by their high-school exam performance.[2]
   i. Application Match – Among all applicants for their first preference degree
   ii. Enrollment Match – Among all students that ultimately enrolled in the degree.

(2) **Degree quality:** Calculate the percentile rank of each degree, based on the median student's quality measure, weighted by the number of enrolled students, to account for capacity constraints.
   i. Application Match – Among all applicants for their first preference degree
   ii. Enrollment Match – Among all students that ultimately enrolled in the degree.

(3) **Calculate Match:** Degree Mismatch = Degree Quality - Student Quality

This provides two continuous measures of match for each student. Any differences between the two allow us to ascertain if the Portuguese application process exacerbates any gender or FIF match gaps. Table 1 shows the summary statistics for the analysis sample, and by gender and FIF.

[2] We have tested robustness to alternate measures of student quality including high school GPA, scaled average of high school exams and high school GPA with equal weights, and the entrance grade (with the actual weights used by the institutions). Our preferred measure is the average end-of-high school exams for two reasons. First, this is more objective than a high-school GPA. Second, entrance grades are not constant across degrees, which complicates the ranking of students.

Table 1: Summary Statistics

Variables	Total	Gender		FIF	
		Female	Male	Yes	No
High School GPA (0-200)	151.86	153.02	150.23	147.87	158.69
	(19.90)	(19.79)	(19.95)	(18.51)	(19.82)
Admission exams score (0-200)	140.18	140.48	139.76	136.29	147.31
	(25.86)	(25.48)	(26.38)	(24.75)	(26.05)
Student applied quality (percentile)	50.06	50.34	49.66	45.94	58.06
	(28.97)	(28.69)	(29.36)	(28.01)	(28.69)
Student enrolled quality (percentile)	50.04	50.45	49.45	45.77	57.93
	(28.96)	(28.61)	(29.44)	(28.02)	(28.69)
Degree applied quality (percentile)	49.90	50.38	49.22	43.99	57.66
	(28.85)	(28.60)	(29.27)	(27.82)	(28.61)
Degree enrolled quality (percentile)	49.98	50.97	49.18	44.73	58.81
	(28.91)	(28.61)	(29.27)	(27.75)	(28.47)
No. students	259,272	151,237	108,035	122,157	91,683
No. degrees	1,279	1,243	1,246	1,264	1,205
No. Applications	1,574,862	916,576	658,286	694,951	527,514
No. Applicants	333,460	193,232	140,228	145,361	105,749

Sources: Direção-Geral do Ensino Superior (DGES) and Registry of student enrollment and graduation in higher education (RAIDES, DGEEC). Standard deviations in parentheses.

2.5 Methods

To understand the nature of student matching we first show a simple plot of students' achievement decile against average degree quality for all students in that decile. If all students were perfectly matched to their degrees, this line would be straight and at a 45-degree angle. At the extremes of the quality distribution, this would require no variation in student quality within a degree, e.g. perfect segregation by qualifications. This does occur for some courses, e.g. in 2019 for the top-ranked degree of Aerospace Engineering at the University of Lisbon, 99 percent of students have perfect scores in the required subjects. The extent to which a point is above a 45-degree line indicates how overmatched these students are on average, and similarly, the distance below the 45-degree line reveals the extent of undermatch. Plotting this match-line for different types of students allows us to study inequalities in the match at any point in the achievement distribution.

Our second approach is to estimate FIF and gender gaps in the match, unconditionally at different points in the achievement distribution using the following regressions:

$$M_i = \alpha + \beta FIF_i + \gamma female_i + \varepsilon_i\,,\ \Delta a \tag{1}$$

Where M_i is our measure of the match of individual i, β represents our estimated FIF gap in the match, and γ is our estimated gender gap in the match. Given that achievement is used to define match, there will be ceiling and floor effects; it is impossible for the lowest-ranked students to undermatch or the highest-ranked student to overmatch. We therefore estimate the models separately, across deciles of achievement (a), allowing us to estimate the precise size of the match gap at different points in the distribution. To account for the role high-school GPA plays in admissions, we also condition on a flexible polynomial of student achievement (P_i).

We then explore the role of market failures and student preferences as potential drivers of gender and FIF gaps in the match. We consider a range of different mechanisms by including a series of additional covariates X_i in the model (1) representing different sources of potential market failure and preferences and interpret how this affects the estimated gaps.

$$M_{isma} = \alpha_a + \ \beta_a FIF_i + \gamma_a female_i + \pi_a P_i + \omega_{sa} + \mu_{ma} + \varepsilon_{isma},\ \Delta a \tag{2}$$

To explore the role of school-level factors in driving FIF and gender inequalities in the match, we control for school fixed effects (ω_s). This captures all factors constant at the school level, such as information, and sorting to schools. To explore the importance of preferences we consider the role of geography and degree subject. Our mismatch metric assumes no subject or geographic preferences. In reality, they will exist, so some of the students who we determine to be mismatched may not be, from the student's perspective. Note that the existence of market failures, such as misinformation, may cause students revealed preferences to not match their true normative preferences (Beshears et al 2008).

The inclusion of 25-degree major categories in the model (2) (μ_m) accounts for the average mismatch of students studying a certain subject area. Any reduction in the gender or FIF parameters represents the correlation between that characteristic and sorting into the subjects taken. Any within-subject gap remaining in this specification can be interpreted as institutional-driven inequalities.

3. Results

3.1 Raw Plots

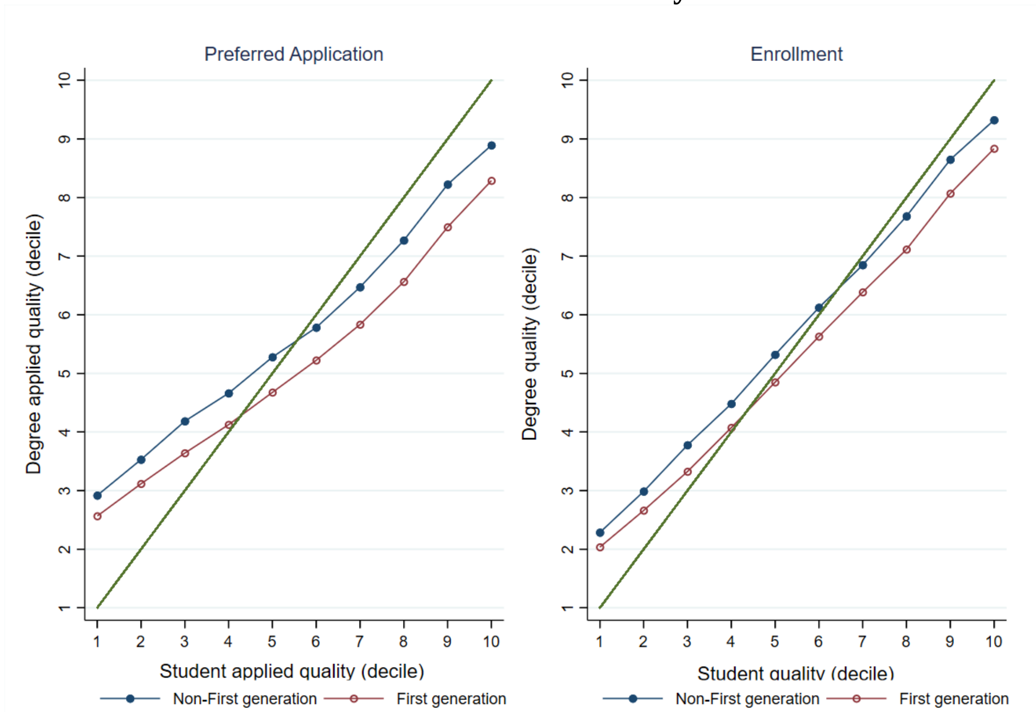
Figure 1 presents four panels plotting student quality against degree quality measures. The first column presents the application match, and the second column displays the enrollment match. The top row plots match rates by FIF, while the bottom row plots match rates by gender. In all panels, the 45-degree line reflects perfect matching, and the gradients of the plots reflect the actual matching between students and degrees they applied to as their first preference (column 1) or enrolled in (column 2). All panels show the relationship is approximately linear and is flatter than 45 degrees, meaning that low-attainers are more likely to overmatch and high-attainers are more likely to undermatch. Part of this will be mechanical, as it is impossible for the lowest-ranked students to undermatch or the highest-ranked students to overmatch.

Comparing the application match gradients to the enrollment match gradients, we see that enrollment relationships are closer to the 45-degree line meaning that the application system ameliorates the initial mismatch in student application patterns. To compare the Portuguese match rates to others found in the literature, we use the binary definition of mismatch from Dillon and Smith (2017) where the mismatch is +/- 20 percentiles from the perfectly matched degree. They find around 50 percent of US post-secondary students to be mismatched according to their composite college-input-quality measure.[3] Campbell et al (2022) using the same definition found 32 percent of enrolled students in England to be academically mismatched. In comparison, we find only 21.5 percent of Portuguese students are mismatched in terms of enrollment.

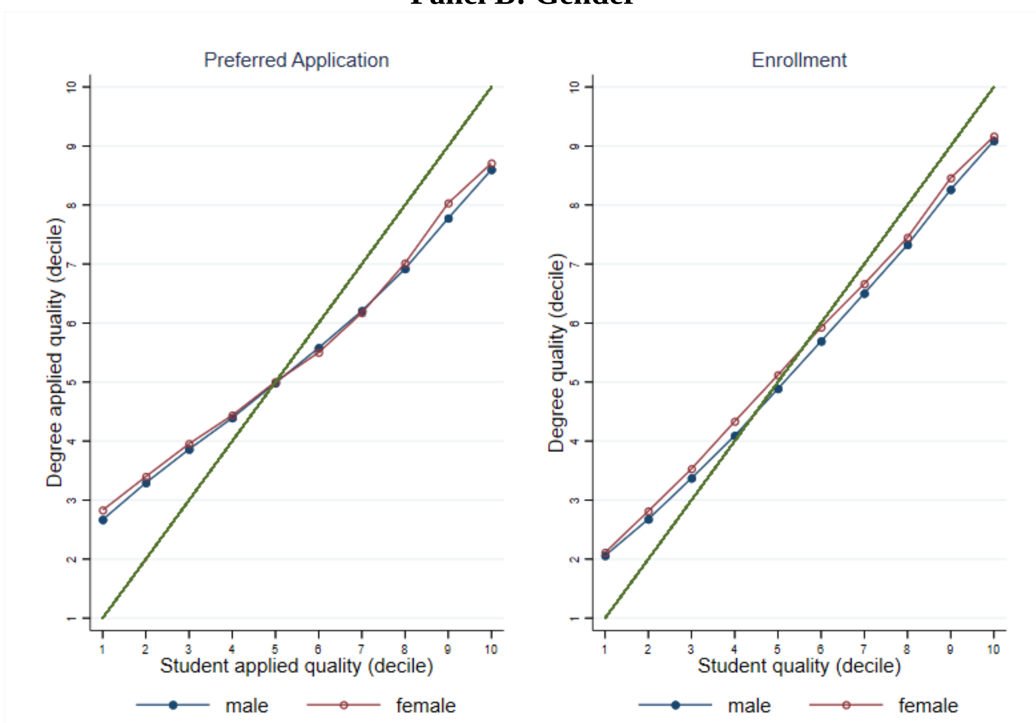
[3] Dillon and Smith's college quality measure comprises 4 measures of quality – the mean SAT score (or ACT score converted to the SAT scale) of entering students, the percent of applicants rejected, the average salary of all faculty engaged in instruction, and the undergraduate faculty-student ratio.

Figure 1: Student Quality Rank on Measures of Degree Quality Rank, By Gender and FIF

Panel A: First in Family



Panel B: Gender



Notes: Student and degree quality are as defined in Section 2.3 'Student-to-degree match'

While we cannot draw concrete conclusions from these international comparisons, there are three principal reasons why Portugal has better match rates. First, unlike the US, the centralized admission system makes it easier for students to apply to a range of institutions. Second, unlike the UK (which also has a centralized system) students apply to university after knowing their grades and so are more informed. Third, different from both the US and the UK, universities do not get to select students.

Focusing on the systematic difference in match by gender and FIF status, consider the vertical distance between the series in each panel. The top panels show FIF students consistently undermatch compared to non-FIF students. Importantly, it is not just the highest achieving FIF students undermatching, but at every decile FIF students undermatch on average compared to non-FIF students. Much of the literature on undermatch has focused on high-attaining low-SES students (Hoxby and Avery, 2012; Black, Cortes and Lincove, 2015, Dynarski 2022), our finding implies such focus underestimates the extent of mismatch.

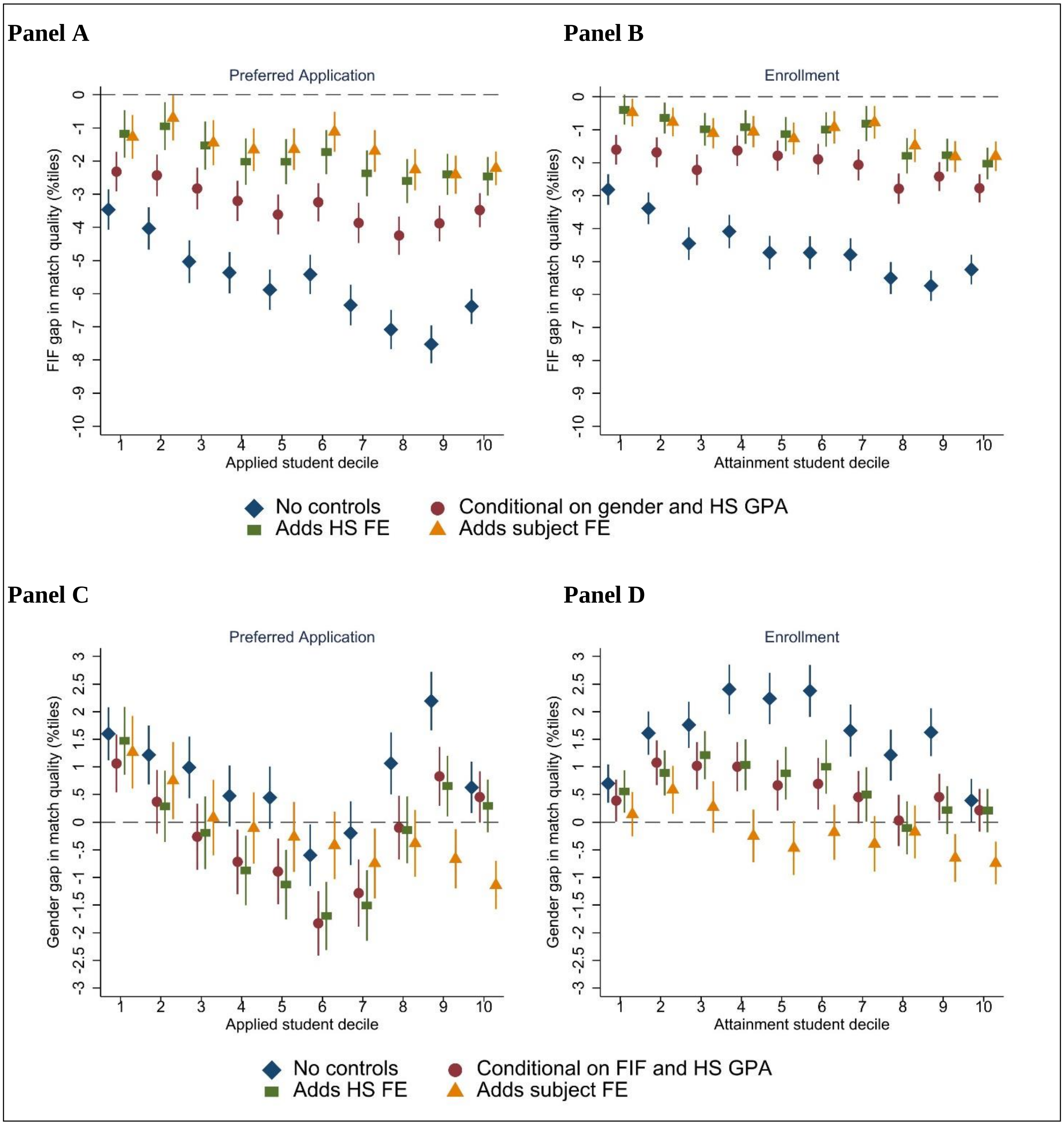
In contrast, the bottom panels of Figure 1 show there is very little difference in the application pattern by gender. In terms of enrollment females consistently overmatch compared to males, albeit by a small margin. This means that women with the same end-of-high school exam grades as men are enrolling in degrees with slightly higher peer quality, consistent with Campbell et al (2022).

3.2. Conditional Plots

Figure 2 presents estimates of the unconditional match gaps for each student type, and the gaps conditional on three additional sets of student characteristics: high school GPA, school effects, and subject effects. Each point represents the match gap at a decile of students and comes from a separate regression for the respective student type.

We begin by considering the top two panels of Figure 2 which plot the FIF student match gaps for applications and enrollment. The unconditional match gaps (diamond series) represent the raw match gap between groups and are equivalent to the raw gap between groups found in Figure 1. We see that the FIF student match gap in applications is steadily increasing through the achievement distribution from 3.5 percentiles to 7.5 percentiles. Interestingly, the match gap for enrollment starts smaller (2.8), and increases at a smaller rate, culminating in a maximum FIF student match gap of 5.7 percentiles.

Figure 2: FIF and Gender Conditional Match Inequalities



Notes: Student and degree quality are as defined in Section 2.3 ‘Student-to-degree match’

The second series additionally conditions on gender and high school GPA (circle series), which is used as part of the admissions process. Accounting for GPA reduces the FIF student application match gap by around 40% (top left panel), with the largest proportional reductions occurring for the higher test-score achievement deciles. This implies that for a given test-score decile FIF students have higher GPAs. With regards to enrollment match-gap conditioning on GPA has a greater proportional effect, reducing the FIF student match gap by 55% on average. As GPA is used in admissions, it would be reasonable to expect that conditioning on GPA closes the enrollment gap by more than the application gap. Despite this, we continue to find that FIF students are enrolling in degrees 2.1 percentiles lower in quality on average.

The bottom two plots in Figure 2 present the gender match gaps for applications and enrollment. The unconditional gender gaps (diamond series) are different from the unconditional FIF gaps in three key ways. First, the scale of the gender match gaps is considerably smaller than the FIF match gaps. At their largest, 9th decile of achievement, the gender match gap is 2.2 percentiles, which is overshadowed by the corresponding FIF gap of 7.5 percentiles.

The second observation is that the gender match gaps are not monotonically increasing in student achievement. Broadly speaking, for the unconditional application match, women in the bottom and top thirds overmatch relative to males, while they tend to match to similar degrees in the middle of the distribution. Therefore, on average, women are applying to degrees with higher quality peers than their male counterparts.

A likely determinant of this pattern are the differences in subject preferences by gender, and the relative availability of such degrees in the quality distribution. The most popular topic for female applicants is health, and for males it is engineering. Health degrees are concentrated at the very top and bottom half of the course quality distribution, while engineering courses mostly appear in the top half of the quality distribution. The consequence of this is that females at the 8th and 9th decile attempt to overmatch by applying to health courses at the 10th decile, while males at the same deciles can apply to engineering courses which they match. Given the lack of health degrees in the upper half of the distribution females at the 6th-7th deciles are consequently applying to courses below their qualifications and so are relatively undermatching compared to males.

The third key difference is that the patterns for applications and enrollment are different. The u-shaped pattern of the application match gap is inverted into a hump for the enrollment gap. Now the gender gap is smallest for low and high-achieving students, and largest for median students, meaning that the application to enrollment mapping is different by gender. Note that the degree quality for enrollment is based on students who ultimately enroll in the degree, while the application match is based on the qualifications of students who put it as their first preference. The median female student preferred application to degrees for which they are as qualified as male applicants but ultimately enrolled in courses for which they are overmatched (underqualified) compared to males.

Conditioning on high-school GPA (circle series), reduces female overmatch for both applications and enrollment. This is reflective of females having higher GPAs conditional on final test scores. This means that only women in the bottom and top two deciles continue to overmatch relative to males in terms of applications. We now see that women are undermatching in terms of applications in the middle of the test score distribution by 1.8 percentiles, i.e. women with the same test score and GPA as men are applying to degrees with lower attaining peers than their male counterparts.

Similar to application gaps, conditioning on GPA reduces the female enrollment overmatch, but they continue to overmatch compared to their male counterparts. Women in the middle of the achievement distribution are applying to degrees with lower quality peers compared to men as their first preference but ultimately enroll in degrees with higher quality peers. There is no gender enrollment match gap for the top-performing students.

We now explore the potential drivers of these gaps through the inclusion of high-school fixed effects ω_s (square series) to proxy for some market failures and subject fixed effects μ_m (triangle series) to proxy for preferences.

High school fixed effects account for any factor that is constant at the high school level. This will include any geographic factors such as distance to universities, and so will pick up transportation costs and potentially informational constraints. The school fixed effects will also account for any constant socio-economic factors such as the average income of families attending the school, and so may account for credit constraints, as well as informational constraints.[4] The gap parameters now represent the average difference in application and enrollment choices by students with the same high school test scores, GPA and who attended the same school.

[4] Appendix Figure A1 shows how the estimates change when geographic and socio-economic school factors are included separately. These factors are: Region; % of the high school population that is FIF; % of the high school population that attends prestigious university; % of the high school population that specializes in science; % of the high school population that is female. The factor that reduces the FIF gap by the largest extent is the proportion of FIF students enrolled at the high school, followed by the region.

The inclusion of school fixed effects reduces the magnitude of the FIF application gap by a further 50 percent for all but the top-performing students (Panel A). This implies that factors correlated with the school attended play a large role in the type of degree students apply to. Still, the FIF application gap remains between 1 and 2.5 percentiles. Similar reductions occur with the enrollment gap (reduced by 55 percent), which leaves no difference for the lowest-achieving students.

In contrast to the FIF gap, the gender match gap for applications or enrollment is unaffected by the inclusion of school fixed effects (Panels C & D). This implies that any market failures at the school level do not exacerbate the gender gap. This should not come as a shock, as, unlike FIF where the concentration varies across schools, all public schools in Portugal are not unisex, and so the school effects would impact both genders equally. This does not rule out the possibility that geographic market failures are not playing a role. The fact that a gender gap continues to exist may imply that some market failures impact women more than men, such as differential distance costs.

The role of subject preferences has the potential to be a large determinant of mismatch. Thus far we have made a strong assumption that students have no preferences and would simply want to attend the degree that most closely matches their national achievement rank. If we observe a reduction in the FIF or gender parameters with the inclusion of subject effects, then the gaps are partially driven by students sorting into subjects due to preferences. Any gap remaining will be due to the group attending lower-quality institutions, regardless of the subject studied.

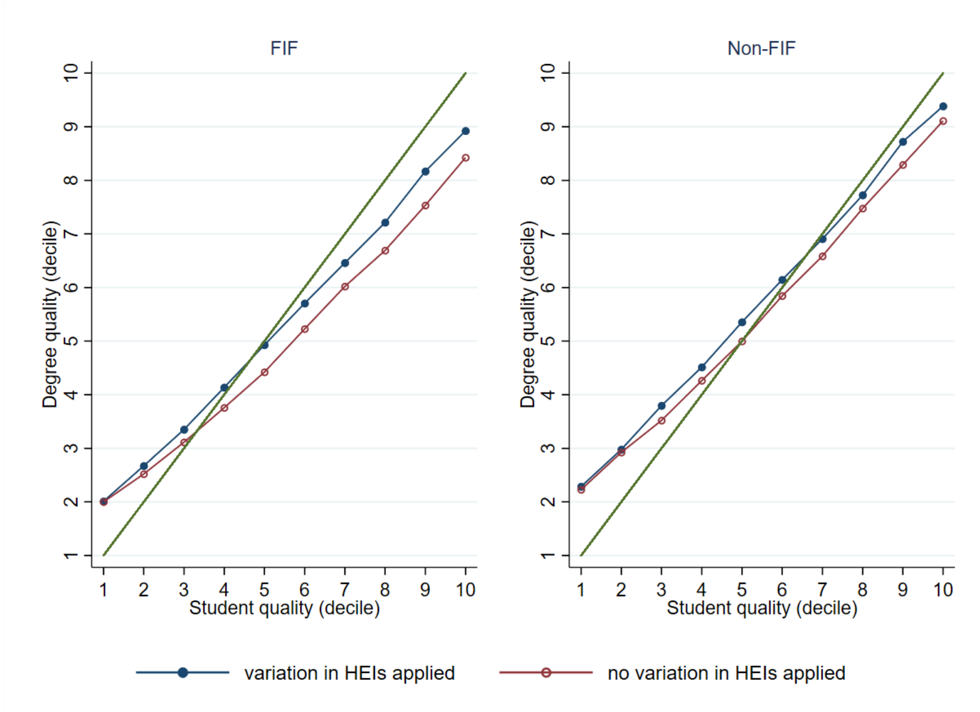
The FIF match gap for application and enrollment is largely unaffected by the inclusion of school effects, but the gender match gap for both measures is greatly reduced. The implication is that the undermatching by FIF students is not driven by the subjects they choose but by the universities. First-generation students with the same qualifications and subject preferences as non-FIF students tend to attend institutions with lower-achieving peers. This could reflect preferences, e.g. big fish in a small pond, or market failures, e.g. geographical constraints.

In contrast, once first subject preference is accounted for, no gender gap remains for most of the attainment distribution. This implies that geographic market failures are not playing a role in the gender gap. The exception is at the top, where females from the top two deciles systematically undermatch to degrees, e.g. women are enrolling in institutions with relatively lower-quality peers. The gender gap only occurring at the top of the achievement distribution conditional on the subject is consistent with the estimates for England (Campbell et al 2022). We contribute to their finding by establishing that this pattern occurs at the application stage, rather than being a consequence of the student-to-degree match system, and that it continues to occur in a different institutional setting. Again, this could be reflective of differences in preferences, or women facing a higher cost of attending an institution further from home.

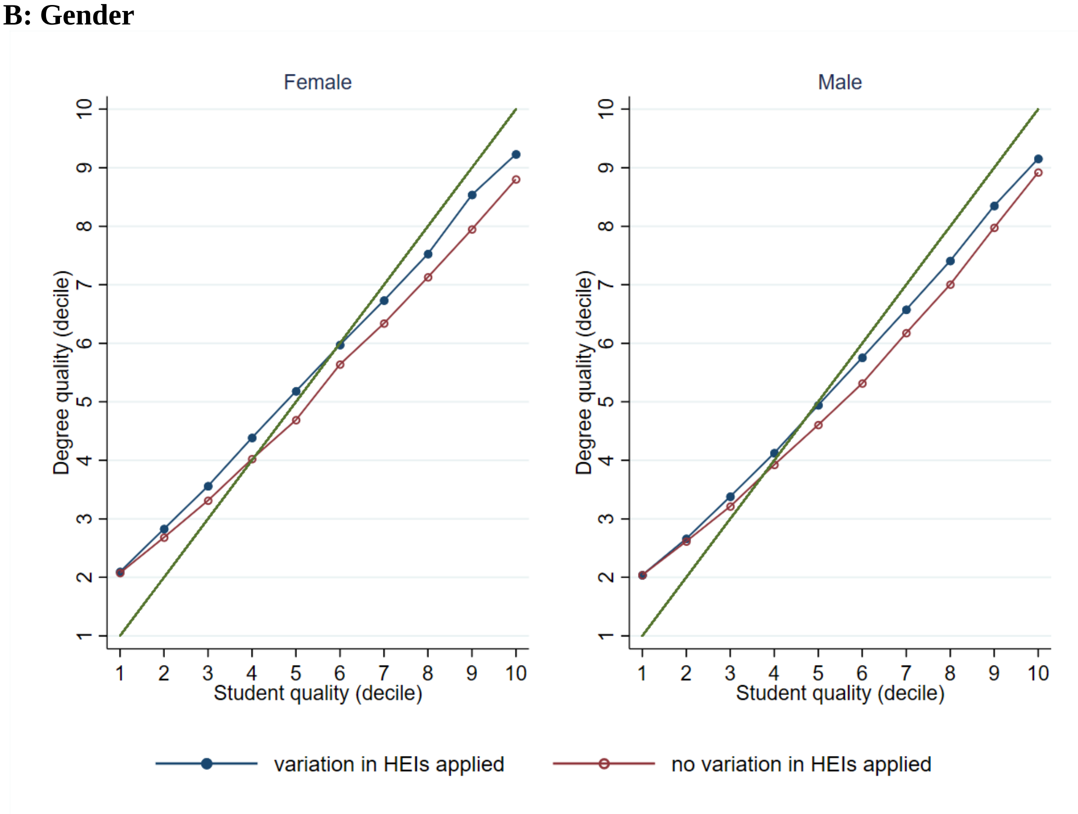
To speak to this, Figure 3 shows how the enrollment match rate changes by different sets of preferences at the application stage. A student can have up to six degree applications. The figure shows enrollment for students that had variation in their institutions among their applications and those that did not, by FIF status (Panel A), and gender (Panel B). Note, that we found extremely similar match rates for students who only applied for one subject, compared to students who applied for multiple, and so have not presented these figures. Consequently, the diversity (or lack of) of subjects chosen is likely not a driver of the gender-match gap. In contrast, students who only applied to one institution are more likely to undermatch compared to students with less narrow preferences. Considering the first FIF (Panel A) we see that FIF students undermatch more and the gap increases by student achievement (lower intercept, shallower gradient). Moreover, we see students typically enroll in a course of lower quality if they only apply to one institution, with this disparity being 0.15 percentiles larger for FIF students. With regards to gender (Panel B), female students have better match rates, as seen by the gradient being closer to 45 degrees, but the reduction in course quality associated with applying to only one institution is also on average 0.15 percentiles larger for women. For both panels, we observe that the extent of undermatch relative to students who apply to more than one institution increases throughout the ability distribution. A potential explanation for this effect is that lower-qualified students are able to find matching degrees at their local institution, while students in the top decile may have to travel to other regions. This pattern is consistent with the gender match gap conditional on preferred major. Moreover, comparing the differences in location preferences by gender, females have stronger institutional revealed preferences than men. Only 15 percent of female applicants apply to more than one institution, compared to 18.5 percent of male students. We consider differential transport costs by gender to be a prime candidate for the gender match gap.

Figure 3: Limiting Choice Set and Match Rate in Applications to Institutions (HEI)

Panel A: FIF



Panel B: Gender



Notes: Student and degree quality are as defined in Section 2.3 'Student-to-degree match'

4. Conclusions

A large proportion of Portuguese students are enrolled in degrees to which they are mismatched. The extent of mismatch in Portugal, however, is considerably smaller than in the UK or US. We highlight three reasons why this might be the case. First, unlike the US, Portugal has a centralized admission system. Second, unlike the UK (which also has a centralized system) students apply to university after receiving their grades and are therefore better informed. Third, different from both the US and the UK, universities do not get to select students. By employing new data on student application patterns, we find that the Portuguese Gale-Shapley matching algorithm reduces the extent of mismatch found in initial student applications.

Despite this, we still find significant inequalities in the match by the type of student. Students who are the first in their family to attend university are more likely to undermatch throughout the achievement distribution. This has large implications for the effectiveness of post-secondary education in improving intergenerational mobility. A policy that only focuses on increasing participation in post-secondary education will not reflect the differences in the college experience of similarly qualified FIF and non-FIF students. Such enrollment gaps are attenuated to a significant degree with the inclusion of high school effects, pointing to the role of factors, such as peers, parental sorting, and geography as likely key drivers for improving FIF student-to-degree match.

With regards to the gender match gap, unconditionally women are overmatching to degrees, but when accounting for their higher GPA women are matching to similar or worse degrees than men in terms of peer quality. Subject choice plays a key role in closing this gap. However, small gaps remain for high-attaining women, meaning that even when they enroll in a similar field as men, they still appear to study at lower-quality institutions. We find evidence that women are more limited in their geographic choices than men which could be driving these differences. This finding has implications for the gender pay gap, suggesting that higher education plays an important role in this much-studied issue.

References

Arcidiacono, P., & Lovenheim, M. (2016). Affirmative action and the quality–fit trade-off. *Journal of Economic Literature*, *54*(1), 3-51.

Beshears, J., Choi, J. J., Laibson, D., & Madrian, B. C. (2008). How are preferences revealed?. *Journal of public economics*, *92*(8-9), 1787-1794.

Bettinger, E. P., Long, B. T., Oreopoulos, P., & Sanbonmatsu, L. (2012). The role of application assistance and information in college decisions: Results from the H&R Block FAFSA experiment. *The Quarterly Journal of Economics*, *127*(3), 1205-1242.

Black, S. E., Cortes, K. E., & Lincove, J. A. (2015). Academic undermatching of high-achieving minority students: Evidence from race-neutral and holistic admissions policies. *American Economic Review*, *105*(5), 604-610.

Bleemer, Z. (2022). Affirmative action, mismatch, and economic mobility after California's Proposition 209. *The Quarterly Journal of Economics*, *137*(1), 115-160.

Campbell, S., Macmillan, L., Murphy, R., & Wyness, G. (2022). Matching in the Dark? Inequalities in student to degree match. *Journal of Labor Economics*, *40*(4), 807-850.

Carneiro, P., & Heckman, J. J. (2002). The evidence on credit constraints in post-secondary schooling. *The Economic Journal*, *112*(482), 705-734.

Cotton, C. S., Hickman, B. R., & Price, J. P. (2022). Affirmative Action and Human Capital Investment: Evidence from a Randomized Field Experiment. *Journal of Labor Economics*, 40(1), 157-185.

Dillon, E. W., & Smith, J. A. (2017). Determinants of the match between student ability and college quality. *Journal of Labor Economics*, *35*(1), 45-66.

Dillon, E. W., & Smith, J. A. (2020). The consequences of academic match between students and colleges. *Journal of Human Resources*, *55*(3), 767-808.

Dynarski, S., Libassi, C. J., Michelmore, K., & Owen, S. (2021). Closing the gap: The effect of reducing complexity and uncertainty in college pricing on the choices of low-income students. *American Economic Review*, *111*(6), 1721-1756.

Dynarski, S., Nurshatayeva, A., Page, L. C., & Scott-Clayton, J. (2022). *Addressing non-financial barriers to college access and success: Evidence and policy implications* (No. w30054). National Bureau of Economic Research.

Hoxby, C. M., & Avery, C. (2012). *The missing" one-offs": The hidden supply of high-achieving, low income students* (No. w18586). National Bureau of Economic Research.

Hoxby, C. M., & Turner, S. (2015). What high-achieving low-income students know about college. *American Economic Review*, *105*(5), 514-517.

Kirkeboen, L. J., Leuven, E., & Mogstad, M. (2016). *Field of study, earnings, and self-selection.* The Quarterly Journal of Economics, 131(3), 1057-1111.

Lochner, L. J., & Monge-Naranjo, A. (2011). The nature of credit constraints and human capital. *American economic review*, *101*(6), 2487-2529.

McGuigan, M., McNally, S., & Wyness, G. (2016). Student awareness of costs and benefits of educational decisions: Effects of an information campaign. *Journal of Human Capital*, *10*(4), 482-519.

Mountjoy, J. (2022). Community colleges and upward mobility. *American Economic Review*, *112*(8), 2580-2630.

Mountjoy, J., & Hickman, B. R. (2021). *The returns to college (s): Relative value-added and match effects in higher education* (No. w29276). National Bureau of Economic Research.

Murphy, R., & Wyness, G. (2023). Testing Means-Tested Aid. *Journal of Labor Economics*, *41*, 3, 687-727.

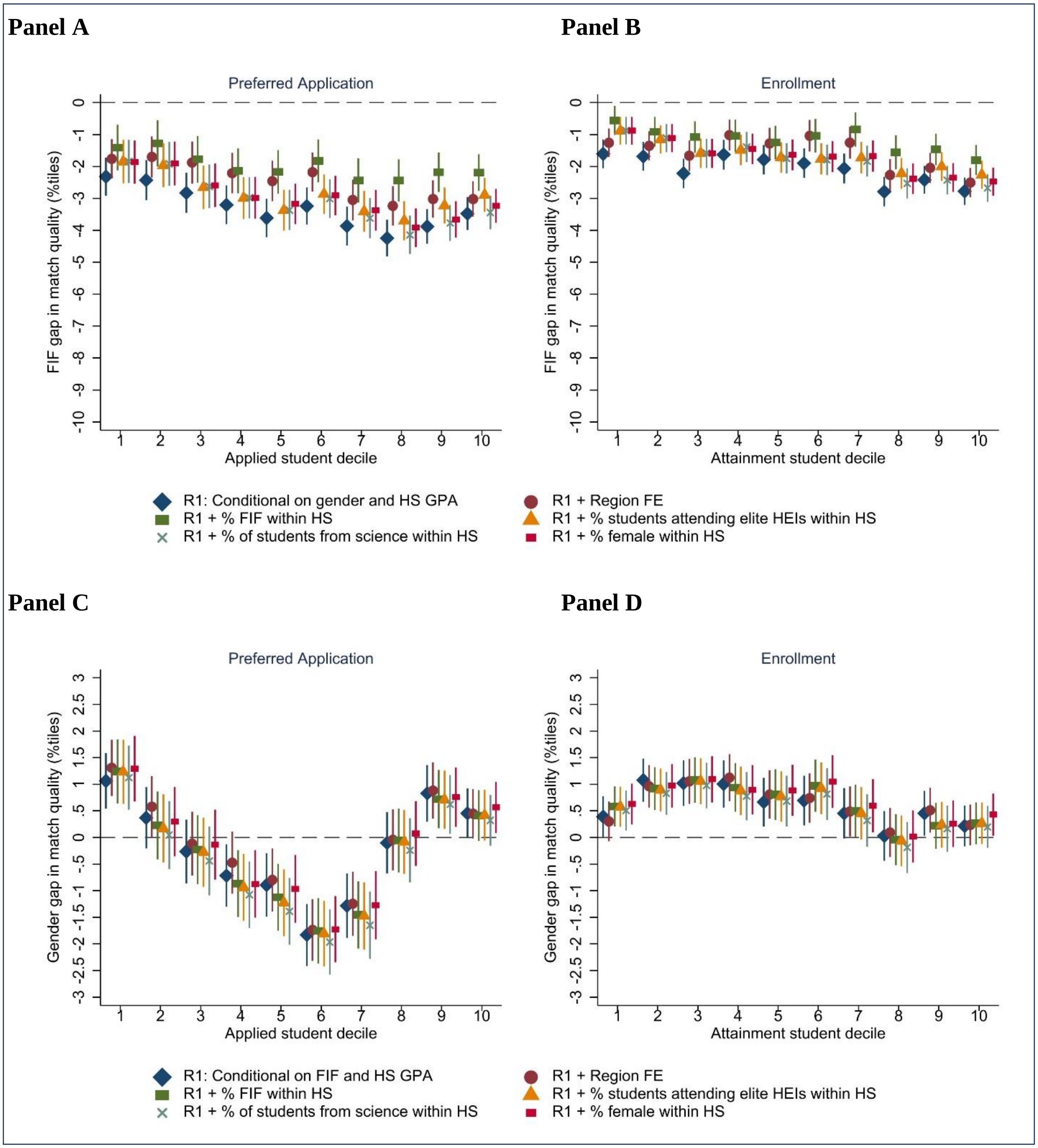
Zimmerman, S. D. (2019). Elite colleges and upward mobility to top jobs and top incomes. *American Economic Review*, *109*(1), 1-47.

Appendix

Table A1: Summary Statistics of Application and Enrollments for the Period 2013-2019

	Applications		Enrollments		
By ISCED-F Narrow	**% of applicants**	**% of applicants in 1st preference**	**Total (%)**	**% Female**	**% FIF**
Education	1.62%	1.56%	2.05%	90.92%	77.38%
Arts	6.21%	7.51%	6.21%	67.22%	57.97%
Humanities (except languages)	1.55%	1.11%	1.67%	50.93%	67.59%
Languages	3.42%	3.59%	3.59%	74.62%	69.11%
Social and behavioural sciences	11.74%	10.88%	9.72%	64.53%	58.03%
Journalism and information	2.94%	3.10%	2.15%	73.77%	63.83%
Business and administration	16.36%	15.65%	14.63%	58.90%	64.25%
Law	4.54%	6.00%	4.54%	71.55%	56.19%
Biological and related sciences	4.74%	4.32%	5.04%	65.23%	55.23%
Environment	0.12%	0.08%	0.14%	46.65%	70.39%
Physical sciences	1.98%	1.58%	2.11%	41.98%	48.04%
Mathematics and statistics	1.20%	0.96%	1.16%	55.17%	48.31%
IT and Communication	1.60%	1.17%	1.63%	25.30%	52.85%
Engineering and engineering trades	15.21%	14.96%	16.31%	25.10%	46.04%
Manufacturing and processing	0.69%	0.36%	0.67%	47.52%	53.99%
Architecture and construction	2.10%	1.95%	2.83%	45.98%	45.56%
Agriculture	0.39%	0.39%	0.58%	48.81%	50.23%
Forestry	0.05%	0.02%	0.05%	32.56%	52.78%
Veterinary	0.88%	1.22%	1.06%	79.99%	44.95%
Health	15.68%	15.73%	16.17%	77.54%	56.70%
Welfare	1.61%	1.55%	1.83%	90.22%	83.64%
Personal services	5.02%	5.97%	5.40%	46.11%	63.90%
Hygiene and Occupational Health	0.19%	0.11%	0.28%	63.09%	68.15%
Security services					
Transport services	0.14%	0.21%	0.18%	19.53%	40.82%
Unknown or transversal	0.02%	0.02%	-	-	-
Total	1,574,862	333,460	259,272	259,272	213,840

Figure A1: Additional Regression on FIF and Gender Conditional Match Inequalities



Notes: Blue diamonds represent the FIF and Gender gaps conditional on gender, FIF and High School GPA. Each other series represents the gap additionally conditioning on a school level covariate: Region of origin FE; FIF proportion; Higher Education elite status; the proportion of Science Technology; the proportion of female.